SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE BOURGOGNE

Intégration de systèmes d'acquisition de données spatiales et spectrales haute résolution, dans le cadre de la génération d'informations appliquées à la conservation du patrimoine

Camille SIMON CHANE

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE BOURGOGNE

N° X X X

THÈSE présentée par

Camille SIMON CHANE

pour obtenir le

Grade de Docteur de
l'Université de Bourgogne

Spécialité : **Instrumentation et Informatique de l'Image**

Intégration de systèmes d'acquisition de données spatiales et spectrales haute résolution, dans le cadre de la génération d'informations appliquées à la conservation du patrimoine

Unité de Recherche :
Laboratoire le2i

Soutenue le mars 2013 devant le Jury :

Kadi BOUATOUCH	Rapporteur	Université Rennes I, France
Alain TRÉMEAU	Rapporteur	Université Jean Monnet, France
Markku HAUTA-KASARI	Examinateur	University of Eastern Finland, Finlande
András JUNG	Examinateur	Cubert GmbH, Allemagne
Marcello PICOLLO	Examinateur	Istituto di Fisica Applicata Nello Carrara, Italie
Robert SITNIK	Examinateur	Warsaw University of Technology, Pologne
Frank BOOCHS	Directeur de thèse	Fachhohchschule Mainz, Allemagne
Franck MARZANI	Directeur de thèse	Université de Bourgogne, France

Abstract

The concern and interest of this PhD thesis is the registration of featureless 3D and multispectral datasets describing cultural heritage objects. In this context, there are few natural salient features between the complementary datasets, and the use of targets is generally proscribed. We thus develop a technique based on the photogrammetric tracking of the acquisition systems in use.

A series of simulations was performed to evaluate the accuracy of our method in three configurations chosen to represent a variety of cultural heritage objects. These simulations show that we can achieve a spatial tracking accuracy of 0.020 mm and an angular accuracy of 0.100 mrad using four 5 Mpx cameras when digitizing an area of 400 mm $\times$ 700 mm. The accuracy of the final registration relies on the success of a series of optical and geometrical calibrations and their stability for the duration of the full acquisition process.

The accuracy of the tracking and registration was extensively tested in laboratory settings. We first evaluated the potential for multiview 3D registration. Then, the method was used for to project of multispectral images on 3D models. Finally, we used the registered data to improve the reflectance estimation from the multispectral datasets.

Key-words: 2D–3D registration, close range photogrammetry, optical calibration, 3D digitization, multispectral imaging, cultural heritage

Résumé

Cette thèse s'intéresse au recalage de données issues de capteurs 3D et multispectraux pour l'étude du patrimoine. Lorsque l'on étudie ce type d'objet, il y a souvent peu de points saillants naturels entre ces jeux de données complémentaires. Par ailleurs, l'utilisation de mires optiques est proscrite. Notre problème est donc de recaler des données multimodales sans points caractéristiques. Nous avons développé une méthode de recalage basé sur le suivi des systèmes d'acquisition en utilisant des techniques issues de la photogrammétrie.

Des simulations nous ont permis d'évaluer la précision de la méthode dans trois configurations qui représentent des cas typiques dans l'étude d'objets du patrimoine. Ces simulations ont montré que l'on peut atteindre une précision du suivi de 0,020 mm spatialement et 0,100 mrad angulairement en utilisant quatre caméras 5 Mpx lorsque l'on numérise une zone de 400 mm × 700 mm. La précision finale du recalage repose sur le succès d'une série de calibrations optiques et géométriques, ainsi que sur leur stabilité pour la durée du processus d'acquisition.

Plusieurs tests ont permis d'évaluer la précision du suivi et du recalage de plusieurs jeux de données indépendants ; d'abord seulement 3D, puis 3D et multispecrales. Enfin, nous avons étendu notre méthode d'estimation de la réflectance à partir des données multispectrales lorsque celles-ci sont recalées sur un modèle 3D.

Mots-clés : recalage 2D–3D, photogrammétrie, calibrations optiques, numérisation 3D, imagerie multispectrale, étude du patrimoine

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem statement

There is an increasing use of imaging techniques to study cultural heritage objects. These non-destructive, contactless, optical techniques may not provide as accurate answers as chemical analysis based on spot techniques or sampling, but they provide data regarding large surfaces, even the full object. What they lack in discrimination they compensate in quantity of data acquired. The strategy is thus to multiply the acquisitions on comparable objects with complementary techniques.

We are particularly interested in two techniques that are independently gaining widespread use for the study of cultural heritage: 3D imaging and multispectral imaging. The first if these techniques models and records the 3D surface of the object under study. These models can be used to observe the surface structure of an object without manipulating it. Art scholars can easily examine fine brushstrokes on paintings and chisel marks of statues. 3D models can also be used for communication purposes to create virtual museums or as virtual archives of the object. Multispectral imaging aims at recording the object reflectance properties. Common applications include producing more faithful color reproductions, guiding pigment identification and increasing readability of covered surfaces.

Combining these two methods provides 3D models with multispectral texture that are most useful for analysis purposes, highlighting spectral and structural correlations. These integrated models benefit conservators who can make insights on causation relationships in spectral and structural alterations. Such models also have clear advantages for pedagogical purposes. However, the necessary registration of multispectral and 3D data is not an easy task, particularly when dealing with data describing cultural heritage.

On one hand, the main advantage of these techniques which has prompted their widespread use is the fact that they are contactless. When studying cultural heritage, it is often forbidden not only to touch the object under study, but also to approach it closer than 20 cm or 50 cm. Though in some specific cases it may be possible to use targets to guide the registration, these defeat the purpose of using contactless acquisition systems. However, registration methods that rely on the structure of the data itself are often not adapted: the very fact that these techniques

are complementary means that there may be few or no corresponding features between the multimodal datasets. The problem can be not only that of finding *corresponding* salient points, but of finding *any* salient point. Smooth surfaces covered with a flat tint, which are not unusual in paintings, present no salient point, be it 3D or multispectral.

Our application requires a contactless registration technique that does not depend on the quality or content of the data. That is, a method to register smooth and featureless multimodal datasets. If these integrated models are to be used for analysis purposes a high registration accuracy is essential to avoid drawing false conclusions.

Tracking the acquisition systems in use moves the focus away from the object and onto the acquisition systems — which we do not mind manipulating. Optical techniques are the path of predilection for contactless acquisitions and we build on these strengths to devise a registration technique based on photogrammetry. This means we will need a set of camera systems in addition to the acquisition systems, to perform the tracking. Though this may seem cumbersome, photogrammetry is a highly flexible technique which can provide accurate results.

This interdisciplinary work relies on many distinct fields in image processing (3D data acquisition and processing, multispectral acquisition and processing, registration, photogrammetry) while keeping a strong sensitivity to the needs and requirements of the conservation community.

1.2 Context

This work stems from a long and on-going collaboration between the le2i laboratory in Dijon (France) and i3mainz laboratory in Mainz (Germany). Since 2004 these two laboratories have been developing common projects along the thematics of 3D data processing, multispectral imaging, cultural heritage and knowledge management. This collaboration is marked by many exchange students (in particular bachelor and masters internships), several co-advised PhDs, and research stays of several month for post-docs. These exchanges have resulted in a variety of common publications (among which [Ben Hmida *et al.*, 2012; Karmacharya *et al.*, 2010a,b; Li *et al.*, 2008, 2009]).

As a co-advised PhD student myself, I spent time both in Dijon and in Mainz, in roughly six month alternations. A few additional visits from France to Germany were necessary, since all acquisitions were performed in Mainz. Support for multispectral data acquisition and processing was provided by the le2i, while I relied on i3mainz's expertise on 3D digitization and photogrammetry. Insights for cultural heritage were provided by Adrian Heritage, formally from the Cologne Institute of Conservation Science (Cologne, Germany) and by our collaborators from the IFS (Institut für Steinkonservierung — Institute for stone conservation, Germany).

1.3 Motivation

I3mainz laboratory has strong ties with many local cultural institutions such as RGZM (Römish-Germanishes Zentralmuseum — Roman-German Central Museum, Mainz, Germany) and IFS. As part of the collaboration with IFS, several stone artifacts have been digitized by i3mainz for monitoring purposes, in particular a sandstone sarcophagus from the late 3rd century A.D.

Figure 1.1: 3D rendering of the crypt and the sarcophagus.

and a series of wall paintings from the 16th century. The objects, the conservation problems they pose, and past acquisitions are quickly presented to highlight the need for a registration method adapted to this type of data.

1.3.1 Objects of interest

Sarcophagus

This sarcophagus was discovered by archaeologists about fifty years ago in a crypt under the Friedhofs Chapel of theSt. Matthias Abbey, Trier (Germany). Unfortunately, the microclimate in the crypt is uncontrolled and the fluctuating relative humidity and air flow around the sarcophagus is damaging the stone and its fragmentary remains of polychromy. This is particularly the case on the area of the sarcophagus facing the entrance of the chamber (see figures 1.1 and 1.2). This side of the sarcophagus is the most subject to erosion as air enters from the stairway and flows out through openings over the sarcophagus. Traces of polychromy on the surface of the sarcophagus are flaking while the stone itself erodes.

An area of approximately 400 mm × 700 mm — the left half of the most degraded face — has been digitized and studied. 3D models of this area were acquired in December 2008, December 2009 and January 2011. A few adhesive targets were stuck to the surface to guide the 3D registration of the multiple views and to permit the comparison of the 3D models over time. However, the same process that is damaging the sarcophagus surface is also causing these targets to fall, making them unreliable global targets for temporal registration. In 2009, seven multispectral acquisitions were also performed. The need to precisely register the multispectral data with the 3D model stems naturally from the will to understand the correlations between the structural and spectral surface degradation.

Figure 1.2: Front face of the sarcophagus. Notice the remains of polychromy, the surface flaking and the uncoded targets surrounding the area under study.

Wall paintings

The group of 16^{th} century wall paintings we study is located in the Brömser Hof in Rüdesheim (Germany) — a small castle that still belongs to the Brömser family and currently houses a regional museum of mechanical instruments. Two adjacent rooms of this castle are entirely covered with wall paintings: the chapel (Hauskapelle) and the Ancestor Hall (Ahnensaal). Three inscriptions on the wall paintings date them to 1558 and 1559 and they are attributed to Hans Ritter known as Döring and a student of Lucas Cranac the Elder. The wall paintings were discovered and freed in 1898 – 1900. They are globally well preserved except for a few small losses. Unfortunately, they are in unstable conditions.

Due to damage to the building roof during the second world war, the wall paintings were exposed to weathering for several years. In the 1980s the roof was repaired and a horizontal barrier was installed to stop the penetration of moisture. However, the effects of past humidity are still felt today, causing the surface to flake as salts resurface. In particular, this has caused the formation of an overlying casein film on the wall paintings. Not only does this film have a strong visual impact per se, as it turns yellow with aging, but it also binds highly to dust and dirt, further tarnishing the readability of the painting. Moreover, this film becomes brittle and cracks as it ages, ripping off the paint layer. This process is exacerbated by the climate variability in the building. The humidity has more impact on the coating than on the wall, causing the two to dissociate.

In the 1990s small-scale consolidation experiments were performed on the paint layer without removing the heterogeneous casein film. Unfortunately it became evident in 2005 that these experiments had no lasting effects. These wall paintings were again partially cleansed in 2009

Figure 1.3: Flaking wall painting representing a bird on a balcony. Coded and uncoded targets are scattered around the area of interest.

to try to remove this casein film. Our goal is to monitor changes in these wall paintings and to detect any differences in the alteration of the restored and unrestored surfaces.

We have studied three areas in the past. The first is area, from the chapel room, measures approximately 500 mm × 500 mm and represents a bird on a balustrade (see figure 1.3). The flaking surface was digitized using a 3D scanner in October 2008. A series of adhesive coded and uncoded targets was once again used to guide the 3D registration. Two large multispectral acquisitions were also performed on this area, as well as three smaller ones, in February 2009. The small acquisitions were repeated in March 2010.

The second wall painting we are interested in is a full wall of the Ancestor Hall representing Jonah and the whale in the Rhine river, in front of the city of Mainz (see figure 1.4). It is the first known representation of Mainz. The right half of this wall was cleansed in 2009. Multispectral acquisitions of both the cleansed and non-cleansed side were performed in February 2009 and repeated in March 2010. However, no 3D digitizations were performed of this large area (measuring more than 1.5 m × 2 m). This would have required a high number of targets which would have greatly impaired the readability of the wall painting.

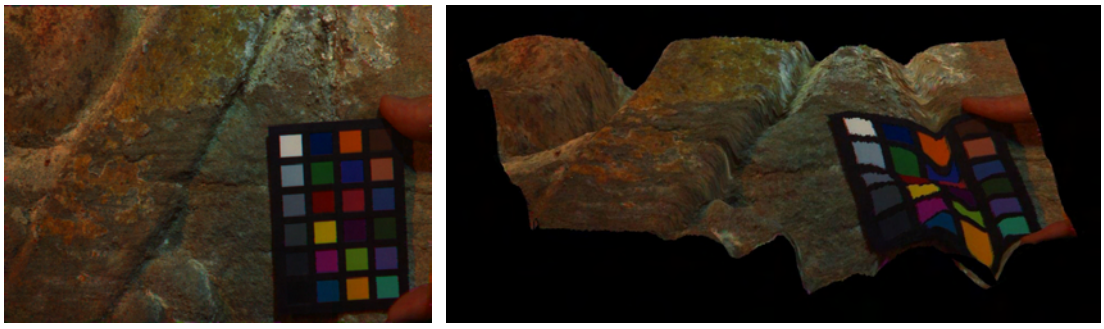
The third area we study is a panel near the window of the Ancestor Hall. This last area, also presents visible flaking (see figure 1.5). Three multispectral acquisitions were performed in February 2009 and repeated in March 2010.

1.3.2 Previous registration technique

We acquired both 3D and multispectral data for two areas of interest : the sarcophagus and the first wall painting. Preliminary registration tests were based on the manual input of cor-

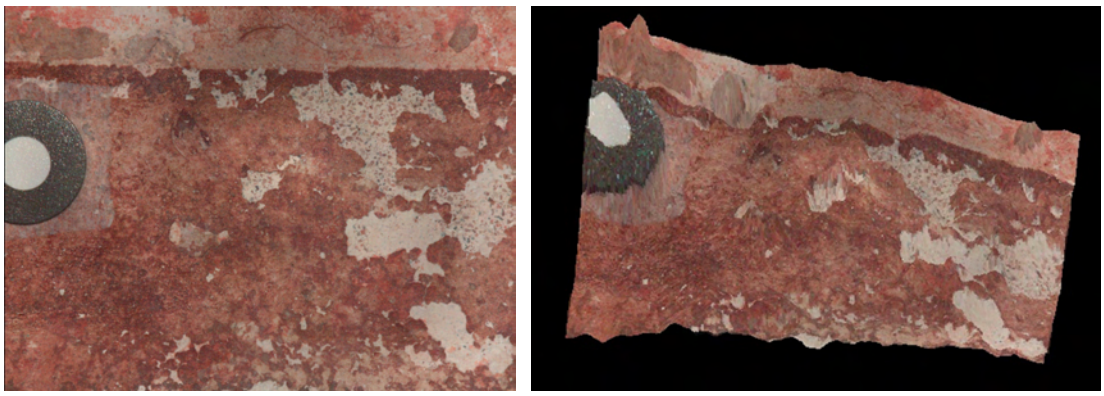
Figure 1.4: Jonas and the whale in the Rhine, in front of the city of Mainz. Wall painting partially cleansed in 2009 (right side).

Figure 1.5: Third wall painting under study.



(a) (b)

Figure 1.6: Acquisitions of the surface of the sarcophagus. (a) Color rendering of a multispectral acquisition (b) Multispectral acquisition mapped to a 3D model.



(a) (b)

Figure 1.7: Acquisitions of the surface of the first wall painting. (a) Color rendering of a multispectral acquisition (b) Multispectral acquisition mapped to a 3D model.

responding points between a Digital Elevation Model (DEM) and a multispectral acquisition. These points were used to calculate the affine transformation to apply to the multispectral model to map it on the 3D data.

Manually selecting control points is not only a long and tedious task, it also proved to be difficult: in the case of the wall paintings there are few common salient features between the multispectral and 3D data. A pair of resulting mappings are shown figures 1.6 and 1.7. Each of these mappings was created by manually selecting five corresponding points. However, the residuals of these control points are not satisfactory: they are between 1.4 pixels and 7 pixels in the first case and reach a maximum of 2 pixels in the second case. This preliminary work has been published in [Simon *et al.*, 2010].

Challenges

These prefatory acquisitions and processing pinpoint the following problems and difficulties:

- We can only use markers in rare and specific cases.

- These markers were sufficient in number and size for the registration of multiple 3D views
- These markers were not sufficient in number to register the 3D and multispectral data.
- These markers tarnish readability.
- For monitoring purposes, markers have to stay in place over a long time period, but they fall off as the surface degrades.
- Hand picking salient points is time-consuming and not robust.
- Sometimes there are few/ no good salient points between different data representations.

A review of the literature taught us that our problems are common when treating with cultural heritage and that no straightforward solution exists for the multimodal registration of featureless datasets. We use contactless acquisition systems to avoid damaging the object, as such we also need a contactless registration technique that works for smooth and featureless multimodal datasets.

1.4 Contributions

We develop our own registration method based on the photogrammetric tracking of the acquisition systems. The strengths and contributions of this method are:

- A method suitable for registering data with no salient features.
- A registration precision independent from the content of the acquired data.
- A registration method which works for different optical sensors.
- A flexible solution suitable for different applications.

So far this method has only been evaluated on test objects in laboratory settings. It was essential to solve the practical and logistical difficulties in controlled settings before using the method in situ. It was also important for us to assess the potential of the method for general registration purposes, and not only in the two specific configurations presented here.

Some of these tests were performed on objects with much greater spectral and structural variability than the objects we developed it for. This enabled a quick visual assessment of the accuracy of the final registration. Other tests were performed on featureless objects which we covered with targets, using these targets only for the evaluation of the tracking accuracy. Finally the method was applied to an object which presents similar spatial and spectral variations as the sarcophagus. These important steps permitted to assess the strengths and the accuracy of our method before applying it to the study of cultural heritage.

1.5 Overview

A state of the art of 3D and multispectral acquisition and registration techniques adapted for the study of cultural heritage is given chapter 2, a version of which has been published as [Simon Chane *et al.*, 2013a]. This review highlights the need of a registration method for multimodal featureless datasets for the study of cultural heritage. This chapter also shows that the optical tracking of the acquisition systems can be a suitable technique on which to base the registration. The achievable accuracy of tracking the acquisition systems using photogrammetry was evaluated through a series of simulations, described in chapter 3 and partially published in [Simon *et al.*, 2012].

This technique was first tested for the registration of featureless 3D datasets in chapter 4. In this chapter we also evaluate the overall tracking accuracy and the stability of the systems in play. Chapter 5 presents the registration of 3D and multispectral data using our method: some of these results were published in [Simon Chane *et al.*, 2013b]. Finally, chapter 6 shows a possible application of these registered 3D/ multispectral datasets. A conclusion presents future work and perspectives in chapter 7.

Chapter 2

State of the Art

Cultural heritage is a favorite application domain of the image processing and computer graphics community. We enjoy developing tools to study and analyze these unique and precious surfaces. We strive to create accurate virtual representations of cultural heritage objects, even though this might not be the goal of conservators. When studying highly two dimensional objects such as paintings, multispectral imaging is the path of predilection. For statues, we favor 3D digitizations.

Of course it is not uncommon to create 3D models of paintings, and multispectral imaging is sometimes used on non-planar objects. But the growing interest is in combining these complementary datasets to create 3D models with multispectral texture. Such models can be used for virtual reality applications such as virtual museums. Most importantly, 3D models with multispectral texture are an improvement on 3D models for the analysis and study of works of art. Conservators can gain much insight from these augmented models.

However, multispectral cameras and 3D digitization systems are built on different concepts. There are difficulties associated with both developing integrated multispectral 3D digitization systems and registering independent datasets via post-processing. Creating multispectral 3D models is not a straightforward task.

This review details acquisition systems and algorithms that can be used to obtain multispectral 3D models of cultural heritage objects. The applications of each technique are meant to highlight the potential of multispectral 3D models for the analysis of works of art. My scope is limited to the study of cultural heritage *objects* as opposed to full buildings, archaeological sites or cities. The term "object" is used to describe anything from the smallest archaeological fragment to 5 m high statues and wall paintings that cover full rooms. Systems and algorithms optimized for the digitization of full cities or archaeological sites are only mentioned to the extent that they can be adapted to smaller volumes.

The accent is put on non-contact acquisition systems. Contact acquisition systems are hardly ever used for the study of cultural heritage, as they risk damaging the object under study. As such, they are only mentioned in passing. Portable systems that can be used directly in museums, monuments or any other non-laboratory site are put forward. We also highlight flexible solutions that can be adapted to the variability of the objects under study. In heritage

settings, objects of similar size, shape and material have an individual history that can influence the acquisition process and systems with variable field-of-view and resolution are most appreciated.

This chapter is organized as follows: We first introduce multispectral acquisitions and multispectral data processing, with an accent on systems developed for the analysis of cultural heritage objects and the possible uses of such data. Then 3D digitization techniques are explored. Great 3D digitization campaigns and the various uses of 3D models for conservation purposes are described. We then present the most common methods used to obtain 3D data with multispectral or color texture, using either integrated devices or registration algorithms. Finally, we show the potential of photogrammetry-based optical tracking for multimodal registration.

2.1 Multispectral imaging

2.1.1 Limitations of color imaging

The perceived color of an object is a function of the spectral reflectance of its surface, the spectral distribution of the illumination and the spectral sensitivities of the cones in the eye. A recorded color image is thus both device and illuminant dependent. When a surface is acquired under a given illuminant, it is impossible to estimate the surface color accurately under another illuminant in the absence of additional information on the spectral reflectance. In conventional color imaging, each pixel is characterized by three components such as red, green and blue. These three components are necessary and sufficient to synthesize any color from a colorimetric point of view, but imaging systems based on three primary colors present a number of limitations:

- Conventional RGB chips violate the Luther rule which states that a sensor can distinguish the same colors as the human eye only if the spectral responses of this sensor can be obtained by a linear combination of the S, M and L eye cone responses.
- Due to metamerism, two surfaces with strongly different spectra can appear identical under a given illuminant and completely different under another.
- The use of color information in image databases is limited by the fact that recorded RGB signals depend on the acquisition devices, the illuminants and image preprocessing.

These drawbacks result from the weak spectral resolution of conventional color imaging systems. A high spectral resolution allows a better estimation of the surface reflectance. This in turn permits a better characterization of its intrinsic physical properties regardless of the acquisition conditions (illuminant, acquisition device). It is possible to acquire color-accurate images with conventional trichromatic sensors by working in adapted color spaces such as CIELAB [Berns, 2001]. However, systems that acquire more than three spectral components permit a better estimation of the reflectance. Such systems are called "multispectral".

It is common to acquire 31 bands between 400 nm and 700 nm (that is, one channel every 10 nm) but Berns *et al.* [2005] state comparable spectral precision can be obtained using less

than 10 bands. There is thus a great variability in the number of bands acquired by multispectral acquisition systems. When more than 100 bands are acquired the term "hyperspectral" is used while "ultraspectral" describes devices with thousands of bands. During the last two decades, multi- and hyperspectral imaging systems have shown great potential in domains as varied as remote sensing, agronomy, industrial quality control, metrology, medical imaging and artwork digitization.

2.1.2 Multispectral acquisition systems

As opposed to spectrophotometers which acquire spectra with nanometric precision in a single spot, multispectral imaging systems acquire data along two spatial dimensions and one spectral dimension. This is usually referred to as a multispectral cube. This data is generally acquired along two dimensions at a time and scanned across the third dimension. There are thus two acquisition paradigms: In the first case successive images of a given area are acquired at varying wavelengths. In the second case spatio-spectral images are acquired and there must be some motion of the acquisition sensor to acquire data along the other spatial dimension [Cotte and Dupouy, 2003].

Multispectral scanners and imagers are generally categorized based on the filtering technology used:

Optical filters These filters (most often interference filters) are generally mounted on a rotating wheel or another mechanical device. The filters are sequentially positioned in front of a panchromatic sensor to acquire a single image in the range of the filter [Brauers *et al.*, 2008; Mansouri *et al.*, 2005c; Pelagotti *et al.*, 2008; Yamaguchi *et al.*, 2008]. We deal with the case in which the filters are placed between the light source and the object in the paragraph titled "tunable light sources".

Various algorithms can be used to compensate geometric distortion [Brauers *et al.*, 2008; Mansouri *et al.*, 2005c], longitudinal aberrations [Brauers and Aach, 2008a] and ghosting [Brauers and Aach, 2008b] introduced by the filter-wheel. The use of optical filters has the disadvantage of creating a fix setup, with cumbersome filter-wheels. The size of the filter wheel limits the number of spectral bands. The long acquisition times can be reduced by using a synchronized flash [Brauers *et al.*, 2009], though such strong lights are not adapted to study of cultural heritage.

Electronically tunable filters LCTFs (Liquid Crystal Tunable Filters), AOTFs (Acousto-Optic Tunable Filters) and adapted Fabry-Perot devices allow to electronically select the filtering band. This greatly speeds up the acquisition process compared to filter-wheel systems [Berns, 2005a; Hardeberg *et al.*, 2002; Hardeberg, 1999; Miyazawa *et al.*, 2001; Novati *et al.*, 2005]. Devices based on such filters are also very flexible since the number and width of bands can be changed programmatically.

Poger and Angelopoulou [2001] have compared the characteristics of these three types of electronically tunable filters. Most importantly, AOTFs have a high transmission rate (98 %) but they require collimated light while LCTFs and Fabry-Perot devices suffer from a low transmission rate (lower than 50 %).

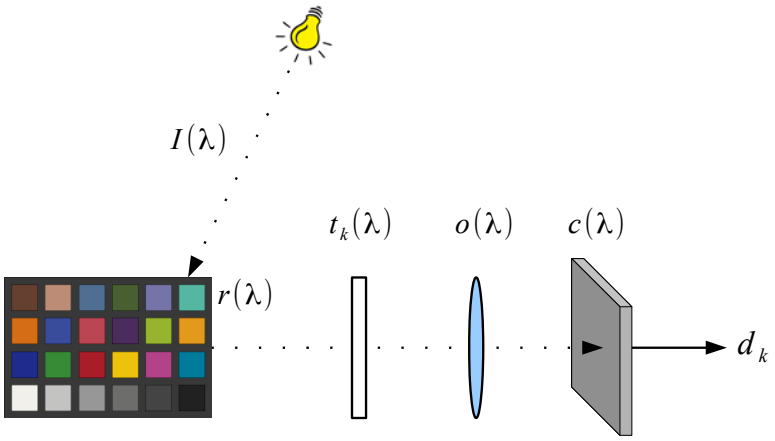


Figure 2.1: Spectral model of emission filtering multispectral acquisition systems. $I(\lambda)$ represents the spectral radiance of the illuminant, $r(\lambda)$ is the spectral reflectance of the surface, $t_k(\lambda)$ denotes the spectral transmittance related to the k^{th} filter, $o(\lambda)$ is the spectral transmittance of the optical system, $c(\lambda)$ is the spectral sensitivity of the camera.

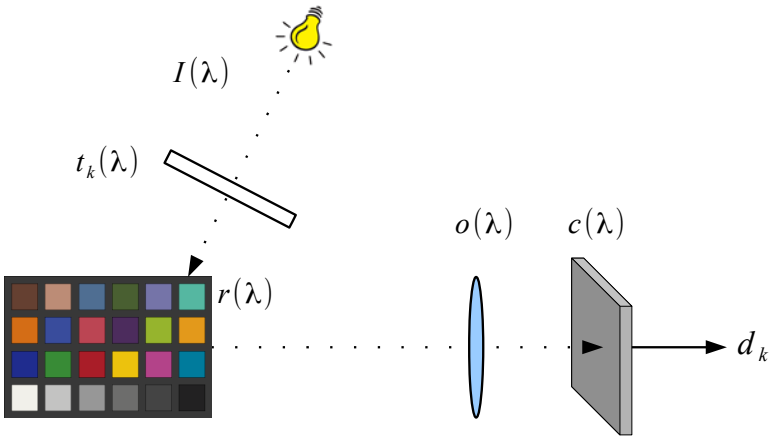


Figure 2.2: Spectral model of excitation filtering multispectral acquisition systems. Notations are the same as those used in figure 2.1.

Extension of Color Filter Arrays (CFAs) Instead of positioning filters in front of the sensor, extended mosaicking of CFAs is used to obtain more than three color bands [Berns *et al.*, 2005; Brauers and Aach, 2006; Miao *et al.*, 2004]. The practical use of this type of multispectral system is still being investigated.

Tunable light sources Instead of filtering the light reflected from the object (emission filtering), it is possible to successively project light at different wavelengths on the object (excitation filtering). Emission and excitation filtering are illustrated figures 2.1 and 2.2, respectively.

A common strategy in excitation filtering is to use multiplex multicolor LED illumination as a tunable light source [Bouchard *et al.*, 2009; Everdell *et al.*, 2009; Park *et al.*, 2007]. This solution is gaining popularity with the increasing availability of LEDs in an increasing variety of colors. Other systems based on excitation filtering rely on optical filters [Jolivot

et al., 2011; Martinez *et al.*, 2002; Tominaga and Tanaka, 2008] or digital light processing [Zuzak *et al.*, 2009] to create a tunable light source.

To avoid contaminating the acquisition with stray light, multispectral acquisition systems based on excitation filtering must be used either in contact with the object, or in a dark environment. Because of these constraints, few multispectral imagers for cultural heritage are built on excitation filtering.

Additionally, spatially varying filters [Schechner and Nayar, 2002] or prisms [Du *et al.*, 2009] can be used in scanning devices to acquire wide-field of view multispectral images or multispectral videos.

Among the recent advances in multispectral imaging, Kawakami *et al.* [2011] combined a high spatial resolution RGB sensor with a low resolution MS camera to obtain high spatial resolution spectral images. The results are promising though the method produces some errors and is thus not mature enough to be used for the analysis of cultural heritage.

2.1.3 Multispectral imaging for cultural heritage

Many multispectral acquisition systems have been developed for the study of cultural heritage. We present a selected few, starting with the most well known and large projects and finishing with lab-scale projects and the applications of multispectral imaging for cultural heritage. Table 2.1 summarizes the technical characteristics of existing acquisition systems developed for the study of cultural heritage.

In the early 1990's the VASARI project [Martinez *et al.*, 2002] focused on the development of a 12 band multispectral acquisition system based on a filter-wheel tunable light source. Using a 3000×2000 pixel CCD and a gantry, the system captured $20\,000 \times 20\,000$ pixel acquisitions. At the time, such high spectral resolution came at a high cost : it took about 3 hours to digitize a $1\,\text{m} \times 1\,\text{m}$ painting. However, this was a remarkable project that opened the path to multispectral imaging of paintings for conservation, material identification and multispectral printing.

The European project CRISATEL [Ribés *et al.*, 2005] also aimed at developing a multispectral acquisition system for the analysis of paintings. The device is based on a linear CCD (12 000 pixel) which is mechanically displaced to acquire 20 000 vertical lines. 13 interference filters are used (10 in the visible and 3 in the near infrared) to cover the 380 nm to 1000 nm range. This multispectral camera is used in conjunction with a dedicated synchronized lighting system. With a comparable spatial resolution to the VASARI camera, this device digitized the surface of the Mona Lisa ($78\,\text{cm} \times 82\,\text{cm}$) in an hour and a half [Cotte and Dupraz, 2006b]. This system is also less bulky than the one developed by the VASARI team.

Both the VASARI and the CRISATEL systems were developed within large projects, but it is possible to develop small-scale multispectral acquisition systems with reasonable means [Berns *et al.*, 2008; Novati *et al.*, 2005; Pelagotti *et al.*, 2008; Tominaga and Tanaka, 2008]. These four systems are image-based and they acquire much smaller areas than the previous two systems. However, their acquisition time is less than a few minutes and tiling can be used to acquire larger areas. The system developed by Pelagotti *et al.* [2008] is particularly impressive as it

Table 2.1: Multispectral systems developed for cultural heritage study and analysis. Given in chronological order.

References	Filtering type	Spectral bands	Spectral range (nm)	Image size (pixel)	Dynamic range (bits)
VASARI project [Martinez *et al.*, 2002]	Tunable light	12	$450 - 1000$	20000×20000	8
CRISATEL project [Ribés *et al.*, 2005]	Interferential	13	$400 - 1000$	12000×30000	12
Novati *et al.* [2005]	LCTF	31	$400 - 700$	1392×1040	12
Berns *et al.* [2008]	CFA	6	$380 - 750$	22 000 000	16
Pelagotti *et al.* [2008]	Interferential	15	$380 - 1130$	3072×2048	14
Tominaga and Tanaka [2008]	Tunable light	6	$400 - 700$	1636×1236	10

acquires visible reflectance, infrared reflectography and UV-fluorescence, a combination of data most adapted for the study of paintings.

Applications

Multispectral data alone lacks the spectral resolution necessary to distinguish artist pigments [Martinez *et al.*, 2002] but it can be used as a complement to spot techniques for material identification [Delaney *et al.*, 2005; Pelagotti *et al.*, 2008] or to distinguish metamers [Colantoni *et al.*, 2006; Delaney *et al.*, 2005], an important step before restoration campaigns. Multispectral imaging has been used to increase the readability of palimpsests using pseudocoloring [Bloechl *et al.*, 2010; Easton Jr. *Et al.*, 2003; Rapantzikos and Balas, 2005]. In the same vein, virtually restoring the original colors of paintings [Berns, 2005b] can increase their readability. Multispectral images can be used for the analysis of cracks and the study of varnish aging [Cotte and Dupraz, 2006a,b]. This is both important to understand the history of the painting and to guide the virtual restoration. The use of multispectral data for virtual archives is also a common application [Martinez *et al.*, 2002; Tominaga and Tanaka, 2008].

The possibility to view the digitized surface under arbitrary illuminants [Ribés *et al.*, 2005; Tominaga and Tanaka, 2008] is a real strength of multispectral data. This can be used to view digitized versions of the artworks under broad daylight when such conditions could damage fragile pigments such as those of watercolors. It is also possible to simulate the illumination environment before exhibiting a work of art. Multispectral acquisitions are also the first necessary step for multispectral printing using 6 or 7 color inks [Berns *et al.*, 2008; Imai *et al.*, 2001].

2.1.4 Reflectance estimation from multispectral data

Multispectral imaging aims at acquiring the reflectance of the surface of the scene rather than its color. Increasing the number of acquisition channels eases this process but reflectance estimation from the acquired images is not trivial and requires appropriate models. The most used spectral model of the acquisition chain in color and multispectral systems is illustrated figure 2.1.

Using the notations introduced in this figure, and with η_k representing the spectral noise for the k^{th} channel with $k = 1 \dots K$, the camera output d_k, related to the channel k for a single pixel of the image is given by:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} I(\lambda)\, r(\lambda)\, o(\lambda)\, c(\lambda)\, t_k(\lambda)\, d\lambda + \eta_k. \tag{2.1}$$

If we assume that the noise is removed by preprocessing [Mansouri *et al.*, 2005a,c], and assume a linear optoelectronic transfer function, we can replace $I(\lambda)$, $c(\lambda)$, $o(\lambda)$ and $t_k(\lambda)$ by the spectral sensitivity $S_k(\lambda)$ of the k^{th} channel. Equation 2.1 thus becomes:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} S_k(\lambda)\, r(\lambda)\, d\lambda. \tag{2.2}$$

By regular sampling of the spectral range at N wavelengths, equation 2.2 can be written in matrix notation as follows:

$$d_k = \mathbf{S}_k^T(\lambda)\,\mathbf{r}(\lambda), \tag{2.3}$$

where $\mathbf{S}_k^T(\lambda) = [s_k(\lambda_1)\, s_k(\lambda_2) \ldots s_k(\lambda_N)]$ is the vector which describes the spectral sensitivity of the k^{th} channel of the acquisition system, $\mathbf{r}(\lambda) = [r(\lambda_1)\ r(\lambda_2) \ldots r(\lambda_N)]^T$ is the vector of the sampled spectral reflectances of the scene and ^{T} is the transpose vector operator. If we consider a system with all channels, equation 2.3 can be written as:

$$\mathbf{d} = \mathbf{S}^T\mathbf{r}, \tag{2.4}$$

where $\mathbf{d}$ is the vector containing all d_k camera outputs and $\mathbf{S} = [s_1\ s_2 \ldots s_K]$ is the matrix containing the channels' spectral sensitivities $\mathbf{S}_k$. The final goal is to recover $\mathbf{r}$ from the camera output given by equation 2.4. This is obtained by finding an operator $\mathbf{Q}$ that satisfies $\tilde{\mathbf{r}} = \mathbf{Q}\,\mathbf{d}$. Many methods permit this estimation [Cheung *et al.*, 2005; Day, 2003; Haneishi *et al.*, 2000; Imai *et al.*, 2000; Imai *et al.*, 2003; Shi and Healey, 2002; Zhao and Berns, 2007] , most of which can be categorized depending on how matrix $\mathbf{S}$ is determined.

- If $\mathbf{S}$ is obtained by a direct physical system characterization [Imai and Berns, 1999; Shen *et al.*, 2007; Tominaga, 1996; Zhang and Xu, 2008], the operator $\mathbf{Q}$ is the inverse of $\mathbf{S}$. However, since $\mathbf{S}$ is usually not a square matrix, its inverse does not exist. Only a pseudo-inverse or another regularized inverse such as Wiener can be calculated. Thus $\mathbf{Q} = \mathbf{S}^+$ where $^+$ is the pseudo-inverse operator. The main drawback of this technique is that it is highly noise sensitive and the matrix inversion further amplifies the noise.

- $\mathbf{S}$ can also be obtained indirectly by matching a set of M patches of known theoretical reflectance with a multispectral acquisition of these patches captured by the camera [Imai and Berns, 2002; Mansouri *et al.*, 2005b]. We then have a set of corresponding pairs (d_m, r_m), for $m \in [\![1, M]\!]$, where d_m is a vector of dimension K containing the camera responses, and r_m is a vector of dimension N representing the spectral reflectance of the m^{th} patch. The r_m reflectances are gathered in matrix $\mathbf{R}$ and the camera outputs for the M patches are gathered in matrix $\mathbf{D}$. The operator $\mathbf{Q}$ is obtained directly by calculating this match. Any optimization method can fulfill this aim (neural networks, least squares regression, etc.). $\mathbf{Q}$ is obtained from $\mathbf{R} = \mathbf{Q}\,\mathbf{D}$ through the pseudo-inverse of $\mathbf{D}$: $\mathbf{Q} = \mathbf{R}\,\mathbf{D}^+$.

- The third paradigm for spectral reflectance estimation consists in directly interpolating the camera outputs d_k [Bianco *et al.*, 2008; Connah and Hardeberg, 2005]. No knowledge about the matrix $\mathbf{S}$ is required. However, rigorous conditions about the shape and number of filters make this technique ineffective for reflectance estimation in the general case.

As part of the reflectance estimation based on a diffuse light model, it is also necessary to separate specular components of the images [Bajcsy *et al.*, 1996; Lin and Shum, 2001; Tan and Ikeuchi, 2005; Wolff, 1990]. The first and third paradigms also require the separation of the illumination and surface reflectance [Chang and Hsieh, 1995; Drew and Finlayson, 2007; Ho

et al., 1990; Maloney and Wandell, 1986; Ohta and Hayashi, 1994; Tominaga and Wandell, 1989].

2.1.5 Perspectives

Multispectral acquisition systems with high spatial and spectral resolutions are currently restricted to large laboratories and projects with considerable means. There are still many aspects of multispectral imaging that require deeper investigation for its widespread use in conservation. It is possible to develop multispectral cameras with high spectral resolution and somewhat limited spatial resolution with a reasonable budget. The spatial resolution of such systems is increasing as bigger sensors become widely available. Though somewhat portable, these systems remain cumbersome and require a computer to pilot them. It is thus necessary to develop fully integrated, portable multispectral acquisition systems with high spatial and spectral resolution.

Also, the spectral acquisition range of existing systems is often limited to the visible. Though it is difficult to find sensors with a high responsivity from the UV to the IR, multispectral cameras which can acquire data over larger ranges are needed for the analysis of cultural heritage. UV fluorescence and IR reflectography are common imaging tools for conservators. The former is mainly used to study varnishes while the later permits the study of charcoal and pencil underdrawings. Improving these techniques' spectral resolution results in useful and easily grasped data for the conservation community.

Characterization and calibration algorithms currently used for light/reflectance separation and spectral reconstruction are based on simplified light/matter interaction models. Acquisition systems and algorithms adapted for the digitization of non-planar, glossy, transparent, surfaces must still be developed. This is of utmost importance since such surfaces are often found on cultural heritage objects: glossy varnishes on paintings, metal and glass artifacts, bas-reliefs, etc.

For the widespread adoption of multispectral imaging by conservators it is also necessary to improve existing visualization and analysis tools [Jordan and Angelopoulou, 2010; Kim *et al.*, 2010]. The amount of data acquired by multispectral imaging systems can be overwhelming for the non expert and it is difficult both for computers and humans to interpret such high-dimensional data. It is also important to develop intuitive interfaces adapted to users with no computer science background. Image processing algorithms generally used for the analysis of cultural heritage [Barni *et al.*, 2005] must be extended to multispectral imaging. This includes Specific algorithms that exploit the wealth of data acquired by multispectral systems must also be developed. Such tools have to be created hand in hand with the conservation community to provide them with methods they need.

2.2 3D digitization

2.2.1 3D digitization as a conservation tool

The reproduction and documentation of works of art is traditionally based on photography, even for three dimensional objects such as statues. Color photos are part of condition assessment

reports which are elaborated by museums prior to lending an object of their collection to another museum. They illustrate the current condition of the object and are relied on if it is necessary to prove that the object was damaged during the loan. The use of multispectral acquisitions can make this process more rigorous by recording the spectral reflectance though, like photos, multispectral acquisitions only record two spatial dimensions.

There are obviously serious limitations to using two spatial dimensions to describe a three dimensional object. A collection of pictures (or a video) can only represent the object from predefined viewpoints and not as a comprehensive whole. No matter how methodical and professional, there is always the risk that a photographic campaign does not record with sufficient detail an area which will undergo spatial damage that may need a quantitative evaluation. In comparison, a digital 3D model can record the surface condition of an entire object at a given moment.

This section first gives a concise overview of the existing 3D digitization systems. We then present various large-scale 3D digitization projects for the study and conservation of cultural heritage. Other uses of 3D models for the study of cultural heritage are given at the end of this section.

2.2.2 Principles and devices

3D digitization techniques are generally categorized in active and passive techniques. Active digitization systems integrate a light emitting system which is used to highlight surface structures and fine details, whereas passive techniques depend on ambient light.

Passive 3D digitization techniques

Historically, photogrammetry is the passive digitization technique most used for the study of cultural heritage. Photogrammetry is the science of measuring the position and shape of objects using photography. It is often used in surveying by processing satellite images or aerial photographs. When the camera is relatively near the subject (a few meters), as in the case when studying cultural heritage, the technique is called close range photogrammetry and relies on different geometric assumptions. In the rest of this document the word "photogrammetry" denotes close range photogrammetry.

The advent of digital cameras has prompted the widespread use of photogrammetry in the cultural heritage community and elsewhere. This flexible and well documented technique [Luhmann, 2010b; Remondino *et al.*, 2008] can be performed with consumer cameras. Though practice makes for more reliable measurements, the technique is easy to apprehend by non experts since it consists in simply taking pictures of the scene. Not only is photogrammetry a contact-less digitization technique, it is also performed at very safe distance from the object and large volumes can be digitized quickly [Yilmaz *et al.*, 2008].

In stable and well calibrated setups, off-line photogrammetry systems, reach a potential measurement precision of up to 1:500,000 with respect to the largest object dimension [Luhmann, 2010a]. To achieve such precision, the intrinsic camera parameters must be determined with great accuracy. Such intrinsic parameters are also called the camera interior orientation

(I.O.) and include sensor resolution, focal length, lens distortion, principle point offset, pixel ratio, pixel skew, etc. The exterior orientation (E.O.) of a sensor is its position and orientation in a fixed system. These parameters, as well as the object coordinates, are estimated by recognizing corresponding points in the images.

Lately, several algorithms have emerged which automatically create 3D models from uncalibrated images [Agarwal *et al.*, 2009], for example using Multi-View Stereopsis [Furukawa and Ponce, 2010]. This type of techniques is well adapted to the documentation of archaeological sites [Pollefeys *et al.*, 2003]. To encourage the use of such techniques by conservators, Vergauwen and Van Gool [2006] have developed a web-based 3D reconstruction service for cultural heritage. The user must simply upload a set of photos and is notified when the 3D reconstruction is done. He can then download the original images, the camera calibrations, quality parameters and depth maps. The authors also provide an interactive viewer to analyze the data.

Other passive digitization techniques include shape from shading [Zhang *et al.*, 1999], shape from silhouette [Forbes *et al.*, 2006], as well as shape from edges, -texture, -focus. Though generally easy to setup, these techniques suffer from low accuracy [Pavlidis *et al.*, 2007].

Active 3D digitization techniques

The most used active digitization techniques are time of flight (TOF) and triangulation laser scanners as well as systems based on the projection of structured light. Time of flight laser scanners digitize a scene point by point based on the time a laser beam takes to hit an obstacle and reflect back. This technique is suitable to digitize large object — even full archaeological sites [Guidi *et al.*, 2008] — with a resolution of a few millimeters.

Triangulation laser scanners project a line of light on an object and derive the shape of the object from the line deformation. These systems are adequate to scan small [Seulin *et al.*, 2006] and medium-size objects with an accuracy of a few micrometers.

The projection of structured light is another triangulation-based 3D digitization technique and achieves comparable accuracy as triangulation laser scanners. A projector illuminates the surface with predetermined patterns. These images are captured with one or more cameras and the deformation of the pattern is used to evaluate the shape of the object. This technique is the most used for the digitization of cultural heritage objects. It is relatively easy to develop a lab prototype [Rocchini *et al.*, 2001a], and many commercial systems exist. The most widely used for conservation purposes are those developed by GOM, Breuckmann, and Steinbichler. Fraunhofer has developed a compact, cordless, handheld device [*Fraunhofer, Kolibri*].

State-of-the art active digitization techniques include shape from polarization [Ferraton *et al.*, 2009] using a multispectral light source or shape from heating [Bajard *et al.*, 2011; Eren *et al.*, 2009].

2.2.3 From 3D acquisitions to a 3D model

3D digitization is but a single step in creating 3D models, the full 3D processing pipeline is described in [Bernardini and Rushmeier, 2002]. Since my focus is on data registration, I will examine these techniques in greater detail. Even if we only acquire 3D data, a registration phase

is often necessary to create a complete 3D model. When digitizing but the smallest objects, multiple acquisitions must be performed and stitched together. The registration process highly influences the accuracy of the end model and is thus a critical step in digitizing cultural heritage artifacts.

Algorithms

In the case of highly spatially-structured objects such as statues, the registration may profit from the spatial characteristic of the object itself and use local shape information expressed in grouped points, which are then transformed by algorithms like the Iterative Closest Point algorithm (ICP) [Besl and McKay, 1992]. Given an initial positioning of two 3D point clouds, the algorithm converges by minimizing the root mean square derivation of the overlapping areas. However, the initial positioning must be quite precise otherwise the algorithm converges towards a local minimum. Also, this algorithm requires a $30-40\%$ overlap between contiguous views to ensure a reliable registration. This algorithm is generally not adapted to register views of flat objects such as painting or wall paintings, though some variants tackle this issue [Rusinkiewicz and Levoy, 2001].

There is great focus on automating the registration process to boost the use of 3D models by the conservation community [Andreetto *et al.*, 2004; Guarnieri *et al.*, 2003]. Andreetto *et al.* [2004] describe a fully automatic registration pipeline for 3D models of cultural heritage objects which takes into account both shape and texture information and is thus particularly useful for symmetrical objects with a distinct texture, such as painted vases.

Mechanical tracking

When scanning small objects, the registration process can be easily sped up by placing the object on a turntable and relying on its positioning information. In this case another technique must be used to register top and bottom views of the object. Often targets are placed on the turntable to be used as corresponding points between views.

Many commercial handheld scanners rely on the deformation of a six degrees of freedom (6DoF) arm to evaluate the relative position of each stripe acquisition [*Faro, Edge*]. In this case, the surface that can be digitized is limited by the range of the arm.

Optical tracking

A new generation of handheld laser scanners relies on optical tracking of the scanning probe to register the flow of acquisitions. In this setup active or passive targets are fixed to the acquisition system. A fixed laser tracker [*Leica, T-Scan*] or a fixed photogrammetric setup [*Creaform, MetraScan*; *NDI, ScanTrak*; *Steinbichler, T-Scan*] determines the position and orientation of the sensor at each moment.

A variation on the use of optical tracking for range registration is presented in [Kawasaki and Furukawa, 2004]. Active targets are fixed to a laser projector. The object under study is placed on a turnable and a fixed video camera observes the full scene (projector, turntable and

object). The video stream provides both the relative position of the object and projector and the shape of the object given the deformation of the projected stripe.

This type of optical tracking can also be performed by fixing targets around the scene and attaching the tracking sensors to the acquisition system. This is the configuration used in [Valkenburg *et al.*, 2006]: Six cameras are fixed to a handheld TOF laser scanner and observe the surroundings while the scanning occurs. Since active targets have been placed in the scene, it is possible to compute the position and orientation of the scanner relative to the targets. This data is in turn used to register the range data.

An example of optical tracking for 3D registration in the context of cultural heritage study is given by Blais *et al.* [2005]. They scanned a painting with both a high resolution color laser scanner and a lower-resolution laser scanner. The lower-resolution scanner acquired the full painting in a single scan and was also used to project optical markers on the surface of the painting, defining the sub-areas to scan with the high-resolution scanner. White spheres were mounted on the high resolution scanner and the third tasks of the low resolution scanner was to track the position and orientation of this second scanner while in use.

2.2.4 3D digitization of cultural heritage

The following are a few remarkable projects where the previously described 3D digitization techniques were applied to study cultural heritage. These projects illustrate the variety of uses of 3D models for conservation purposes.

In 1998–99, Levoy *et al.* [2000] digitized ten Michelangelo statues and over a thousand fragments of a marble map as part of "The Digital Michelangelo Project". The David statue alone was made of about 2 billion polygons. The extent of this dataset pushed the authors to develop an algorithm adapted to the global registration of large 3D datasets [Pulli, 1999]. The model of the statue of David was used during its restoration to index and visualize data, to evaluate the exposure of the statue to the fall of contaminants and to perform measures (distances, surface, volume) that can not be performed on a statue [Callieri *et al.*, 2004]. A similar project of smaller scope consisted in monitoring the restoration of a statue of Minerva by digitizing it before and after restoration [Rocchini *et al.*, 2001b].

The focus of the Great Buddha Project [Ikeuchi *et al.*, 2007] was the digital preservation and restoration of large outdoor objects, namely the Nara Buddha statues and the temple they sit in. This project demonstrated one of the great advantages of 3D models: they can be manipulated in a way real objects can not, altering their proportions, removing additions to increase readability [Bernardini *et al.*, 2002], restoring the object's original color, reconstructing missing information [Li *et al.*, 2010]. In the same vein, it is also possible to interact with archaeological fragments to reconstruct broken objects, either automatically or based on algorithms [Koller *et al.*, 2006; Zheng *et al.*, 1998] while digitized wooden stamps [Seulin *et al.*, 2006] can be used for virtual printing.

3D models also provide very rich communication tools. For example, visitors can digitally approach a 5 m high statue with dedicated interactive kiosks in a museum [Levoy, 2002], or immerse themselves in larger settings through virtual displays [Li *et al.*, 2010].

2.2.5 Perspectives

Passive techniques are generally easy to set up. Though they do not provide very accurate results, they are well adapted to visualization purposes.

For applications which include surface analysis and metrology, active sensors are better suited. These are mature techniques that have been widely used to digitize heritage and many commercial systems exist. However, they do not perform well in daylight, as the resulting contrast is insufficient.

Triangulation laser scanners have a large acquisition range and provide 3D data which is automatically registered. They are thus easy to use. Digitization techniques based on the projection of structured light have the advantage of being somewhat flexible: by changing the sensors and performing a calibration, the field of view and resolution can be changed. The same system can thus be used to acquire both small artifacts (of the order of ten centimeters) and larger ones (half a meter or more). The need to register multiple views to acquire most objects is a drawback (using a turntable is only possible for the smallest ones). Not as intuitive as triangulation laser scanners, some training and technical background is necessary to use them.

Active sensors are not suitable for the digitization of glass and metal surfaces, though these materials are common in cultural heritage. In industrial settings, this limit is overcome by covering the surface with a fine powder with cooperative reflective properties but this is not possible when dealing with unique and fragile objects. The use of heating to scan metallic [Bajard *et al.*, 2011] and glass surfaces [Eren *et al.*, 2009] has been proposed, but this solution is also unadapted for cultural heritage applications.

There is ongoing research to develop digitization systems based on both active and passive sensors. The integration of several sensors requires a calibration procedure, but the complementary strengths of each technique then provide more accurate 3D models. For example one or more cameras can be used to improve the depth maps provided by a TOF sensor [Chan *et al.*, 2010; Zhu *et al.*, 2008].

Hybrid systems based on the integration of normal data to enhance the geometry have been developed, but such systems are cumbersome since they include fixed light sources [Nehab *et al.*, 2005; Rushmeier and Bernardini, 1999]. Recently, comparable results have been achieved using a simpler system based on a projector, a low resolution camera and a high resolution camera [Lu *et al.*, 2010].

Generally speaking, the digitization of cultural heritage often involves large objects that must be scanned with high precision [Guidi *et al.*, 2006; Levoy *et al.*, 2000]. To be used in situ, the acquisition systems must be portable. Ease of use and automatic post-processing are important if this technique is to gain acceptance by conservators and be of widespread use for archiving applications, and not merely reserved to a few objects of great fame.

2.3 Integrated 2D – 3D measurements

The usefulness of 3D digitization and multispectral acquisitions of cultural heritage objects has been well proven by past projects. Since these two methods of studying and archiving object properties provide complementary data, it would be most useful for art historians and

conservators to have both datasets present in an integrated frame. Such need is highlighted in [Lahanier *et al.*, 2005] where the authors successively perform multispectral acquisitions of "La dame en prière" and two 3D digitizations of its surface but they rely on independent viewers to explore each dataset.

2.3.1 Integrated acquisition systems

Multispectral imaging is seldom used in combination with 3D digitization, though there is an abundant literature on 2D to 3D registration. Since multispectral imaging is an extension of color imaging these registration methods can often be adapted to registration of multispectral texture on 3D models. This section first details the integration of color and 3D models, using algorithms or unique sensors before surveying the existing integrated multispectral – 3D acquisition systems.

Integrated color – 3D acquisition devices with a single sensor

The acquisition of RGB 3D models has been a topic of research for longer than the past decade. When using a single sensor, shape and texture are automatically registered. The National Research Council of Canada, in particular, has been very active in developing an RGB laser scanner and using it to study various works of art, of which the Mona Lisa [Beraldin *et al.*, 2000; Blais *et al.*, 2007; Blais and Beraldin, 2006; Blais *et al.*, 2005; Taylor *et al.*, 2003]. Fringe projection systems with a color sensor also exist [Rocchini *et al.*, 2001a] as do integrated laser stripe projection and a color camera [Zheng *et al.*, 1998].

Integrated color – 3D acquisition devices with different sensors

When different sensors are used to acquire the two types of data we lose the natural correspondence between color and 3D data. We must thus find the correct transformation to project the 2D data on the 3D model. When the sensors are attached to one another, a simple calibration procedure can determine the internal parameters and the fixed relative position and orientation of the sensors. A few sensors based on this technique were developed to acquire 3D models of cultural heritage [Bernardini *et al.*, 2001; Ikeuchi *et al.*, 2007; Levoy *et al.*, 2000].

Integrated multispectral – 3D acquisition devices

A few integrated 3D/multispectral acquisition devices have been developed over the last decade, mostly for cultural heritage applications, but not solely. They are described in table 2.2. Mansouri *et al.* [2007] and Sitnik *et al.* [2010b] have each developed an integrated acquisition system based on an interference filter-wheel, a projector and a single sensor to acquire both 3D and multispectral data. The system developed by Sitnik *et al.* has the particularity of also integrating a flash lamp (for the multispectral acquisitions) and eleven LEDs to acquire additional images which permit the estimation of the surface BRDF. The system developed by Tonsho *et al.* [2001] also acquires a combination of multispectral, 3D and gonio-photometric data using a triangulation laser scanner, a multispectral camera and a tungsten lamp in seven successive positions. A turntable is used to survey the full object.

Table 2.2: Selected characteristics of integrated 3D digitization and multispectral acquisition systems.

References	Number of sensors	Filtering type	Spectral bands	Spectral range (nm)	3D digitization system
Mansouri *et al.* [2007]	single	interference filters	7	$400 - 1000$	projection of structured light
Sitnik *et al.* [2010b]	single	interference filters	20	$420 - 740$	projection of structured light
Tonsho *et al.* [2001]	multiple	optical filters	5	$380 - 780$	triangulation laser scanner
Brusco *et al.* [2006]	multiple	spectrograph	60	$400 - 1000$	time of flight laser scanner

Although [Mansouri *et al.*, 2007; Sitnik *et al.*, 2010b; Tonsho *et al.*, 2001] all cite cultural heritage as a possible application, to our knowledge only Brusco *et al.* [2006] have developed an integrated 3D/multispectral acquisition system and used it in-situ. Compared to the previous systems developed for the study of small objects (vases, small statues, etc), this system is aimed at digitizing frescoes: large, planar objects. It is based on a commercial spectrophotometer and a time-of-flight laser scanner. A rotating mirror and a rotating stage are used to scan the surroundings respectively vertically and horizontally. The calibration parameters are calculated using the correspondence between projected spots from the laser scanner and their image in multispectral datasets.

2.3.2 2D – 3D registration

When studying cultural heritage we value the possibility of choosing the 3D digitization system and the multispectral acquisition system independently. However, when separate devices are used, the registration procedure is no longer trivial. The internal parameters may be the same for a group of acquisitions but the relative position and orientation of the sensors must be determined independently for each view.

Using corresponding points

The traditional approach is to use homologous points in the 2D and 3D data to retrieve the unknown intrinsic and extrinsic camera parameters using the Tsai camera calibration method [Pulli, 1997; Rocchini *et al.*, 2002; Tsai, 1987]. The main defect of this method is the difficulty to identify corresponding points between the 2D and 3D data, be it manually or automatically. Color discrepancies do not necessarily correspond to structural discrepancies and vice versa.

Targets may be used to guide the registration process, but they are usually not adapted to cultural heritage applications where we want to minimize the disturbance to the object. Depending on the target resolution and registration accuracy, many targets may be necessary, partially occluding the object. However, there have been efforts to minimize user intervention [Franken *et al.*, 2005] in this type of registration setup.

Mutual Information methods

The need to find corresponding points is altogether eliminated if one uses fully automated maximization of mutual information methods [Corsini *et al.*, 2009; Maes *et al.*, 1997; Palma *et al.*, 2010b; Remondino *et al.*, 2009; Viola and Wells, 1997] for the registration of 2D on 3D. Mutual information is a statistical measure of similarity between two images. It is used to compare the 2D data to be mapped with a rendering of the 3D model. Many renderings have been used (depth map [Remondino *et al.*, 2009], gradient map [Palma *et al.*, 2010b], silhouette map, reflection map and other illumination-based renderings [Corsini *et al.*, 2009]). The camera parameters are iteratively optimized and a new rendering is created until the registration is achieved. The precision of the ensuing registration is of the order of a few pixels, though the success of such methods greatly depends on the rendering strategy [Corsini *et al.*, 2009]. This

method provides good results for visualization purposes, even using low quality images, as in [Remondino *et al.*, 2009].

Registration based on known sensor position

There is no need to estimate the camera parameters from the data if their position is known with sufficient precision at the time of each acquisition. In theory, coordinate measurement machines can be used, but they are too cumbersome to be used in situ. Another possibility is to use magnetic tracking [Paperno *et al.*, 2001; Raab *et al.*, 1979] to derive the position and orientation of the sensor in use. However, even recent sensors [Sherman *et al.*, 2007] are not sufficiently precise: they have over a millimeter of error in positional accuracy and 4° in angular accuracy. Furthermore, surrounding metals in the acquisition space increase this error to the point of rendering the measures useless [Bernardini *et al.*, 2002].

2.3.3 Perspectives

Integrated multispectral/3D digitization systems can acquire an impressive amount of data. The absence of a registration procedure (or ease of such) is a strong advantage of integrated color/multispectral and 3D acquisition systems, be they single sensor or multisensor. However, these systems lack flexibility. They are often built for a specific application and, especially in the case of single sensor devices, it is impossible or difficult to adapt the acquisition system to objects of a different scale. It may also be difficult to acquire clean data since 3D digitization and multispectral (or color) acquisitions often require different illumination setups or poses.

When relying on algorithms to register data from independent sensors, the accuracy of the registration greatly depends on the quality of the data acquired. To perform well, these algorithms need structured surfaces or salient points. Cultural heritage does not always present natural salient points and the use of artificial targets is frowned upon.

Registration is greatly simplified when the position of the sensor at the time of the acquisition is known. However, mechanical structures are cumbersome while magnetic tracking does not reach the necessary accuracy.

The strength and flexibility of tracking resides in the fact that it is independent from the acquisition system in use. Section 2.2.3 showed that optical tracking can greatly aid the 3D – 3D registration process [Blais *et al.*, 2005; *Creaform, MetraScan*; Kawasaki and Furukawa, 2004; *Leica, T-Scan*; *NDI, ScanTrak*; *Steinbichler, T-Scan*; Valkenburg *et al.*, 2006]. Similar techniques can be used to extract the position and orientation of any acquisition system for multimodal registration.

Breuckmann and Metronor have developed a multisensor digitization system, the NaviSCAN [*Breuckmann, NaviSCAN*], based on two photogrammetric cameras, a structured light projection scanner, a frame to fix on the digitization system and a probing unit. The NaviSCAN can measure a 1 m stick 6 m away from the tracking cameras with a 0.03 mm (1σ) accuracy. The photogrammetric cameras track the position and orientation of the reference frame which is fixed to the digitization system. Acquisitions from multiple views are thus automatically stitched together and data from the probing unit are seamlessly integrated in the same coordi-

nate system.

This type of setup is based on close range photogrammetry, and can be extended to register multispectral data on 3D models. Close range photogrammetry is a mature field which is widely used in industrial settings [Luhmann, 2010a] and has long been used for real-time applications [Maas, 1997a]. In industrial settings photogrammetry is often used to track robot end effectors. This tracking is used to carry out factory calibration and is increasingly being investigated as a method to monitor and calibrate a robot while in use on factory floor. Two main setups are possible: either the robot end effector is equipped with one or more cameras that detect targets in the workspace to derive its position and orientation [Clarke and Wang, 2000; Hefele, 2002; Hefele and Brenner, 2001] or cameras survey the workspace to obtain the position and orientation of the effector, which is enhanced with a target object [Hefele and Brenner, 2001; Maas, 1997b; Schütze *et al.*, 2009]. The Optopose system developed at i3mainz, Institute for Spatial Information and Surveying Technology (Germany), based on this second setup, tracks the position and orientation of a robot effector with a precision of 0.05 mm [Boochs *et al.*, 2009].

Our work extends this type of setup to track multiple optical acquisition systems in cultural heritage settings to permit a precise integration of multispectral and 3D data. We believe the use of optical tracking presents a good compromise in terms of flexibility, portability and end precision. The following chapter presents the method devised and describes the simulations performed to evaluate the achievable tracking accuracy of such setup.

Chapter 3

Simulations

3.1 Overview

Our goal is to devise a flexible technique to register featureless multispectral and 3D datasets describing cultural heritage objects. We want to achieve a registration with a sub-pixel accuracy, more accurate than mutual information methods (for which the accuracy is a couple of pixels). We thus define a registration accuracy goal of half a pixel. We saw in the previous chapter that tracking the acquisition system is a practical technique to register 3D data. We extend such methods for the registration of 3D and multispectral datasets. Our method relies on close range photogrammetry techniques to track the acquisition systems in use.

3.1.1 Method description

The proposed acquisition setup is illustrated figure 3.1: A set of cameras observe the acquisition systems while they successively digitize the surface under study from various positions. These cameras are subsequently called "tracking cameras" to differentiate them from the other camera systems mentioned throughout this dissertation. The tracking cameras are fixed for the duration of the acquisitions and must survey the full volume that each acquisition system will occupy.

By processing the images acquired by the tracking cameras we can measure the position and orientation of the acquisition systems in a coordinate system defined by the tracking cameras. The subsequent registration procedure is straightforward: the acquired datasets are projected in a world coordinate system, using the known position and orientation of each acquisition system for each acquisition.

To ensure a precise tracking, all optics and objects in play must be carefully calibrated. We introduce the following coordinate systems, linked to the materials in use and illustrated figure 3.2:

- $C_{S_i}, (O_{S_i}, \vec{x}_{S_i}, \vec{y}_{S_i}, \vec{z}_{S_i})$ is the coordinate system linked to acquisition system i.
- $C_{C_j}, \left(O_{C_j}, \vec{x}_{C_j}, \vec{y}_{C_j}, \vec{z}_{C_j}\right)$ is the coordinate system linked to tracking camera j. O_{C_j} is the optical center of the camera; $(\vec{x}_{C_j}, \vec{y}_{C_j})$ define the image plane; $\vec{z}_{C_j}$ is collinear to the optical axis.

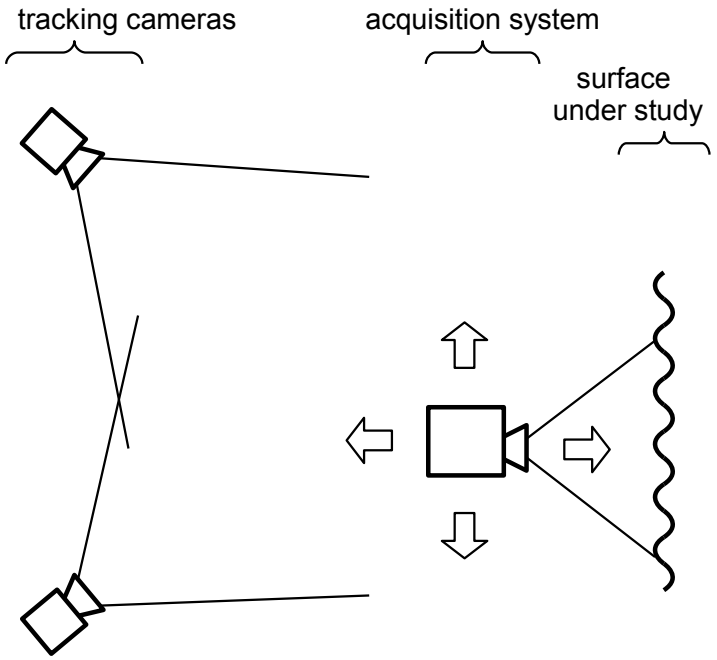


Figure 3.1: Overview of the tracking procedure.

- $C_0, (O_0, \vec{x}_0, \vec{y}_0, \vec{z}_0)$ is the world coordinate system.

We now describe the necessary calibrations and how they are performed. The following notations will be used throughout: $A|_{C_U}$ are the homogeneous coordinates $(x_A, y_A, z_A, 1)$ of point A in coordinate system C_U. We define T_{C_U,C_V} the transformation matrix between two coordinate systems C_V and C_U such that for all points A, $A|_{C_U} = T_{C_U,C_V} \cdot A|_{C_V}$.

Internal orientation of the tracking cameras The calibration of the tracking cameras is performed by taking close to a hundred images of a calibration plate from various points of view. The calibration plate is covered with coded and uncoded targets and two distances are precisely known. From this we can measure the internal camera parameters such as focal length, principle point offset and lens distortions. Several tracking cameras can be calibrated together if they are placed side by side and observe the same area. Previous experience has taught us that the internal orientation can stay stable for over a week if the cameras are handled with care during this time frame. We often perform the interior orientation a few days before the acquisitions.

Internal orientation of the acquisition systems It is necessary to know the distortions introduced by the acquisition system to project the acquired data correctly. When using multispectral cameras, the procedure is the same as that for the tracking cameras. Commercial 3D digitization systems generally have a specific calibration procedure that must be performed so that the output data is automatically corrected by the acquisition software.

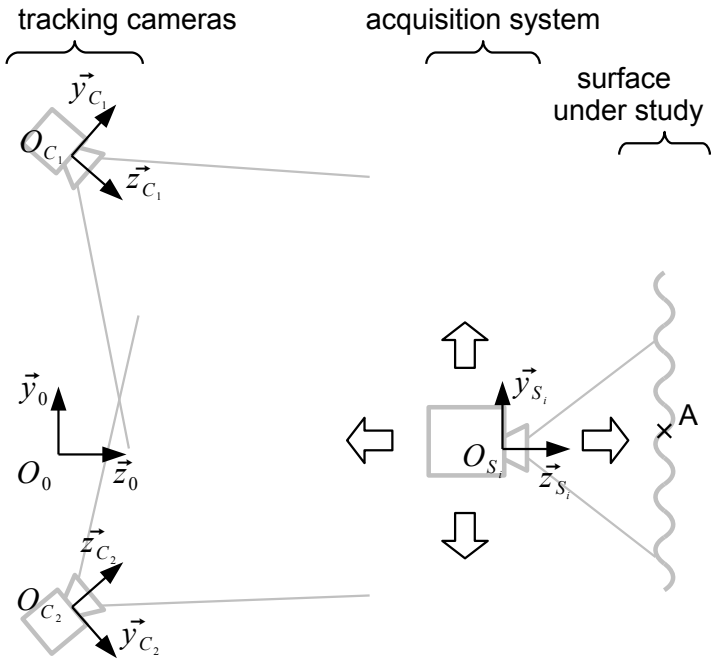


Figure 3.2: Coordinate systems defined for the tracking procedure.

Exterior orientation of the tracking cameras Once the tracking cameras have been positioned to observe the area in front of the surface under study, we can measure their relative position and orientation. This is done by acquiring approximately eighty images of a scale bar in various positions and orientations, simultaneously by all calibrated tracking cameras. The position and orientation of the cameras are measured with respect to the first. We generally define the coordinate system of the first tracking camera to be the world system: $C_0 = C_{C1}$.

Calibration of the acquisition systems The accuracy of any photogrammetric setup is greatly improved if well defined artificial targets are used instead of relying on natural features. A variety of photogrammetric targets are thus placed on the acquisition systems. It is necessary for us to know the position of these targets in the coordinate system defined by the acquisition system. This requires an additional calibration. It is essential to note that we have moved the disturbance away from the object under study and on to the acquisition system, which we do not mind manipulating.

This calibration is performed by taking between fifty and one hundred images of the acquisition surrounded by a scale bar and additional targets, while it digitizes another target-covered object. The photogrammetric processing of these images provides us with the coordinates of all targets in a common coordinate system C_{temp}. These targets include those fixed to the acquisition system, those on the target-covered object, those on the scale bar and any additional targets in the scene. The 3D or multispectral acquisitions of the target-covered object provide us with the coordinates of the targets in C_{S_i}. It is thus possible to calculate $T_{S_i,temp}$ and thus to evaluate the position of the targets describing the acquisition system in C_{S_i}. The accuracy of this calibration is improved if it is performed for varying relative positions of the target-covered

object and the acquisition system.

Data processing

We rely on two pieces of software for all photogrammetric image processing: Tritop [*Gom, Tritop*] and i3AxOriGui, a lab-developed software based on the AXOri library [*Axios3D, Axori*].

Tritop is used to print coded targets of the dimensions we need. This software also recognizes the coded and uncoded points in the images and computes a first assessment of the position of the cameras. This data is then exported to i3AxOriGui in which we have more flexibility and control on what we want to compute given our input parameters.

i3AxOriGui performs the bundle adjustment of the data both for the calibrations and the tracking. Each parameter can be set either as an unknown or an input parameter, depending on what we want to calculate.

For example, we can estimate the target coordinates or the object position by assuming the tracking cameras' interior and exterior orientation are known. It is also possible to assess the exterior orientation of the tracking cameras assuming the interior orientation and the object coordinates are accurate. If we assume the tracking camera exterior orientation and the object coordinates are known then we can determine the tracking cameras interior orientation. The bundle adjustment results also provide us with an internal accuracy measure of the parameters calculated.

Each acquisition provides us with the coordinates of a group of surface points in a given sensor system, $A|_{C_{S_i}}$. The simultaneous tracking provides us with $T_{C_0,C_{S_i}}$ for each acquisition. The known interior orientation of the tracking cameras and their relative orientation ensures that $T_{C_0,C_{S_i}}$ is sufficiently accurate. We can thus calculate $A|_{C_0}$, the coordinates of the surface points in the world system using:

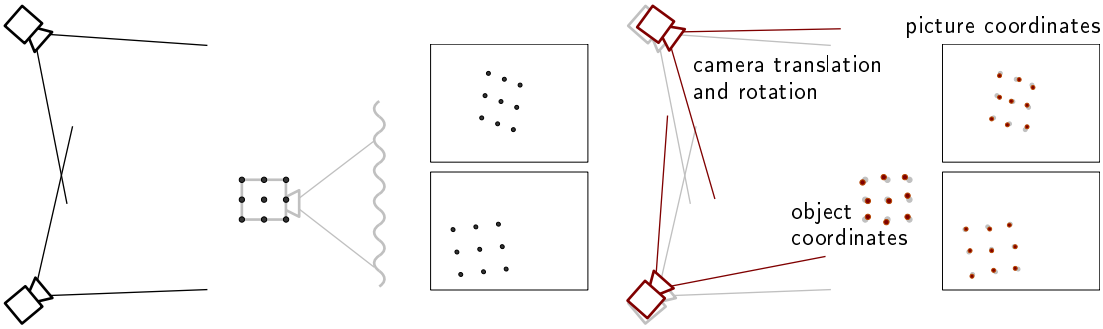
$$A|_{C_0} = T_{C_0,C_{S_i}} \cdot A|_{C_{S_i}} \quad . \tag{3.1}$$

3.1.2 Simulation procedure

The tracking accuracy depends on many parameters such as the number, focal length, position, and sensor characteristics of the photogrammetric cameras; the dimensions of the acquisition area; the dimensions of the acquisition system, etc. We rely on a series of simulations to evaluate how accurately we can track the position and orientation of a given acquisition system.

These simulations are used to evaluate the accuracy with which we can determine the exterior orientation of the tracking cameras, as well as the accuracy of tracking of the acquisition systems. The input parameters are the characteristics of the tracking cameras, their position and orientation, the position and orientation of the acquisition systems, the accuracy of the calibration of the acquisition systems. These simulations enable us to easily test many configurations, to detect the stable setups that will provide the registration accuracy we aim to attain and to discard unstable setups. Though they only concern the tracking and part of the calibrations, these simulations enable us to evaluate the accuracy of the global registration.

Our four stage simulation pipeline is based on the following steps:



(a) Scene creation (b) Image generation (c) Noise addition

Figure 3.3: First three simulation steps.

1. Scene creation (figure 3.3a).
2. Image generation (figure 3.3b).
3. Noise addition (figure 3.3c).
4. Bundle adjustment.

Scene creation A 3D representation of the simulation scene is created using 3ds Max [*Autodesk, 3ds Max*]. This scene must contain the tracking cameras and the acquisition devices in the positions we want to track them. We can vary the field of view of the tracking cameras to match our needs. It is important that the cameras survey the full volume that the acquisition systems will successively occupy. We call this volume the acquisition system "bounding box". The dimension of each bounding box depends on the dimensions of the surface under study, the acquisition distance and the acquisition system volume.

The scene can also represent the object under study to help guide the positioning of the acquisition systems. Each acquisition system is modeled by a box defined by a variable number of points. These points represent the targets that we will fix to the acquisition device. The scene is exported as a *.WRL file which includes the tracking cameras' position and orientation, their field of view and the coordinates of the points representing the tracked object.

Image generation This *.WRL file is read by a lab-developed software, jbb-i3mainz. The input scene parameters are used to calculate the images captured from each camera. The sensor dimensions must be entered manually and it is possible to provide lens distortion parameters. This software can also remove the points that are hidden by other surfaces.

Noise addition Jbb-i3mainz is also used to add Gaussian noise to the following four scene parameters:

- Picture coordinates: the coordinates of the targets in the images taken by the tracking cameras. These are usually known with an accuracy better than $1/10^{\text{th}}$ of a pixel. In

favorable conditions (sufficient contrast and focus, well resolved targets) the picture coordinates can be resolved with an accuracy of $1/30^{\text{th}}$ of a pixel or even $1/50^{\text{th}}$ [Luhmann, 2010a].

- Object coordinates: the coordinates of the points that define the tracked object. The calibration of the acquisition systems should provide these coordinates with an accuracy better than 0.05 mm.
- Camera translation: the position of the tracking cameras (X, Y, Z coordinates). The accuracy of these values depends on the camera setup and on the results of the calibration procedures. We can usually resolve the exterior camera orientation with an accuracy better than 0.05 mm spatially.
- Camera rotation: the orientation of the tracking cameras (Ω, Φ , K coordinates). We usually know these coordinates with an accuracy better than 0.05 mrad though this also depends on the success of the exterior orientation.

The noisy scene and image parameters are then exported as an *.axo file.

Bundle adjustment This *.axo file is treated by i3AxOriGui, which performs the bundle adjustment of the simulation data just as it would treat real data. The accuracy results provided by the bundle adjustment are those given in the simulation results.

3.1.3 Simulation configurations

The simulations are performed for predefined configurations that include three acquisition systems owned by our laboratories and three objects of interest partially defined by our motivating objects.

Acquisition systems

The simulations are based on the characteristics of two multispectral cameras and an optical 3D digitization system owned by our institutions. Two of these acquisition systems have been used in the past to digitize our objects of interest and all three are adapted for the study of cultural heritage objects.

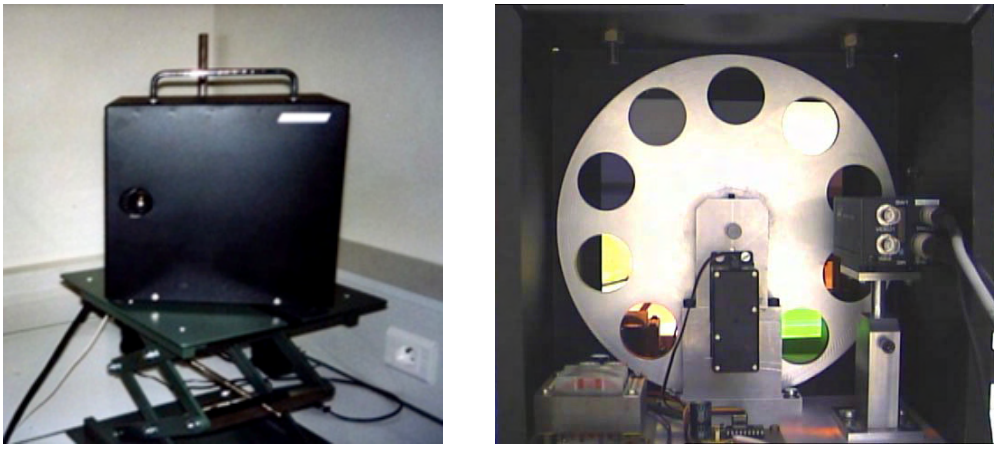
The two multispectral cameras are a commercial camera from FluxData and a lab-designed multispectral camera. A few characteristics of both cameras are given in table 3.1. The lab-designed multispectral camera [Mansouri, 2005; Mansouri *et al.*, 2005c], shown figure 3.4, is based on interference filters mounted on a filter-wheel and has been used in the past to document the Trier sarcophagus and the Rüdesheim wall-painting. Careful calibration and a neural network algorithm provide us with a reflectance spectra for each pixel [Sanchez *et al.*, 2005].

The FluxData FD-1665-MS camera [*FluxData*], shown figure 3.5, is based on a 3 CCD system which provides simultaneous data for each spectral band. This camera acquires less spectral bands than the lab-designed multispectral camera, but the bigger sensor and pixel size will allow us to register the data more precisely. The output data can also be used in conjunction

Table 3.1: Characteristics of the lab-designed and FluxData multispectral cameras.

Multispectral camera	Sensor size W × H (mm × mm)	Sensor size W × H (pixels × pixels)	Cell size (µm)	Focal length (mm)
Lab-designed	8.9 × 6.7	1392 × 1040	6.45	25
FluxData	7.5 × 6.2	659 × 494	9.9	25

Multispectral camera	Filtering technology	Number of bands	Acquisition range (nm)
Lab-designed	interference	9	400 – 1000
FluxData	3 CCD	7	380 – 1000



(a) (b)

Figure 3.4: Exterior (a) and interior (b) of the lab-developed multispectral camera. The camera, filter wheel and controllers are visible in the interior view.

© FluxData

Figure 3.5: FluxData multispectral camera.

Table 3.2: External characteristics of the acquisition systems.

Acquisition system	External dimensions W × H × D (mm × mm × mm)	Weight (kg)
Lab-designed multispectral camera	270 × 320 × 180	5
FluxData multispectral camera	92 × 112 × 187	1.51
Gom Atos III	490 × 170 × 300	7.5

with a neural network algorithm for reflectance estimation. The FluxData multispectral camera is more compact and lighter than our lab-designed camera (see table 3.2). This camera is better suited to the in situ study of cultural heritage objects.

We perform all 3D digitizations using a commercial fringe projection system, the Atos III, manufactured by GOM [*Gom, Atos*] and shown figure 3.6. The Atos is composed of two 4 Mpx cameras and a projector. Different acquisition configurations can be built by varying the position, orientation and lens of the cameras and projector. The resolution of the system is proportional to the dimensions of the field of view. Among the many possible configurations, we use the one that provides a 500 mm × 500 mm field of view. This setup entails a 0.24 mm resolution and a fixed distance of 760 mm between the acquisition system and the object under study.

Simulation configurations

Our goal is to show that this registration technique is flexible and adapted to a variety of cultural heritage applications. Three simulation configurations were defined to evaluate the potential of the method for the study of a variety of cultural heritage objects. In addition to the objects which motivated this study, we imagine studying a small statue. Our three case studies are thus:

- an area of 0.4 m × 0.7 m, corresponding to the study of the Trier sarcophagus;

Figure 3.6: Gom Atos III 3D digitization system.

- an area of 2 m×1.5 m, corresponding to the analysis of the largest Rüdesheim wall painting;
- a cylinder 1 m high and with a 0.3 m radius, representing a small statue.

The remainder of this chapter successively presents the simulations performed in these three configurations. The simulations performed in the sarcophagus configuration were used to fine tune many tracking parameters. As such, they are presented in greater detail. For the wall painting and statue simulations we assume we use the tracking material defined during the sarcophagus simulations. In these last two setups, the simulations are primarily used to optimize the number and position of the tracking cameras to reach our tracking accuracy goal.

3.2 Sarcophagus configuration

3.2.1 Configuration overview

The first simulations are performed in a setup that corresponds to the digitization of a small area of the surface of the Trier sarcophagus. The sarcophagus is approximately 1.5 m high, 1 m wide and 2.3 m long. It occupies a quasi-central position in a fairly small crypt. The crypt is 6 m at its widest, 6.6 m deep and 3.8 m high. The ground around the sarcophagus is set 8 cm lower than the rest of the crypt. Since the area we are interested in faces the entrance, the stairs also limit the space that is available to position the tracking cameras. Such spatial constraints are not unusual in cultural heritage settings. The acquisition systems and the tracking cameras must fit in the two meters between the sarcophagus and the bottom steps, as illustrated figure 3.7.

In the current configuration, the area of interest is 0.4 m × 0.7 m and is very close to the floor, as shown in figure 3.8. This means that our acquisition systems must sit very low to digitize the lower areas of the sarcophagus. The Gom Atos III measuring distance is 760 mm for the field of view we have chosen. If the Atos is perfectly perpendicular to the surface it can digitize the complete area of interest by moving in a volume that is approximately 500 mm wide, 370 mm high and 300 mm deep, as illustrated figure 3.9.

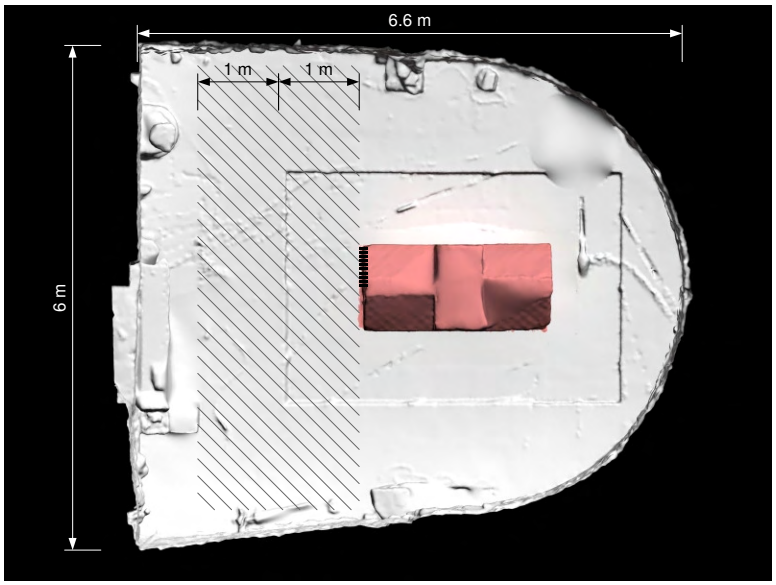


Figure 3.7: Crypt interior, top view. The sarcophagus is shown in the center of the room. The entrance of the crypt is at the top of the stairs, on the left. The hatched area represents the space that is available to position the tracking cameras and the acquisition systems.

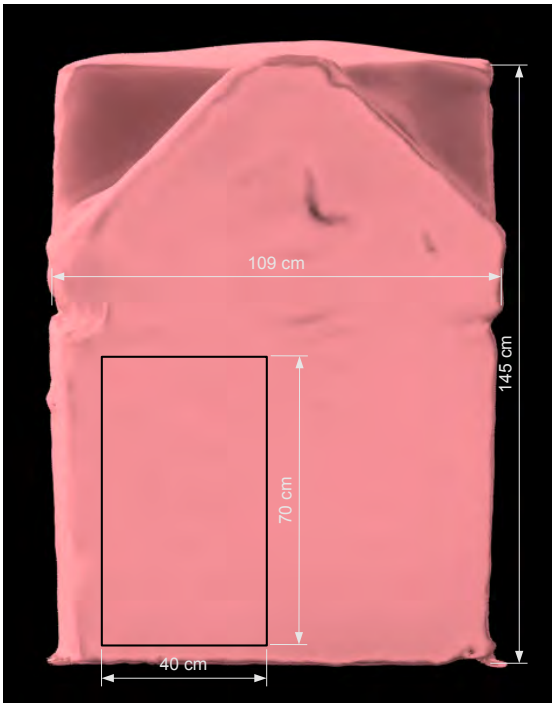
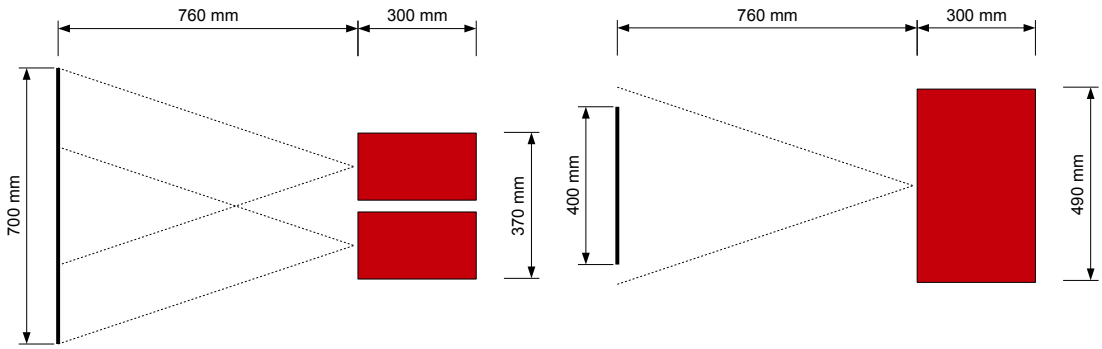


Figure 3.8: Dimensions of the front face of the sarcophagus. The black rectangle represents the approximate position of the area of interest.



(a) Side view (b) Top view

Figure 3.9: Top and bottom position of the Gom Atos III acquiring an area of 400 mm × 700 mm.

Table 3.3: Characteristics of the multispectral acquisitions at varying distances.

Multispectral camera	On object	Distance 500	600	700	800	(mm)
Lab-designed	image width	178	213.6	249.2	284.8	(mm)
	image height	134	160.8	187.6	214.4	(mm)
	pixel size	0.129	0.154	0.181	0.206	(mm)
	required images	3×6	2×5	2×4	2×4	(W × H)
FluxData	image width	130	156	182	208	(mm)
	image height	96	115.2	134.4	153.6	(mm)
	pixel size	0.198	0.238	0.277	0.317	(mm)
	required images	4×8	3×7	3×6	2×5	(W × H)

To evaluate the bounding box of each multispectral camera, we must first settle on an acquisition distance. The measuring distance for the multispectral cameras is a compromise between the desired level of detail and the number of acquisitions necessary to cover the full area of interest. Table 3.3 gives the image and pixel dimensions and the number of necessary acquisitions to cover the sarcophagus area of interest for various acquisition distances. We usually use the lab-designed multispectral camera at approximately 500 mm from the surface. In this case however, we will be able to track the acquisition systems with greater accuracy if they are all at the same distance from the sarcophagus since this limits the total tracking volume. We thus assume the multispectral cameras are 700 mm from the sarcophagus. At this distance we can acquire the full area of interest with eight images using the lab-designed multispectral camera and eighteen images using the FluxData multispectral camera. We also retain a resolution that is under 0.3 mm.

A 3D rendering of the three resulting bounding boxes is shown figure 3.10; their dimensions are given table 3.4. There is approximately 1 m available between the Gom Atos III bounding box and the bottom step of the stairs to position the tracking cameras.

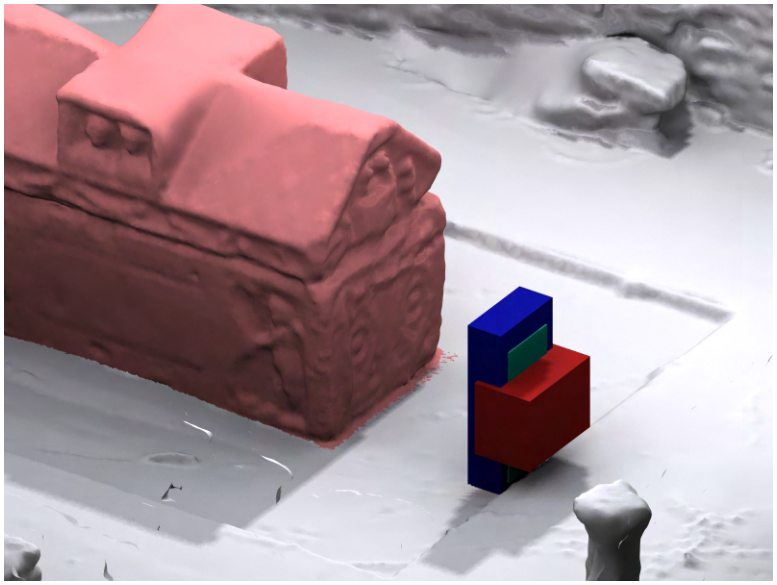


Figure 3.10: Bounding boxes of the Gom Atos III (red), the lab-designed multispectral camera (blue) and the FluxData multispectral camera (green) in front of the sarcophagus.

Table 3.4: Bounding box dimensions to digitize an area of 0.4 m × 0.7 m.

Acquisition system	Bounding box W × D × H (mm × mm × mm)	Acquisition distance (mm)
Lab-designed multispectral camera	421 × 832 × 180	700
FluxData multispectral camera	310 × 678 × 187	700
Gom Atos III	490 × 370 × 300	760

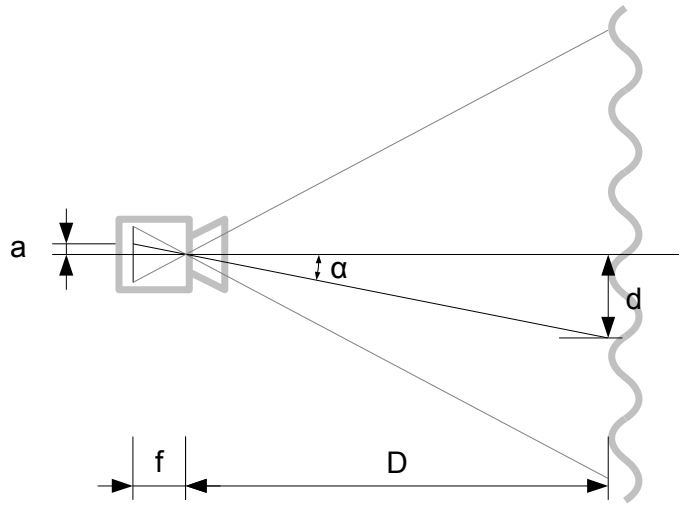


Figure 3.11: Notations used to calculate the simulation target accuracy.

Table 3.5: Target accuracy for each acquisition system.

Acquisition system	Acquisition distance D (mm)	Target accuracy spatial d (mm)	Target accuracy angular α (mrad)
Lab-designed multispectral camera	700	0.090	0.128
FluxData multispectral camera	700	0.139	0.198
Gom Atos III	760	0.120	0.158

Simulation goal

Our goal is to track the acquisition systems with an accuracy better than half an image pixel. The target accuracy is calculated with simple geometric considerations. If f is the camera focal length, D the acquisition distance, a the target accuracy of the sensor and d the spatial target accuracy on the object, α the target angular accuracy (see figure 3.11), we have:

$$\alpha \approx \tan\alpha = \frac{a}{f} = \frac{d}{D} \ .$$

For the multispectral cameras a is half the pixel size, and it is easy to calculate α. Because of the difference in sensor dimensions, this translates into a value which is lower — and thus harder to achieve — for the lab-designed multispectral camera, compared to the FluxData multispectral camera. In the case of the Gom Atos, the value we have is the resolution on the object ($2d = 0.24\,\text{mm}$) so the target spatial accuracy is half the resolution. The target spatial and angular accuracy for each acquisition system is given table 3.5. Note that for a given acquisition system, only the spatial target accuracy depends on the acquisition distance.

3.2.2 Positioning the tracking cameras

We position the cameras based on general guidelines in close range photogrammetry, such as those given by Remondino and El-Hakim [2006, p. 275]:

(a) the accuracy of a network increases with the increase of the base to depth (B:D) ratio and using convergent images rather than images with parallel optical axes;

(b) the accuracy improves significantly with the number of images in which a point appears. But measuring the point in more than four images gives less significant improvement;

(c) the accuracy increases with the number of measured points per image. However, the increase is not significant if the geometric configuration is strong and the measured points are well defined (like targets) and well distributed in the image;

(d) the image resolution (number of pixels) influences the accuracy of the computed object coordinates: on natural features, the accuracy improves significantly with image resolution, while the improvement is less significant on well-defined, large, resolved targets.

Item (b) motivates us to perform our first simulations with four tracking cameras. As much as possible, we try to position them at 90° angles with respect to one another.

To ensure the most accurate tracking achievable, each photo must be filled as much as possible with targets. Since each acquisition system occupies only a portion of its bounding box, the cameras must be as close as possible to the bounding boxes, while still observing them entirely.

3.2.3 Choice of sensor and lens

The limited space in front of the sarcophagus constrains our choice of material. The tracking cameras must be used in conjunction with lenses that provide a wide field of view. The relationship between a cameras' angle of view Ω, its focal length f and its sensor dimension s is:

$$\Omega = 2 \times arctan\left(\frac{s}{2 \times f}\right) \ . \tag{3.2}$$

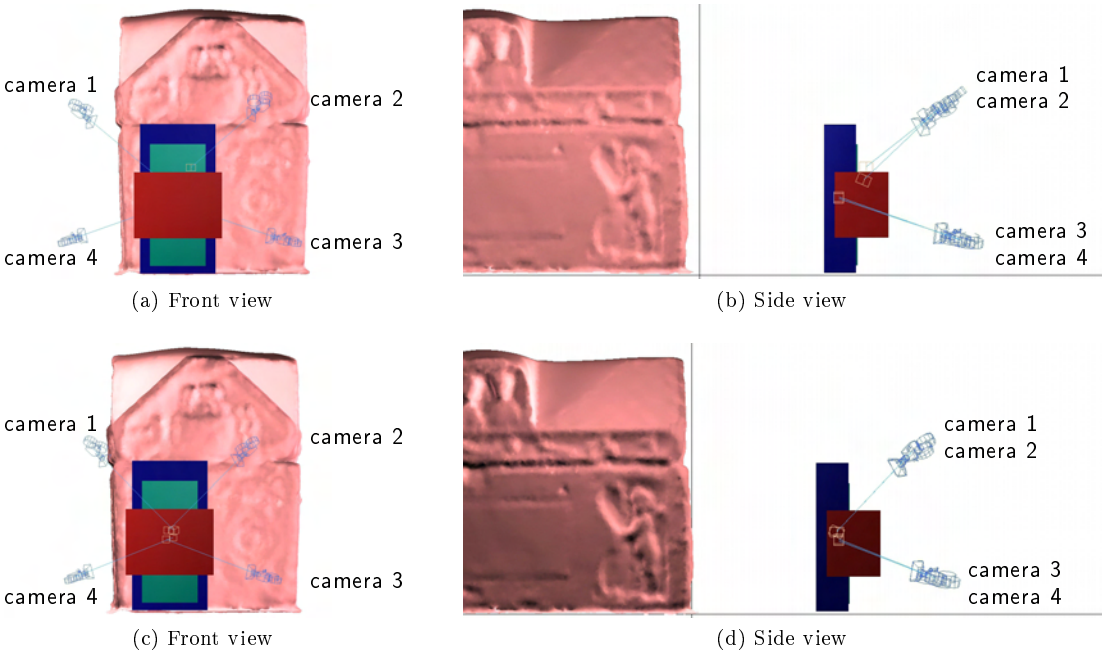
The smaller the focal length, the larger the field of view. We opt for 6 mm lenses so the tracking cameras can see the full bounding box from a short distance.

The first goal of the simulations is to determine whether 2 Mpx tracking cameras can be used or if we need 5 Mpx cameras to reach our target tracking accuracy. These first simulations were based on the sensor characteristics of two Baumer cameras, given in table 3.6. These cameras are representative of the 2 Mpx and 5 Mpx cameras available from other manufacturers.

Equation 3.2 shows that the field of view depends not only on the focal length, but is also a function of the camera sensor dimensions. The 2 Mpx cameras used with the 6 mm lens have a smaller field of view and must be positioned further from the sarcophagus than the 5 Mpx cameras with the same lens. The resulting camera arrangements are illustrated figure 3.12.

Table 3.6: Characteristics of the cameras and lens used in the simulations.

	Sensor size (W × H) (mm × mm)	Sensor size (W × H) (pixels × pixels)	Cell size (µm)	Ω_{horz}	Lens (mm)
Baumer TXF20	7.1 × 5.4	1624 × 1236	4.40	61.2°	6
Baumer TXF50	8.4 × 7.1	2448 × 2050	3.45	70.0°	6



(a) Front view (b) Side view

(c) Front view (d) Side view

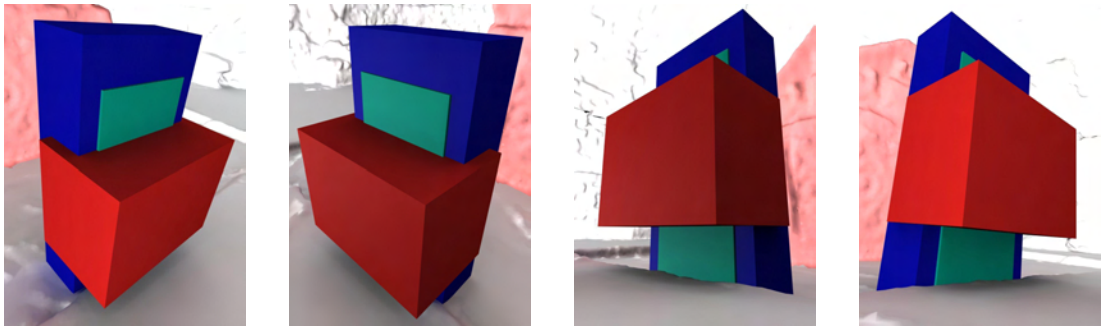
Figure 3.12: Tracking cameras and acquisition system bounding boxes. (a-b) 2 Mpx cameras. (c-d) 5 Mpx cameras.

Even though the camera characteristics are different, it is possible to obtain equivalent views, as shown figure 3.13.

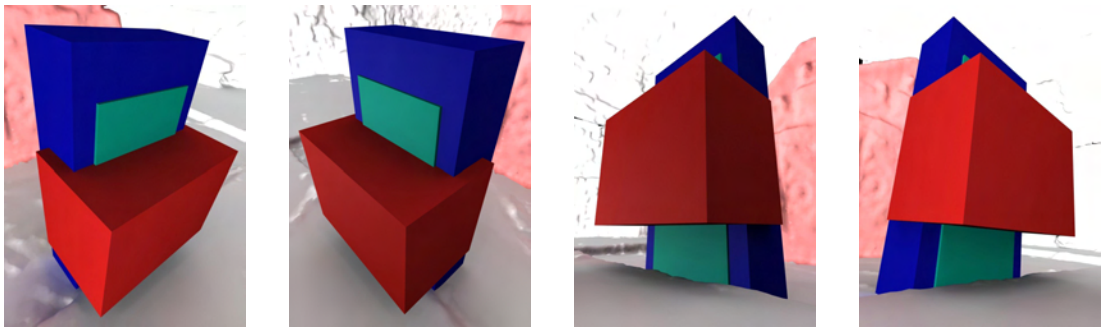
We do not place the tracking cameras exactly at 90° angles because of the constraints of the room. The multispectral camera bounding boxes are in front of a small step and the bottom cameras can not be positioned much lower — they are ten centimeters above the ground. It is also not recommended to position the top cameras higher since this would result a lower number of targets visible from all four cameras, which would hurt the achievable tracking accuracy.

The acquisition systems are represented in the simulations by a box the size of their outer dimensions. This is an accurate representation in the case of the lab-designed multispectral camera. If the objects are smaller and more intricate than their representation — as is the case for the FluxData and Atos acquisition systems — then the simulations will overestimate the achievable tracking accuracy.

Each box is itself represented by a given number of points, which corresponds to the number of targets that are placed on its surface. Figure 3.14 shows the boxes representing our three ac-



(a) Camera 1 (b) Camera 2 (c) Camera 3 (d) Camera 4



(e) Camera 1 (f) Camera 2 (g) Camera 3 (h) Camera 4

Figure 3.13: View of the bounding boxes from the tracking cameras. (a-d) 2 Mpx cameras. (e-f) 5 Mpx cameras.

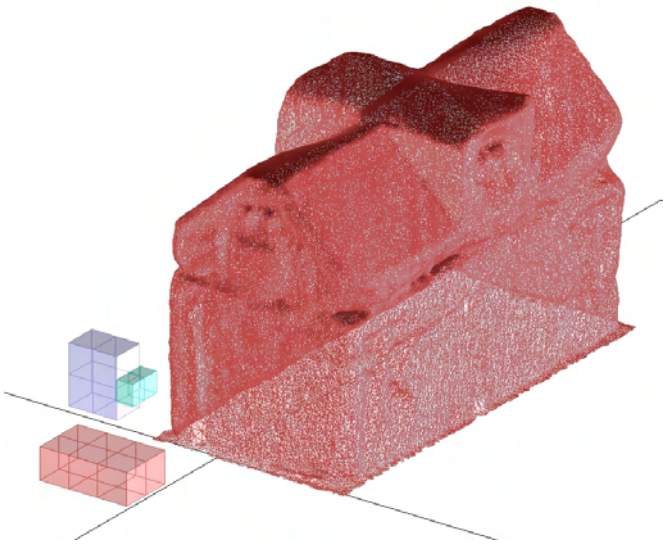


Figure 3.14: Representation of the acquisition systems for the simulations: lab-designed multispectral camera (blue), FluxData multispectral camera (cyan) and Atos digitization system (red).

quisition systems. The points defining each acquisition system are the corners and intersections of the wireframe view overlaid on the boxes. The number of points per acquisition system is chosen based on the dimensions of the box and thus the surface available to fix photogrammetric targets.

The lab-designed multispectral camera is represented by eighteen points, the FluxData multispectral camera by twelve and the Atos digitization system by twenty four. Once again, this is a realistic estimate for the lab-designed multispectral camera but less so for the other two acquisition systems. Very little surface is available to fix targets on the FluxData multispectral camera and the Atos. We would not place any targets on the lens of the FluxData multispectral camera but they occupy almost half of the volume of the complete camera.

The simulations are run for several positions of each acquisition system in both tracking camera configurations. In the case of the multispectral cameras five positions are used: the four corners and the center of the corresponding bounding box. Since the Atos bounding box is quite small and as large as the acquisition system itself, only three positions are used: the top, center and bottom of the bounding box.

For all positions we add a Gaussian noise of standard deviation 0.345 µm to the picture coordinates. This corresponds to $1/10^{\text{th}}$ of the pixel size of the 5 Mpx sensors and approximately $1/13^{\text{th}}$ of the pixel size of the 2 Mpx sensors. The simulation results are presented table 3.7. This table gives the maximum value of the spatial and angular tracking inaccuracy reached by a given acquisition system for each configuration. The full results are available in the appendix (tables A.2 to A.8, page 145). Unless otherwise noted, all accuracy results presented throughout this thesis are given at $2\,\sigma$, that is with a 95 % confidence interval.

We notice that the tracking the accuracy is always better in the 5 Mpx configuration than in the 2 Mpx configuration, even though this simulation protocol does not take into account the

Table 3.7: Simulation results, tracking the acquisition systems successively with 2 Mpx and 5 Mpx cameras.

Acquisition system	Points	2 Mpx Spatial (mm)	2 Mpx Angular (mrad)	2 Mpx See table	5 Mpx Spatial (mm)	5 Mpx Angular (mrad)	5 Mpx See table
Lab-designed MS camera	18	0.020	0.292	A.2	0.017	0.238	A.6
FluxData MS camera	12	0.025	0.536	A.3	0.022	0.536	A.7
Gom Atos III	24	0.013	0.168	A.4	0.011	0.142	A.8

sensor resolution. The difference in tracking accuracy between the two configurations is thus only due to the difference in sensor size. The results would be even more distant if the resolution were taken into account. Since the 5 Mpx cameras appear to provide a tangible improvement on the achievable accuracy the subsequent simulations are run with the characteristics of the 5 Mpx cameras.

The most accurately tracked acquisition system in both configurations is the Gom Atos III. This is not surprising since it is the biggest object, defined by the highest number of targets and placed in the center of the tracking volume.

In all configurations the attained spatial accuracy is better than our most constraining target value of 0.09 mm. The difficulty resides in reaching our target angular accuracy of 0.128 mrad in the most restrictive case (for the tracking of the lab-designed multispectral camera). Even when tracking the Atos with the 5 Mpx cameras, which is the most favorable configuration, we do not attain this target angular accuracy.

3.2.4 Target frame necessity

The tracking accuracy can be improved by increasing the number of tracking cameras that survey the bounding box. However, using more than four cameras to track the acquisition systems when the area of interest is 0.4 m × 0.7 m seems disproportionate. We thus explore another strategy to better the achievable tracking accuracy: increasing the dimensions of the tracked object.

The angular accuracy of the tracking strongly depends on the object size and less so on the number of targets. We would obtain better results if the acquisition systems were bigger. In practice this can be achieved by fitting a 2D or 3D frame around the acquisition systems. This virtually increases the dimensions of the acquisition system. Photogrammetric targets cover the frame instead of the acquisition systems and it is the frame which is the object of the tracking. The modified acquisition overview is illustrated figure 3.15.

We design a target frame such that all three acquisition systems fit in it. The frame is 0.5 m wide to fit the width of the Atos, 0.4 m high to fit the height of the lab-designed multispectral camera and 0.3 m deep to accommodate the depth of the Atos.

By increasing the dimensions of the acquisition system we also increase the bounding box dimensions. The cameras must be slightly repositioned to account for this accretion. Since all acquisition systems can fit in the target frame we position the tracking cameras based on a single bounding box. The tracking accuracy no longer depends on the dimensions of the

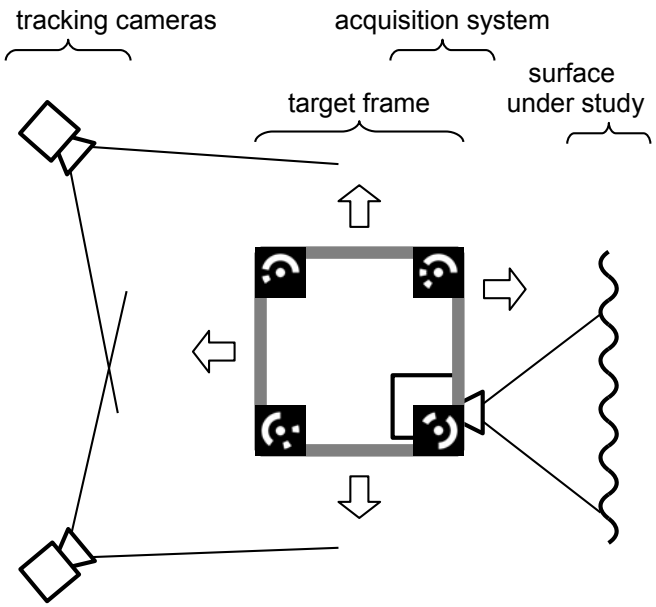


Figure 3.15: Overview of the tracking procedure using a target frame.

Table 3.8: Simulation results, tracking a target frame of varying dimensions.

	Target frame dimensions W × H × D (m × m × m)	Points	Output accuracy Spatial (mm)	Angular (mrad)	See table
(a)	0.5 × 0.4 × 0.3	26	0.015	0.144	A.10
		56	0.012	0.124	A.11
(b)	0.5 × 0.5 × 0.5	26	0.017	0.120	A.12
		56	0.012	0.088	A.13

acquisition system but on those of the target frame. Any acquisition system that is attached to the target frame will be tracked with the same accuracy.

The target frame is bigger than the acquisition systems and can be covered by a greater number of targets. We alternatively represent it by a total of twenty six and fifty six targets. Some of these targets are useless since they face the sarcophagus and not the cameras. Once again we run the simulations for five positions of the target frame (the four corners and the center of the bounding box). Section (a) of table 3.8 shows the maximum inaccuracy values obtained spatially and angularly.

There is no improvement in the achieved tracking accuracy if only twenty six points are used to define the frame. This is due to the increase of the bounding box dimensions: the target frame occupies close to the same proportion of the image than the acquisition system did previously. However, when fifty six points are used to define the target frame, there is a tangible improvement in the achievable angular tracking accuracy. We can track this target frame with an angular accuracy of 0.124 mrad while our target angular accuracy is 0.128 mrad. Unfortunately, this improvement is not sufficient since these preliminary simulations only take

into account the picture coordinate inaccuracy. Future complete simulations will also take into account the noise on the object coordinates and on the tracking cameras' position and orientation. With these added values the tracking inaccuracy will necessarily exceed 0.128 mrad.

We now run the simulations with a larger target frame, a cube of side 0.5 m, represented successively by twenty six and fifty six points. Section (b) of table 3.8 shows that the achievable tracking accuracy when using fifty six targets seems compatible with our target accuracy. All the following sarcophagus simulations are thus performed on a tracking frame of dimensions a $0.5\,\mathrm{m} \times 0.5\,\mathrm{m} \times 0.5\,\mathrm{m}$ represented by fifty six targets in the five positions defined previously.

Modified calibration and data processing

The use of a target frame modifies the necessary calibration steps: we no longer calibrate the acquisition systems, since they are not covered with targets. Instead, we must calibrate the target frame and estimate the orientation between each acquisition system and the target frame. This could be done in an integrated setup as was the case for the calibration of the acquisition systems. However, we choose to calibrate the target frame independently since it will be used for several acquisition systems.

We introduce a new coordinate system $C_F, (O_F, \vec{x}_F, \vec{y}_F, \vec{z}_F)$, linked to the target frame.

Calibration of the target frame We take over one hundred images of the target frame surrounded by a scale bar and additional targets. We can then define C_F and know the position of each target in this coordinate system.

Orientation of the acquisition system with respect to the target frame To know the position of the acquisition system in the system defined by the target frame, we proceed in three steps:

1. Fix the acquisition system to the target frame.
2. Use the acquisition system to digitize another target-covered 3D object.
3. Take over fifty photos of the target frame and the additional 3D object.

We associate a coordinate system C_{temp} to the 3D object. Step 2 provides us with $T_{C_{S_i},C_{temp}}$, which describes the position and orientation of the 3D object in C_{S_i}. Similarly, step 3 provides us with $T_{C_F,C_{temp}}$, the position and orientation of the 3D object in C_F previously defined.

We can thus easily calculate the transformation between C_{S_i} and C_F:

$$T_{C_F,C_{S_i}} = T_{C_F,C_{temp}} \cdot \left(T_{C_{S_i},C_{temp}}\right)^{-1}.$$

The simultaneous tracking now provides us with T_{C_0,C_F} (we track the frame, and no longer the acquisition system). We now calculate $A|_{C_0}$, the coordinates of the surface points in the world system using:

$$A|_{C_0} = T_{C_0,C_F} \cdot T_{C_F,C_{S_i}} \cdot A|_{C_{S_i}} \ .$$

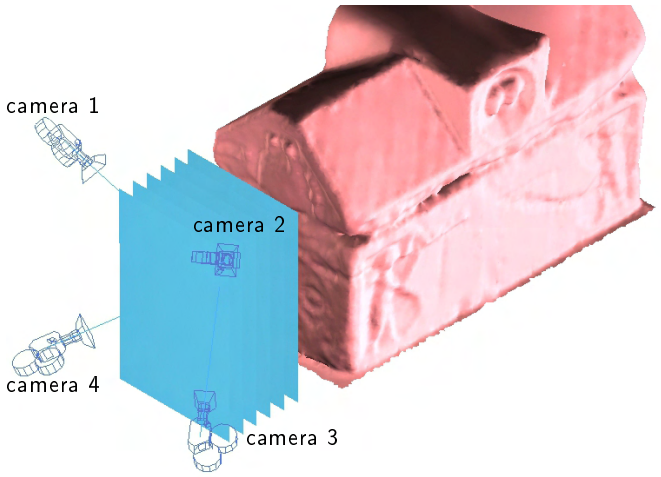


Figure 3.16: Overview of the orientation scene for the simulations.

3.2.5 Complete simulations

Up to now, the simulations were run with noise only on the picture coordinates. A thorough simulation must also take into account the inaccuracy of the object coordinates and the accuracy with which we know the position and orientation of the tracking cameras. For the object coordinate accuracy we will use the conservative estimate of 0.05 mm. The tracking camera exterior orientation accuracy greatly depends on the setup. We thus perform a set of simulations to estimate the camera translation and rotation values to use for the comprehensive simulations.

Orientation simulations

The estimation of the relative position and orientation of a group of cameras is performed by acquiring a series of simultaneous acquisitions of the same target-covered object. For the simulations, this target plane is modeled by a 0.8 m × 1 m plane represented by 99 points and positioned at 0.1 m intervals between the tracking cameras and the sarcophagus, as illustrated figure 3.16.

The orientation simulations are run with noise only on the picture coordinates. We assume the target plane is perfectly calibrated; the camera positions and orientations are our unknowns. We define two levels of noise: in the realistic scenario we assume that the picture coordinates are resolved with an accuracy of $1/10^{\text{th}}$ of a pixel; in the best-case scenario we assume the picture coordinates are resolved with an accuracy of $1/30^{\text{th}}$ of a pixel.

The resulting orientation accuracy is given in table 3.9. These results are used to define a realistic amount of noise to add to the camera translation and rotation. Since the orientation simulations result in an 0.023 mm spatial accuracy and 0.058 mrad angular accuracy in the realistic scenario, we will use 0.03 mm and 0.07 mrad respectively as input values. Similarly, we add a small margin to the best case scenario simulation results, yielding an input noise of 0.02 mm and 0.03 mrad.

Table 3.9: Orientation simulation results, sarcophagus configuration.

Input noise					Output accuracy		See table
Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
1/10	0.345	0.00	0.00	0.00	0.023	0.058	A.15
1/30	0.115	0.00	0.00	0.00	0.009	0.020	A.16

Table 3.10: Comprehensive simulation results, sarcophagus configuration.

Input noise					Output accuracy		See table
Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
1/10	0.345	0.05	0.03	0.07	0.023	0.176	A.17
1/30	0.115	0.05	0.02	0.03	0.016	0.116	A.18

Comprehensive simulations

It is now possible to simulate the complete tracking pipeline using the tracking frame defined in section 3.2.4 and the orientation parameters from the previous simulations. The comprehensive simulations are also run both with realistic and best-case scenario levels of noise for five positions of the target frame. The maximum output inaccuracy for each noise sequence is given table 3.10. In both cases we largely reach our target spatial accuracy of 0.09 mm but our target angular accuracy of 0.128 mrad is only attained in the best-case scenario.

3.2.6 Simulations with actual material characteristics

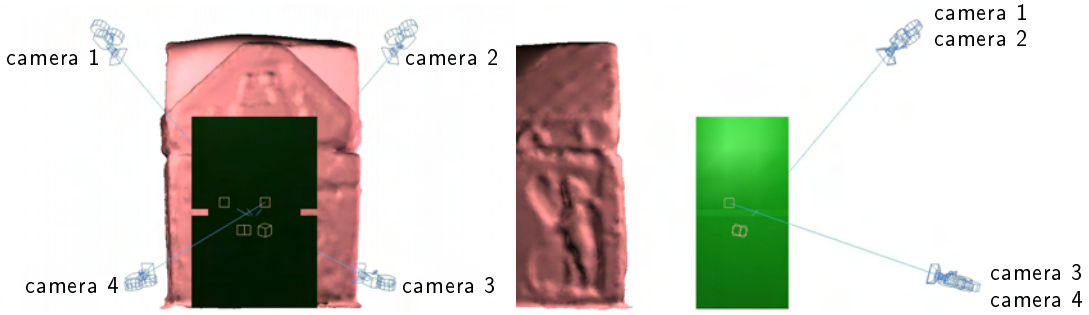
For practical reasons we bought AVT Stingray cameras instead of the Baumer cameras used during the simulations. The AVT camera sensors are only slightly bigger, as can be seen in table 3.11. There is a more significant difference between the lens used in the previous simulations and those bought for the experiments: no 6 mm C-Mount lenses were available for 2/3 " sensors so we bough 8 mm lenses (Pentax C814-5M). These camera and lens parameters are used for the remaining simulations.

We repeat the simulations to take into account these new parameters. Since the focal length of the lens is longer, the tracking cameras must be positioned slightly further from the target frame bounding box to maintain an equivalent view. Figure 3.17 shows this slightly modified configuration, which fits in the space available in front of the sarcophagus. Once again, we simulate tracking the 0.5 m × 0.5 m × 0.5 m frame covered with fifty six targets in five positions.

Section (a) of table 3.12 shows the achieved tracking accuracy both when we add noise only on the picture coordinates and when we use the comprehensive level of noise defined for the Baumer cameras with 6 mm lens. The achieved tracking accuracy is comparable to that obtained previously.

Table 3.11: Characteristics of the AVT Stingray cameras used for the tracking. Characteristics of the Baumer sensors used for the first simulations are given for comparison.

	Sensor size (W × H) (mm × mm)	Sensor size (W × H) (pixels × pixels)	Cell size (µm)	Ω_{horz}	Lens (mm)
Baumer TXF20	7.1 × 5.4	1624 × 1236	4.40	61.2°	6
Baumer TXF50	8.4 × 7.1	2448 × 2050	3.45	70.0°	6
AVT Stingray F-504B	8.5 × 7.1	2452 × 2056	3.45	56.0°	8



(a) Front view (b) Side view

Figure 3.17: Tracking camera and target frame positions for the simulations based on actual material characteristics. The five superimposed green boxes represent the five target frame positions. The yellow squares on the box represent the direction pointed at by the tracking cameras.

Moving the cameras away from the sarcophagus has the positive effect that there is more space available between the tracking cameras and the sarcophagus to perform the orientation. The orientation is now more accurate, as shown in section (b) of table 3.12. We use these output values to define the amount of noise to add to the orientation parameters for the new comprehensive simulations. The results of these comprehensive simulations is given in section (c) of table 3.12. The final tracking accuracy is slightly improved in the best-case scenario but we still do not reach our target angular tracking accuracy in the realistic scenario.

We have found a stable tracking setup that enables us to reach our tracking accuracy goal in the sarcophagus configuration. This setup uses four 5 Mpx cameras with 8 mm lens and a cubic target frame of side 0.5 m covered with fifty six targets. These cameras, lens and target frame will be used for the remaining simulations. Our setup also requires that we ensure a picture coordinate accuracy of $1/30^{\text{th}}$ of a pixel.

3.3 Painting configuration

The next simulations are based on the tracking of the target frame when the acquisition systems digitize the Rüdesheim wall paintings. The area of interest in now 2 m × 1.5 m but the target accuracy (table 3.5) does not change: our most constraining case is tracking the lab-designed

Table 3.12: Simulation results, sarcophagus configuration. Tracking based on actual material characteristics.

Setup	Input noise					Output accuracy		See table
	Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
	(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
(a)	1/10	0.345	0.00	0.00	0.00	0.015	0.104	A.19
	1/30	0.115	0.00	0.00	0.00	0.005	0.032	A.20
	1/10	0.345	0.05	0.03	0.07	0.028	0.198	A.21
	1/30	0.115	0.05	0.02	0.03	0.016	0.112	A.22
(b)	1/10	0.345	0.00	0.00	0.00	0.021	0.030	A.23
	1/30	0.115	0.00	0.00	0.00	0.007	0.010	A.24
(c)	1/10	0.345	0.05	0.03	0.04	0.024	0.172	A.25
	1/30	0.115	0.05	0.01	0.02	0.014	0.100	A.26

(a) Preliminary simulations.
(b) Orientation simulations.
(c) Comprehensive simulations.

multispectral camera, for which we must obtain an accuracy better than 0.09 mm spatially and 0.128 mrad angularly. The sensor characteristics are those of the Stingray cameras used in the experiments and the object we track is the target frame defined previously. This target frame moves in a plane that is 0.7 m from the painting.

3.3.1 Preliminary simulations

Tracking with four tracking cameras

We optimize the positioning of four cameras tracking the target frame in front of a 2 m × 1.5 m area. This time we can position the cameras at exactly 90° with respect to one another, without being bothered by the floor. The cameras are positioned at 2 m from the painting to fully observe the tracking volume.

As previously, the first simulations are performed with noise only on the picture coordinates. Even by varying the orientation of the cameras we are unable to reach the target angular accuracy with no noise on the object coordinates nor on the orientation parameters (see results table 3.13). Four cameras seem insufficient to track our target frame with the desired accuracy in this configuration. Because the area under study is quite large, it does not seem exaggerated to increase the number of tracking cameras. On the other hand, increasing the dimensions of the target frame would result in an object that is difficult to manipulate in situ.

Tracking with six cameras

We try to reach our target tracking accuracy using six cameras. Since we know that using more than four cameras to observe the same points does not greatly increase the tracking accuracy, we divide the bounding box in two overlapping reduced bounding boxes (see figure 3.18). Each reduced bounding box is now observed by three cameras.

Table 3.13: Preliminary simulation results, painting configuration. The bounding box is surveyed by four cameras. The camera positions vary slightly between each setup.

Setup	Input noise					Output accuracy		See table
	Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
	(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
(a)	1/10	0.345	0.00	0.00	0.00	0.028	0.196	A.28
(b)	1/10	0.345	0.00	0.00	0.00	0.026	0.186	A.29
(c)	1/10	0.345	0.00	0.00	0.00	0.029	0.214	A.30

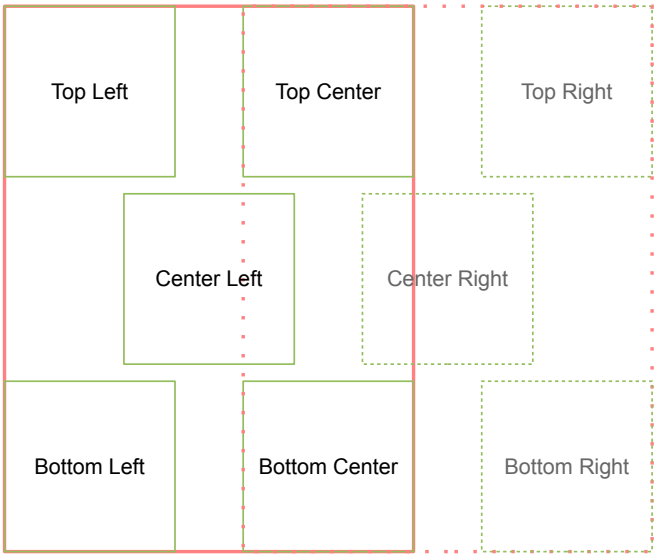


Figure 3.18: Reduced bounding boxes and target frame positions defined for the painting simulations.

Table 3.14: Preliminary simulation results, painting configuration. The reduced bounding box is surveyed by three cameras. The camera positions vary slightly between each setup.

Setup	Input noise					Output accuracy		See table
	Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
	(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
(a)	1/10	0.345	0.00	0.00	0.00	0.025	0.174	A.31
(b)	1/10	0.345	0.00	0.00	0.00	0.026	0.170	A.32
(c)	1/10	0.345	0.00	0.00	0.00	0.023	0.162	A.33
(d)	1/10	0.345	0.00	0.00	0.00	0.025	0.172	A.34
(e)	1/10	0.345	0.00	0.00	0.00	0.024	0.160	A.35
(f)	1/10	0.345	0.00	0.00	0.00	0.026	0.172	A.36

We first perform the simulations for five frame positions in one of the reduced bounding boxes surveyed by three tracking cameras. We slightly vary the position and orientation of these three cameras to optimize this configuration. The results are given table 3.14. The angular accuracy is slightly improved compared to the four camera setup: we reach 0.160 mrad in the best-case instead of 0.186 mrad previously.

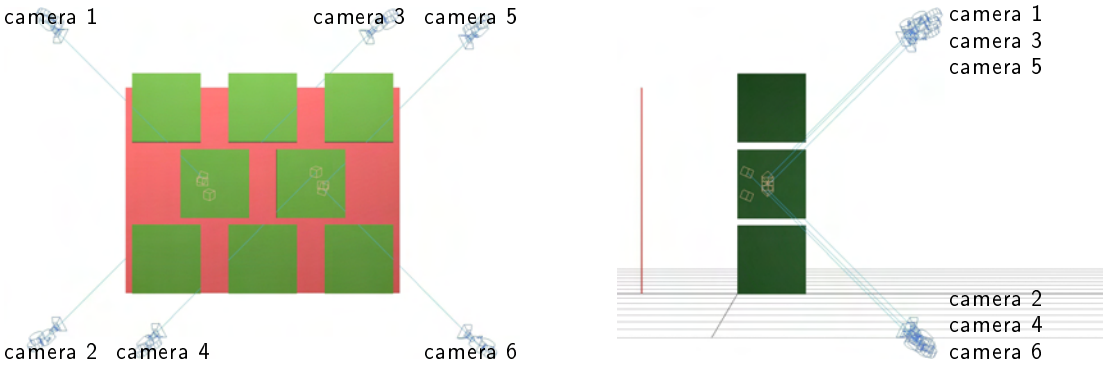
There is not much variation in the achieved accuracy among the different three camera arrangements. Since moving the cameras does not greatly modify the output accuracy, we have found a strong geometric configuration. The two best configurations are those described row (c) and (e). Row (c) we reach the best spatial accuracy and the second best angular accuracy while the second best spatial accuracy and the best angular accuracy are reached row (e). Since the bottleneck is reaching the target angular accuracy we choose the configuration corresponding to row (e) to be used as a basis for the full six camera configuration.

The next step is to perform the simulations with the other three cameras positioned symmetrically to observe the other sub-area of interest. We expect the simulation results to be improved by the presence of these other three cameras because the two sub-areas overlap. Figure 3.19 shows the position and orientation of all six cameras in front of the painting while figure 3.20 shows the view from these tracking cameras. We perform the simulations when the target frame is successively in all eight positions defined previously. The results are summarized in section (a) of table 3.15. There seems to be sufficient margin on the accuracy values to add the noise corresponding to a practical situation.

3.3.2 Complete simulations

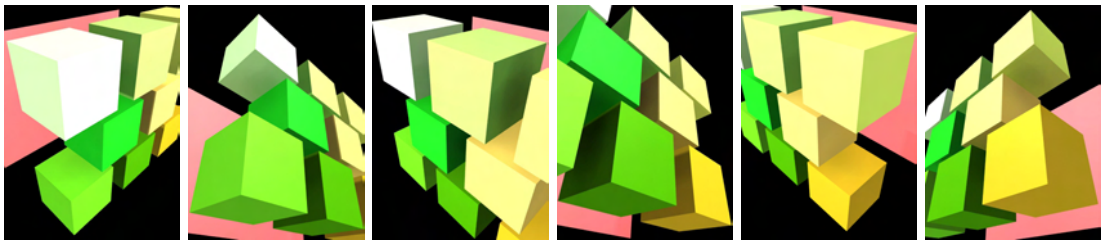
Orientation simulations

As in the sarcophagus configuration, we now try to evaluate how accurately we can obtain the tracking camera exterior orientation parameters. The procedure is the same as previously. Since the tracking area is larger, we use a 3 m × 2 m plane represented by 7 × 5 points. This plane is positioned in ten positions every 0.1 m between the cameras and the painting, starting at 0.5 m from the painting to avoid disturbing it (in practice we would use a smaller plane in more positions.) The simulations are run both with a picture coordinate accuracy $1/10^{\text{th}}$ of a



(a) Front view (b) Side view

Figure 3.19: Tracking cameras and the target frame positions in the painting configuration.



(a) Camera 1 (b) Camera 2 (c) Camera 3 (d) Camera 4 (e) Camera 5 (f) Camera 6

Figure 3.20: View of the target frame positions from the tracking cameras for the painting simulations.

Table 3.15: Simulations results, painting configuration. The bounding box is surveyed by six cameras.

Setup	Input noise					Output accuracy		See table
	Picture coordinates (pixel)	Picture coordinates (µm)	Object coordinates (mm)	Orientation translation (mm)	Orientation rotation (mrad)	Spatial (mm)	Angular (mrad)	
(a)	1/10	0.345	0.00	0.00	0.00	0.016	0.116	A.37
	1/30	0.115	0.00	0.00	0.00	0.005	0.038	A.38
(b)	1/10	0.345	0.00	0.00	0.00	0.052	0.050	A.39
	1/30	0.115	0.00	0.00	0.00	0.017	0.016	A.40
(c)	1/10	0.345	0.05	0.06	0.06	0.032	0.216	A.43
	1/30	0.115	0.05	0.02	0.02	0.015	0.106	A.45

(a) Preliminary simulations.
(b) Orientation simulations.
(c) Complete simulations.

pixel and with a picture coordinate accuracy $1/30^{\text{th}}$ of a pixel. The results (given in section (b) of table 3.15) are roughly linear: increasing the picture coordinate accuracy three times, we can evaluate the position and orientation of the cameras three times more accurately.

To determine a realistic amount of noise to add to the orientation parameters of the next simulations, we add a small margin to the highest values obtained. In the case of an accuracy of $1/10^{\text{th}}$ of a pixel on the picture coordinates the maximum values are 0.052 mm spatially and 0.050 mm angularly. Thus, adding a Gaussian noise of 0.06 mm standard deviation on the orientation translation and 0.06 mrad on the orientation rotation seems reasonable. These values can be lowered to 0.02 mm and 0.02 mrad if we assume the picture coordinates accuracy is $1/30^{\text{th}}$ of a pixel.

Comprehensive simulations

The noise parameters previously defined are applied to the six camera configuration for eight positions of the cube. The full results are summarized in section (c) of table 3.15. There is no difficulty in attaining the target spatial accuracy: our most constraining goal is 0.09 mm while the worst achieved accuracy is 0.027 mm. However, to ensure that we reach our angular accuracy goal of 0.128 mrad we must once again ensure a pixel precision of $1/30^{\text{th}}$ of a pixel.

3.4 Statue configuration

We now assume the object under study is a 1 m high statue with a radius of approximately 0.3 m. The distance between the object and the acquisition systems is still 0.70 m to 0.76 m. The target accuracy stays the same, the constraining target value is 0.128 mrad angular accuracy. We still assume the object we track is the previously defined tracking frame. With the acquisition system in the center of the frame, this yields a distance of at least 0.5 m between the frame and the object under study. This is a reasonable stand-off distance when studying cultural heritage objects.

To observe the full statue, the acquisition system must be positioned all around the statue at different heights, and observe it from above and below. We define five heights for these simulations numbered row 0 to row 4, bottom up. This column of positions is wrapped around the statue in eight positions, as illustrated figure 3.21. The global size of the tracking volume can be enclosed in a box of dimensions 2.3 m × 2.3 m × 2.5 m.

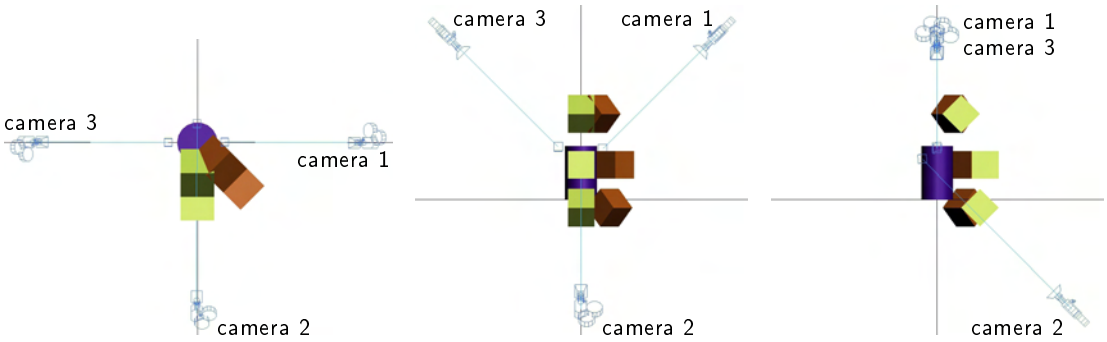
3.4.1 Preliminary simulations

Tracking with four cameras

We first base the tracking on four cameras placed at 90° angles. We take advantage of the symmetry of the configuration to simplify the first simulations: we only consider two of the eight columns and three of the four cameras. Furthermore we consider only three rows per column. This simplified configuration is illustrated figure 3.22. The view of from these cameras is shown figure 3.23.

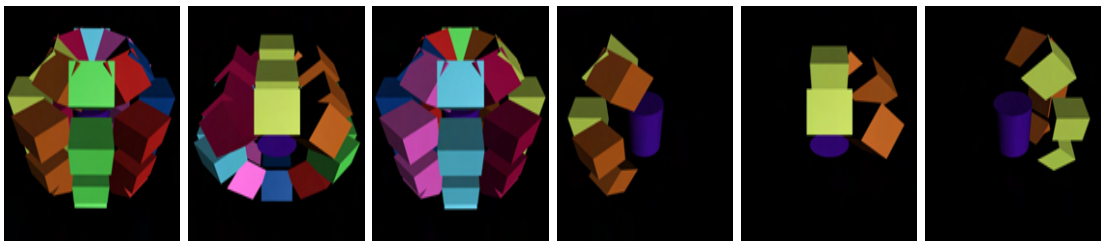
(a) (b) (c)

Figure 3.21: Target frame positions defined for the statue simulations. Purple cylinder represents the statue. Row 2 is not represented, it is in the middle of rows 1 and 3. (a) Side view of a single column. (b) Top view of all columns. (c) Perspective view of all columns.



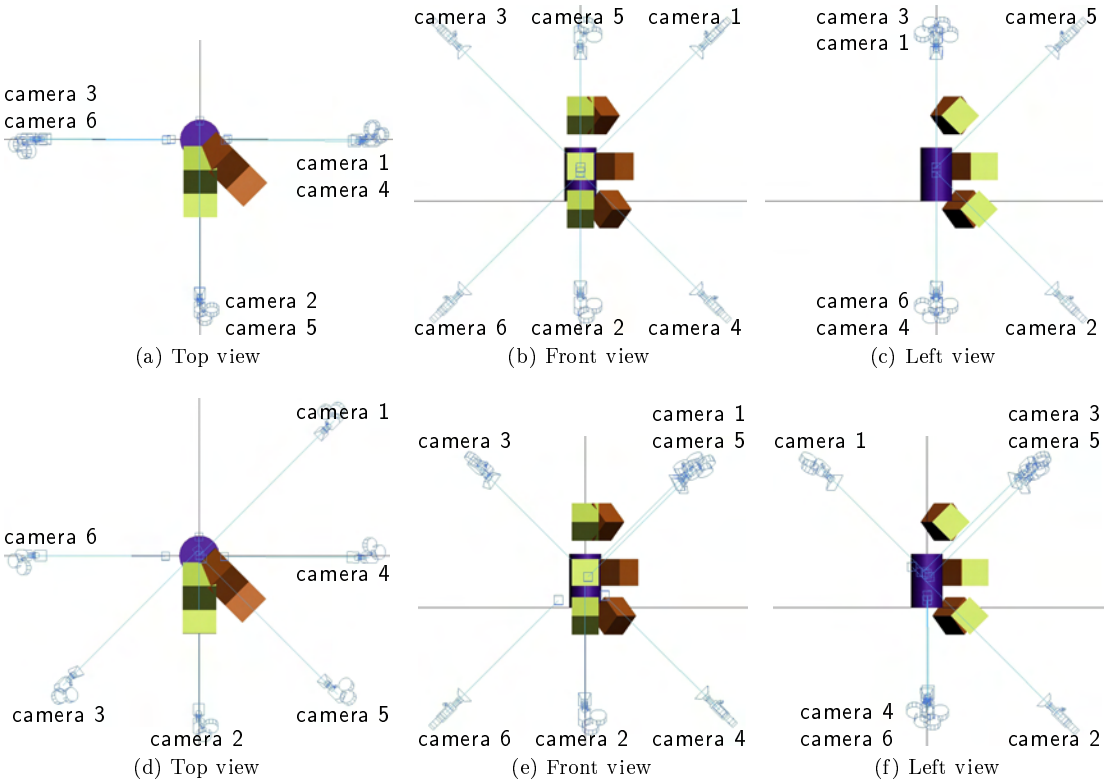
(a) Top view (b) Front view (c) Side view

Figure 3.22: Subset of cameras and target frame positions used in the simulations when the full volume of interest is surveyed by four cameras.



(a) Camera 1 (b) Camera 2 (c) Camera 3 (d) Camera 1 (e) Camera 2 (f) Camera 3

Figure 3.23: View of target frame positions from the tracking cameras when the full volume of interest is surveyed by four cameras. (a-c) All target frame positions. (d-e) Subset used in simulations.



(a) Top view (b) Front view (c) Left view

(d) Top view (e) Front view (f) Left view

Figure 3.24: Subset of six cameras and six target frame positions used in the simulations when the full volume of interest is surveyed by eight cameras. (a-c) Cameras are vertically aligned two by two. (d-f) Cameras are staggered at 45° angles.

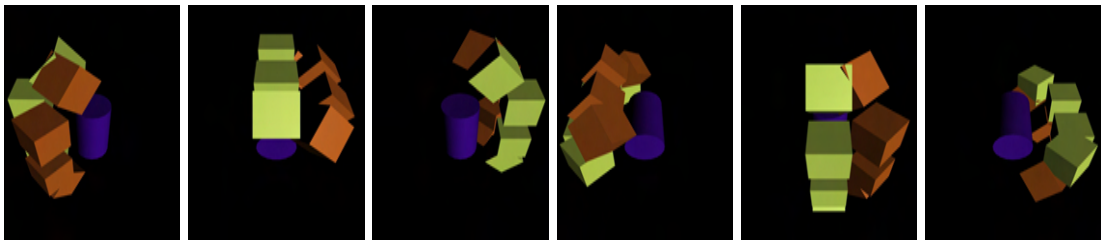
For these preliminary simulations we only add noise to the picture coordinates. Even with a low amount of noise ($1/10^{\text{th}}$ of a pixel), we can only track the target frame with an accuracy of 0.056 mm spatially and 0.308 mrad angularly (see table 3.16, row (a)). The angular accuracy is more than two times worse than than our target value of 0.128 mrad. It is not surprising that the accuracy reached is very low since many frame faces are seen only by a single camera. For example all the bottom faces are seen only by camera 2. Four cameras are insufficient to track the target cube in this configuration.

Tracking with eight cameras

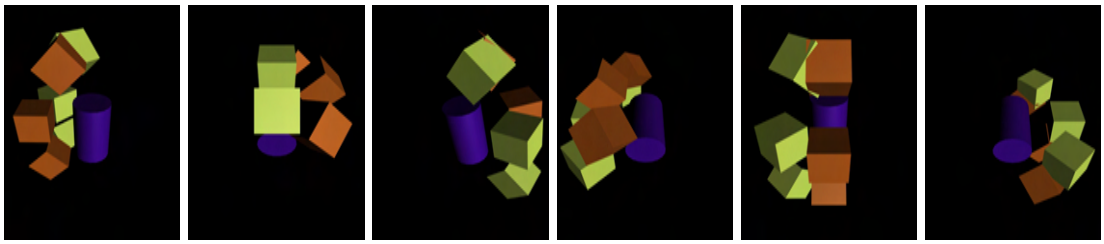
We double the total number of tracking cameras to eight. Four cameras look down from above the statue and four look up from below. Two different configurations are tested: either the top and bottom cameras are vertically aligned, or they are staggered at 45° angles. Only the position of the top cameras (number 1, 3 and 5) differ between the two camera arrangements.

As previously, we take advantage of the symmetry of the configuration: we use a subset of six cameras and the same six acquisition positions as before. The camera positions are illustrated figure 3.24 and the view from the cameras is shown figure 3.25.

We run the simulations with noise only on the picture coordinates. The results show that



(a) Camera 1 (b) Camera 2 (c) Camera 3 (d) Camera 4 (e) Camera 5 (f) Camera 6



(g) Camera 1 (h) Camera 2 (i) Camera 3 (j) Camera 4 (k) Camera 5 (l) Camera 6

Figure 3.25: View of target frame positions from the tracking cameras defined for the simulations when the full volume of interest is covered by eight cameras. Each row presents a variation of the camera arrangement. (a-f) Cameras are vertically aligned two by two. (g-l) Cameras are staggered at 45° angles.

the two configurations are equivalent (see rows (b) and (c) of table 3.16). We thus settle on the setup with the vertically aligned cameras, since it will be easier to build in practice.

3.4.2 Complete simulations

Orientation parameters

Because of the opposite positions of the cameras and the fact that there is an object in the center, all eight cameras can not be oriented in a single pass: they can not all view the same plane. A suitable approach is to calibrate them in overlapping pairs of four, considering at a given moment only two neighbors from the top and the two corresponding bottom cameras.

This is somewhat equivalent to the orientation procedure for the painting configuration where half of the cameras point towards the left and the other half towards the right. On one hand the real results should be less accurate since there is less space available between the cameras and object to perform the orientation. On the other hand each camera will be oriented twice, improving the accuracy of the results. We thus use the parameters defined by the painting orientation simulations for the statue simulations. These values are 0.06 mm and 0.06 mrad for the realistic scenario (picture coordinate accuracy of $1/10^{\text{th}}$ of a pixel) and 0.02 mm and 0.02 mrad for the best-case scenario (picture coordinate accuracy of $1/30^{\text{th}}$ of a pixel).

Table 3.16: Simulation results, statue configuration.

Setup	Input noise					Output accuracy		See table
	Picture coordinates		Object coordinates	Orientation translation	Orientation rotation	Spatial	Angular	
	(pixel)	(µm)	(mm)	(mm)	(mrad)	(mm)	(mrad)	
(a)	1/10	0.345	0.00	0.00	0.00	0.056	0.308	A.47
(b)	1/10	0.345	0.00	0.00	0.00	0.031	0.210	A.48
(c)	1/10	0.345	0.00	0.00	0.00	0.031	0.196	A.49
(d)	1/10	0.345	0.05	0.06	0.06	0.057	0.362	A.51
(e)	1/30	0.115	0.05	0.02	0.02	0.019	0.122	A.52

(a) Four cameras are used to track the full scene, only three for this subset.
(b) Eight cameras are used to track the full scene, only six are used in the simulations.
(c) Same as (b) with cameras vertically aligned instead of staggered.
(d) Same as (c) with realistic noise.
(e) Same as (c) with best-case scenario noise.

Comprehensive simulations

These simulations are still performed with a subset of six out of eight cameras but realistic noise is applied to all positions considered. The last section of table 3.16 shows that we reach our target accuracy in the best-case scenario (row (e)) but not in the realistic scenario (row (d)).

Once again, the angular accuracy goal is the the hardest to reach. As previously, we must thus ensure a pixel accuracy of $1/30^{\text{th}}$ of a pixel. Though we have managed to design a configuration where we reach our target accuracy, it must be noted that this configuration will be harder to set up in practice than the sarcophagus or wall painting configurations. Eight cameras are a lot to manage, their presence may disturb the surrounding artwork, and will complicate the manipulation of the target frame and acquisition systems.

3.5 Conclusion

This chapter described the photogrammetric tracking procedure developed for the registration of multimodal datasets. The potential accuracy of such procedure was explored with a wide range of simulations. Material parameters were based on the available acquisition systems while the object parameters were defined by three case studies. Two of these case studies were based on existing objects we are interested in. The constraints posed by these setups were taken into account in the choice and positioning of the tracking cameras. These simulations were also used to evaluate the characteristics of the tracking material best adapted to these three configurations and to determine a coherent amount of noise to apply to the final simulations.

The first simulations showed the advantages of using 5 Mpx cameras instead of 2 Mpx cameras. These first simulations also established that the acquisition systems were too small to be accurately tracked as is. A second batch of simulations showed that fixing a target frame to the acquisition systems improved the achievable tracking accuracy. This target frame was defined to be a cube of side 0.5 m covered with fifty six targets.

The following simulations were run with the real material parameters based on the sensors and lens bought for the experimental tracking. When the acquisition systems are fixed to the target frame and survey an area of $0.4\,\mathrm{m} \times 0.7\,\mathrm{m}$, the simulations show that they can be tracked with a spatial accuracy of $0.020\,\mathrm{mm}$ and an angular accuracy of $0.100\,\mathrm{mrad}$ by four cameras. This would enable us to digitize the area of interest of the Trier sarcophagus and register the data from our acquisition systems with an accuracy better than half an image pixel.

Equivalent tracking accuracy can be obtained using six cameras when the acquisition systems digitize an area of $2\,\mathrm{m} \times 1.5\,\mathrm{m}$. If the object under study is a $1\,\mathrm{m}$ high statue with a $0.3\,\mathrm{m}$ radius, then the target frame can be tracked with comparable spatial accuracy and an angular accuracy of $0.122\,\mathrm{mrad}$. This, however, requires eight tracking cameras and would be harder to achieve in practice.

These promising results are based on assumption that we can resolve the picture coordinates with an accuracy of $1/30^{\mathrm{th}}$ of a pixel. This entails that our camera calibrations will have to be very accurate and stable. We must also make sure our tracking cameras acquire images with well resolved, sharp targets with good contrast. These experimental details are explored thoroughly in the next chapter, where we validate the simulation results and use the tracking protocol to register 3D datasets.

Chapter 4

Tracking Accuracy and 3D Registration

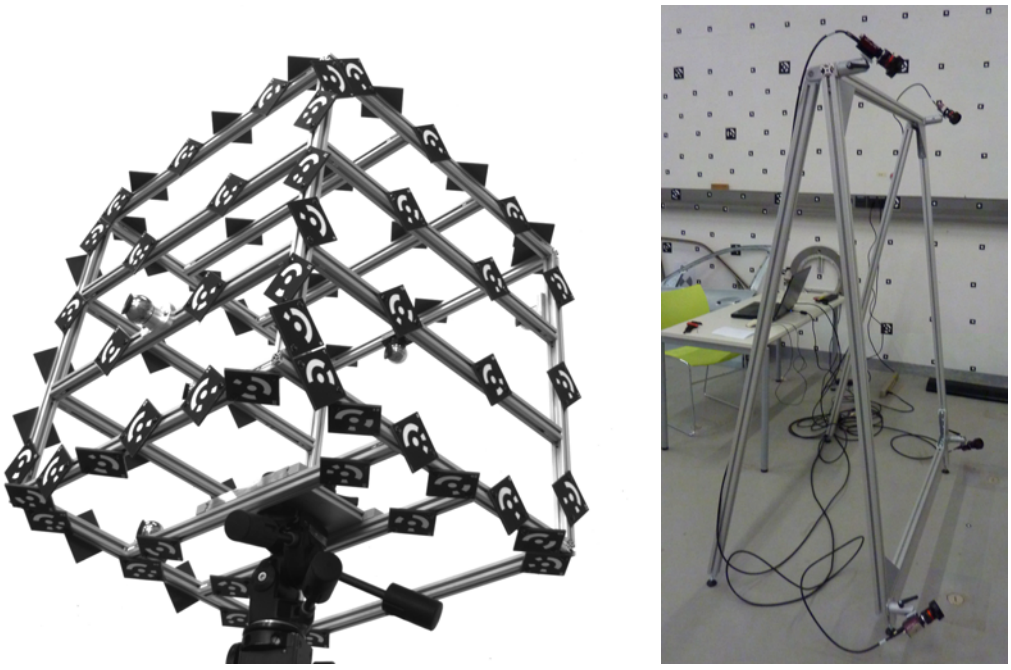
In this chapter we present a series of laboratory experiments used to evaluate our method. These experiments were performed both to validate the simulation results and to assess the accuracy of our technique to register a set of 3D meshes. After quickly presenting the design choices made to set up an experiment that matches the simulation configurations, we evaluate the achievable accuracy by simultaneously tracking the target frame with a commercial laser tracker and using our method. In section 4.2 we evaluate the accuracy of the registration of thirteen 3D meshes representing a car door. The digitization of this smooth surface is performed while tracking our target frame, with the Gom Atos III attached to it using four cameras, as in the sarcophagus configuration. In the following section we evaluate the stability of each component of the full setup. Finally, we assess the accuracy of the tracking achieved in the wall-painting configuration, using six cameras.

Practical considerations

Part of the tracking material has already been described in the previous chapter. The tracking cameras are 5 Mpx AVT Stingray cameras used with 8 mm Pentax lenses. The target frame is defined in the simulations as a cube of side 500 mm uniformly covered by fifty six targets. In practice, different target frames made of aluminum profiles were used throughout the experiments. These frames were covered with 48 to 80 targets and are described in greater detail in appendix B.1, page 167. The final prototype is illustrated in figure 4.1, as well as the system used to fix the tracking cameras in the positions described by the simulations.

4.1 Laser tracker measurements

In addition to the tracking parameters, the processing of our acquisition and calibration data provides us with internal accuracy measurements. Though these values are a good indication of the tracking accuracy, they are not sufficient to validate our method. We thus simultaneously



(a) (b)

Figure 4.1: Target frame (a) and tracking camera supports (b).

track our target frame with a laser tracker, which will provide the ground truth, and our method.

The laser tracker we use is a Leica LTD500. The targets used to reflect the laser beam back to the base are Corner Cube Reflector (CCR) targets, shown figure 4.2 in a magnetic target holder. We fix five of these CCR targets to the target frame in addition to the photogrammetric targets and perform the simultaneous tracking of the frame with the photogrammetric cameras and the laser tracker. Holes are drilled in the target holders so they can be securely screwed to the target frame.

The laser tracker can be used in two modes: either it follows a single CCR target that is positioned successively in all sockets, or we use several CCR targets that stay in place and we manually point the laser tracker to the successive targets. The first mode provides a positional accuracy of 0.003 mm at $1\,\sigma$ (based on our own measurements given in appendix B.2.1, page 170). The second setup provides a positional accuracy of approximately 0.020 mm.

We performed three tracking experiments with simultaneous laser tracker measurements. All these experiments were run in the four camera configuration by tracking an empty target frame, with no acquisition system in it. The reduced acquisition protocol is composed of the following steps:

- Tracking cameras interior orientation (performed a few days earlier).
- Target frame calibration.
- Tracking cameras exterior orientation.
- Simultaneous tracking of the target frame with the laser tracker and the tracking cameras.

Figure 4.2: CCR target used with the laser tracker.

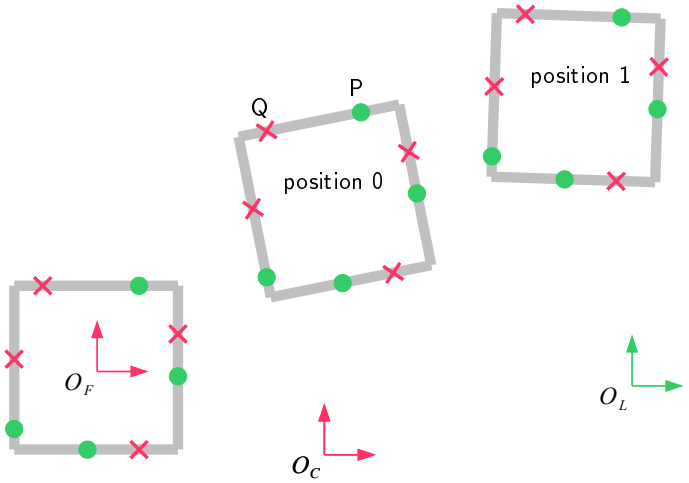


Figure 4.3: Simultaneous tracking of the target frame using our method and a laser tracker. The red crosses represent the photogrammetric targets and the green dots the CCR targets.

4.1.1 Laser tracker data processing

Comparing the data provided by the photogrammetric tracking and by the laser tracker is not a straightforward task. Unfortunately, the coordinates of the photogrammetric targets, noted Q, are available only in the target frame coordinate system C_F while the CCR targets' coordinates, noted P, are available only in the laser tracker coordinate system $C_L, (O_L, \vec{x}_L, \vec{y}_L, \vec{z}_L)$, as illustrated figure 4.3. For each acquisition, the laser tracker provides us with the coordinates of the CCR targets in the laser tracker coordinate system, while the photogrammetric tracking provides us with the transformation matrix between the target frame coordinate system C_F and the position of the target frame.

Before comparing the data, we start by surveying the accuracy of the photogrammetric tracking, given by the internal accuracy of the bundle adjustment. As in the previous chapter, these values are provided at $2\,\sigma$. We also evaluate the accuracy and the stability of the laser tracker data both by calculating the point to point distance between all targets for every acquisition, and by evaluating the accuracy of the transformation matrix estimated between

every pair of positions, as explained below.

Point to point distance

The laser tracker provides us with $P_k|_{Li}$, the coordinates of CCR target k ($k \in [\![1, K]\!]$) in the laser tracker coordinate system for position i ($i \in [\![1, I]\!]$) expressed in homogeneous coordinates. $K = 5$ for our acquisitions. The relative position of the CCR targets on the target frame is fixed during every series of acquisitions. This means the distance between any two CCR targets of the target frame should be the same for every acquisition. We calculate the difference between every pair of target points $(k, l) \in [\![1, K]\!]^2$ for every acquisition i:

$$P_k|_{Li} - P_l|_{Li} \ . \tag{4.1}$$

This distance should be constant when i varies for a fixed pair of points (k, l) so we also calculate its standard deviation. We expect this standard deviation to be less than twice the laser tracker punctual accuracy. If it is higher, it means the targets have moved with respect to one another of a distance that is detectable with the laser tracker.

Transformation accuracy

For every pair of positions (i, j) we calculate $T_{Li,Lj}$ such that

$$P_k|_{Li} = T_{Li,Lj} \cdot P_k|_{Lj} \tag{4.2}$$

for every point P_k using a Singular Value Decomposition (SVD). Then, we calculate the average of the norm of the difference vector between the measured points and the points calculated using the estimated transformation matrix:

$$\frac{1}{K} \cdot \sum_{k=1}^{K} \left\| P_k|_{Li} - T_{Li,Lj} \cdot P_k|_{Lj} \right\|_2 . \tag{4.3}$$

This result can be interpreted as a distance, and tends towards zero as the accuracy of the laser tracker measurements increases.

Comparing the photogrammetric and laser tracker measurements

The simultaneous tracking of our target frame with a laser tracker and our photogrammetric setup provides us with the following data:

- The coordinates of the CCR targets for each acquisition position, $P_k|_{Li}$, which is our ground truth.
- The transformation matrix between each of these positions, $T_{Li,Lj}$, which we estimate from the ground truth.
- The transformation matrix between the target frame in its coordinate system and each position of the target frame measured by our photogrammetric setup, $T_{Ci,F}$.

Since the laser tracker data and the photogrammetric measurements are not in the same coordinate system we can not compare them easily. It is possible to calculate $T_{Ci,Cj}$ from the various $T_{Ci,F}$, but this is different from $T_{Li,Lj}$.

For every acquisition i we have:

$$P_k|_{Li} = T_{Li,Ci} \cdot T_{Ci,F} \cdot P_k|_F \ ,$$

where $P_k|_F$ are the coordinates of CCR target k in the frame coordinate system and $T_{Li,Ci}$ is the transformation between the tracking cameras coordinate system and laser tracker system. Since the transformation between these two systems is fixed, $T_{Li,Ci} = T_{Lj,Cj} = T_{L,C}$ for every pair of positions (i, j).

Using all target positions P_k and a variety of $T_{Ci,F}$ it is possible to estimate $T_{L,C}$ and $P_k|_F$ using a least-squares fitting. We have 27 unknowns: 12 for $T_{L,C}$ (this is a 4×4 matrix with the last row equal to [0 0 0 1]) and 3×5 for the coordinates of $P_k|_F$. Each position provides with 3×5 equations so theoretically two positions are sufficient to estimate $T_{L,C}$ and $P_k|_F$. In practice, we need at least three positions to obtain a matrix of sufficient rank.

We can use all available positions to estimate the coordinates of the CCR targets in the target frame coordinate system. This simply enables us to check that the estimated CCR target coordinates are consistent with their position on the frame. However, we can not use the same $T_{Ci,F}$ values to estimate the parameters and to evaluate the tracking accuracy. We have to partition our dataset and use a subset of the values of $T_{Ci,F}$ to estimate $T_{L,C}$ and $P_k|_F$ and the rest to evaluate the accuracy of the tracking.

We use all but one position (position i) to estimate $T_{L,C}$ and $P_k|_F$. To compare the photogrammetric tracking results with the laser tracker results we calculate $P_k|_{Ci}$ in two different ways. Using the laser tracker data and estimated $T_{L,C}$ we have:

$$P_k|_{Ci} = (T_{L,C})^{-1} \cdot P_k|_{Li} \ . \tag{4.4}$$

On the other hand, the photogrammetric tracking results and the estimated $P_k|_F$ enable us to calculate

$$P_k|_{Ci} = T_{Ci,F} \cdot P_k|_F \ . \tag{4.5}$$

These two values should be equal, with equation 4.4 representing the ground truth from the laser tracker and 4.5 representing the measured values. We estimate the accuracy of the photogrammetric tracking for target k and position i with:

$$\left\| (T_{L,C})^{-1} \cdot P_k|_{Li} - T_{Ci,F} \cdot P_k|_F \right\|_2 \ . \tag{4.6}$$

4.1.2 Target frame fixed to a theodolite tripod

For these first laser tracker measurements we used a target frame covered with 40 targets (target frame A from appendix B.1) and fixed it to a theodolite tripod. The target frame is tracked in four successive positions which correspond to the four corners of the bounding box in the sarcophagus configuration: two positions are close to the floor, while the other two are

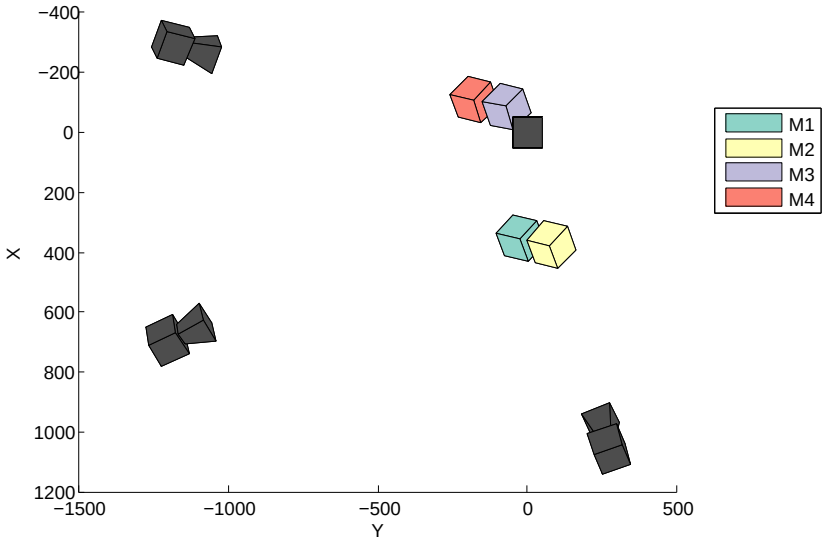


Figure 4.4: Relative position of the target frame (M1 to M4) and tracking cameras (dark grey) during the laser tracker measurements with the target frame fixed to a theodolite tripod.

approximately 50 cm higher (see figure 4.4). The five CCR targets are attached to the side of the target frame that faces the laser tracker, opposite from the tracking cameras.

Photogrammetric tracking

The spatial and angular internal accuracy of the phogrammetric tracking for all four positions is illustrated figure 4.5. It is compared to the simulation results and to the target accuracy values. We notice that the spatial accuracy is much worse than the simulation results, even in the less constraining "realistic" case. However, for all three acquisition systems considered the spatial accuracy is better than the target value. The angular accuracy, however, is not only worse than the simulation results, but also worse than the target values for all acquisition systems and all positions. These disappointing results prompted us to change the target frame design.

Laser tracker acquisitions

We evaluate the stability of the target frame from the laser tracker data by calculating the point to point distance between the five targets for all positions. The standard deviation of this point to point distance for each pair of targets is given row (a) of table 4.1 (all distances are given in appendix B.2.3, page 171). The cells of this table are colored based on a linear RGB scale that goes from green for a standard deviation of 0 mm to magenta for values greater than 0.04 mm. 0.04 mm is twice the positional accuracy of the laser tracker when it is used with fixed CCR targets.

In this first configuration, the standard deviation of the point to point distance between every pair of CCR targets varies between 0.005 mm and 0.259 mm. Since we use the laser tracker by repositioning a single CCR target, we expect the standard deviation to be less than 0.006 mm,

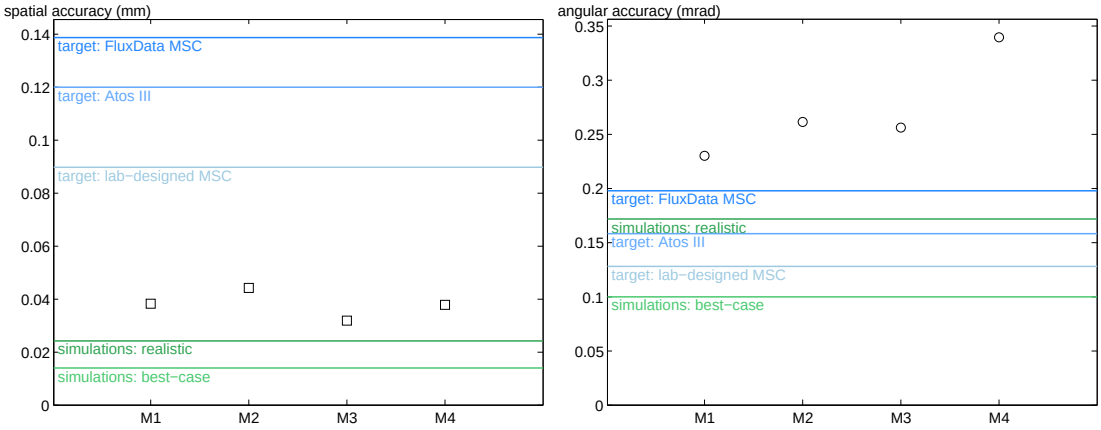


Figure 4.5: Spatial and angular accuracy of the photogrammetric tracking during the laser tracker measurements with the target frame fixed to a theodolite tripod, compared to the simulation results and target accuracy.

Table 4.1: Standard deviation of the point to point distances between every pair of CCR targets for all laser tracker measurements. Values greater than 0.040 mm are colored magenta.

	Standard deviation of point to point distance (mm)										See
	1 to 2	1 to 3	1 to 4	1 to 5	2 to 3	2 to 4	2 to 5	3 to 4	3 to 5	4 to 5	table
(a)	0.005	0.258	0.026	0.032	0.254	0.029	0.022	0.006	0.150	0.008	B.5
(b)	0.460	0.018	0.162	0.339	0.398	0.314	0.139	0.197	0.387	0.508	B.6
(c)	0.009	0.019	0.014	0.035	0.012	0.013	0.030	0.037	0.042	0.023	B.7

(a) Target frame fixed to theodolite tripod, four positions.
(b) Target frame fixed to marble table, five positions.
(c) Target frame with fixed CCR target, ten positions.

which is the case only between two pairs of targets. For all other pairs, the standard deviation is much greater. This means the distance between the CCR targets varies noticeably. During the measurements, we could feel the target frame vibrate when we detached and reattached the CCR targets to its magnetic holders. Though we waited for the vibrations to die out before each measure, they hurt the accuracy of the laser tracker measurements.

If this data is to be used as the ground truth against which we compare our photogrammetric tracking, it should be more accurate than the photogrammetric tracking data. The spatial accuracy of the photogrammetric tracking is better than 0.05 mm. A comparable accuracy of the laser tracker is represented by a standard deviation of less than 0.10 mm, which is reached for only three positions. This laser tracker data can not be used as a ground truth for evaluation of our photogrammetric tracking.

4.1.3 Target frame fixed to a marble table

To remedy the vibrations of the target frame, we clamp the frame to a stable marble table (see figure 4.6). We also use an improved target frame, which is much lighter and covered with 78 targets instead of 40 (target frame B2). This target frame is tracked in five positions, illustrated figure 4.7. There is not much diversity in these positions since we are limited by the height and rail of the table. We vary the inclination of the target frame by setting it on wooden blocks of varying heights before clamping.

Photogrammetric tracking

The tracking accuracy provided by our method is shown figure 4.8. We have hardly improved the tracking accuracy compared to the previous configuration: once again the spatial accuracy is sufficient for the registration purposes, even though it is worse than the values expected from the simulations. The angular accuracy is acceptable for the registration of data from the FluxData multispectral camera for only two of the five positions (M3 and M4). These two positions are strikingly better than the other three: the angular tracking accuracy is close to two times better. Nothing obvious differentiates these two positions from the lot, except the fact that they are very close to one another (see figure 4.7).

Looking at the photo acquisitions from the tracking cameras, it appears clear that the globally disappointing tracking accuracy is due to the fact that the marble table hides many of the targets from the bottom two cameras. Though we have added close to 40 targets to the target frame since the previous measurements, the marble table hides approximately the same amount. The two positions with the best tracking accuracy are the positions where the target frame slightly protrudes from the table (see figure 4.9). The difference seems small but it also appears to highly influence the tracking accuracy. This is not due to the increase in targets visible from the tracking cameras, since these values are very close (as shown in the appendix, table B.3). The accuracy of the target detection in the input images, however, is better for positions M3 and M4 (shown in the appendix, table B.4). This is not the single factor that explains the improved accuracy, or position M5 would be more accurately tracked. However, this is the factor which seems to have the single highest contribution.

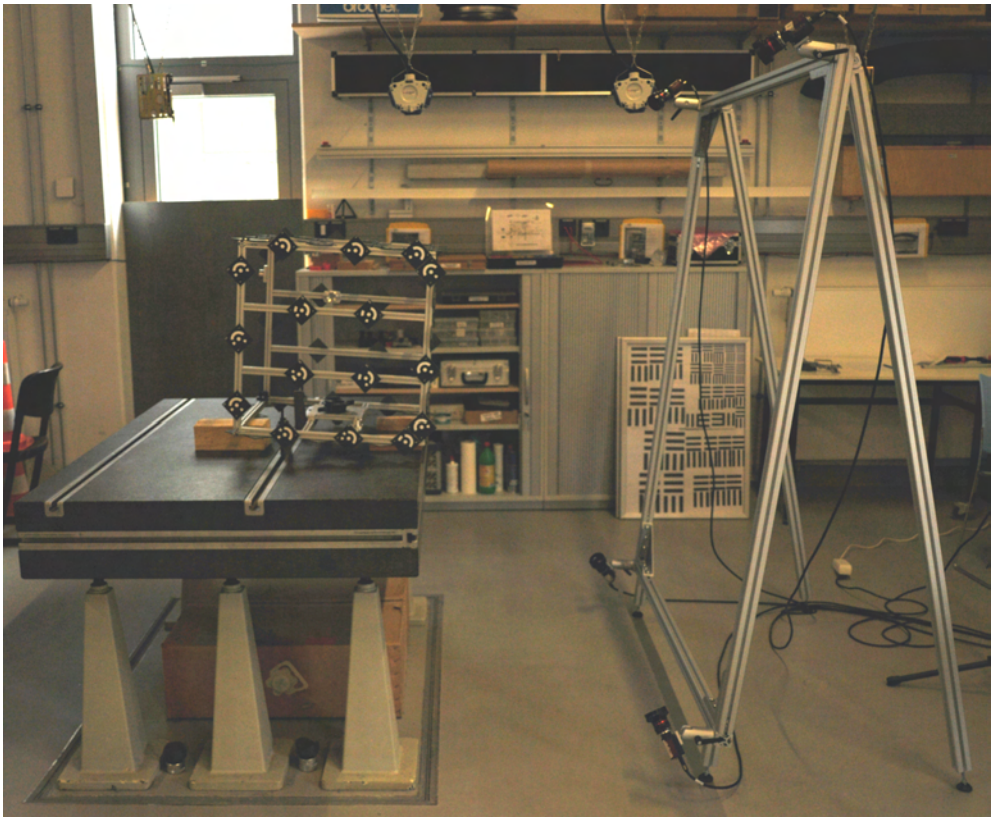


Figure 4.6: Target frame clamped to a marble table.

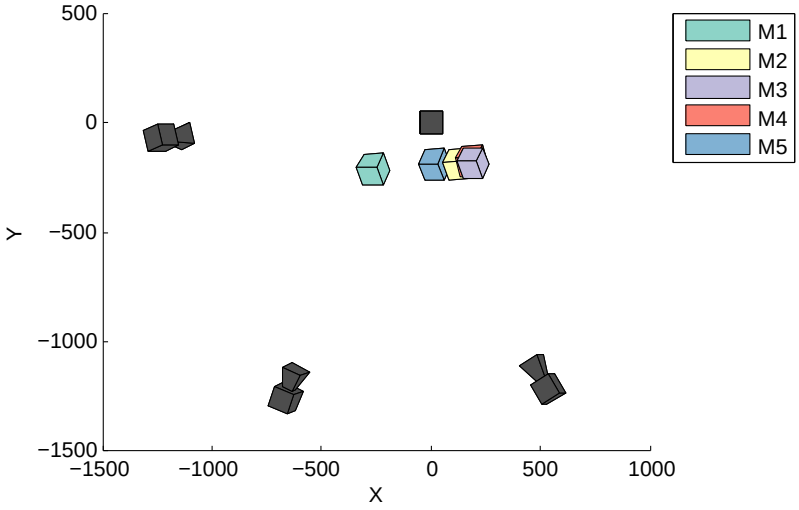


Figure 4.7: Relative position of target frame and tracking cameras during the laser tracker measurements with the target frame clamped to a marble table.

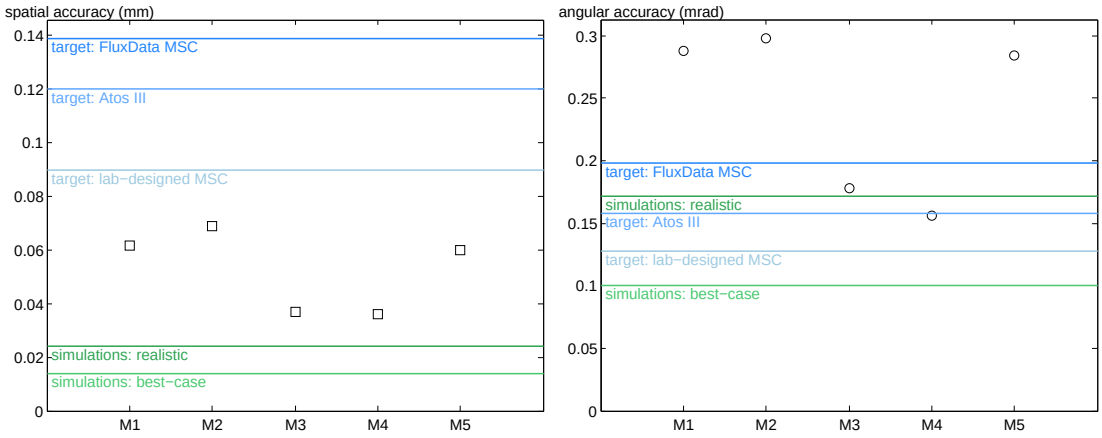


Figure 4.8: Spatial and angular accuracy of the photogrammetric tracking during the laser tracker measurements with the target frame clamped to a marble table, compared to the simulation results and target accuracy.

There are at least two possible explanations for the variation in pixel accuracy: First, the "limited" depth of view of the tracking cameras can result in the inadequate positioning of these cameras with respect to the marble table, given their fixed focus. This drawback is due to a change in calibration procedure: instead of performing the interior and exterior calibration in a single run (as was the case for the first experiment), all four cameras were calibrated beforehand and their focus fixed for the duration of the series of experiments. The exterior orientation was then calibrated using the known interior orientation. This procedure can greatly improve the accuracy of the exterior and interior tracking camera calibration. In turn, it can improve the tracking accuracy, though it requires the object under study to be correctly positioned with respect to the predefined focus. It is also possible that the illumination conditions were more favorable for measurements M3 and M4, resulting in sharper targets.

Laser tracker acquisitions

We expect the accuracy of the laser tracker measurements to be improved by this setup. Instead, row (b) of table 4.1 shows us that it is much worse than previously. The standard deviation is now between 0.139 mm and 0.508 mm for all but one pair of targets. Clamping the target frame to the table has bent the frame, causing a deformation of the cube and a variation in the distance between the CCR targets. This laser tracker data can not be used as a ground truth to evaluate the accuracy of the photogrammetric tracking.

4.1.4 Target frame with fixed CCR targets

To remedy the vibration and deformation problem, we use the laser tracker with five CCR targets that are fixed to the target frame and remain in place for the duration of the measurements. With this setup the accuracy of the laser tracker measurements is reduced, but the stability is improved. We use target frame B3, which is covered with 80 photogrammetic targets. The

(a) (b)

Figure 4.9: Target frame on the marble table seen from a bottom tracking camera. (a) representative of the badly tracked positions M1, M2 and M5. (b) representative of better tracked positions M3 and M4.

target frame is fixed to a tripod so that photogrammetric targets are easily visible from the tracking cameras and all CCR targets can be reached by the laser tracker. The simultaneous tracking with the laser tracker and the tracking cameras is performed for ten positions numbered M1 to M10. The relative position of the target frame and tracking cameras is illustrated figure 4.10. There is not much variation in the orientation of the target frame, since all CCR targets must at all times face the laser tracker and we were reluctant to move them during the acquisitions to ensure the most stable configuration possible. The target frame positions are confined to an area that is approximately 1.2 m from the tracking cameras, 0.6 m wide and 0.6 m high.

Photogrammetric tracking

The spatial and angular accuracy of the target frame tracking using our method is given figure 4.11. The results are much improved from the previous acquisitions. The spatial tracking accuracy is within the realistic simulation results and thus noticeably better than all target values. The angular orientation is also better than the target accuracy for the registration of data from the Gom Atos III and the FluxData multispectral camera. Five of the ten positions are borderline for the registration of data from the lab-designed multispectral camera and one is approximately ten percent worse.

Laser tracker acquisitions

The standard deviation of the point to point distance between the coordinates of the CCR targets measured by the laser tracker is between 0.009 mm and 0.042 mm (see row (c) of table 4.1). The laser tracker acquisitions are much more stable than previously. However, we wish to validate a tracking accuracy better than 0.025 mm, which is the maximum value of the internal

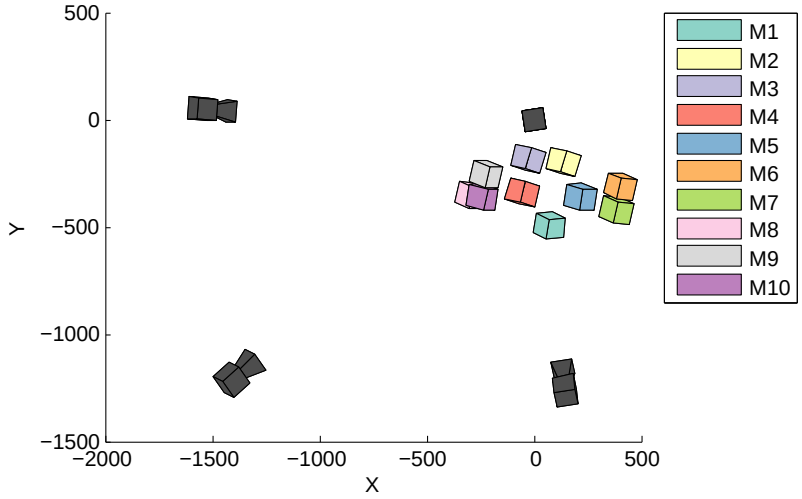


Figure 4.10: Relative position of the target frame and tracking cameras during the laser tracker measurements with fixed CCR targets.

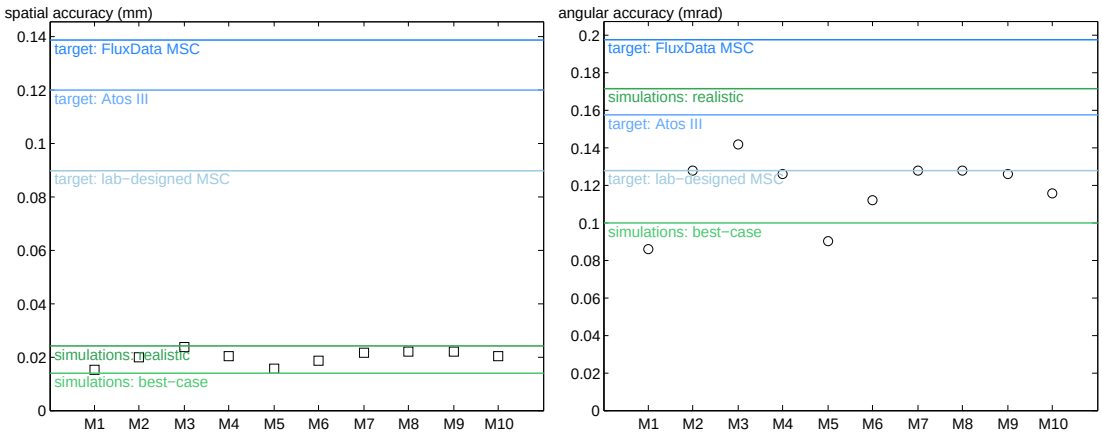


Figure 4.11: Spatial and angular accuracy of the photogrammetric tracking during the laser tracker measurements with fixed CCR targets compared to the simulation results and target accuracy.

Table 4.2: Accuracy of the transformation estimated from the laser tracker data with the fixed CCR targets for every pair (i, j) of acquisition positions. Calculated using equation 4.3, all values in mm.

i, j	1	2	3	4	5	6	7	8	9	10
1		0.064	0.057	0.063	0.021	0.033	0.025	0.027	0.031	0.025
2	0.064		0.040	0.036	0.051	0.045	0.045	0.044	0.043	0.049
3	0.057	0.040		0.013	0.040	0.042	0.038	0.036	0.034	0.043
4	0.063	0.036	0.013		0.046	0.048	0.042	0.040	0.036	0.044
5	0.021	0.051	0.040	0.046		0.021	0.020	0.020	0.023	0.021
6	0.033	0.045	0.042	0.048	0.021		0.027	0.023	0.029	0.030
7	0.025	0.045	0.038	0.042	0.020	0.027		0.012	0.009	0.016
8	0.027	0.044	0.036	0.040	0.020	0.023	0.012		0.015	0.014
9	0.031	0.043	0.034	0.036	0.023	0.029	0.009	0.015		0.015
10	0.025	0.049	0.043	0.044	0.021	0.030	0.016	0.014	0.015	
Avg.	0.038	0.046	0.038	0.041	0.029	0.033	0.026	0.026	0.026	0.029

tracking accuracy of our photogrammetric measurements. Unfortunately, our ground truth is not as accurate as what we try to measure. Nevertheless, we try to compare our datasets, as these are the most accurate laser tracker measurements we managed to perform.

We estimate the transformation matrix for every pair of positions from the laser tracker data. The accuracy of these estimates is give table 4.2. The minimum value (highest accuracy) is 0.009 mm between positions 7 and 9. The highest value is 0.064 mm between positions 1 and 2. These values are slightly worse than the point to point distances, but of the same order.

Globally, the average accuracy per position varies between 0.026 mm and 0.046 mm. We notice that the least well tracked position is position 2, while the most accurately tracked positions are 7, 8 and 9. These inaccuracies are mostly a result of the inaccurate laser tracker measurements.

Comparing datasets

Using all but the current position of interest we estimate the transformation matrix and the target frame coordinates. The accuracy of this estimate is given table 4.3. Globally we see that target 5 presents the most errors and targets 3 and 4 the least.

The average value per position of this estimated external accuracy of the photogrammetric measurements is compared with the average errors for each position of the laser tracker transformations in table 4.4. In both cases the least well tracked position is position 2 while the most accurately tracked position is position 7.

The calculation of the external photogrammetric accuracy is based on the data from the laser tracker. Since the laser tracker has an accuracy of the order of 0.030 mm, the estimated photogrammetric tracking accuracy can not be lower than 0.030 mm. The high inaccuracy of the laser tracker data makes it difficult to evaluate what fraction of the inaccuracies are inherent to the photogrammetric setup. It can be noted that, globally, the estimated external accuracy photogrammetric measurements is approximately the sum of the laser tracker inaccuracies and the internal photogrammetric accuracy. Though this tends to validate our internal accuracy

Table 4.3: Photgrammetric tracking accuracy for each position and each CCR target during the laser tracker measurements with the fixed CCR targets. Calculated using equation 4.6. All values in mm.

Position	Target					Average
	1	2	3	4	5	
1	0.069	0.043	0.035	0.046	0.073	0.053
2	0.117	0.093	0.071	0.030	0.215	0.105
3	0.116	0.073	0.044	0.040	0.140	0.083
4	0.062	0.041	0.036	0.054	0.051	0.049
5	0.059	0.051	0.069	0.055	0.083	0.063
6	0.028	0.021	0.037	0.035	0.079	0.040
7	0.057	0.029	0.012	0.026	0.043	0.033
8	0.057	0.069	0.035	0.010	0.068	0.048
9	0.026	0.036	0.033	0.066	0.033	0.039
10	0.044	0.063	0.071	0.075	0.060	0.062
Average	0.064	0.052	0.044	0.044	0.084	0.058

measurements, it is in no way a confident interpretation of the data. We can not precisely evaluate how the laser tracker inaccuracies influence the photogrammetric accuracy estimation and it is always possible that our internal accuracies are but a minimized estimate of the global tracking accuracy.

Though using the laser tracker in the mode where it follows the CCR target would provide better ground truth data, we were unable to devise a configuration that is stable enough to perform those measurements. Another improvement would be to use spherical photogrammetric targets that provide the same centerline value as a CCR target during the photogrammetric measurements. This would greatly ease the comparison and processing of the two datasets and eliminate errors that come from the estimation of $T_{L,C}$ and $P_k|_F$. However, we would still need to perform the laser tracker measurements with several fixed CCR targets, resulting in a punctual accuracy of approximately 0.020 mm.

4.2 Tracking and 3D registration

We now present a tracking and registration test performed with a single acquisition system in the target frame. This experiment is still performed in the four camera configuration, using a target frame covered by 78 targets (target frame B2). The Gom Atos III is used to digitize a car door that measures approximately 1100 mm × 600 mm. Since the simulation configuration was optimized for an area of 400 mm × 800 mm, we only digitize a portion of the car door measuring close to 800 mm × 600 mm.

The smooth surface of the car door makes it a good test object: there are few salient features and this is typically the type of object where conventional registration algorithms fail. We no longer use the laser tracker to perform simultaneous acquisitions, but we cover the car door with un-coded photogrammetric targets that are used to evaluate the accuracy of our registration procedure. These targets are not use to guide the tracking nor the registration.

Table 4.4: Average tracking accuracy for each position during the laser tracker measurements with the fixed CCR targets. The laser tracker data is from table 4.2 while the photogrammetry data is an average of each line of table 4.3. The internal spatial tracking accuracy shown figure 4.11 is given for comparison.

Position	Tracking accuracy (mm)		
	Laser tracker	Photogrammetry	
	Estimated	Estimated	Internal
1	0.038	0.053	0.015
2	0.046	0.105	0.020
3	0.038	0.083	0.024
4	0.041	0.049	0.020
5	0.029	0.063	0.015
6	0.033	0.040	0.019
7	0.026	0.033	0.021
8	0.026	0.048	0.022
9	0.026	0.039	0.022
10	0.029	0.062	0.020
Average	0.033	0.058	0.020

The bulk of the calibrations were performed the same day as the acquisitions. The necessary acquisitions and calibrations are the following:

- Tracking cameras interior orientation (a few days in advance).
- Tracking cameras exterior orientation.
- Target frame calibration.
- Gom Atos III interior orientation.
- Target frame to Gom Atos III orientation.
- Simultaneous digitization of the car door and tracking of the target frame.

We first present the accuracy achieved for the individual calibrations. Then, we analyze the accuracy with which we track the target frame and finally we evaluate the accuracy of the global registration.

4.2.1 Individual calibrations

The accuracy of the individual calibrations is given in table 4.5. When available, they are compared with the corresponding simulation parameters and results from chapter 3.

The accuracy of the tracking cameras interior orientation is between the results for the best-case and realistic simulations. Since the interior orientation accuracy directly influences the accuracy of the tracking cameras exterior orientation, it is not surprising that this last value is also between the best-case and realistic simulation results.

The target frame calibration is performed with much higher accuracy than expected but the target frame to acquisition system orientation introduces a non negligible amount of errors.

Table 4.5: Accuracy of the individual calibrations during the car door digitizing, compared to the accuracy expected from the simulations.

Calibration	Experiment	Simulations		Unit
		Realistic	Best-case	
Tracking cameras I.O.	0.056	0.100	0.033	pixel
Target frame calibration	0.015	0.050		mm
Target frame to Gom	0.025	—		mm
Atos III orientation	0.054	—		mrad
Tracking cameras E.O.	0.017	0.030	0.010	mm
	0.030	0.040	0.020	mrad

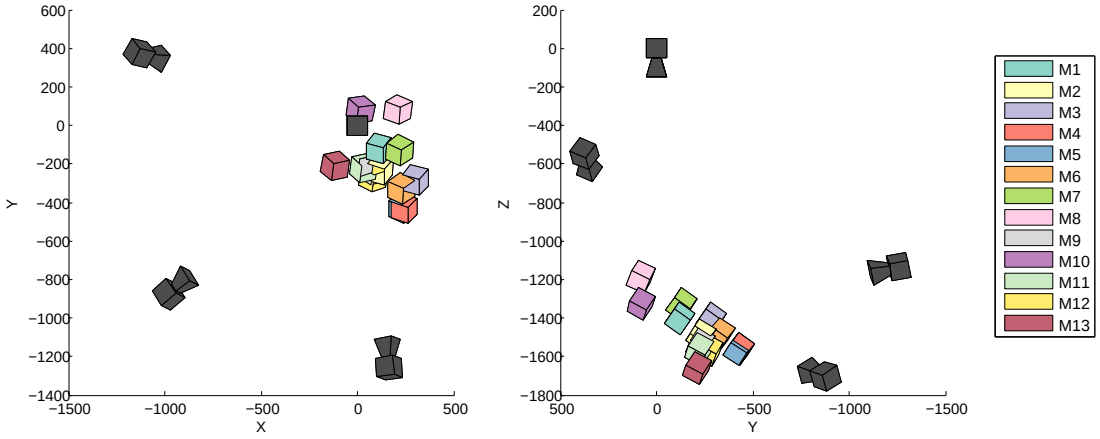


Figure 4.12: Relative position of the target frame and tracking cameras during the car door digitization.

A more accurate frame to acquisition system calibration can be executed by performing the calibration with several relative position of the frame and acquisition system with respect to the calibration target-covered object, instead of one as is the case here.

4.2.2 Tracking the target frame

Thirteen acquisitions were performed with the 3D digitization system; they are numbered M1 to M13. The relative position and orientation of the four tracking cameras and the target frame for each measurement is illustrated figure 4.12.

This subsection examines the accuracy with which we evaluate the position of the target frame in the world coordinate system for each measurement. The accuracy of the tracking of the target frame depends on the accuracy of the tracking cameras interior and exterior orientation calibrations, as well as the target frame calibration. Given the results of the previous section, we can not expect the frame tracking accuracy to reach the levels of the best-case scenario simulations.

Table 4.6 shows the position and orientation of the frame given by the photogrammetric

Table 4.6: Tracking results and accuracy during the car door digitization. Position and orientation values are significantly rounded.

	Position			Orientation			Accuracy	
	X (mm)	Y (mm)	Z (mm)	Ω (rad)	Φ (rad)	K (rad)	Spatial (mm)	Angular (mrad)
M1	133	−95	−1440	0.589	−0.445	−0.007	0.026	0.132
M2	139	−204	−1525	0.589	−0.444	−0.007	0.024	0.118
M3	321	−259	−1440	0.591	−0.445	−0.008	0.030	0.146
M4	267	−404	−1595	0.591	−0.512	−0.014	0.023	0.114
M5	266	−405	−1594	0.590	−0.682	−0.033	0.023	0.110
M6	261	−306	−1515	0.589	−0.681	−0.033	0.022	0.110
M7	250	−108	−1362	0.589	−0.684	−0.033	0.027	0.140
M8	243	95	−1222	0.394	−0.684	−0.091	0.028	0.176
M9	117	−219	−1567	0.350	−0.848	−0.163	0.028	0.148
M10	57	88	−1355	0.297	−0.974	−0.237	0.025	0.138
M11	74	−214	−1590	0.296	−0.973	−0.238	0.029	0.158
M12	117	−258	−1600	0.370	−0.772	−0.128	0.025	0.134
M13	−73	−203	−1687	0.282	−0.993	−0.253	0.026	0.140

tracking, as well as the accuracy of these parameters. In figure 4.13 these accuracy parameters are compared with the simulation results and with the target tracking accuracy for the registration of data from our three acquisition systems. Although only the Gom Atos III was used here, this permits a better evaluation of achieved tracking accuracy.

This figure shows that the spatial accuracy of the tracking is hardly never as good as expected from the simulation results, though it is not much worse. However, we are more than three times better than the target spatial accuracy for the registration of data from the lab-designed multispectral camera, the most constraining case. The angular accuracy more closely matches the simulation results. Though the accuracy is always worse than the best-case simulation results, it is always better than the realistic simulations, except for position M8. In this case, the angular accuracy is less than 10 % worse than expected.

The angular accuracy is always sufficient to register data from the Gom Atos III and the FluxData multispectral camera, with the exception, once again, of position M8. It is not surprising that M8 is the least well-tracked position: as figure 4.12 shows, this position of the frame is slightly off-center compared to the other positions and thus partially out of the simulation bounds.

4.2.3 3D registration

The registration of all 13 positions using the tracking data is illustrated figure 4.14. When we interact with these meshes, the registration seems seamless. Our goal however, is to create a 3D model that is not only visually satisfying, but that can also be used for metrology purposes.

The Gom Atos III acquisition software recognizes any coded or uncoded target present in the scene. In our case the full surface is covered with over 48 uncoded targets of diameter 6 mm. We export the list of targets visible in each mesh, this number varies between 8 and 16 per mesh. This provides us with $A_{k_i}|_{C_S}$ the coordinates of target point k from mesh i in the

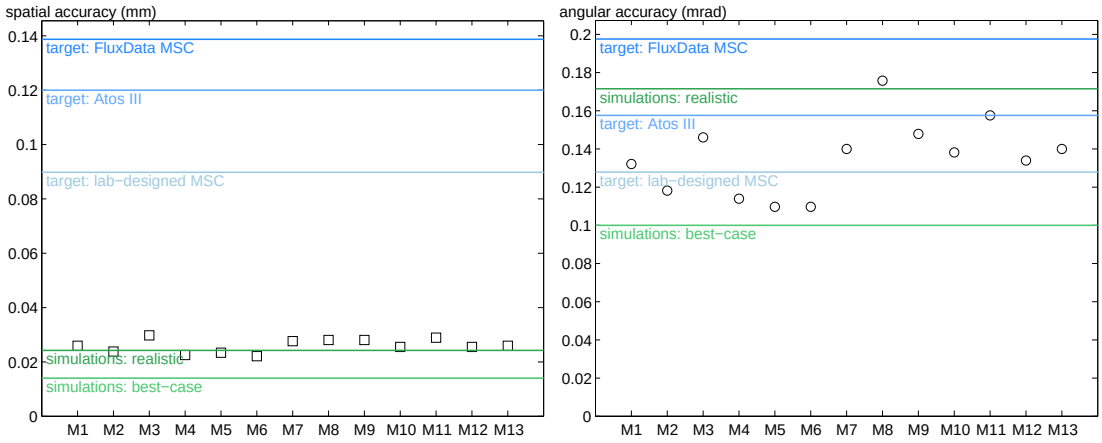
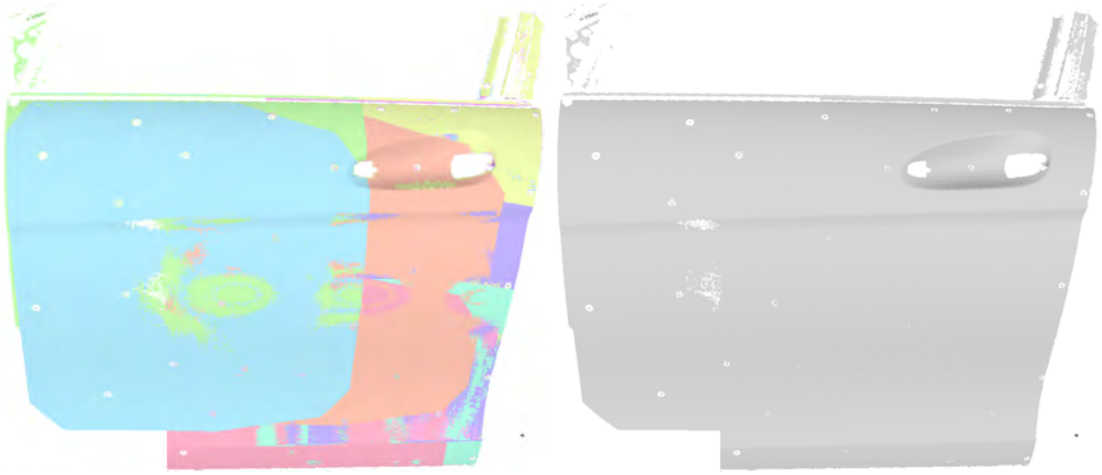


Figure 4.13: Spatial and angular tracking accuracy during the car door digitization, compared to the simulation results and to the target accuracy.



(a) (b)

Figure 4.14: All 13 meshes representing the car door registered in a single model. (a) One color per mesh. (b) All meshes in grey.

acquisition system coordinate system.

We apply the same transformation to these points as we applied to the meshes and obtain $A_{k_i}|_{C_0}$ for all targets of all meshes. For every target k and for every pair of meshes (i, j) where target k is visible, we now calculate

$$A_{k_i}|_{C_0} - A_{k_j}|_{C_0} \ .$$

If the registration were perfect, the result of this subtraction would be a null vector. This provides us with a measure of the accuracy of the registration. Table 4.7 shows the average over k of

$$\left\| A_{k_i}|_{C_0} - A_{k_j}|_{C_0} \right\|_2$$

for all pairs of meshes. This value varies between 0.097 mm and 2.376 mm with an average of 0.889 mm.

We reach our target registration accuracy of 0.12 mm for only two mesh-pairs: between M1 and M2 (0.10 mm) as well as between M5 and M6 (0.12 mm). This is partially due to the fact that M1, M2, M5 and M6 are among the most accurately tracked measurements (see figure 4.13). However, a high tracking accuracy for two positions does not necessarily result in a high registration accuracy. The pairwise registration between M1 and M5 on one side and M2 and M6 on the other size is more than seven times worse than M1 to M2 or M5 to M6.

Table 4.6 shows that the frame orientation hardly changes between M1 and M2, as well as between M5 and M6, though there is a translation of over 100 mm in both cases. The accurate tracking in these two cases suggests that the acquisition system and target frame are not as tightly fixed as we need them to be: changing the frame orientation slightly changes the position of the acquisition system in the frame, resulting in an imprecise registration. The fact that the registration accuracy between M1–M5, M2–M5, M1–M6 and M2–M6 is almost the same (0.86 mm or 0.87 mm) confirms this interpretation. The bulk of the registration error in this subset seems to be a fix error resulting from the unstable position of the acquisition system in the tracking frame.

This also explains why succeeding measurements generally have a better registration accuracy: the position of the acquisition system has not moved as much in the frame. The exceptions are always cases in which the orientation changes significantly between two consecutive measurements (M11 to M12, M12 to M13 and M7 to M8).

We also notice that a low number of shared points, which denotes positions with little overlap, does not necessarily correspond to a low tracking accuracy: positions M4 and M7, as well as positions M5 and M7 respectively share three and two points. Nevertheless, they are tracked with an accuracy of 0.33 mm, which places them among the ten most accurately tracked pairs of positions.

Though it is possible to register 3D datasets with an accuracy of 0.12 mm using our setup, this is confined, in this experiment, to very specific cases. To generalize this registration accuracy, the overall stability of the setup must be improved. In particular, the acquisition system must be more tightly fixed to the target frame. The hexagonal fixtures used ensure that the acquisition system does not move during each acquisition, but not that it stays in the same

Table 4.7: Average point to point distance between the targets on the car door for every pair of positions, in mm. The number of points shared by each pair is given in parenthesis. Cells are colored on a linear RGB scale with green for 0 mm and magenta for values greater than 2 mm.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
M1		0.12 (9)	0.59 (5)	0.68 (3)	0.86 (1)	0.87 (3)	0.83 (3)	0.81 (2)	1.64 (2)	0.91 (2)	1.17 (2)	0.58 (2)	1.60 (2)
M2	0.12 (9)		0.65 (10)	0.67 (6)	0.86 (2)	0.87 (4)	0.87 (4)	0.79 (2)	1.59 (3)	0.92 (3)	1.10 (3)	0.68 (5)	1.48 (5)
M3	0.59 (5)	0.65 (10)		0.57 (9)	0.65 (4)	0.61 (8)	0.47 (7)	0.36 (5)	1.37 (7)	0.79 (6)	0.99 (7)	1.20 (8)	1.50 (8)
M4	0.68 (3)	0.67 (6)	0.57 (9)		0.16 (6)	0.20 (6)	0.33 (3)	— (0)	0.97 (8)	0.54 (3)	0.65 (8)	1.41 (10)	0.92 (10)
M5	0.86 (1)	0.86 (2)	0.65 (4)	0.16 (6)		0.10 (5)	0.33 (2)	— (0)	0.81 (8)	0.51 (2)	0.57 (8)	1.64 (8)	0.79 (8)
M6	0.87 (3)	0.87 (4)	0.61 (8)	0.20 (6)	0.10 (5)		0.25 (6)	0.82 (2)	0.84 (10)	0.50 (5)	0.57 (10)	1.60 (6)	0.88 (7)
M7	0.83 (3)	0.87 (4)	0.47 (7)	0.33 (3)	0.33 (2)	0.25 (6)		0.64 (10)	1.01 (5)	0.61 (13)	0.74 (5)	1.51 (2)	1.13 (2)
M8	0.81 (2)	0.79 (2)	0.36 (5)	— (0)	— (0)	0.82 (2)	0.64 (10)		1.44 (1)	0.69 (9)	1.00 (1)	— (0)	— (0)
M9	1.64 (2)	1.59 (3)	1.37 (7)	0.97 (8)	0.81 (8)	0.84 (10)	1.01 (5)	1.44 (1)		0.85 (5)	0.60 (13)	2.38 (9)	0.73 (10)
M10	0.91 (2)	0.92 (3)	0.79 (6)	0.54 (3)	0.51 (2)	0.50 (5)	0.61 (13)	0.69 (9)	0.85 (5)		0.30 (5)	1.55 (2)	1.06 (2)
M11	1.17 (2)	1.10 (3)	0.99 (7)	0.65 (8)	0.57 (8)	0.57 (10)	0.74 (5)	1.00 (1)	0.60 (13)	0.30 (5)		1.85 (9)	0.81 (10)
M12	0.58 (2)	0.68 (5)	1.20 (8)	1.41 (10)	1.64 (8)	1.60 (6)	1.51 (2)	— (0)	2.38 (9)	1.55 (2)	1.85 (9)		2.08 (12)
M13	1.60 (2)	1.48 (5)	1.50 (8)	0.92 (10)	0.79 (8)	0.88 (7)	1.13 (2)	— (0)	0.73 (10)	1.06 (2)	0.81 (10)	2.08 (12)	

position in its socket for successive acquisitions. The stability we require is very high, of the order of a few micrometers. These are difficult constraints to reach for a camera fixture that must support 7.5 kg (weight of the Atos III) in varying orientations (ideally even sideways and upside down).

Part of the inaccuracies may also come from a deformation of the target frame, due to its manipulation or to temperature changes that result in an expansion or contraction of the aluminum bars. This hypothesis will be tested in the following section.

Barone *et al.* [2012a,b] have developed a similar tracking technique for the registration of 3D meshes. In their case, the tracking is performed by stereo vision and the registration is the first step of a two-step registration process: the coarse registration supported by the optical tracking is used as an input for an ICP optimization. The use of only two cameras makes their setup more easily transportable that ours, but it yields a tracking and registration accuracy that must be completed by an optimization step. Their setup is better adapted to registration of highly three-dimensional objects, such as statues, while our method is targeted for the registration of smooth, featureless meshes.

4.3 Stability of the individual calibrations

We assess the stability of the individual calibrations by comparing measurements performed in the same configuration over a period of time that varies from a single day to a few weeks.

4.3.1 Target frame stability

During one of our campaigns of laboratory experiments, we performed twelve calibrations of the target frame (target frame B3) over the course of two weeks. The internal accuracy of each of these calibrations is given in table 4.8. All calibrations were far more accurate than our target 0.05 mm from the simulations. The least accurate calibrations, G and H, are more accurate than 0.03 mm.

We compute all point to point distances between the 80 photogrammetric targets for all pairs of calibrations. For every pair of calibrations, we then calculate the average over all targets of the absolute difference between the point to point distances. The results are given table 4.9. This is a symmetric matrix but all data is given to improve readability. Groupings show calibrations that were performed the same day.

As with the internal accuracies, calibrations G and H stand out. This is probably due to a bad image quality and an overall unstable internal accuracy of the camera during the acquisitions. If we ignore these two acquisitions, the stability of the calibration is quite satisfactory: it varies between 0.02 mm and 0.06 mm.

Though the internal accuracy values give us an indication of the accuracy of the calibration, they are not sufficient: Looking at the RMS of the target residuals it seems that H is the worst calibration. The comparison of the average point to point distance between the targets shows that, though calibration H is bad, calibration G is far worse. Also, looking at the RMS of the target residuals, G seems only slightly worse than L, when in reality L is a good calibration.

Table 4.8: Internal accuracy of the target frame calibrations. Groupings show calibrations that were performed the same day. RMS of the residuals of the target coordinates.

	RMS of residuals x (mm)	y (mm)	z (mm)	RMS of number of rays per point	Number of images
A	0.018	0.018	0.021	24.5	113
B	0.013	0.013	0.014	46.5	133
C	0.014	0.015	0.017	50.0	127
D	0.010	0.010	0.011	48.5	119
E	0.013	0.013	0.016	45.7	119
F	0.011	0.011	0.012	50.2	128
G	0.019	0.019	0.020	49.2	123
H	0.023	0.023	0.026	53.8	139
I	0.013	0.014	0.017	51.3	130
J	0.011	0.012	0.014	51.8	129
K	0.010	0.010	0.011	53.4	145
L	0.018	0.018	0.020	38.1	101

Table 4.9: Average absolute difference in the point to point distances between the photogrammetric targets for every pair of target frame calibrations, in millimeters. Cells are colored on a linear RGB scale that goes from green for 0 mm to magenta for 0.1 mm.

	A	B	C	D	E	F	G	H	I	J	K	L
A		0.04	0.05	0.05	0.04	0.05	0.10	0.07	0.05	0.05	0.05	0.06
B	0.04		0.03	0.03	0.03	0.03	0.08	0.06	0.04	0.04	0.05	0.04
C	0.05	0.03		0.03	0.03	0.04	0.10	0.07	0.02	0.04	0.04	0.04
D	0.05	0.03	0.03		0.04	0.02	0.07	0.05	0.04	0.04	0.05	0.04
E	0.04	0.03	0.03	0.04		0.05	0.10	0.08	0.03	0.04	0.04	0.05
F	0.05	0.03	0.04	0.02	0.05		0.07	0.05	0.05	0.04	0.06	0.04
G	0.10	0.08	0.10	0.07	0.10	0.07		0.04	0.10	0.07	0.09	0.07
H	0.07	0.06	0.07	0.05	0.08	0.05	0.04		0.07	0.05	0.06	0.05
I	0.05	0.04	0.02	0.04	0.03	0.05	0.10	0.07		0.04	0.04	0.04
J	0.05	0.04	0.04	0.04	0.04	0.04	0.07	0.05	0.04		0.03	0.03
K	0.05	0.05	0.04	0.05	0.04	0.06	0.09	0.06	0.04	0.03		0.04
L	0.06	0.04	0.04	0.04	0.05	0.04	0.07	0.05	0.04	0.03	0.04	
Average	0.06	0.04	0.04	0.04	0.05	0.04	0.08	0.06	0.05	0.04	0.05	0.05

Our target object coordinate accuracy is of 0.05 mm. In practice, the target frame calibration accuracy is generally better than 0.02 mm. We thus have a tolerance of 0.03 mm for inaccuracies due to the limited stability of the target frame. If we only compare calibrations performed the same day, this value is reached only for target frame calibrations B and C. For all other pairs of calibrations performed a given day, the average absolute difference in the photogrammetric targets is of 0.04 mm. To ensure an optimal tracking accuracy, the stability of target frame should be improved, for example by designing a target frame made of carbon, which is less prone to deformations due to temperature changes than aluminum.

4.3.2 Tracking cameras internal orientation stability

Three cameras were calibrated twice the same week, on Monday and Friday. The difference in the estimation of the focal length is less than 0.001 mm, which is an absolute percentage error of approximately 0.01 %.

There is a greater difference in the principle point offset, particularly in the x direction. The absolute error is less than 0.085 mm, but this represents 68 % of the original value. However, it is difficult to evaluate the impact of these inaccuracies that can be compensated by the accurate calibration of the exterior orientation of the cameras as well as by the calibration of the target frame.

4.3.3 Tracking cameras exterior orientation stability

We compare the stability of ten exterior orientation camera calibrations (A to J) of four tracking cameras, performed over a two week period. The internal accuracy provided by the bundle adjustment is given table 4.10. The calibrations are once again grouped when they were performed the same day.

Compared to the simulation results, the exterior orientation accuracy is quite satisfying: we are always twice as good as the expected values in the realistic scenario (0.03 mm spatially and 0.04 mrad angularly). Compared to the best-case simulation result, we are always better than the angular target value. The experimental spatial accuracy is always worse or equal to the spatial accuracy used for the simulations. As the angular accuracy is generally harder to reach, this may not have too much of a negative impact on the tracking accuracy.

The orientation of the tracking cameras changed between each group of acquisitions, so that the cameras faced the evolving area of interest. All calibrations can not be compared with one another, only the calibrations performed in a single position can be compared over time. For every set of calibrations performed in a given configuration we calculate for every camera the difference in its position between every pair of configurations. The absolute average of these three values (since one camera is set as the origin) is calculated for all pairs of configurations. This data is given table 4.11.

The stability of a given configuration is generally better for calibrations that were performed the same day, but the data varies threefold between 0.06 mm and close to 0.18 mm. There are two weak points of our setup that can explain these variations. First, the camera support, like the target frame, is made of aluminum bars that are sensitive to temperature changes. Though

Table 4.10: Exterior orientation accuracy reached when tracking with four cameras, compared to the simulation values.

Configuration		Max. std. deviation Spatial (mm)	Angular (mrad)
Experiments	A	0.014	0.016
	B	0.013	0.018
	C	0.014	0.018
	D	0.010	0.012
	E	0.011	0.014
	F	0.014	0.018
	G	0.011	0.012
	H	0.011	0.014
	I	0.010	0.012
	J	0.011	0.010
Simulations	Realistic	0.030	0.040
	Best-case	0.010	0.020

Table 4.11: Average absolute displacement of the cameras for every pair of exterior orientation calibrations, in millimeters. Cells are colored on a linear RGB scale that goes from green for 0 mm to magenta for 0.2 mm.

	A	B	C	D	E	F	G	H	I	J
A		0.175	0.111							
B	0.175		0.075							
C	0.111	0.075								
D					0.081	0.090	0.085	0.121		
E				0.081		0.110	0.136	0.172		
F				0.090	0.110		0.115	0.139		
G				0.085	0.136	0.115		0.060		
H				0.121	0.172	0.139	0.060			
I										0.080
J									0.080	
Avg.	0.143	0.125	0.093	0.094	0.125	0.114	0.099	0.123	0.080	0.080

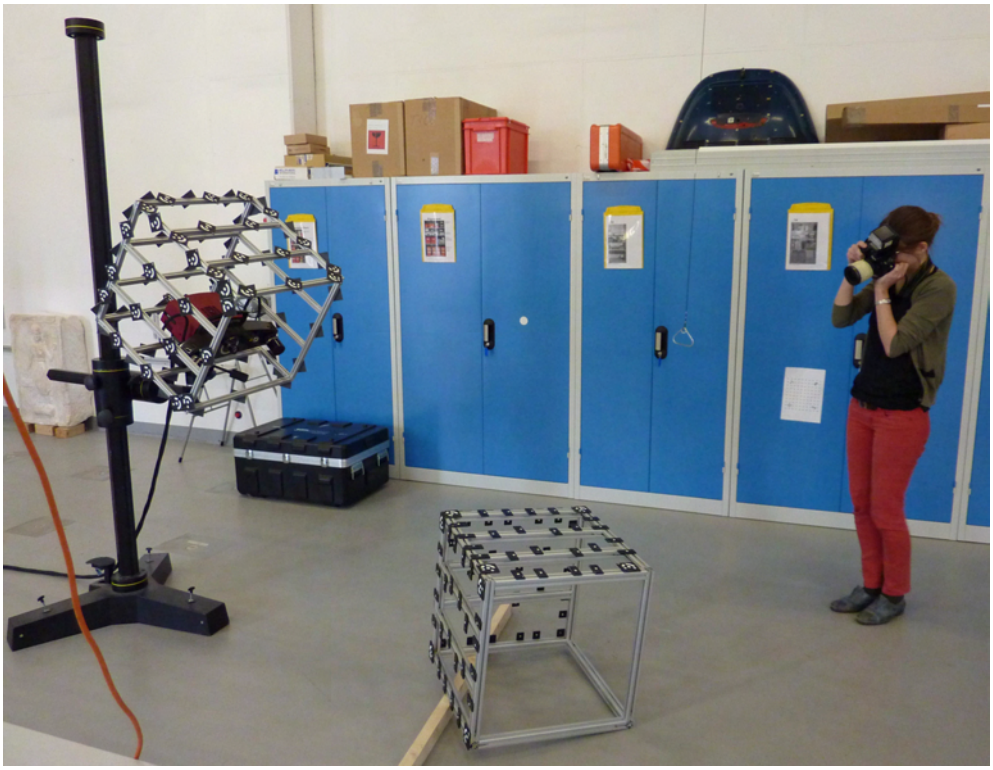


Figure 4.15: Calibration between the target frame and the Gom Atos III. The target-covered object used for the calibration is shown on the floor.

these bars are thicker than those used for the target frame, they are also much longer, varying between 1.4 m and 1.7 m. A temperature change can thus produce measurable differences.

The ball-joints used to orient and support the tracking cameras were not photo-quality. They may have resulted in the cameras barely tilting down over time.

4.3.4 Target frame to Atos III stability

The calibration between target frame B3 and the Gom Atos III was performed three times. For calibrations A and B, we used two relative positions of the target frame and target-covered object (the setup is shown figure 4.15) to measure the position of the acquisition system in the target frame. These relative positions are numbered a1, a2, b1, b2 and processed as individual calibrations to compare with the global calibration made by processing the data in pairs. A glitch during the third calibration resulted in data from only one relative position of the target frame and target-covered object adequate to be used for the calibration, labeled 'c'.

The calibration results and associated accuracy are given table 4.12. The differences in the spatial positioning are of the order of a few tenths of a millimeter while the difference in the orientation angles are of the order of a few tenths of a milliradian. Combining two positions seems to improve the calibration accuracy by approximately 25 %. The integration of the data from two different relative positions does not necessarily result in values that are between the values calculated with the individual positions. For example the Z coordinate for A is greater

Table 4.12: Results and accuracy of the target frame to Gom Atos III calibrations.

	Position			Orientation			Accuracy	
	X (mm)	Y (mm)	Z (mm)	Ω (rad)	Φ (rad)	K (rad)	Spatial (mm)	Angular (mrad)
a1	18.235	−319.384	846.368	0.2450	−0.0400	−0.0457	0.040	0.074
a2	18.066	−319.294	846.302	0.2450	−0.0402	−0.0457	0.045	0.080
A	18.160	−319.382	846.395	0.2450	−0.0401	−0.0457	0.029	0.056
a1	17.885	−319.564	846.031	0.2469	−0.0393	−0.0452	0.053	0.098
a2	18.191	−319.611	846.177	0.2472	−0.0391	−0.0452	0.046	0.082
B	18.037	−319.611	846.173	0.2471	−0.0392	−0.0452	0.036	0.068
c	18.238	−319.578	846.263	0.2465	−0.0398	−0.0453	0.066	0.106

than both the Z coordinates for a1 and a2.

There seems to be as much variation between two positions that make up a single calibration as between measurements performed on different days. This is difficult to establish with great confidence without performing a statistical analysis that would require more data than what is available here. However, the variation between different calibrations is also greater than the accuracy of the measurements. This tends to corroborate our hypothesis that the target acquisition system moves in the target frame to an extent that hurts the final registration accuracy.

4.4 Tracking with six cameras

All tests presented so far were performed using four cameras. We now test our method in the six camera configuration. For this, we track the target frame in sixteen positions (M1 to M16). The relative position and orientation of the target frame and cameras for these acquisitions is illustrated figure 4.16.

4.4.1 Individual calibrations

Since no acquisitions are performed and we only evaluate the tracking, only three calibrations are necessary:

- Tracking cameras interior orientation (performed a few days earlier).
- Target frame calibration.
- Tracking cameras exterior orientation.

The accuracy of the individual calibrations is given table 4.13. The calibration of the tracking cameras follows the same protocol as previously, though it takes longer because of the increase in the number of cameras. We expect the accuracy of the interior orientation to be of the same order as that reached in the four camera configuration. In practice we have improved the accuracy of this calibration from worse than $1/20^{\text{th}}$ of a pixel to better than $1/30^{\text{th}}$ of a pixel

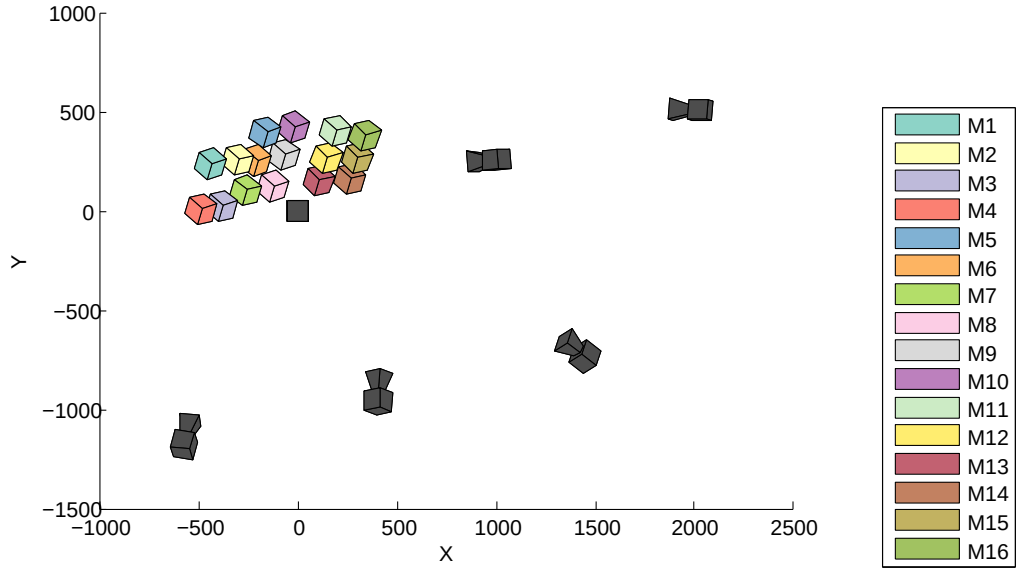


Figure 4.16: Relative position of the target frame and tracking cameras when tracking with six cameras.

Table 4.13: Accuracy of the individual calibrations when tracking with six cameras, compared to the accuracy expected from the simulations.

Calibration	Experiment	Simulations		Unit
		Realistic	Best-case	
Tracking cameras I.O.	0.029	0.100	0.033	pixel
Target frame calibration	0.014	0.050		mm
Tracking cameras E.O.	0.031	0.030	0.010	mm
	0.034	0.040	0.020	mrad

for all six cameras. This last result is better than the best-case simulation constraints. The improvement in calibration accuracy is due to a better control of the illumination conditions during the internal orientation acquisitions.

The target frame calibration procedure is exactly the same as previously and was part of the results presented in section 4.3.1. Two calibrations, B and C, were performed the same day. Since calibration B is slightly more accurate than calibration C, we use calibration B for the data processing. The RMS of the residuals is better than 0.014 mm, which is more than twice as good as the values used for the simulations.

The exterior orientation of the tracking cameras, is longer and more complicated now that we have six cameras. This is because they do not all observe the same area, and there is only a small overlap in their common view. The setup is that of the wall painting simulations and is shown figure 4.17). As predicted by the simulations, we expect the accuracy of these calibrations to be slightly worse than in the four camera configuration.

Two exterior orientations were performed. Their accuracy is compared to the simulation results in table 4.14. The experimental results do not reach the best-case scenario simulation

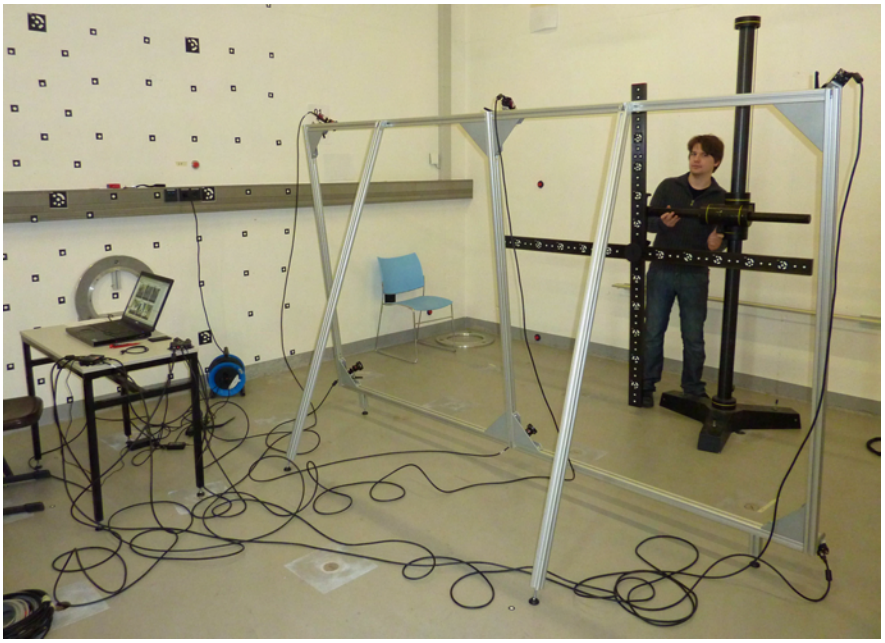


Figure 4.17: Calibration of the exterior orientation of six tracking cameras.

Table 4.14: Exterior orientation accuracy reached when tracking with six cameras, compared to the simulation values.

		Spatial (mm)	Angular (mrad)
Experiments	A*	0.045	0.048
	B*	0.031	0.034
Simulations	Realistic	0.060	0.060
	Best-case	0.020	0.020

results, but they are always better than the realistic simulation results. We process the tracking with the results from exterior orientation B*, since it is the most accurate of the two (table B.8 page 174 in the appendix shows that this indeed results in more accurate tracking results). Given the exterior orientation accuracy, we do not expect the tracking results to reach the best-case scenario results, though they should reach the realistic tracking accuracy.

4.4.2 Tracking accuracy

The tracking accuracy of all sixteen positions is given figure 4.18 and compared with the target and simulation values. Looking first at the spatial accuracy we notice that only a few positions (M11 and M16) are tracked with the accuracy expected from the best-case simulations. And even then, they are just slightly better than the target value. However, all positions are tracked with the spatial accuracy expected from the realistic simulations. Most importantly, all positions are tracked with a spatial accuracy many times better than that necessary for the registration of the 3D data from the Atos III and the multispectral data from both cameras

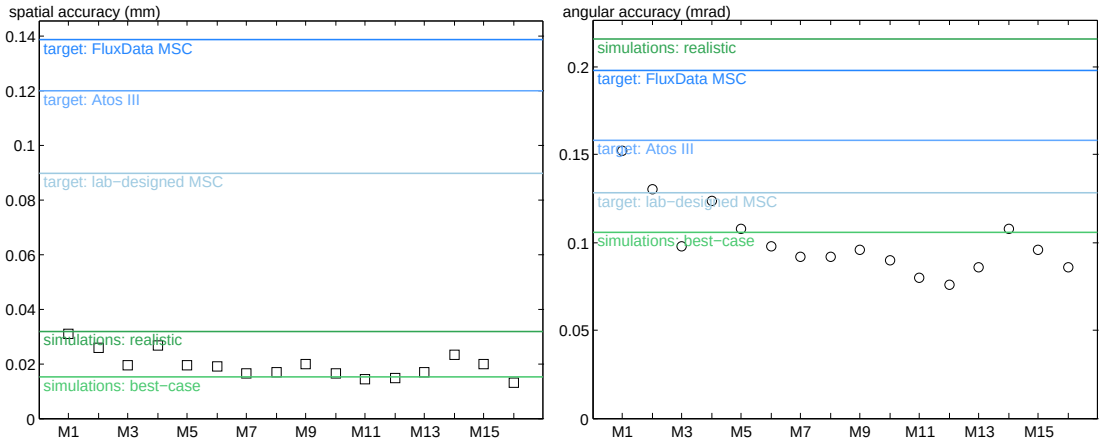


Figure 4.18: Spatial and angular tracking accuracy when using six cameras, compared to the simulation results and to the target accuracy.

that we use.

All positions are tracked with an angular accuracy that is better than the realistic simulation results, the target value for the FluxData multispectral camera and the target value for the Gom Atos III. Two positions (M1 and M2) are tracked with less accuracy than the lab-designed angular target and five positions (M1, M2, M4, M5 and M14) are tracked with less accuracy than the best-case scenario simulations.

Though the least well tracked positions are all in somewhat extreme positions (highest, lowest, most left) this can not solely account for their disappointing tracking accuracy since other extreme positions such as M3, M10 and M15 are well tracked. Other factors must also cause this inaccuracy for example the orientation of the target frame with respect to the cameras.

We are satisfied with the tracking accuracy in the six camera configuration and believe it can be used to accurately register 3D and multispectral datasets. Globally, these tracking results are close to the best-case simulation values, leaving a wide margin for the registration of data from the Gom Atos III digitization system and the FluxData multispectral camera.

4.5 Conclusion

All the tracking results of this chapter are grouped in figure 4.19. We defined an experimental setup that corresponds to two of the three simulation configurations. We extensively tested the tracking of a target frame of dimensions 500 mm × 500 mm × 500 mm using four cameras, in a setup that corresponds to the sarcophagus simulations.

The first tests of our photogrammetric setup were performed simultaneously with a laser tracker, to provide an external measurement of the tracking accuracy of our system. The different constraints to obtain accurate measurements by both the photogrammetric setup and the laser tracker made it difficult to obtain satisfactory data with the two systems in a given configuration. The data processing and comparison was complicated by the fact that no common coordinate system exists between the two types of measurements. Nevertheless, the data seems

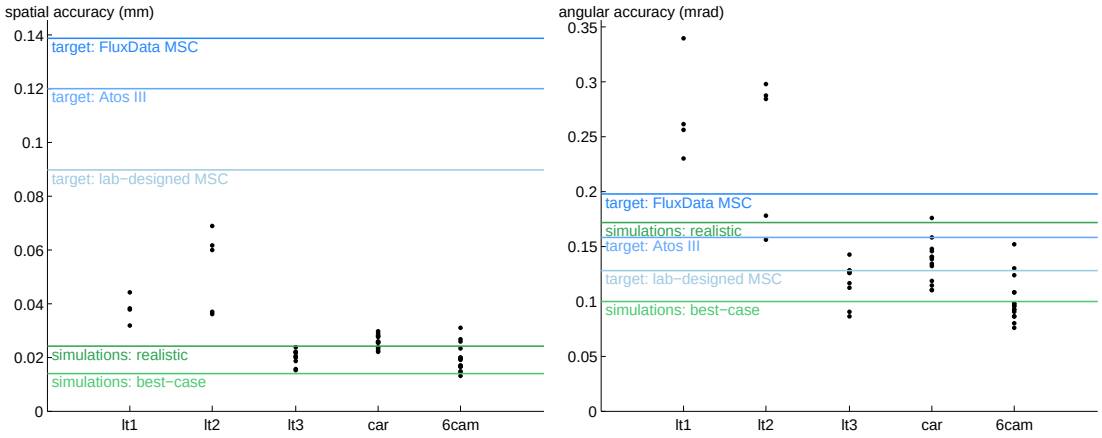


Figure 4.19: Spatial and angular accuracy of all photogrammetric tracking experiments from this chapter. 'lt1', 'lt2' and 'lt3' represent, respectively, the first, second and third laser tracker measurements. 'car' represents the digitization of the car door and '6cam' the tracking using six cameras.

consistent with our tracking accuracy of approximately 0.02 mm. The best tracking accuracy was obtained during the third laser tracker measurements, performed with the final version of the target frame. The use of spherical photogrammetric targets that provide the same centerline value as a CCR and can be placed in the CCR sockets could help the processing. However, this would not provide more accurate laser tracker measurements in a configuration that is favorable for the photogrammetric tracking.

The first registration test based on our method was performed with only 3D data. The registration of these thirteen independent 3D meshes representing a smooth surface highlighted a weak point of our setup: the attach between the target frame and the acquisition system. When the orientation between two successive positions was stable, we were able to reach our registration accuracy goal of 0.12 mm. For the other positions, the accuracy of the registration was of the order of a millimeter.

The stability of the target frame over time is globally satisfying, as is the interior orientation stability. The exterior orientation stability could be improved, but is currently sufficient for acquisitions performed during a single day. The calibration between the target frame and acquisition system should be improved, though this would require a better fixture between the two.

We also test the tracking in the six camera configuration. The accuracy is slightly reduced compared to the four camera configuration, but these results are quite satisfying given the increased difficulty of maintaining a stable configuration when the setup is scaled up. The calibration of an increased number of cameras is more accurate than the realistic simulation results, thought the best-case accuracy is not reached. In most cases, the final tracking accuracy is adequate to register data from our three acquisition systems.

The tracking and registration tests have yeilded promising results. Our method can now be applied to the registration of 3D and multispectral data.

Chapter 5

Integration of 3D and Multispectral Data

In this chapter, our tracking method is applied to a multispectral camera and a 3D digitization. The digitizations are executed with the Gom Atos III, as previously, while all multispectral acquisitions rely on the FluxData multispectral camera. The acquisitions from both acquisition systems are registered using the data provided by the tracking.

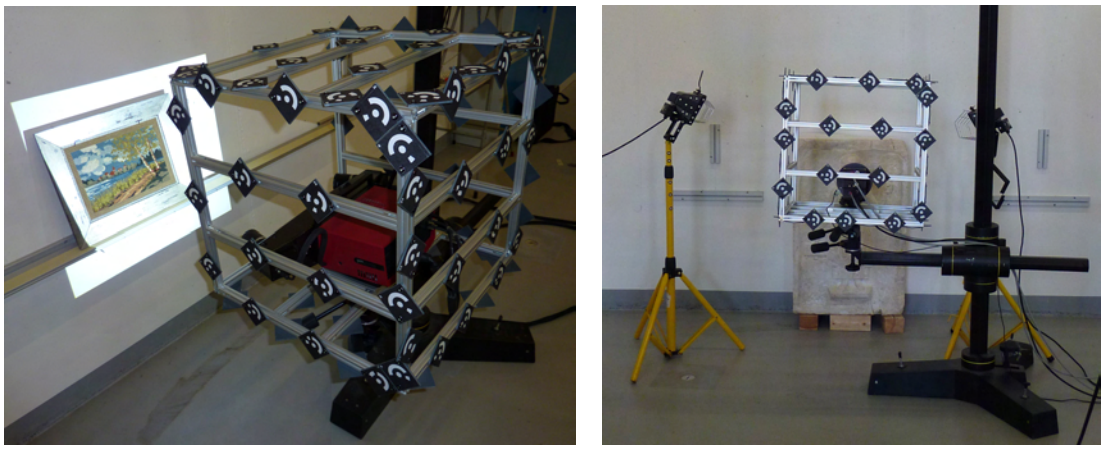
This experiment is performed for two objects, a cross-stitch canvas and a bas-relief. In both cases, four tracking cameras survey the scene, as in the sarcophagus configuration. The tests presented here were performed in laboratory settings to validate the method in controlled settings. The first section of this chapter presents the two objects digitized and the corresponding acquisition setup. We then describe the multispectral data processing. Finally, we present the tracking and registration results for the two configurations.

5.1 Configuration overview

5.1.1 Cross-stitch canvas

The frame of the cross-stitch canvas measures 45 cm × 36 cm and the cross-stitch itself covers an area of 32 cm × 23 cm. Though it is smaller than the area of interest of the sarcophagus, the cross-stitch canvas is a good test object for the registration of 3D and multispectral data. It contains both spatial surface variations on the frame and in the stitches, as well as reflectance variations in the color of the thread. During the digitizations, the frame sat on a small metallic shelf and the four tracking cameras were positioned to observe the acquisition systems while they digitized the surface. Figure 5.1a shows the object under study being digitized by the Atos III fixed to the target frame. The last iteration of the target frame (target frame B.3 described in appendix B.1, page 167), which is covered with 80 targets, is used for this and all subsequent experiments.

We performed five 3D digitizations and seventeen multispectral acquisitions of the cross-stitch, which were largely sufficient to digitize the full canvas. The 3D datasets are labeled G1



(a) (b)

Figure 5.1: Multisensor acquisitions using the target frame. (a) Detail of the cross-stitch canvas 3D digitization. (b) Detail of the plaster bas-relief multispectral acquisition. Even though the ring-light is still fixed to the multispectral camera, only the two halogen lights are used during the acquisitions.

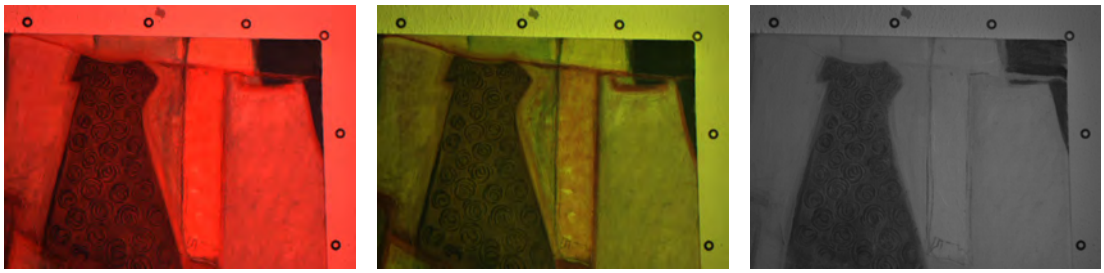
to G5, while the multispectral acquisitions are labeled FD1 to FD17. These were our first 3D – multispectral registration tests. As such, we acquired overlapping multispectral acquisitions to better evaluate the final registration accuracy.

The multispectral acquisitions must be performed with an external light source (the 3D digitizations do not require any external light — they are actually not as accurate in bright environments). An LED ring light was fixed to the multispectral camera to uniformly light the surface of the cross-stitch canvas. This ring light provides a fairly uniform illumination in the visible range. Unfortunately, it does not emit in the near infrared, so these first multispectral acquisitions are composed of only six channels acquired in the visible spectrum. Since the current object of interest is smaller than the sarcophagus, the multispectral camera was positioned closer to the surface under study, at approximately 500 mm from the canvas.

5.1.2 Bas-relief

The bas-relief we study is a painted plaster provided by the RGZM (Römish-Germanishes Zentralmuseum — Roman-German Central Museum, Mainz, Germany). This bas-relief measures approximately 500 mm × 800 mm and is slightly bigger than the sarcophagus area of interest (400 mm × 700 mm). Surface and color variations are comparable to those present on the sarcophagus. The goal of this digitization was to closely mimic the acquisitions that were previously perform on the sarcophagus in Trier: the full surface of interest is acquired by the Gom Atos III while specific sub-areas of interest are acquired by the multispectral camera.

The bas-relief sits close to the floor, on a few pieces of wood to ensure its stability. The tracking setup is similar to the one used for the cross-stitch digitization. In this case, twelve 3D acquisitions (G1 to G12) and four multispectral acquisitions (FD1 to FD4) were performed. The multispectral camera is now approximately 700 mm away from the surface under study, as



(a) CCD A (b) CCD B (c) CCD NIR

Figure 5.2: Sample image acquired by each CCD.

was the case during the simulations. We no longer use the ring light to illuminate the scene for the multispectral acquisitions. Instead, a halogen light is placed on each side of the bas-relief, as shown figure 5.1b. These halogen lights also illuminate the NIR, so all seven bands of the multispectral acquisitions can be used.

5.2 Multispectral data representation

The FluxData multispectral camera is made of two RGB CCDs that acquire data in the visible (six channels) and one greyscale CCD that acquires data in the Near InfraRed (NIR). An example of the individual acquisitions of the three CCDs is given in figure 5.2. Using this sample image we will quickly present the processing performed on the multispectral acquisitions from the FluxData camera to obtain reflectance data and color representations from the acquired data. The spectral response of each channel is plotted in figure 5.3. This figure clearly shows that CCD A has a high sensitivity mapped to the red channel, while CCD B has a higher sensitivity assigned to the green channel. The better equilibrium between the channels of CCD B explains why, despite its green tinge, figure 5.2b is more representative of the true colors of the original painting.

The acquisitions from seven individual channels of the sample acquisition are shown figure 5.4. In image 5.4e we see that the red channel from CCD A has such high sensitivity that it is saturated, while the other channels from the same sensor (figures 5.4a and 5.4c) are under exposed. Similarly, the blue channel of CCD B (figure 5.4b) is clearly underexposed, while the other two (figures 5.4d and 5.4f) are satisfactory. It is difficult to set an acquisition time per CCD that is adapted to all three channels of a given sensor. This is further complicated by the fact that, during the acquisition, the software only shows us the images in figure 5.2, making it difficult to detect when individual channels are saturated.

To recover reflectance information with physical significance from these images, a series of spectral calibration steps are necessary. We first present how we compensate nonuniform light conditions. We then briefly describe the spectral reconstruction. Finally, we show a few RGB representations created from the spectral data. This section describes widespread techniques in dealing with multispectral data, as well as previous work done in our institutions. We revert to presenting original work in section 5.3.

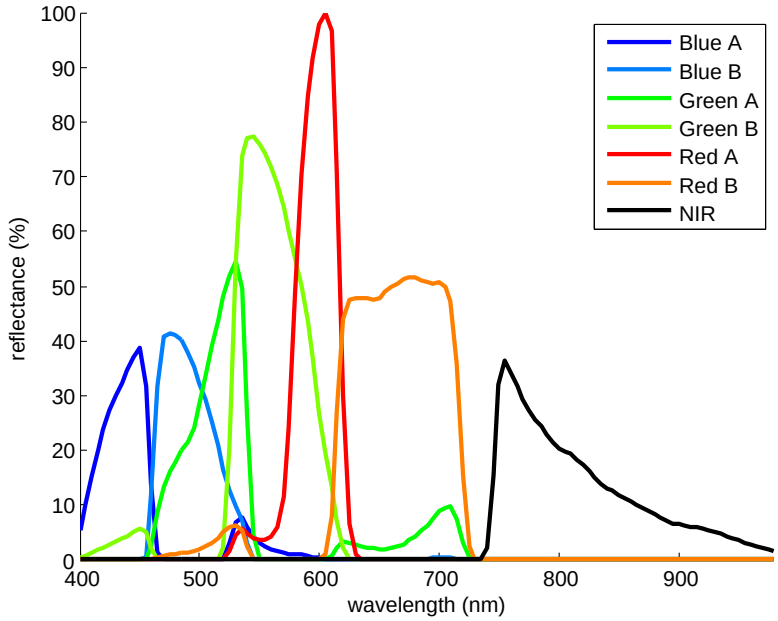


Figure 5.3: FluxData multispectral camera spectral response, as measured by the manufacturer.

(a) Blue A (b) Blue B (c) Green A (d) Green B (e) Red A (f) Red B (g) NIR

Figure 5.4: The seven individual spectral bands of the sample image.

5.2.1 Multispectral white balance

We start by leveling the images to compensate the spatially uneven light conditions and the varying sensitivities of the channels. For this, we acquired a plane white surface in several positions in the measurement acquisition conditions. From the white acquisitions we create a mean image for each spectral band. This white multispectral acquisition represents the light conditions. The averaging of the several takes smoothens out any irregularity that may describe the surface itself.

This white multispectral data is used to normalize the acquisitions by applying the following simple algorithm; where A denotes an 8-bit multispectral acquisition and W the corresponding white multispectral acquisition:

for every spectral plane k **do**
  for every image row i **do**
    for every image column j **do**
      $A(i,j,k) \leftarrow A(i,j,k) \times 255/W(i,j,k)$
      if $A(i,j,k) > 255$ **then**
        $A(i,j,k) = 255$
      end if

(a) Blue A (b) Blue B (c) Green A (d) Green B (e) Red A (f) Red B (g) NIR

Figure 5.5: The seven individual spectral bands of the sample image after the white balance.

end for
end for
end for

The result for our test image is shown figure 5.5. The individual channels are globally better balanced, though the saturated channel still stands out.

In dark images, this normalization introduces some noise: a pair of pixels with similar values can be mapped to two values with a large gap. This creates a grainy image. This effect can be compensated by applying a 3×3 mean filter to each normalized channel. By smoothening the features this can attenuate some fine details in the representation, but it also reduces the graininess of the image.

When digitizing several areas of a given objects, a set of white acquisitions per position is necessary to compensate the uneven illumination over the full object. However, when using a ring-light and a fixed acquisition distance and orientation, the acquisition of a single set of white surfaces is neccessary, since the illumination is stable.

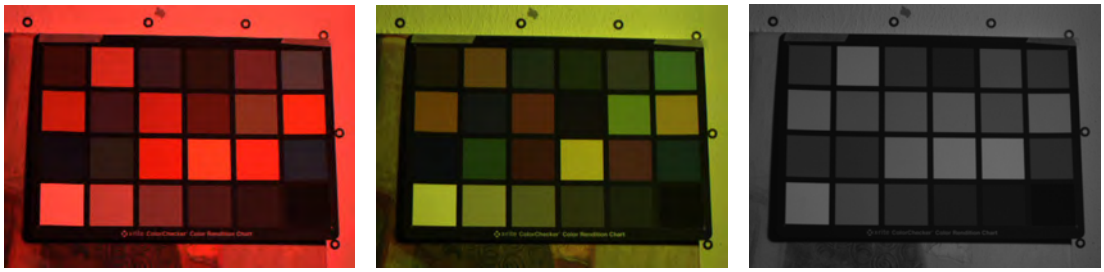
5.2.2 Reflectance estimation

As explained in chapter 2, a series of multispectral images does not directly represent the surface reflectance. The spectral estimation presented here is based on previous work at the Le2i laboratory, in particular Alamin Mansosuri's PhD [Mansouri, 2005]. This reflectance estimation is based on a heteroassociative neural network algorithm (see [Mansouri, 2005, p.109] as well as [Mansouri *et al.*, 2005b; Sanchez *et al.*, 2005]).

The spectral calibration requires the acquisition of several patches of varying and known reflectance. We use the GretagMacbeth ColorChecker (shown in figure 5.6) to obtain 24 diffuse reference patches. We place the ColorChecker in the scene in the same acquisition conditions as the object under study. The resulting acquisitions are shown in figure 5.7. We also perform the white balance presented previously on these acquisitions. Then, we compute the average multispectral value of the 24 patches in these acquisitions.

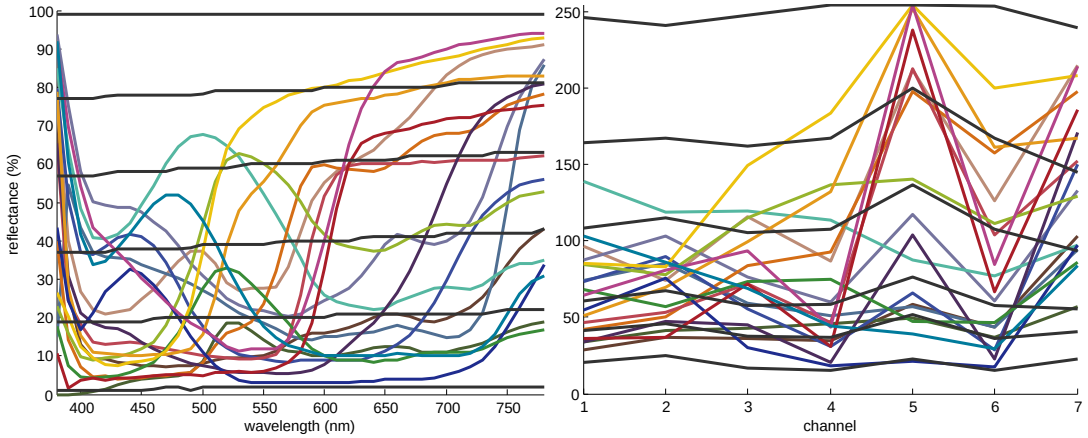
On the other hand, we obtain the theoretical reflectance of each patch using a portable spectrometer (JETI specbos 1200). This provides us with the reflectance spectrum of each patch from 380 nm to 1060 nm with a 1 nm step. We average this data by 10 nm steps between 380 nm and 780 nm. The averaging reduces the noise of the theoretical spectra while the truncation reflects our chosen output spatial range. The reference reflectance of the 24 patches is compared with the mean multispectral acquisition values of the ColorChecker for our sample image in figure 5.8. Notice the exaggerated values and saturation of channel 5.

Figure 5.6: Gretag Macbeth ColorChecker.



(a) CCD A (b) CCD B (c) CCD NIR

Figure 5.7: ColorChecker acquired by each CCD. A and B in the visible, C in the near infrared.



(a) (b)

Figure 5.8: ColorChecker acquisitions. (a) Reflectance measured by a spectrometer and smoothened over 10 nm. (b) Values from the 8-bit multispectral acquisition shown figure 5.7.

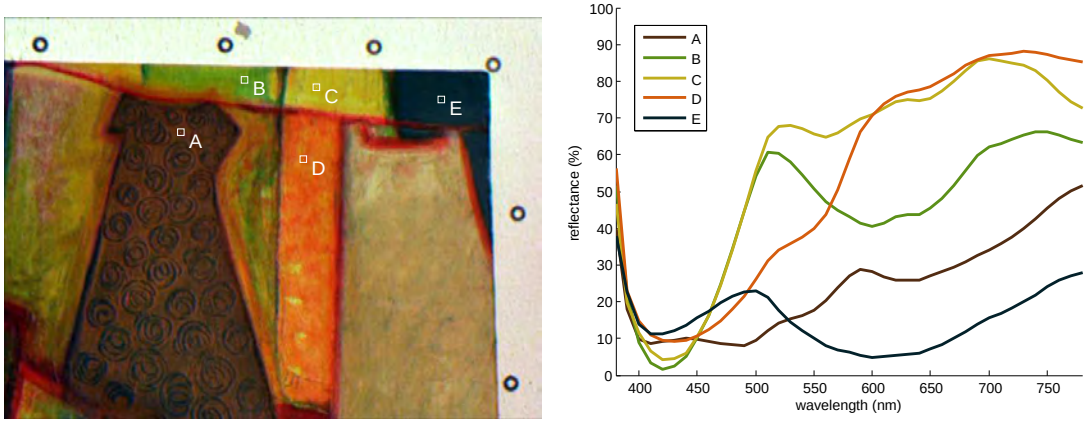


Figure 5.9: A few reflectance spectra reconstructed from the sample image.

These datasets are used to create a neural network that can then process our input data, as explained in chapter 2 (section 2.1.4, page 17). This provides us with a reflectance cube storing the reflectance of each pixel sampled every 10 nm between 380 nm and 780 nm. Figure 5.9 shows a few of these spectra. Though this data is very useful for fine data analysis, it is not easy to apprehend it its integrity. We now present various color representations of this reflectance data.

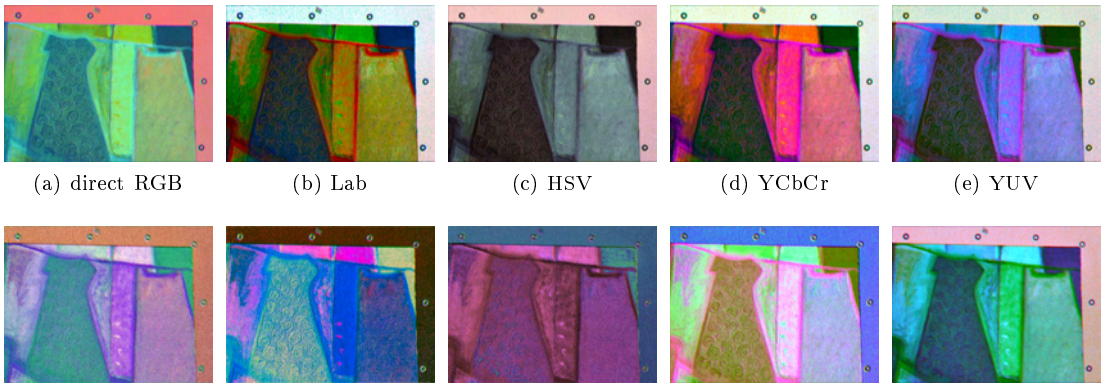
5.2.3 Color representations

Using the estimated reflectance of each pixel, it is possible to create color representations that are accurate from a colorimetric point of view. Each spectrum is multiplied by the spectrum of a given illuminant and by a standard observer. This provides us with a triplet of values in the XYZ color space. This triplet is then covered to RGB for representation purposes. The necessary colorimetric data is available on the website of the International Commission on Illumination (CIE — Comission Internationale de l'Éclairage [*CIE*]). Using the 1931 observer and standard illuminant D65, which simulates daylight, we obtain the image used in figure 5.9.

It is also possible to create false color representations that can ease certain interpretations of the spectral data. Such representations are often based on either a Principle Component Analysis (PCA) or an Independent Component Analysis (ICA) of the spectral data. The first three resulting bands are used to create an image. Many different renderings can be produced depending on which color space is used to interpret the output of the PCA/ICA. The three channels can be directly assigned as RGB channels, but it is also possible to interpret them as values from another color space. A few of examples are shown figure 5.10.

5.3 3D – multispectral registration

We now present the tracking and registration accuracy achieved when both the Atos III and the FluxData multispectral camera are used to digitize our two test objects. The target registration accuracy for both acquisition systems imposed by the acquisition distance and by our goal of a half-pixel tracking accuracy is given table 5.1. For the cross-stitch canvas, the reduced



(a) direct RGB (b) Lab (c) HSV (d) YCbCr (e) YUV

(f) direct RGB (g) Lab (h) HSV (i) YCbCr (j) YUV

Figure 5.10: False color representations of the sample image. (a-e) based on a PCA. (f-j) based on an ICA.

Table 5.1: Target accuracy for each acquisition configuration.

Acquisition system	Acquisition distance (mm)	Acquisition dimensions (mm×mm)	Target accuracy Spatial (mm)	Angular (mrad)
FluxData multispectral camera	500	97×130	0.099	0.198
	700	137×183	0.139	0.198
Gom Atos III digitization system	760	500×500	0.120	0.158

acquisition distance when using the multispectral camera has an influence on the target spatial accuracy.

This section successively presents the accuracy of the individual calibrations; the accuracy of the tracking of the target frame; the 3D registration and the projection of the multispectral acquisitions on the 3D meshes.

5.3.1 Individual calibration accuracy

Now that we use two acquisition systems, the number of necessary calibrations increases. These calibrations and acquisitions are the following:

- Interior orientation of the tracking cameras
- Interior orientation of the multispectral camera*
- Calibration of the 3D digitization system
- Target frame calibration
- Exterior orientation of the tracking cameras
- Multispectral camera to target frame orientation*
- Multispectral acquisitions with simultaneous tracking*

Table 5.2: Accuracy of the individual calibrations, compared to the simulation values.

Calibration	Experiments		Simulations		Unit
	Cross-stitch	Bas-relief	Realistic	Best-case	
Tracking cameras I.O.	0.029		0.100	0.033	pixel
Multispectral camera I.O.	0.035	0.072	—		pixel
Target frame calibration	0.011	0.017	0.050		mm
Tracking cameras E.O.	0.014	0.011	0.030	0.010	mm
	0.016	0.012	0.040	0.020	mrad
Target frame to Gom	0.029	0.066	—		mm
Atos III orientation	0.056	0.106	—		mrad
Target frame to multispectral	0.924	0.126	—		mm
camera orientation	3.156	0.428	—		mrad

- 3D digitization system to target frame orientation
- 3D digitizations with simultaneous tracking

An asterisk highlights the three steps introduced by the use of an additional acquisition system. This section examines the corresponding calibration results in greater detail. The accuracy of the individual tracking and calibration procedures is given in table 5.2. When available, the equivalent values from the simulations are given for comparison.

Sensor calibrations

The same interior orientation of the tracking cameras was used for the two digitizations, which were performed the same week. As previously, this calibration is more accurate than the values used for the best-case scenario simulations.

The focus of the multispectral camera is adapted to each set of acquisitions, so we must repeat the calibration. The three CCDs of the FluxData multispectral camera are independently calibrated using a much smaller calibration plate than that used for the calibration of the tracking cameras. Table 5.2 gives the worst accuracy of all three sensors. There is quite a lot of variation in the accuracy of the calibration of the varying sensors, in particular for the calibrations performed for the digitization of the bas-relief. In this case CCD B is the least accurately calibrated (value given in table of 0.072 pixel) but the infrared sensor is much more accurately calibrated (0.019 pixel). The accuracy of the calibration of CCD A is between these two values (0.052 pixel). These variations denote that some of the calibration acquisitions were performed in areas that are in focus for one sensor, and slightly out of focus for the other two.

Though this was not the case for the bas-relief digitization, the calibration of the multispectral cameras can be performed with an accuracy close to that of the tracking cameras.

Calibration of the target frame

The calibration of the target frame is as accurate as in the previous chapter, and much better than the values used during the simulations.

Exterior orientation of the tracking cameras

The exterior orientation of the tracking cameras is much more accurate than the realistic simulation results and almost as good as the best-case simulation results.

Orientation between the target frame and the acquisition systems

The accuracy of the calibration between the Gom Atos III and the target frame was presented in the previous chapter (section 4.3.4). The least accurate calibration is due to the single relative position of the target frame and target-covered object during the calibration.

The orientation between the target frame and FluxData multispectral camera is calculated using three different relative positions of the target frame and calibration object. The discrepancy between the small field of view of the multispectral camera and the large tracking frame are the cause of the disappointing accuracy of the orientation of the tracking frame in the system defined by the multispectral camera. This orientation can be improved if the multispectral camera observes a greater number of targets. During the cross-stitch digitization each image could only see four or five targets of the calibration object. For the digitization of the bas-relief we improved this setup and seven to twelve targets were visible per position, resulting in a much improved calibration.

5.3.2 Tracking accuracy

The relative position of the tracking cameras and the target frame is illustrated figure 5.11 for the cross-stitch digitization and figure 5.12 for the bas-relief acquisitions. The spatial and angular accuracy of the frame tracking for each of these positions is given figure 5.13 in the case of the cross-stitch canvas and figure 5.14 for the bas-relief. The achieved tracking accuracy is compared to the simulation results and to the tracking accuracy goals.

Cross-stitch canvas

We only reach the simulation results for the angular accuracy during the Atos III digitizations. The spatial accuracy is also much better for these five positions, rarely exceeding 10 % more than the spatial accuracy of the best-case simulations. As shown figure 5.11, the target frame is closer to the cameras during the Gom acquisitions. These positions closely resemble those used for the simulations. It is thus not surprising that the tracking accuracy better corresponds to the simulation results.

If we compare the achieved accuracy to the tracking goal we notice that the tracking accuracy is always better than our goal. As expected from the simulations, we have no difficulty reaching our target spatial accuracy. Our worst spatial value (0.026 mm for FD3) is more than three times better than the spatial accuracy goal of 0.099 mm.

The least well tracked positions are FD3, FD4, FD9, FD10, and FD15. The first four of these positions are those of the third column from the left in figure 5.11. For these positions we have a perspective view of all targets on the left and right side of the frame. The ensuing tracking is less accurate, even though there is a comparable number of targets detected by each camera compared to the other positions.

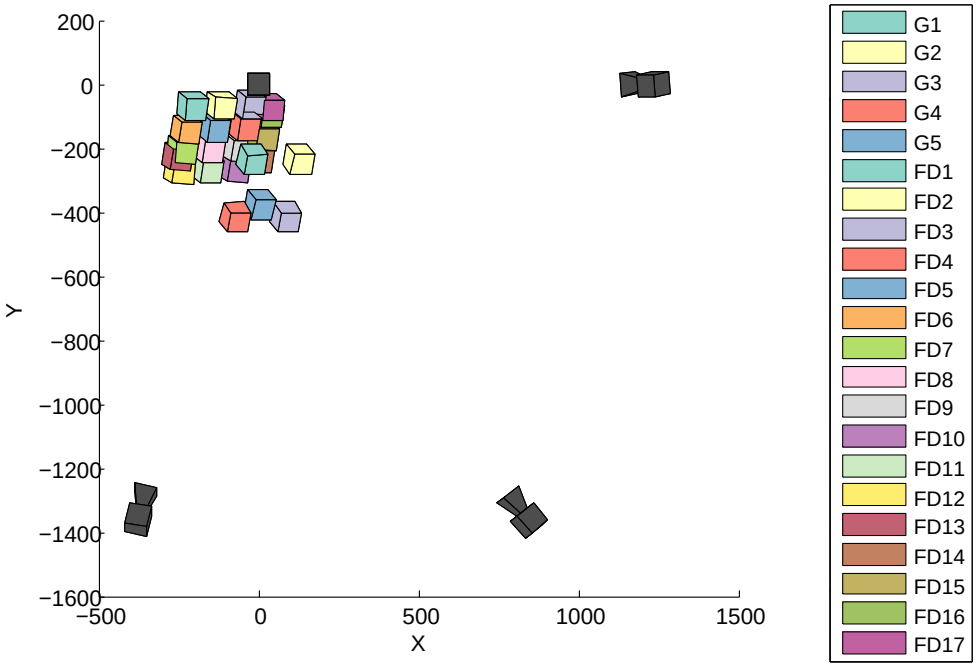


Figure 5.11: Relative position of the tracking cameras and target frame during the digitization of the cross-stitch canvas.

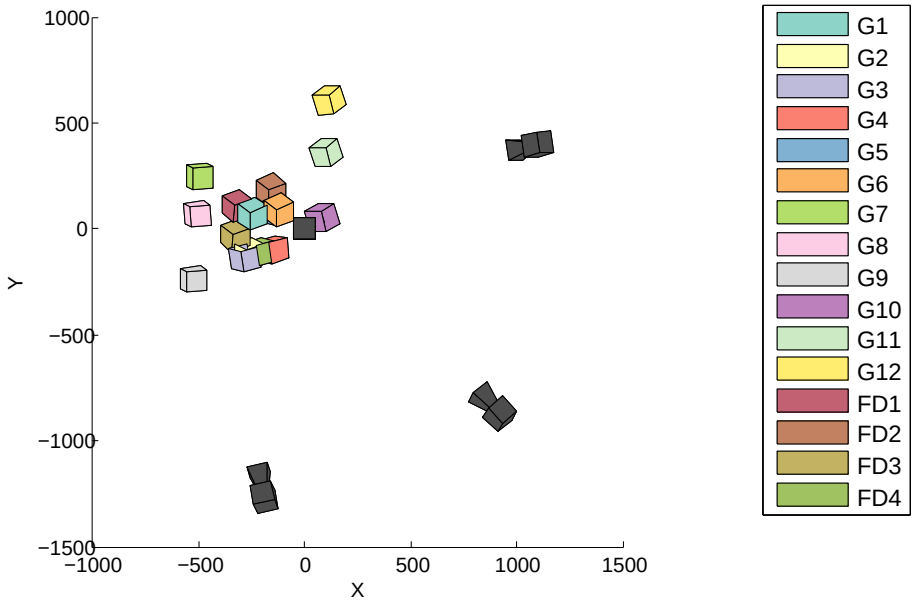


Figure 5.12: Relative position of the tracking cameras and target frame during the digitization of the plaster bas-relief.

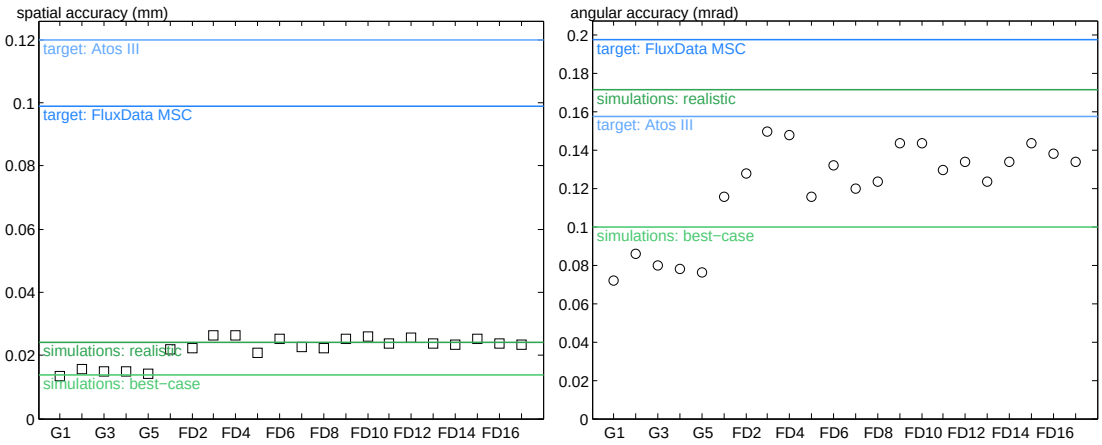


Figure 5.13: Spatial and angular tracking accuracy during the cross-stitch digitization, compared to the simulation results and to the target accuracy. The target FluxData spatial accuracy is lower than it is for all other figures of this type, due to the reduced acquisition distance.

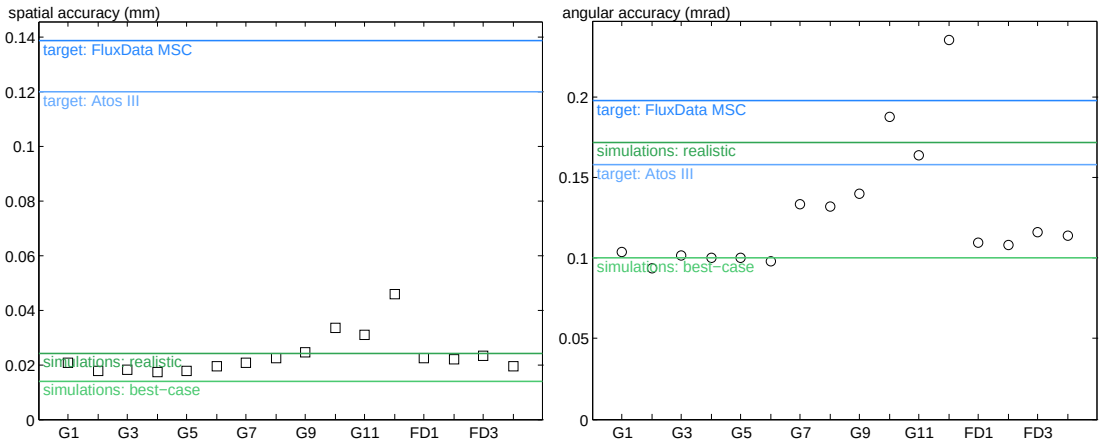


Figure 5.14: Spatial and angular tracking accuracy during the bas-relief digitization, compared to the simulation results and to the target accuracy.

Bas-relief

During the bas-relief digitizations, the positions of the target frame when using the Atos III and FluxData camera are less discernible. The resulting tracking accuracy is also much closer for the two acquisition systems. As previously, all positions are tracked with an accuracy well within the spatial target accuracy.

Figure 5.14 shows that the angular tracking accuracy is satisfying for all but three positions: G10, G11 and G12 are less accurately tracked than the target angular accuracy for the Gom Atos III. These are the three positions in a column on the right of figure 5.12. These three positions are not only off-center, they are also very tilted, creating a perspective view on the targets.

Positions G7, G8 and G9, though satisfying, also stand out in the angular tracking accuracy results. These three tracking positions are the left column of figure 5.12. They are off-center, but not as tilted as the three following positions.

For the multispectral acquisitions, the tracking accuracy is quite stable and is better than the realistic simulation results angularly and close to the best-case simulation results spatially.

5.3.3 3D registration

Cross-stitch canvas

The frame of the cross-stitch canvas is difficult to digitize with the Gom Atos III, due to the shiny paint that covers it. The intricate decorations of the interior of the frame are particularly challenging to digitize and were only partially acquired.

Figure 5.15 shows the registration of the five meshes of the cross-stitch canvas and frame, as well as the final model. There are some holes remaining in this final model, particularly in the area representing the frame, but the cross-stitch canvas is almost fully acquired. There are no visible discontinuities in the 3D model thought these would be easily visible on the exterior and interior edges of the frame.

The theoretical accuracy of the final registration is the sum of the tracking accuracy and the target frame to acquisition system orientation. In the case of the 3D data, the theoretical spatial accuracy is thus the sum of the spatial accuracy of the least well tracked position of the Gom Atos III, position G2, (0.016 mm) and of the spatial accuracy of the target frame to Gom Atos III orientation (0.029 mm). This theoretical spatial accuracy of 0.045 mm is well below our target accuracy of 0.09 mm. The theoretical angular accuracy is 0.158 mrad (0.0860 mrad, angular accuracy of G2, plus 0.072 mrad, spatial accuracy of the target frame to Gom Atos III orientation). This is exactly our target angular accuracy. We reach our target 3D registration accuracy somewhat narrowly and based on the assumption that the calibration of the 3D digitization system introduces negligible errors.

Bas-relief

The registration of all twelve meshes representing the bas-relief is shown in figure 5.16. Figure 5.17a shows a close-up of the areas where defects in the registration are visible. Mesh G5 (in pale blue in figure 5.16a) is not well aligned to the surrounding meshes. There are also several

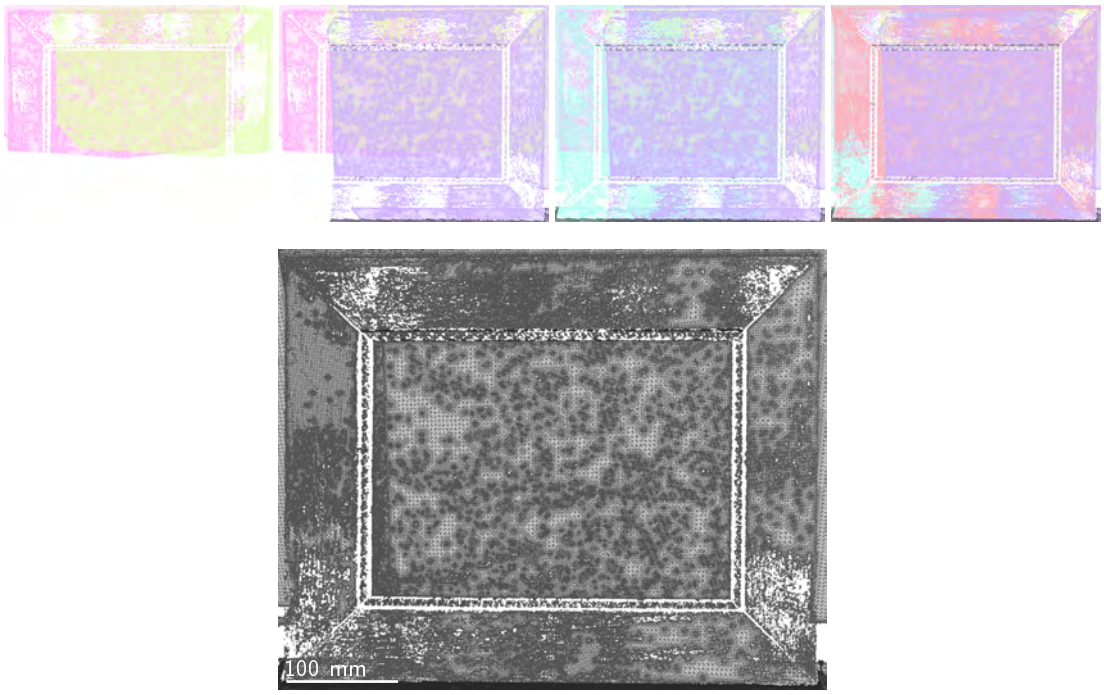


Figure 5.15: 3D registration of five meshes representing the cross-stitch canvas. First row: successive projection of each individually colored mesh. Bottom image: all meshes.

inaccuracies visible on the hand that holds the box. Removing mesh G5 results in an inaccurate alignment of the meshes in the middle of the skirt, as shown figure 5.17b. Also removing meshes G3 and G4 (respectively the lower pink and salmon meshes of figure 5.16a) results in an overall satisfying model, as shown in figure 5.18.

We could expect the meshes we remove to obtain a coherent model to be the least accurately tracked, but it is not the case. What distinguishes these three meshes from the others is their orientation, which is sensibly the same, yet different from all the others (tracking results and accuracy are given in table 5.3). The registration of only these three meshes results in a visually satisfying, though partial, model.This confirms the idea that most of the inaccuracies stem from the unstable fixture between the target frame and the acquisition system.

The registration accuracy of the remaining nine positions can be calculated as previously by summing the tracking accuracy and the accuracy of target frame to Atos III orientation. For the least accurately tracked position (G12) this results in 0.112 mm spatially and 0.342 mrad angularly. The spatial results are still within our target limit, but the angular results are much worse, since the tracking accuracy was already above the limit and our orientation is not particularly satisfying. If we consider the other six remaining positions, G9 is the least accurately tracked and the resulting angular registration accuracy is 0.246 mrad angularly, which is still more than our target. Even if we consider positions G1, the angular registration accuracy is above 0.2 mrad, which is close to thirty percent higher than our target accuracy of 0.158 mrad.

To reach our target registration accuracy, we need an orientation between the target frame and the Atos III that is as accurate as that performed for the digitization of the cross-stitch

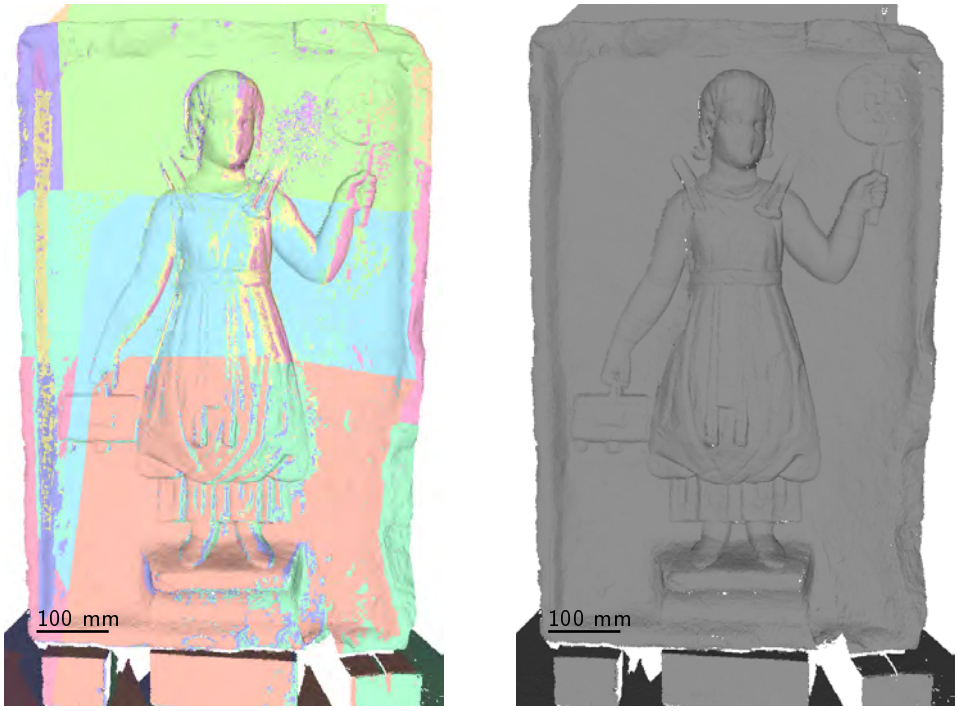
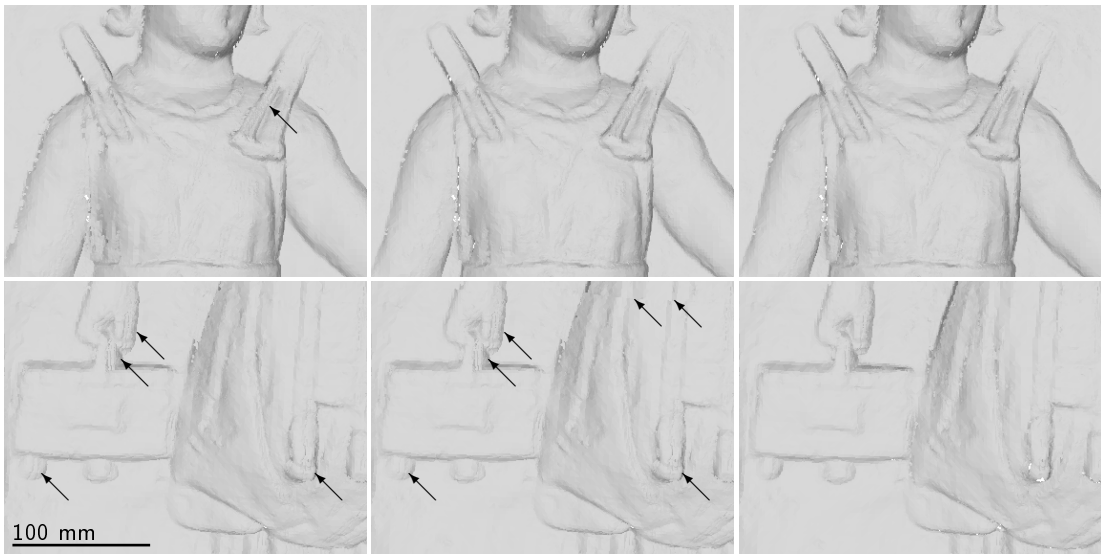


Figure 5.16: 3D registration of the twelve meshes representing the bas-relief. (a) Each mesh is individually colored. (b) All meshes in grey.



(a) (b) (c)

Figure 5.17: Details of the 3D registration meshes representing the bas-relief. Mis-alignments are highlighted by an arrow. (a) All twelve meshes. (b) Without mesh G5. (c) Without meshes G3, G4 and G5.

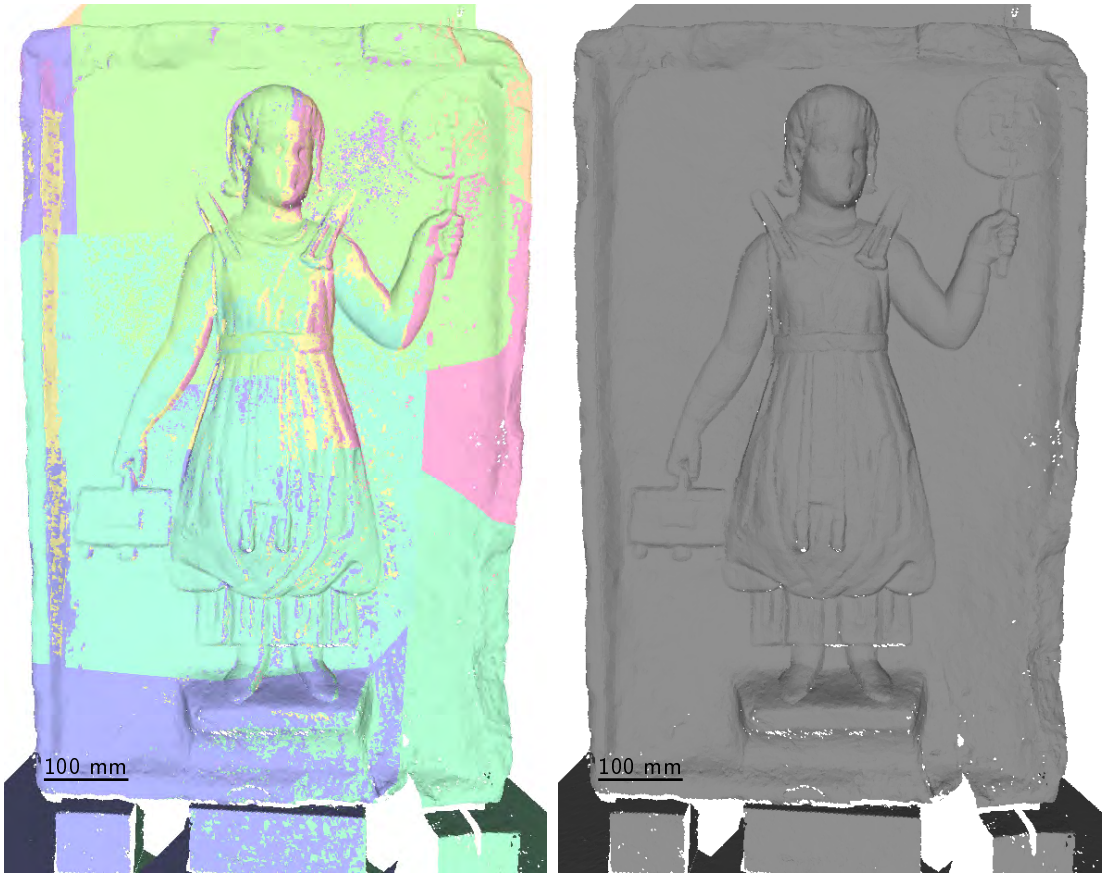


Figure 5.18: 3D registration of nine meshes representing the bas-relief. Meshes G3, G4 and G5 have been removed compared to figure 5.16. (a) Each mesh is individually colored. (b) All meshes in grey.

Table 5.3: Tracking results and accuracy during the bas-relief digitization. Position and orientation values are significantly rounded.

	Position			Orientation			Accuracy	
	X (mm)	Y (mm)	Z (mm)	Ω (rad)	Φ (rad)	K (rad)	Spatial (mm)	Angular (mrad)
G1	−277	90	−1986	0.522	0.649	−0.004	0.021	0.104
G2	−290	−98	−2124	0.453	0.648	0.023	0.018	0.094
G3	−310	−131	−2083	0.133	0.583	0.168	0.018	0.102
G4	−175	−92	−2148	0.135	0.584	0.166	0.017	0.100
G5	−164	91	−2017	0.134	0.583	0.166	0.018	0.100
G6	−155	111	−1978	0.593	0.624	0.005	0.019	0.098
G7	−510	257	−1580	0.179	0.307	0.039	0.021	0.134
G8	−521	78	−1708	0.178	0.307	0.039	0.023	0.132
G9	−539	−229	−1928	0.179	0.307	0.039	0.024	0.140
G10	45	54	−2370	0.006	0.831	0.405	0.034	0.188
G11	63	359	−2152	−0.034	0.879	0.461	0.031	0.164
G12	78	606	−1977	−0.033	0.878	0.461	0.046	0.236
FD1	−350	131	−2043	0.590	0.602	−0.015	0.023	0.110
FD2	−187	205	−2145	0.593	0.602	−0.018	0.022	0.108
FD3	−358	−9	−2145	0.590	0.602	−0.015	0.023	0.116
FD4	−240	−104	−2128	0.130	0.642	0.168	0.020	0.114

canvas and a tracking accuracy that is close to the best-case simulation results.

5.3.4 Multispectral projection

The projection of the multispectral data on the registered 3D meshes is not as straightforward as the 3D registration itself. As stated in chapter 3, the tracking provides us with T_{C_0,C_F}, the transformation between the tracking camera coordinate system and the target frame; while the target frame to multispectral camera calibration provides us with $T_{C_F,C_{S_i}}$. We can thus calculate the coordinates of point A in the tracking camera coordinate system from its coordinates in the acquisition coordinate system using:

$$A|_{C_0} = T_{C_0,C_F} \cdot T_{C_F,C_{S_i}} \cdot A|_{C_{S_i}} \quad . \tag{5.1}$$

We assume that acquisition system i is an imaging system such as a multispectral camera, and not a 3D digitization system. In this case $C_{S_i}, (O_{S_i}, \vec{x}_{S_i}, \vec{y}_{S_i}, \vec{z}_{S_i})$ is defined such that O_{S_i} is the optical center of the camera; $(\vec{x}_{S_i}, \vec{y}_{S_i})$ define the image plane; $\vec{z}_{S_i}$ is collinear to the optical axis. In this setup if A is a point of the surface under study captured by the acquisition system sensor, we do not know its coordinates $A|_{C_{S_i}}$ in the sensor coordinate system. What we do have is A' image of A on the sensor; A' has only two coordinates $x_{a'}$ and $y_{a'}$.

Using the pinhole camera model ([Morvan, 2009, section 2.2], for example) we obtain the

coordinates of point A given $x_{a'}$ and $y_{a'}$ using:

$$A|_{C_{S_i}} = \lambda \underbrace{\begin{bmatrix} & K^{-1} & \\ 0 & 0 & 1 \end{bmatrix}}_{G} \begin{pmatrix} x_{a'} \\ y_{a'} \\ 1 \end{pmatrix} \tag{5.2}$$

with

$$K = \begin{pmatrix} f & \tau & o_x \\ 0 & \eta f & o_y \\ 0 & 0 & 1 \end{pmatrix} .$$

f is the focal length of the camera, τ represents the slant of the pixels ($\tau = 0$ for rectangular pixels), η is the pixel height/length ratio ($\eta = 1$ for square pixels), $(o_x, o_y)^T$ are the coordinates of the principal point offset and λ is a scaling factor. If the distance Z between the sensor and the object is known then $\lambda = Z$.

For each pixel $(x'_a, y'_a)^T|_{C_S}$ we can thus calculate the possible positions of point A in coordinate system C_0 using equations 5.1 and 5.2:

$$A|_{C_0} = \lambda \cdot T_{C_0,C_F} \cdot T_{C_F,C_{S_i}} \cdot G \begin{pmatrix} x_{a'} \\ y_{a'} \\ 1 \end{pmatrix} . \tag{5.3}$$

The camera model can be extended to account for lens distortion, or corrected images can be used directly. The distortion parameters, as well as the precise values of f, τ, η and $(o_x, o_y)^T$ are obtained from the optical calibration of the acquisition system.

In practice the projection is currently performed by tracing the rays that exit the multispectral camera origin and pass through the center of each pixel (ray defined by equation 5.3). The intersection of each ray with the 3D model is matched to the corresponding pixel. For the remaining faces of the model that do not have any pixel correspondence, the intersection between image plane and of the segment joining the center of the face and the camera origin is calculated. When this intersection matches a pixel, this pixel is assigned to the face of the model. To enable different representations of the multispectral data on the 3D model, we assign the coordinates of the pixels to the faces of model, and not directly the pixel values. This way, true color images, single band greyscale representations and false-color PCA renderings of the same multispectral data can all be projected identically on the 3D model.

Cross-stitch canvas

Figure 5.19 shows the full registration of the 3D and multispectral acquisitions. For these acquisitions we only used the six channels in the visible range from the FluxData multispectral camera. The seventeen images are successively projected on the 3D model in the order they were acquired. Areas captured several times are simply hidden by the following acquisitions.

Visual examination of the 3D model presents a horizontal inaccuracy. Manually selecting two points the mesh shows this inaccuracy to be of approximately 0.6 mm between the multispectral and 3D data, highlighted figure 5.20. There is no noticeable registration inaccuracy in the

Figure 5.19: 3D – multispectral registration representing the cross-stitch canvas.

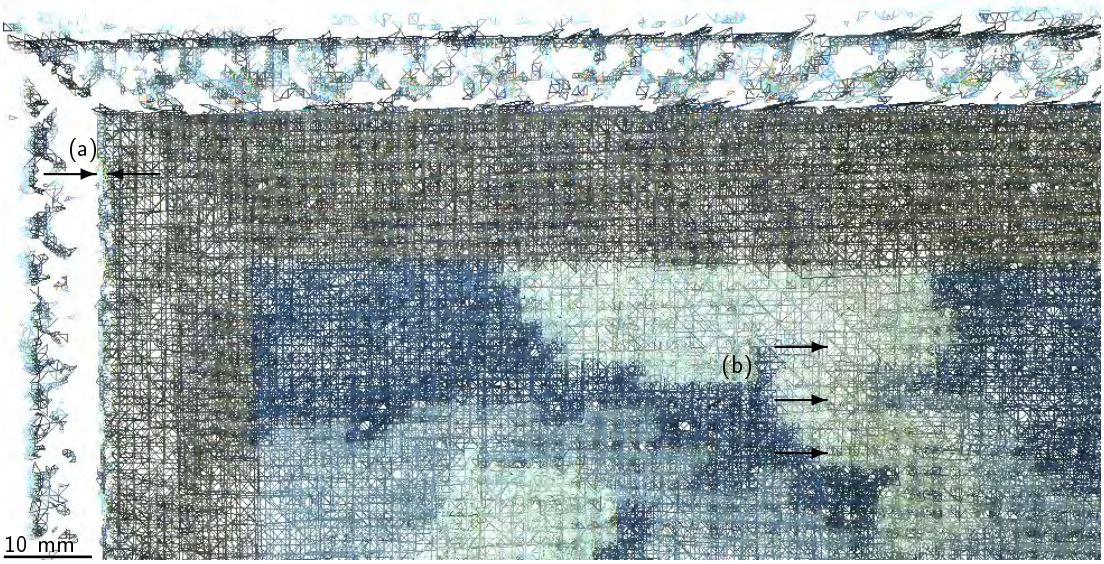


Figure 5.20: Closeup of the multispectral / 3D registration of the cross-stitch canvas datasets. (a) shows the inaccuracy of the 3D to multispectral registration while (b) highlights the seamless multispectral registration.

vertical direction. The 3D model and the images are well aligned both along the edge of the cross-stitch and along the edge of the engraving.

The theoretical spatial accuracy is only 0.950 mm: 0.026 mm, the accuracy of the least well tracked position of the multispectral camera, position FD4, plus 0.924 mm, spatial accuracy of the target frame to multispectral camera orientation. This inaccuracy is thus mostly due to the inaccuracy of the target frame to multispectral camera orientation. We observe the same problem on the theoretical angular accuracy; of the total 3.306 mrad, 3.156 mrad are due to the the target frame to multispectral camera orientation, while only 0.150 mrad come from the tracking accuracy (position FD3). This explains why we have a good multiview registration of the multispectral data, even though the multimodal registration could be improved. The accuracy of the multispectral registration on the 3D model could be greatly improved with a better calibration between the multispectral camera and the target frame.

Another important factor for the accurate registration of the multispectral data on the 3D model is the accuracy of the camera interior orientation. The accuracy of the focal length ensures that the image is correctly scaled (the accuracy of the focal length is better than a micrometer). The principle point offset must also be taken into account, or the image will be translated with respect to 3D model. Though this effect is present here, we believe it is a cause of the movement of the acquisition system in the target frame and not of an inaccurate camera calibration: the principle point offset calibrated with an accuracy of a few micrometers. It is also most important to compensate the distortion of the lens, though this is not a visible problem here.

Figure 5.21: 3D – multispectral registration of the bas-relief data.

Bas-relief

Figure 5.21 shows the projection of the images produced from the multispectral acquisitions, on the 3D model made of nine registered meshes (without G3, G4 and G5). The registration is visually satisfying, though this is easier to assess when interacting with the data. When shading and unshading the model it is clear that the shadows in the images match those in the 3D data.

The theoretical registration accuracy is 0.149 mm spatially (0.023 mm + 0.126 mm) and 0.544 mrad angularly (0.116 mrad + 0.428 mrad). Spatially, we are are only slightly worse than the target registration accuracy of 0.139 mm but angularly we are far from the target 0.198 mrad. The accuracy of the calibration between the target frame and FluxData multispectral camera must be further improved.

5.4 Conclusion

Using a photogrammetric setup based on four tracking cameras we were able to successively track the position and orientation of a fringe projection digitization system and a multispectral camera with an accuracy better than 0.05 mm spatially and, in all but three cases, 0.150 mrad angularly. This tracking was performed during the digitization of a cross-stitch canvas and during the digitization of a bas-relief. Five 3D digitizations and seventeen multispectral acquisitions were performed in the first case while twelve 3D digitizations and four multispectral acquisitions of the second object were performed. The tracking results were used to register the 3D acquisitions to project the multispectral acquisitions on this model.

There are nine necessary calibration and acquisition steps, though some can be performed in advance. Still, it is possible to perform all these steps in a single eight hour working day. To guarantee a registration accuracy within our target values we must ensure that the tracking accuracy is close to the best-case simulation results. It is also important that the target frame to Atos III calibration accuracy is close to that obtained during the cross-stitch digitization. The accuracy of the target frame to FluxData multispectral camera also has to be further improved. An accuracy of 0.070 mm spatially and 0.100 mrad angularly would be sufficient. This calibration accuracy is easily reached for the Atos III to target frame calibration, but is more difficult to achieve for the FluxData camera, given its small field of view.

As in the previous chapter, we have noticed that the fixture between the target frame and multispectral camera is not as stable as we need it to be. Nevertheless, these experimental results show that our tracking method is adapted not only for the registration of 3D datasets, but also for the integration of multispectral texture on 3D models.

Chapter 6

Improved Reflectance Estimation

The spectral reconstruction presented in the previous chapter is based on the assumption that all pixels of a multispectral acquisition are at a fixed angle with respect to the multispectral camera. Since we analyze objects with measurable surface variations, this assumption is false. However, our registration procedure provides us with the orientation of the surface and the position of the acquisition devices. If we also know the position of the light sources, we could use this data to improve the spectral reconstruction, by taking into account the angle between the light and the surface.

This chapter opens by presenting the reflectance correction method we propose and related work, in particular other illumination models. We then describe the experimental configuration and present the corresponding tracking and registration results. Finally, we apply the reflectance correction to this data and present the intermediate and final results.

6.1 Method

6.1.1 Reflectance correction principle

We base the correction of the estimated reflectance on the Phong reflection model [Phong, 1975], an empirical model of local illumination. The Phong reflection model states that the illumination of a surface point is the sum of three components: ambient, diffuse and specular. The illumination I_P of surface point P in the presence of M lights is thus defined as:

$$I_p = \underbrace{r_a i_a}_{\text{ambient}} + \sum_{m=1}^{M} \left(\underbrace{r_d \left(\hat{L}_m \cdot \hat{N} \right) i_{m,d}}_{\text{diffuse}} + \underbrace{r_s \left(\hat{R}_m \cdot \hat{V} \right)^{\alpha_n} i_{m,s}}_{\text{specular}} \right) \tag{6.1}$$

with the notations given in table 6.1 and illustrated in figure 6.1. There is a hidden wavelength dependency for many parameters of this equation.

Our previous reflectance estimation assumes that the ambient term is negligible and that there are no specular highlights in the multispectral acquisitions. This is a common hypothesis when working with reflectance and illumination models, though the specular component is often

Table 6.1: Phong reflectance model parameters and notations

Variable	Description	Dependence
M	Total number of lights	
m	Light index	
$\hat{V}$	Unit vector from surface to eye	
$\hat{L}_m$	Unit vector from surface to light m	
$\hat{N}$	Unit surface normal	
$\hat{R}_m$	Unit reflected vector of light m	$\hat{R}_m = 2\left(\hat{L}_m \cdot \hat{N}\right)\hat{N} - \hat{L}_m$
λ	Wavelength	
$i_{m,s}$	Specular component of light m	
$i_{m,d}$	Diffuse component of light m	
i_a	Ambient component of light m [a]	
r_s	Specular reflection [b]	λ, material
r_d	Diffuse reflection [c]	λ, material
r_a	Ambient reflection [d]	λ, material
α	Shininess constant [e]	material

[a] sometimes computed as a sum of contributions from all light sources
[b] ratio of reflection of the specular term of incoming light
[c] ratio of reflection of the diffuse term of incoming light — the Lambertian reflectance
[d] ratio of reflection of the ambient term present in all points in the scene rendered
[e] large for surfaces that are smooth and mirror-like, in which case the specular highlight is small

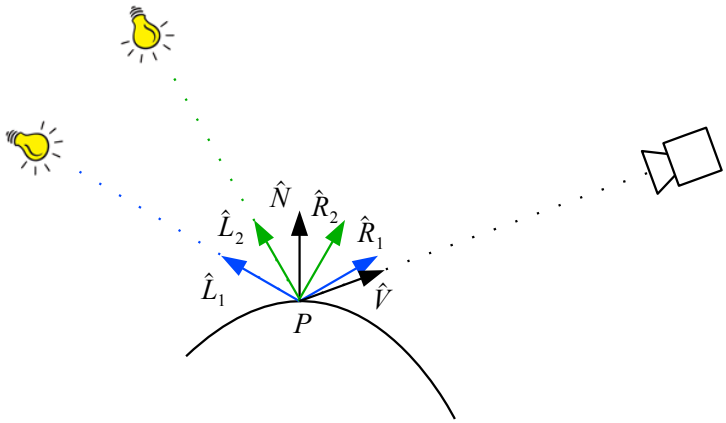


Figure 6.1: Phong reflection model notations illustrated for two light sources.

filtered out using polarizers [Debevec *et al.*, 2000; Malzbender *et al.*, 2001]. We still assume that the ambient term is negligible, but we acknowledged the existence of a specular component. For this, we assume the following data is available in a single coordinate system:

- Multispectral acquisitions of a surface area.
- Corresponding 3D structure.
- Position of light sources.

The specular component of the illumination of surface point P, given by equation 6.1, is:

$$I_{p,\text{specular}} = \sum_{m=1}^{M} \left(r_s \left(\hat{R}_m \cdot \hat{V} \right)^{\alpha_n} i_{m,s} \right)$$

The only variables available to us are $\hat{R}_m$ and $\hat{V}$. However, the specular component generally has a very strong but very localized contribution. It is sufficient that we detect the areas where the specular component is non-negligible, recognizing that we are unable to estimate the surface reflectance in these points using the model described by equation 2.1 from chapter 2.

We thus devise a threshold T close to 1. If $T < \hat{R}_m \cdot \hat{V}$ then the illumination has a strong specular component and the reflectance can not be estimated in this point. If $T > \hat{R}_m \cdot \hat{V}$ then the specular component is negligible and the illumination of point P is given by:

$$I_p = \sum_{m=1}^{M} r_d \left(\hat{L}_m \cdot \hat{N} \right) i_{m,d} \tag{6.2}$$

Highlighting the wavelength dependencies and removing the "diffuse" index, since this is the only remaining component, equation 6.2 can be re-written as:

$$I_p(\lambda) = \sum_{m=1}^{M} r(\lambda) \left(\hat{L}_m \cdot \hat{N} \right) i(\lambda)_m \ . \tag{6.3}$$

Single light source

If there is a single light source, the illumination of point P is:

$$I_p(\lambda) = r(\lambda) \left(\hat{L} \cdot \hat{N} \right) i(\lambda) \ . \tag{6.4}$$

L and N can be calculated while $r(\lambda)$ is what we want to estimate. Integrating equation 6.4 with our multispectral model (equation 2.1 from chapter 2) gives us the response of channel k for a given illumination angle:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} r(\lambda) \left(\hat{L} \cdot \hat{N} \right) i(\lambda) o\,(\lambda)\, c\,(\lambda)\, t_k\,(\lambda)\, d\lambda + \eta_k \ .$$

The corresponding integrated model is illustrated figure 6.2.

As in chapter 2, we assume all the noise is removed by pre-processing and group $i(\lambda)$, $o(\lambda)$, $c(\lambda)$ and $t_k(\lambda)$ in a single term $S(\lambda)$ representing the response of the multispectral camera

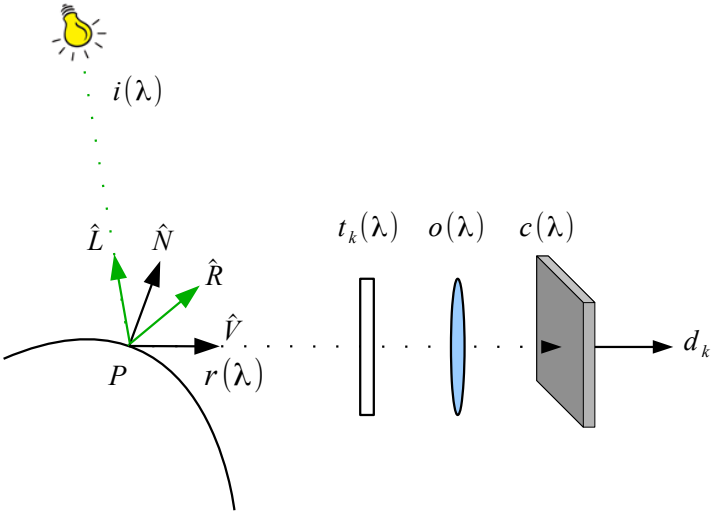


Figure 6.2: Phong reflection model illustrated for a single light source when the reflectance is acquired from a multispectral camera in emission filtering. The notations are those from chapter 2.

system to the current acquisition settings:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} S_k(\lambda) \left(\hat{L} \cdot \hat{N}\right) r(\lambda)\, d\lambda \ .$$

Which, written in matrix form for all acquired channels becomes:

$$\mathbf{d} = \left(\hat{L} \cdot \hat{N}\right) \ .\mathbf{S}^T\mathbf{r}.$$

We want to find a matrix $\mathbf{P}$ such that

$$\tilde{\mathbf{r}} = \frac{1}{\left(\hat{L} \cdot \hat{N}\right)}\mathbf{P}\ \mathbf{d}\ .$$

We can use a neural network (or any other optimization method) to calculate $\mathbf{Q}$ such that $\tilde{\mathbf{r}} = \mathbf{Q}\ \mathbf{d}$ from a set of patches of known reflectance $\mathbf{r}$ and the corresponding measurements from the multispectral camera $\mathbf{d}$.

We recognize that

$$\mathbf{Q} = \frac{1}{\left(\hat{L} \cdot \hat{N}\right)}\mathbf{P}.$$

We do not digitize the surface during the spectral calibration, so we do not know $\hat{N}$. However, the ColorChecker is positioned orthogonally to the multispectral camera so $\hat{N} \approx \hat{V}$. Since the ColorCheceker is flat, and there is over a meter distance between the surface under study and both the light and the multispectral camera, $\hat{L}$ and $\hat{V}$ can be considered constant for every pixel. We calculate the average of $\left(\hat{L} \cdot \hat{V}\right)$ and note this value a. The calibration thus provides

us with:

$$\mathbf{P} = \left(\hat{L} \cdot \hat{N}\right) \mathbf{Q} \approx a\mathbf{Q} \ .$$

The reflectance of each pixel can thus be calculated using:

$$\tilde{\mathbf{r}} = \frac{a}{\left(\hat{L} \cdot \hat{N}\right)} \mathbf{Q} \ \mathbf{d} \ .$$

Multiple light sources

If there are multiple light sources we obtain the reflectance of the k^{th} channel, as acquired by the multispectral camera by integrating equation 6.3 with our multispectral model (equation 2.1):

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} \sum_{m=1}^{M} r(\lambda) \left(\hat{L}_m \cdot \hat{N}\right) i_m(\lambda) o\left(\lambda\right) c\left(\lambda\right) t_k\left(\lambda\right) d\lambda + \eta_k \ .$$

Once again, we assume all noise is removed by processing. We also group $i_m(\lambda)$, $o(\lambda)$, $c(\lambda)$ and $t_k(\lambda)$ in a multiple term $S_m(\lambda)$, one per light source:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} r(\lambda) \sum_{m=1}^{M} \left(\hat{L}_m \cdot \hat{N}\right) S_m\left(\lambda\right) d\lambda \ .$$

Written in matrix form for all acquired channels:

$$\mathbf{d} = \sum_{m=1}^{M} \left(\hat{L}_m \cdot \hat{N}\right) \mathbf{S_m^T r} \ .$$

The calibration procedure is more complicated now. We have to find m matrices $\mathbf{P_m}$ such that

$$\tilde{\mathbf{r}} = \frac{1}{\left(\hat{L}_m \cdot \hat{N}\right)} \mathbf{P_m d_m} \ .$$

This can be done by performing m calibrations, one per light source. The acquisitions must then also be performed with a single light source at a time; we are then in the single light source configuration.

6.1.2 Related work

Sitnik *et al.* [2012, 2010a,b] study cultural heritage using an integrated shape, color and angular reflectivity measurement system. The goal of these acquisitions is to record visual properties of the object through a 3D model that can be visualized in any lighting and viewing condition. In our case, the goal is to correct our reflectance acquisitions to obtain spectra that describe physical properties of the material which are independent from lighting and viewing conditions. Though the end goal differs, they perform a reflectance correction similar to ours, using the angle between the surface normal and the the viewing direction to correct their multispectral acquisitions [Sitnik *et al.*, 2010b].

Rocchini *et al.* [2001a] also perform related work: they design a structured light projection 3D scanner with multiple fixed light sources. They create an unshaded 3D model by determining the diffuse reflection coefficients (i.e. RGB albedo) for each surface point by inverting the illumination model using images from the multiple illumination configurations. However, they do not state which illumination model they use.

A fine differentiation of the diffuse and specular shading from the surface albedo is performed by Neverova *et al.* [2012], using color images and a depth map, with no prior knowledge of the position of the light source. It is also possible to remove specular highlights from RGB images with no geometrical nor illumination information [Koirala *et al.*, 2009; Tan and Ikeuchi, 2005].

Rendering techniques used in computer graphics are based on different characterizations of the surface reflectance properties. These techniques are generally based on a Bidirectional Reflectance Distribution Function (BRDF) [Nicodemus *et al.*, 1977] that defines how light is reflected from an opaque surface. The function takes an incoming and outgoing light direction, both defined with respect to the surface normal, and returns the ratio of reflected radiance exiting along the output direction to the incident radiance on the surface from the input direction. The input and output light directions are parametrized by an azimuth angle and a zenith angle. The BRDF is also wavelength dependent, though in practice independent 4D BRDFs per channel are used.

The Phong [1975] reflectance model is an empirical model that represents the 4D BRDF and one of the first more general shading models in computer graphics. The original Phong model was not presented in the context of physically-based rendering, but Lewis [1994] derived a physically plausible reflectance function from it, a shader that conserves energy. Another commonly used empirical model was developed by Lafortune *et al.* [1997].

Physical models based on geometrical optics were developed by Torrance and Sparrow [1967], and Cook and Torrance [1982]. These models assume the surface is composed of V-shaped grooves, and incorporate masking and self-shadowing effects. The theoretical model developed by Torrance and Sparrow [1967] takes into account surface roughness and explains some inaccuracies of the Phong reflection model. It accounts for the off-specular peaks that occur when the incident light is at a grazing angle relative to the surface normal. Using the specular reflection model from [Torrance and Sparrow, 1967], Blinn [1977] developed a model that is similar to the Phong reflection model.

Whitted [1980] developed a realistic model based on a recursive ray tracing that checks for blockage by shadowing objects. He *et al.* [1991] derived a physically-based analytic model of the BRDF that accounts for masking and shadowing. The model consists of an ideal diffuse component, a directional diffuse component and a specular mirror component. These analytical expressions depend on two surface roughness parameters and a surface refractive index and can be evaluated numerically.

As for any model, BRDF models only approximate the reflectance of real materials. Most analytic reflection models describe only particular subclasses of materials, though much effort has been spent on improving these models by incorporating the relevant aspects of the underlying physics. These analytic models are often dependent on a series of material parameters that are difficult to measure in practice. Often, some version of a gonio-spectro-reflectometer is

used to acquire BRDF samples and the acquired data is fitted to the selected analytical model.

Dana *et al.* [1999] introduced the Bidirectional Texture Function (BTF) to describe materials with spatially varying reflectance. Two additional variables are necessary to parametrize the surface, making the BTF a function of six parameters per wavelength. The original paper provides samples of the BTF from photographs, but many photographs are required to adequately sample the space over the six dimensions. Furthermore, the camera must be calibrated to enable the necessary registration of photos from varying viewing directions. However, there have been developments in the compression and rendering of BTFs [Müller *et al.*, 2003, 2005].

Similar work has been performed by Debevec *et al.* [2000] and Hawkins *et al.* [2001]: a scene is digitized from multiple camera positions under multiple light configurations. Employing high dynamic range imaging, this method is extended to render arbitrary objects under novel illumination conditions.

Polynomial Texture Maps (PTM) defined by Malzbender *et al.* [2001] simplify BTFs by omitting the viewing angles, resulting in a function with four variables. They are based on the assumption that the surfaces described are diffuse, though PTMs can also be used to render specular effects. Multiple images of a static objects are acquired with a static camera under varying lighting conditions, thus sampling the full parameter space. There is no need to calibrate the camera since it is fixed and the resulting images are naturally registered.

Toler-Franklin *et al.* [2007] introduced RGBN images, which store for each pixel the surface normal in addition to the image RGB albedo. The goal is to obtain some of the advantages of 3D models, without the inherent complexity. The data is acquired with a fix camera and varying lighting positions. The readability of the 2D images is improved by incorporating 3D elements that produce a Non-PhotoRealistic (NPR) image. Other such frameworks include that described by Palma *et al.* [2010a]. NPR shading that relies on the existence of a 3D model includes the works of Gooch *et al.* [1998], Rusinkiewicz *et al.* [2006] and Vergne *et al.* [2009].

The illumination model we use is quite simple compared to most of those presented here. It is an empirical model that we further simplify and invert. As such, we do not expect our reflectance correction to fully compensate the effects of the surface variation. However, the goal of these first tests is to show the utility of the method. This method can later be refined with a more complete and adequate model.

6.1.3 New acquisition protocol

As stated previously, to improve the reflectance estimation from our multispectral data we need a multispectral acquisition of a given surface, the corresponding 3D structure and the position of the light sources. For now the correction can only be performed when a single source is used for the acquisitions. The multispectral acquisition and corresponding surface orientation are provided by the optical tracking of the multispectral and 3D acquisition system in the target frame by a set of tracking cameras, as presented in chapter 5.

To obtain the position of the light source, we use the same photogrammetric setup that we use to track the target frame. We simply place a few targets on the light (see figure 6.3) and obtain their position in the tracking camera coordinate system from the acquisitions of the tracking cameras. Targets are also placed on the light stand and the target frame is positioned

Figure 6.3: Acquisition of the position of the light sources.

in the field of view to strengthen the bundle adjustment.

We can then calculate the average of the coordinates of the targets on the light to obtain the approximate punctual coordinates of the light source. These results are inaccurate by construction, since we only place targets on the side of the light that faces the tracking cameras. More accurate results can be obtained by calibrating the position of the targets on the lights, to obtain a better estimate of the coordinates of the center of the light. However, given the distances between the surface under study and the light (over a meter), it is not necessary to compensate a positional error of a few centimeters.

When we project the multispectral acquisition on the 3D model, we store the normal of the surface on which we project ($\hat{N}$), the direction of the vector between the center of this surface and the center of the light source ($\hat{L}$), and $\hat{V}$, the opposite of the projection vector. This is a straightforward task since all data is in the same coordinate system, that defined by the tracking cameras.

Our photogrammetric setup can easily be used to evaluate the position of multiple light sources. However, this technique imposes some constraints on the light positioning: the light must be set in an area that is visible by the tracking cameras without bothering the frame positioning and without casting shadows of the frame on the surface of interest. The advantages of this method to evaluate the position of the light source are compared with those of two other techniques in appendix C.1, page 175.

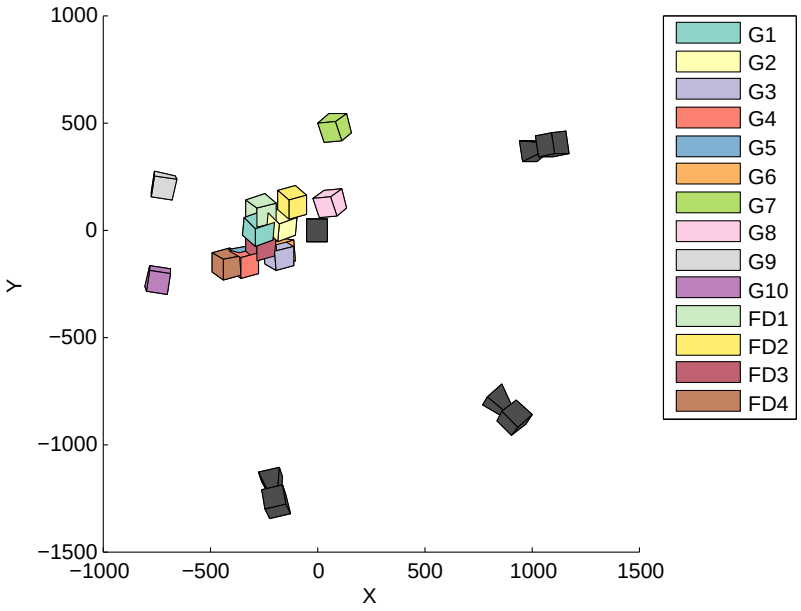


Figure 6.4: Relative position of the tracking cameras and target frame during the digitizations.

6.2 Experimental configuration

6.2.1 Configuration overview

To test the reflectance estimation correction, we repeat the digitization of the bas-relief with the FluxData multispectral camera and the Gom Atos III. We track ten 3D digitization positions (G1 to G10) and four multispectral acquisition positions (FD1 to FD4). The relative positions of the target frame and tracking cameras during these measurements are shown in figure 6.4.

A pair of halogen lights illuminates the bas-relief during the multispectral acquisitions. This is shown in figure 6.3, the second light is positioned roughly symmetrically to the one in the image with respect to the bas relief. As previously, the reflectance estimation necessitates, for every position of the multispectral camera:

- A multispectral acquisition of the ColorChecker.
- Several multispectral acquisitions of a white surface.
- A multispectral acquisition of the surface under study.

We perform these acquisitions three times per position: once with both halogen lights on, once with only the left light on and once with only the right light on. This will enable us to better evaluate the effect of the reflectance estimation correction by comparing corrected views from different light sources. Figure 6.5 shows the twelve resulting color acquisitions, created from the multispectral acquisitions and their corresponding calibration data before the reflectance correction. When both lights are turned on the overall illumination is more regular and lighter shadows are cast on the bas-relief.

Figure 6.5: Color rendering of the twelve multispectral acquisitions. First row: left light on. Second row: right light on. Third row: both lights on.

6.2.2 Tracking and registration accuracy

Individual calibrations

The accuracy of the individual calibrations is given table 6.2. As before, the accuracy of the tracking cameras meets the best-case simulation requirements. The multispectral camera is less accurately calibrated, but the results are still satisfactory. The target frame calibration is, as always, much better than the simulation values. The tracking cameras exterior orientation is only slightly worse than the best-case simulation results spatially, and better than the angular best-case simulation results.

The orientation between the target frame and the Atos III is the most accurate of all those performed. The orientation between the target frame and the FluxData multispectral camera is still disappointing: only two relative positions of the target frame and calibration object are used. At least three relative positions and an improved setup are necessary to increase the accuracy of this calibration.

Tracking accuracy

The tracking accuracy is shown figure 6.6. All but three positions are tracked with the realistic simulation spatial accuracy. These three least-well tracked positions (G7, G8 and G9) are perspective, off-center views though G10, a similar position, is accurately tracked. The spatial accuracy of these three positions is still quite satisfying — more than three times better the necessary value for the Atos III registration.

Table 6.2: Accuracy of the individual calibrations, compared to the simulation values.

Calibration	Experiment	Simulations		Unit
		Realistic	Best-case	
Tracking cameras I.O.	0.029	0.100	0.033	pixel
Multispectral camera I.O.	0.043	—		pixel
Target frame calibration	0.020	0.050		mm
Tracking cameras E.O.	0.011	0.030	0.010	mm
	0.014	0.040	0.020	mrad
Target frame to Gom	0.036	—		mm
Atos III orientation	0.068	—		mrad
Target frame to multispectral	0.243	—		mm
camera orientation	0.556	—		mrad

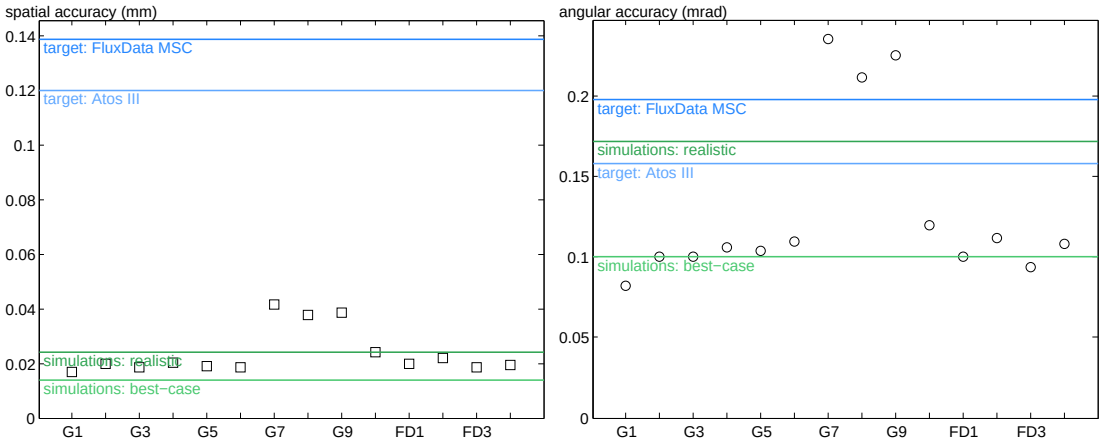


Figure 6.6: Spatial and angular tracking accuracy, compared to the simulation results and to the target accuracy.

The angular tracking accuracy of these three positions, however, is insufficient for the registration of data from both the Gom Atos III and the FluxData multispectral camera. The other eleven positions are tracked with an angular accuracy that is close to the best-case simulation results.

Registration

The integrated 3D model created with these ten meshes in shown figure 6.7. There are no visible discontinuities in the final model. Even the meshes from position G7 to G9 seem correctly registered with the neighboring meshes.

The images projected on the 3D model are the color representations from the multispectral acquisitions performed under both lights. There are no visible discontinuities between the 3D model and these projected images. The slight overlap of the two first image, over the shoulder, is distinguishable only due to the illumination difference.

Figure 6.7: Registration of the ten meshes. (a) One color per mesh. (b) All meshes in grey. (c) Color image from multispectral data projected on the 3D model.

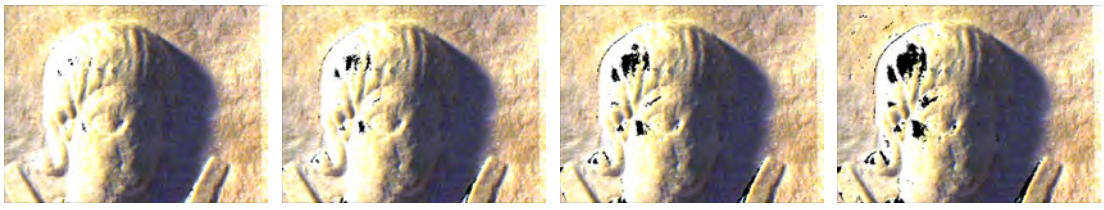
6.3 Results

The reflectance correction results are presented in three subsections. First we evaluate the accuracy of the specular highlight detection. We then show the images produced by calculating the dot product between the light direction and the surface normal. The use of such images to assess the accuracy of a 2D – 3D registration is presented. Finally, we use the light / normal dot product to correct the reflectance cube.

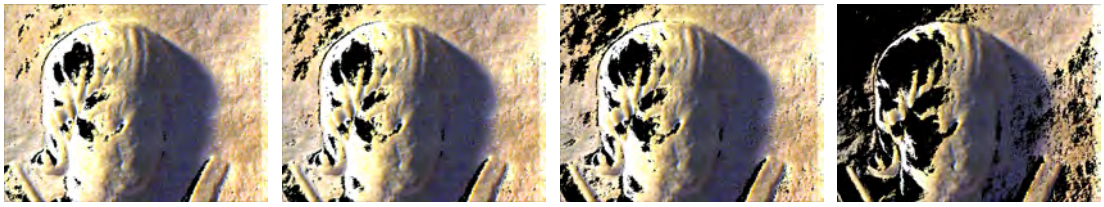
6.3.1 Specular highlights

As explained previously, we detect the specular highlights by calculating $\hat{R} \cdot \hat{V}$ for every pixel of the acquisition and comparing this dot product to a threshold T. Though the result of $\hat{R} \cdot \hat{V}$ is between -1 and 1, T is a positive value. If $T = 1$ no pixel is detected as specular. Figure 6.8 shows the first acquisition with the left light where the specular pixels thus detected are blackened for varying values of T. We also indicate the percentage of blackened pixels in the image, p . Equivalent images for the other eleven acquisitions are given in appendix C.2, page 179.

The choice of an appropriate threshold is non trivial. We would like to find a value that eliminates all the saturated areas of the image, in particular the left side of the face. In practice, even with a low threshold such as $T = 0.5$ not all saturated pixels are detected as specular, though 40 % of the image is blackened. In this case, the non-specular areas of the image also contain many shadows (right side of the face), resulting in few areas where we can perform the reflectance correction. Examining the full results in the appendix also highlights that the threshold value must be chosen independently for each multispectral acquisition, even though



(a) T = 0.99, p = 0.14 % (b) T = 0.95, p = 0.86 % (c) T = 0.90, p = 1.95 % (d) T = 0.85, p = 3.32 %



(e) T = 0.80, p = 5.59 % (f) T = 0.75, p = 9.71 % (g) T = 0.70, p = 14.90 % (h) T = 0.50, p = 40.06 %

Figure 6.8: Specular highlights detected for the first multispectral with the left light for varying values of the threshold T. p represents the percentage of specular pixels in each image.

the acquisitions describe the same material.

The inaccurate localization of the specular reflection is a known drawback of the Phong model [Blinn, 1977; Whitted, 1980], in particular in the case of rough surfaces. We thus examine the effect of the reflectance estimation correction on our multispectral acquisitions without detecting the specular highlights.

6.3.2 Light / normal dot product

The second step of our estimation correction is to calculate the dot product between $\hat{L}$ and $\hat{N}$ for every pixel of the acquisition. The result is a matrix the size of the input image with values theoretically bounded by -1 and 1. In practice, negative values represent shaded areas (the light vector and the surface normal face opposite directions), the lower bound is thus set to 0.

The results can be presented as a greyscale image that represents the relief of the scene for the given light position. This result is shown figure 6.9 for the four multispectral acquisition positions and for both lights separately. The black areas in these images represent areas or pixels for which the multispectral projection provided no 3D data, as well as shaded areas.

Evaluation of the 3D – multispectral registration accuracy

The resulting image can be compared to the multispectral acquisitions to assess the accuracy of the 3D / multispectral registration. It is much easier to compare the similarity of two images, than to evaluate the precision of a 2D mapping on a 3D model.

We perform this comparison by combining the color representation of the multispectral acquisition and a corresponding representation of the light / normal dot product in a single image. The output image is a checkered pattern made of squares of side one hundred pixels taken alternatively from each input image. The color representation of the multispectral acquisition

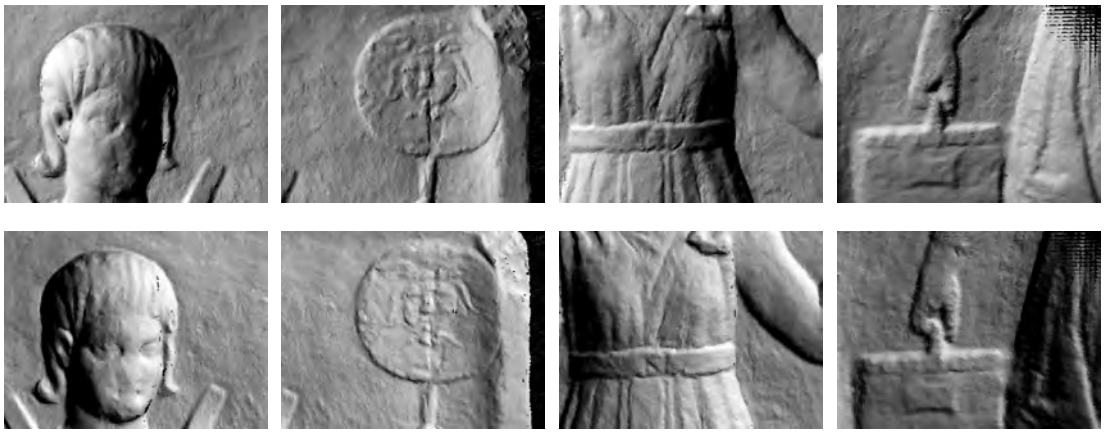


Figure 6.9: Light / normal dot product corresponding to the multispectral acquisitions. First row: left light. Second row: right light.

is converted to greyscale to improve the visual continuity of the output image. The resulting comparison images for all four acquisitions are shown figure 6.10. An equivalent comparison for the previous multispectral acquisitions is given appendix C.3, page 183.

These checkered images show that the registration is globally quite accurate, though some discontinuities are visible. It can seem surprising that some areas of an image are very well registered when discontinuities appear in another area. This is an effect of an inaccurate orientation between the projection and the model, though we saw that the angular tracking accuracy is quite satisfying for the acquisitions.

The two possible causes of these errors have been noted before: the inaccurate angular orientation between the FluxData camera and the target frame, and/or a movement of the multispectral camera in the target frame. The FluxData camera is much lighter than the Gom Atos III, so it is less prone to moving with respect to the target frame when their orientation changes. However, this is still a possible source of inaccuracies, in particular during the FluxData camera to target frame orientation, when more extreme orientations of the frame and camera are used. Furthermore, this would directly impact accuracy of the orientation between the FluxData camera and the target frame.

6.3.3 Reflection correction

The multispectral correction is applied to the reflectance spectra for the eight single-light multispectral acquisitions. The resulting reflectance cubes are once again represented as color images and compared with the color images calculated from the non-corrected cube, in figure 6.11. Black pixels represent areas for which the multispectral projection provided no 3D data.

We can see that the images appear flattened after the correction, suggesting that we do, to some extent, correct the effect of the surface variations. For the fourth acquisition the correction greatly brightens the image, making it somewhat saturated. Since the Phong reflection model does not take into account the presence of shadows and masking, these areas can not be corrected. The masked areas acquire a whitish tinge while the shadows, retain their dark value.

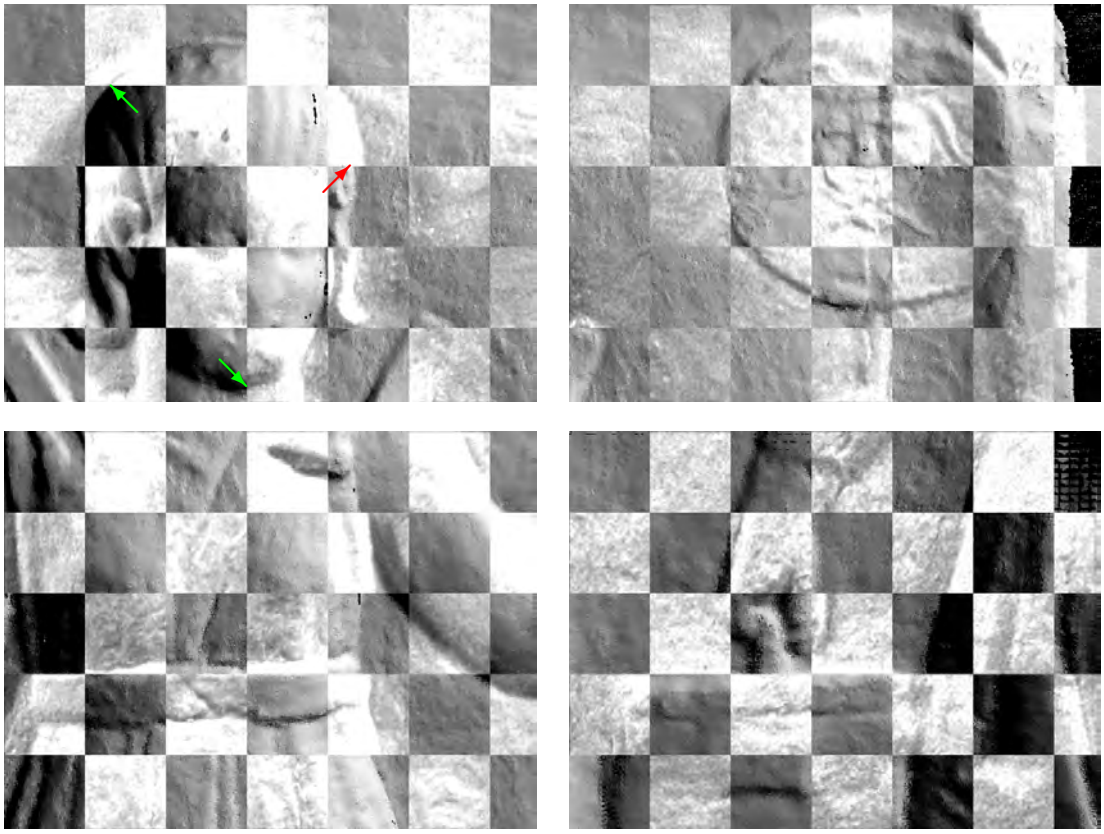


Figure 6.10: Light / normal dot product, compared to the multispectral acquisitions. For all images, the top left corner is from the light / normal product with the right light. In the first figure some satisfactory alignments are pointed by a green arrow while a red arrow shows a misalignment.

Figure 6.11: Effect of the reflectance correction on the color images. First row: right light on, before correction. Second row: right light on, after correction. Third row: left light on, before correction. Last row: left light on, after correction.

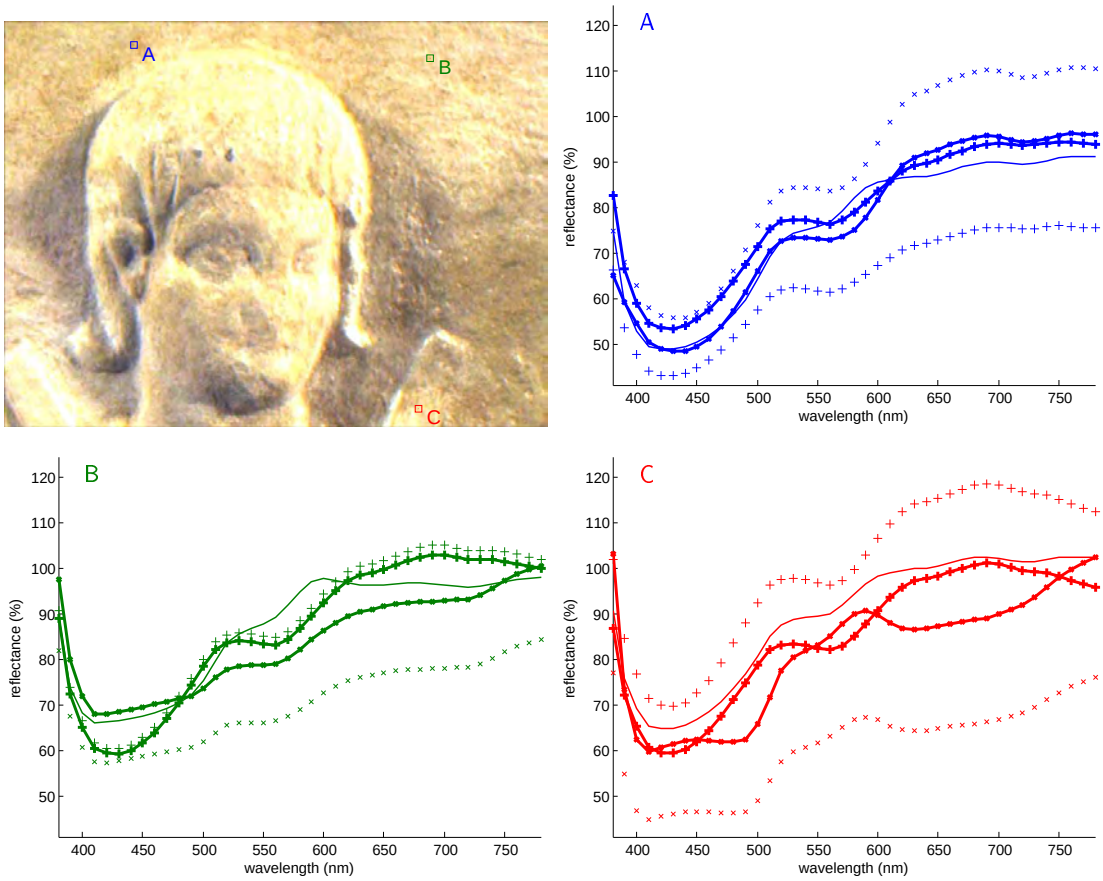


Figure 6.12: Effect of the reflectance correction on selected spectra. 'x' represents the spectra from the left light; '+' represents the spectra from the right light; the thick continuous lines represent the improved spectra, the spectrum from both lights is given as a thin continuous line.

Areas that represent a strong specular component, which were saturated in the original color images, acquire a grey color.

Theoretically, the corrected reflectance of a pixel from an acquisition with the left light should be the same as the corrected reflectance of the same pixel from the same acquisition illuminated from the right. This only holds if the pixel is not in a shadow nor in a saturated area of either image (left light or right light). This leaves only a small portion of each image on which we can perform our comparison.

A few punctual comparisons are shown figure 6.12. The points are chosen in areas where there seems to be no shade nor highlight when either light is cast. The spectra shown are average spectra over an area of 21 pixels $\times$ 21 pixels centered around the selected point. Point A is the one for which the reflectance correction is most effective: the spectra have a similar shape to begin with, and the corrective factor makes them almost superimpose. For the other two positions the correction causes the spectra to be closer, but their shape remains different.

This qualitative observation can be confirmed using two metrics: the Goodness of Fit Coeffi-

Table 6.3: Quantitative evaluation of the similarity of the spectra shown in figure 6.12.

		A	B	C
GFC		0.9985	0.9975	0.9969
RMSE	original	67.0162	49.3788	102.7779
	corrected	11.8890	17.5947	19.3387
	correction %	82.2595	64.3679	81.1840

cient (GFC) and the Root Mean Square Error (RMSE). The GFC is a measure of the similarity of the shape of two spectra $r_1(\lambda)$ and $r_2(\lambda)$, based on the Cauchy-Schwartz inequality:

$$GFC = \frac{\left|\sum_j r_1(\lambda_j) \cdot r_2(\lambda_j)\right|}{\sqrt{\sum_j \left[r_1(\lambda_j)\right]^2} \cdot \sqrt{\sum_j \left[r_2(\lambda_j)\right]^2}} \,.$$

The GFC is closest to 1 the closer the two input spectra, and tends towards 0 as the two spectra differ. Hernández-Andrés *et al.* [2001] define an "accurate" spectral fit as one with GFC > 0.995. For them a "good" fit requires a GFC > 0.999, and a GFC > 0.9999 describes an "excellent" fit.

It is clear from its definition that the GFC is robust to scaling of the spectra. If $r_3(\lambda) = a \cdot r_2(\lambda)$, with a an integer, then $GFC(r_1, r_2) = GFC(r_1, r_3)$. The GFC calculated between reconstructed spectra under the left and right light is thus the same before and after correction. The GFC for the spectra of the three points is given table 6.3. All three pairs of spectra can be considered an accurate fit and point A is, as expected, the one that defines the spectra with the closest shape.

The accurate fit between the pairs of spectra, denotes that they are somewhat proportional. It is thus reasonable to correct these spectra simply by multiplying them each by an adjusting coefficient. To evaluate the effect of this correction, we must use a metric that is sensible to scaling, such as the RMSE.

The RMSE is given by:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N} \left[r_1(\lambda_j) - r_2(\lambda_j)\right]^2}{N}} \,.$$

The closer the two input spectra, the closer the RMSE is to 0. In table 6.3 we compare the spectra under left and right light before the correction (original) and after the correction (corrected). The correction percentage is the percentage by which the correction reduces the RMSE.

For the points that have been chosen in areas that were not subject to specularity nor shadowing, the reflectance correction improves the reflectance by over 60 % and even 80 % in some cases. Though these are carefully chosen points, the use of a more complete illumination model could enable to extend this type of result to a full multispectral acquisition.

6.4 Conclusion

We measured the position of the light sources used during the multispectral acquisitions to correct our reflectance estimation by taking into account the effect of the surface variations. The position of the light sources was measured using the existing tracking setup, by simply placing a few targets on the lights.

We base our correction on the Phong reflection model. Though this produces some inaccuracies in the detection of the specular highlights, these preliminary tests show promising results. The correction of a multispectral cube rendered as a color image appears to flatten the surface in the image. There are numerous regions where the current acquisitions, using the correction based on the Phong reflectance model, can not be compared: shadowed areas, masked areas and areas with a specular component. In areas where the comparison is valid, the reflectance spectra of a given pixel in the left and right light acquisition are closer after the correction. The improvement can reach 80 %.

Better results could be obtained by using a more complex illumination model. Since we are modifying physical properties, a physical illumination model would be more adapted. This would help to better localize the specular component. A shader that takes into account masking and self shadowing would also provide improved results. Even if we can not obtain the reflectance of these dark areas, it could help to localize them, just as we localize the specular component. With an accurate localization of specularity, shadows and masking, multiple single-light acquisitions can be combined to obtain an unshaded reflectance cube. A model that can be inverted in the case of multiple light sources would be even more useful, reducing the number of necessary calibrations. However, as noted by Sitnik *et al.* [2012], fitting sparse responses with a complex model is difficult and error prone, particularly when no a priori knowledge about the surface material is available.

Another approximation was introduced by using area lights, though our model assumes the use of punctual light sources. Future improvement would include using LEDs, fixed to the target frame, as punctual light sources. As is, the results presented in this chapter are a first step in the investigation of the correction of the acquired reflectance to obtain viewer- and illumination-independent spectra.

Chapter 7

Conclusion

7.1 Results

Our goal was to devise a registration technique adapted for the integration of featureless multisensor data, in particular, 3D and multispectral data describing cultural heritage. The work stemmed from the study of two specific objects: a wallpainting in Rüdesheim (Germany) and a bas-relief in Trier (Germany). The registration solution, however, had to be adaptable to other objects and to other situations.

We developed a method based on the photogrammetric tracking of the acquisition systems in use. A set of cameras observe the scene while several acquisition systems successively digitize the object under study. Placing targets on the acquisition systems and calibrating all optics and objects in play enables us to precisely determine the position and orientation of each acquisition system. The data acquired by these acquisition systems can then be projected in a single coordinate system, resulting in an integrated model with a registration accuracy independent from the content of the acquired data.

A series of simulations was performed to evaluate the achievable tracking accuracy and to choose the characteristics of the necessary material. For these simulations, we defined three tracking configurations that represent a variety of cultural heritage setups. The first series of simulations highlighted the necessity of fixing a target frame to the acquisition systems to improve the tracking accuracy by increasing the dimensions of the tracked object. A frame of dimensions $0.5\,\mathrm{m} \times 0.5\,\mathrm{m} \times 0.5\,\mathrm{m}$ and covered with 56 targets proved to be adequate.

When this target frame is fixed to the acquisition systems while they survey an area of $0.4\,\mathrm{m} \times 0.7\,\mathrm{m}$, the simulations show that we can achieve a spatial tracking accuracy of 0.020 mm and an angular accuracy of 0.100 mrad using four 5 Mpx cameras with 8 mm lens. Equivalent tracking accuracy can be obtained using six cameras when the acquisition systems digitize an area of $2\,\mathrm{m} \times 1.5\,\mathrm{m}$, such as a wallpainting. If the object under study is a 1 m high statue with a 0.3 m radius then the target frame can be tracked with comparable spatial accuracy and an angular accuracy of 0.122 mrad. This, however, requires eight tracking cameras and would be harder to achieve in practice. These results correspond to a tracking accuracy better than half a pixel in the predefined configurations and using the acquisition systems from our laboratories.

The method was extensively tested in laboratory settings. It was essential to solve the practical and logistical difficulties in such controlled settings before using the method in situ. The setup was improved throughout these laboratory tests. To assess the experimental tracking accuracy we first tracked the target frame simultaneously using our photogrammetric setup and a laser tracker. Though difficult to compare, the laser tracker results seemed to validate the photogrammetric tracking internal accuracy. We then measured the accuracy of our method for multiview 3D registration. These tests were performed on featureless objects which we covered with targets, using these targets only to evaluate the tracking accuracy. Our registration goal of half a pixel accuracy was reached in specific cases where the orientation of the acquisition system was stable.

The accuracy of the final registration relies on the success of a series of optical and geometrical calibrations: calibration of the target frame, calibration of the interior orientation of the tracking cameras and acquisition systems, calibration of the exterior orientation of the tracking cameras. It is also essential that the calibrated parameters are stable for the duration of the full acquisition process, typically between four and eight hours. We thus evaluated the stability of our setup and calibrations, to ensure that they were satisfactory. We also tested the tracking of the target frame in the larger six camera configuration, and verified that this setup is adapted for multimodal registration.

Our method was then used for the registration of 3D and multispectral datasets. The multispectral data is projected as a texture on a 3D model whose constituent meshes have been previously registered using our technique. The projection parameters are, as before, evaluated from the tracking of the target frame as well as from the optical and geometrical calibrations. The 3D – multispectral registration was first tested on an object with much greater spectral and structural variability than typical cultural heritage objects, enabling a quick visual assessment of the accuracy of the final registration. We noticed once again that the fixture between the target frame and multispectral camera is not as stable as we need it to be. Nevertheless, these experimental results show that our tracking method is adapted not only for the registration of 3D datasets, but also for the integration of multispectral texture on 3D models.

We extended our method to improve the reflectance estimation from the multispectral datasets projected on a 3D model. The photogrammetric setup is used to measure the position of the light sources in addition to tracking the 3D and multispectral acquisition systems. We then apply a correction based on the integration of our reflectance estimation model with the Phong reflectance model. Using the known position of the light sources and of the multispectral camera, as well as the orientation of the surface for each multispectral pixel, we can calculate the pixels that describe a specular reflection. For the remaining areas we correct the diffuse reflectance intensity based on the angle between the incident light and the surface normal. This process provides us with a hyperspectral cube that takes into account the variations in surface orientation for areas where there are no specular reflections, nor any masking or shading.

The final registration method is adapted for the contactless registration of featureless multimodal data. The advantage of our registration method is that the accuracy is independent of the content of the data: if the acquired data represented a smooth plane, it would be registered with the same accuracy. In comparison, other 3D registration methods such as ICP generally

fare badly with smooth and planar data. Furthermore, this eliminates the problem of deviation errors accumulating across the views, particularly when registering open surfaces, that is typical of ICP. Our method also requires little or no overlap between adjacent meshes, compared to 30 % to 40 % for ICP and other feature-based registration techniques.

7.2 Future work

There are many possible improvements of this work that would result in an increased accuracy and easier manipulation. These improvements are important to implement if the method is to be more widely used for cultural heritage digitization campaigns. Such improvements range from simple engineering improvements to conceptual extensions.

The most important engineering improvement that has been mentioned throughout this thesis is the fixation between the target frame and the acquisition system. The current hexagonal plate moves in the fixation. Strengthening the fixation by some screws would keep the relative position of the frame and acquisition fixed, but it remains to be seen how this can be done in practice without drilling holes in the acquisition systems.

The tracking results could also be improved with a newly designed tracking frame: If this frame were spherical instead of cubic, the tracking accuracy would not depend on its orientation with respect to the tracking cameras. Also, the current target frame design is such that the targets hide one another from certain points of view, since they stick out of the tracking frame. This could easily be solved on a spherical target frame.

The use of active targets could also remove the need of a uniform external illumination, increasing the flexibility of the system. If the acquisitions must be performed outdoors, spurious daylight can be suppressed by using active infrared targets. Finally, aluminum is convenient for rapid prototyping but a carbon tracking frame would be more stable and less sensitive to temperature changes that can occur over the course of a day. This would probably result in a lighter frame that is easier to manipulate and to transport.

We have seen that the calibration between the target frame and multispectral camera must be improved. It is important to compensate the difference in scale between the small field-of-view of the multispectral camera and the large tracking frame. Designing a better adapted calibration object would probably be sufficient. To a lesser extent, the target frame to Gom Atos III could also be improved. However, implementing a few of the previously mentioned engineering improvements on the target frame would probably be sufficient to increase the accuracy of this calibration to even more satisfying values.

On-site acquisitions would be eased and sped up if we developed an integrated software tool for photogrammetric processing. Having rapid feedback on the quality of the tracking and registration would help acquire integrated datasets with an overall satisfying registration accuracy. There is a lot of software development associated with this task, in particular to make user-friendly to be used by non-experts, but it relies on known concepts and algorithms.

However, there is much work to be done on the visualization and interaction of models with multiple textures and with textures of high dimensions. The development of tools to perform measurements on 3D models with multispectral textures must be performed with the insight of

conservators to invent adapted tools.

As part of this, the correction of the reflectance estimation must be improved and based on a more sophisticated illumination model, to provide an object reflectance that is truly an object property. We have seen that physically-based models that take into account shading and masking would be better adapted.

An important aspect that we have not tackled is the registration of data acquired at different times for monitoring purposes. Such registration must necessarily be based on the acquired data: variations in cultural heritage occur over large time spans. The tracking setup can not be expected to hold the calibration for so long, not to mention the constraints keeping the object and tracking setup fixed for long periods of time. Perhaps if the multiview and multimodal registration of data from a given time stamp is performed with sufficient accuracy, then registration of data from different acquisition times can be performed with feature-based methods, relying on both the 3D and multispectral features from the integrated datasets. However, there are specific problems to address when registering data for monitoring purposes, such as how to optimize a registration process if one model represents an degraded state of the object.

7.3 Perspectives

Though many possible improvements have been listed, as is the method can be used to study the objects for which it was developed, as well as any object of comparable dimensions. The full setup is transportable and can be used in situ. The necessary acquisitions and calibrations to register data from two acquisition systems can be performed by two people in a single working day. Some of these calibrations may be performed in advance, to reduce the on-site workload.

The method was developed for the study of two specific objects, but the method can easily be scaled and tweaked for the study of other objects. The method is most adapted for the study of objects with a strong 2D component. Though the logistical difficulties are greater, this registration setup can theoretically also be used for the study of free standing 3D objects such as statues. This would require a higher number of tracking cameras, increasing in turn the number of necessary calibrations and the difficulty of maintaining a stable setup. If the object under study is mostly devoted of distinguishable features, then our method still presents clear advantages over feature-based registration methods. If the improvements listed in the previous section are implemented, however, our method would then become advantageous in the context objects with a strong 3D component.

Though our multimodal registration method was developed to respond to the particular challenges of cultural heritage digitization, it is also adapted for quality control in industrial settings. In such fixed and controlled settings, the tracking accuracy can be increased. In this context, the necessary calibrations need only be performed from time to time, instead of once per object. If many measurements are performed, the acquisitions can also be used to correct the calibrations in real-time.

Our registration technique was tested with the acquisition systems present in our laboratories, but it can be used with any 3D digitization system and multispectral camera. Furthermore, it is also adapted for the registration of optical data from other systems. In addition to multi-

or hyperspectral images, data from many techniques that are used for the analysis of cultural heritage can be mapped on their corresponding 3D model using our method: high resolution digital photography, UV fluorescence, IR reflectography, Terahertz imaging [Fukunaga and Picollo, 2010], trans-illumination or trans irradiation of paintings [Cucci *et al.*, 2012].

Several complementary textures can thus be mapped on a given 3D model. Such augmented models can provide insights in material and surface analysis. To better exploit these integrated datasets, it is necessary to develop both multimodal analytical tools and multimodal visualization tools. Such tools must be imagined with the conservation community to provide them with methods and displays that can enrich their research. The high dimensionality and heterogeneity of the datasets makes both these tasks an engaging challenge of the yeas to come.

Appendix A

Simulation Results

The detailed results of the simulations presented in chapter 3 are given in this appendix. All the following simulation accuracy results are given at 2σ. Values in bold indicate the maximum value of the column. Grey backgrounds in the overview tables highlight the simulation results that were not presented in chapter 3.

A.1 Sarcophagus simulations

Table A.1 presents an overview of all sarcophagus simulation results.

Table A.1: Overview of the sarcophagus simulation results.

Tracked object (points)	Input noise: Picture coord. (µm)	Input noise: Object coord. (mm)	Input noise: Orientation translation (mm)	Input noise: Orientation rotation (mrad)	Output accuracy: Spatial (mm)	Output accuracy: Angular (mrad)	See table
2 Mpx Baumer cameras with 6 mm lens							
Lab MSC (18)	0.345	0	0	0	0.020	0.292	A.2
FD MSC (12)	0.345	0	0	0	0.025	0.536	A.3
Atos (24)	0.345	0	0	0	0.013	0.168	A.4
Atos (34)	0.345	0	0	0	0.012	0.160	A.5
5 Mpx Baumer cameras with 6 mm lens							
Lab MSC (18)	0.345	0	0	0	0.017	0.238	A.6
FD MSC (12)	0.345	0	0	0	0.022	0.536	A.7
Atos (24)	0.345	0	0	0	0.011	0.142	A.8
Atos (34)	0.345	0	0	0	0.009	0.128	A.9
Frame[a] (26)	0.345	0	0	0	0.015	0.144	A.10
Frame[a] (56)	0.345	0	0	0	0.012	0.124	A.11
Frame[b] (26)	0.345	0	0	0	0.017	0.120	A.12
Frame[b] (56)	0.345	0	0	0	0.012	0.088	A.13
Frame[b] (56)	0.115	0	0	0	0.004	0.028	A.14
orientation	0.345	0	0	0	0.023	0.058	A.15
	0.115	0	0	0	0.009	0.020	A.16
Frame[b] (56)	0.345	0.05	0.03	0.07	0.023	0.176	A.17
	0.115	0.05	0.02	0.03	0.016	0.116	A.18
5 Mpx AVT cameras with 8 mm lens							
Frame[b] (56)	0.345	0	0	0	0.015	0.104	A.19
	0.115	0	0	0	0.005	0.032	A.20
	0.345	0.05	0.03	0.07	0.028	0.198	A.21
	0.115	0.05	0.02	0.03	0.016	0.112	A.22
orientation	0.345	0.00	0	0	0.021	0.030	A.23
	0.115	0.00	0	0	0.007	0.010	A.24
Frame[b] (56)	0.345	0.05	0.03	0.04	0.024	0.172	A.25
	0.115	0.05	0.01	0.02	0.014	0.100	A.26

Frame[a] target frame of dimensions 0.50 m × 0.40 m × 0.30 m

Frame[b] target frame of dimensions 0.50 m × 0.50 m × 0.50 m

A.1.1 2 Mpx Baumer cameras with 6 mm lens

Tracking the acquisition systems

Table A.2: Tracking the lab-designed multispectral camera represented by 18 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.019	0.015	0.020	0.230	0.278	0.190
Top Right	0.017	0.014	0.018	0.212	0.256	0.172
Center	0.018	0.016	0.019	0.226	0.270	0.190
Bottom Left	**0.019**	**0.017**	**0.020**	**0.240**	**0.292**	**0.200**
Bottom Right	0.019	**0.017**	**0.020**	**0.240**	**0.292**	0.200
Mean	0.019	0.016	0.019	0.235	0.285	0.197

Table A.3: Tracking the FluxData multispectral camera represented by 12 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	**0.023**	**0.019**	**0.025**	**0.444**	**0.536**	**0.602**
Top Right	0.021	0.018	0.023	0.418	0.496	0.556
Center	0.020	0.018	0.021	0.394	0.460	0.530
Bottom Left	0.019	0.017	0.021	0.366	0.438	0.508
Bottom Right	0.021	0.018	0.023	0.404	0.478	0.556
Mean	0.020	0.018	0.022	0.396	0.468	0.538

Table A.4: Tracking the Gom Atos III represented by 24 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top	0.012	0.011	0.012	0.152	0.106	0.104
Center	**0.013**	**0.012**	**0.013**	**0.168**	**0.116**	**0.116**
Bottom	0.013	0.012	0.013	0.162	0.114	0.112
Mean	0.013	0.012	0.012	0.161	0.112	0.111

Table A.5: Tracking the Gom Atos III represented by 34 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top	**0.012**	**0.011**	**0.012**	**0.160**	**0.106**	**0.104**
Center	0.012	0.011	0.011	0.156	0.104	0.102
Bottom	0.012	0.011	0.011	0.152	0.104	0.100
Mean	0.012	0.011	0.011	0.156	0.105	0.102

A.1.2 5 Mpx Baumer cameras with 6 mm lens

Tracking the acquisition systems

Table A.6: Tracking the lab-designed multispectral camera represented by 18 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.012	0.010	0.013	0.148	0.180	0.124
Top Right	0.014	0.011	0.014	0.164	0.198	0.136
Center	**0.016**	**0.015**	**0.017**	**0.204**	**0.238**	**0.174**
Bottom Left	0.015	0.013	0.016	0.192	0.228	0.160
Bottom Right	0.016	0.014	0.016	0.200	0.240	0.168
Mean	0.015	0.013	0.016	0.190	0.226	0.160

Table A.7: Tracking the FluxData multispectral camera represented by 12 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.014	0.012	0.015	0.274	0.324	0.370
Top Right	0.016	0.014	0.017	0.308	0.366	0.418
Center	0.019	0.017	0.019	0.372	0.422	0.498
Bottom Left	**0.020**	**0.018**	**0.022**	**0.388**	**0.458**	**0.536**
Bottom Right	0.019	0.017	0.021	0.370	0.438	0.506
Mean	0.018	0.016	0.020	0.360	0.421	0.490

Table A.8: Tracking the Gom Atos III represented by 24 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top	**0.011**	**0.011**	**0.011**	**0.142**	**0.096**	**0.098**
Center	0.011	**0.011**	0.010	**0.142**	0.092	**0.098**
Bottom	0.010	0.010	0.009	0.130	0.084	0.092
Mean	0.011	0.010	0.010	0.138	0.091	0.096

Table A.9: Tracking the Gom Atos III represented by 34 points in its' bounding box. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top	0.009	**0.009**	0.009	**0.128**	0.078	**0.086**
Center	0.008	0.008	0.008	0.112	0.070	0.074
Bottom	**0.009**	0.009	**0.009**	0.126	**0.080**	0.082
Mean	0.009	0.009	0.008	0.122	0.076	0.081

Tracking a target frame

Table A.10: Tracking a 0.5 m × 0.4 m × 0.3 m target frame represented by 26 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.012	0.011	0.011	0.106	0.092	0.082
Top Right	0.013	0.011	0.012	0.114	0.100	0.086
Center	0.014	**0.014**	0.013	0.126	0.106	0.100
Bottom Left	0.014	0.013	0.014	0.140	0.122	0.102
Bottom Right	**0.014**	0.013	**0.015**	**0.144**	**0.126**	**0.104**
Mean	0.014	0.013	0.014	0.131	0.114	0.098

Table A.11: Tracking a 0.5 m × 0.4 m × 0.3 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.009	0.009	0.008	0.084	0.072	0.064
Top Right	0.009	0.008	0.008	0.086	0.076	0.064
Center	0.010	0.010	0.009	0.096	0.082	0.076
Bottom Left	0.011	0.010	0.011	0.116	0.098	0.078
Bottom Right	**0.011**	**0.011**	**0.012**	**0.124**	**0.106**	**0.082**
Mean	0.010	0.010	0.010	0.106	0.091	0.075

Table A.12: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 26 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.016	0.015	0.013	0.096	0.094	0.098
Top Right	0.014	0.013	0.012	0.086	0.088	0.082
Center	0.015	0.014	0.013	0.092	0.094	0.094
Bottom Left	0.015	0.015	0.014	0.106	0.094	0.100
Bottom Right	**0.017**	**0.017**	**0.015**	**0.120**	**0.106**	**0.114**
Mean	0.015	0.015	0.013	0.101	0.096	0.098

Table A.13: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.011	0.010	0.009	0.070	0.068	0.064
Top Right	0.010	0.009	0.008	0.062	0.066	0.060
Center	0.010	0.009	0.008	0.064	0.064	0.064
Bottom Left	0.010	0.010	0.010	0.078	0.070	0.070
Bottom Right	**0.012**	**0.011**	**0.011**	**0.088**	**0.078**	**0.078**
Mean	0.010	0.010	0.009	0.073	0.070	0.068

Table A.14: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.004	0.004	0.003	0.026	**0.026**	0.024
Top Right	0.003	0.003	0.003	0.022	0.024	0.022
Center	0.003	0.003	0.003	0.022	0.022	0.022
Bottom Left	0.004	0.004	**0.003**	**0.028**	0.024	0.024
Bottom Right	**0.004**	**0.004**	**0.003**	**0.028**	**0.026**	**0.026**
Mean	0.004	0.003	0.003	0.025	0.024	0.024

Orientation simulations

Table A.15: Orientation results. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	**0.020**	0.016	**0.023**	0.040	0.050	0.036
Camera 2	0.019	**0.021**	0.020	**0.056**	**0.058**	**0.044**
Camera 3	0.019	0.015	0.022	0.040	0.050	0.036
Camera 4	0.019	**0.021**	0.020	**0.056**	**0.058**	**0.044**
Mean	0.019	0.018	0.021	0.048	0.054	0.040

Table A.16: Orientation results. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	**0.008**	0.006	**0.009**	0.016	**0.020**	**0.014**
Camera 2	0.006	**0.007**	0.007	**0.018**	**0.020**	**0.014**
Camera 3	**0.008**	0.006	0.009	0.016	**0.020**	**0.014**
Camera 4	0.006	**0.007**	0.007	**0.018**	**0.020**	**0.014**
Mean	0.007	0.007	0.008	0.017	0.020	0.014

Comprehensive simulations

Table A.17: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.03 mm on the camera translation and 0.07 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.020	0.017	0.016	0.126	0.126	0.116
Top Right	0.022	0.021	0.019	0.150	0.156	0.148
Center	0.023	0.022	0.019	0.150	0.152	0.150
Bottom Left	**0.023**	**0.023**	**0.022**	**0.176**	**0.156**	**0.156**
Bottom Right	0.020	0.020	0.019	0.152	0.136	0.136
Mean	0.022	0.021	0.020	0.157	0.150	0.148

Table A.18: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.02 mm on the camera translation and 0.03 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	**0.016**	0.015	0.013	0.110	**0.108**	0.104
Top Right	0.015	0.014	0.013	0.102	0.104	0.100
Center	0.014	0.013	0.012	0.090	0.092	0.090
Bottom Left	0.015	**0.015**	**0.014**	**0.116**	0.104	**0.106**
Bottom Right	0.013	0.012	0.011	0.092	0.084	0.084
Mean	0.014	0.014	0.012	0.100	0.096	0.095

A.1.3 5 Mpx AVT cameras with 8 mm lens

Preliminary simulations

Table A.19: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.011	0.010	0.011	0.076	0.080	0.072
Top Right	0.012	0.010	0.011	0.078	0.084	0.074
Center	0.013	0.011	0.012	0.084	0.090	0.080
Bottom Left	0.014	0.012	0.012	0.090	0.098	0.084
Bottom Right	**0.015**	**0.012**	**0.013**	**0.094**	**0.104**	**0.088**
Mean	0.013	0.011	0.012	0.084	0.091	0.080

Table A.20: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.004	0.003	0.004	0.026	0.028	0.024
Top Right	0.004	0.003	0.004	0.026	0.028	0.024
Center	0.004	0.004	0.004	0.026	0.028	0.026
Bottom Left	0.004	0.004	0.004	0.028	**0.032**	0.026
Bottom Right	**0.005**	**0.004**	**0.004**	**0.030**	**0.032**	**0.028**
Mean	0.004	0.004	0.004	0.027	0.030	0.026

Table A.21: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.03 mm on the camera translation and 0.07 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.023	0.020	0.022	0.154	0.164	0.144
Top Right	0.022	0.020	0.021	0.150	0.160	0.142
Center	0.022	0.020	0.021	0.148	0.160	0.142
Bottom Left	0.023	0.019	0.020	0.144	0.158	0.134
Bottom Right	**0.028**	**0.023**	**0.025**	**0.178**	**0.196**	**0.166**
Mean	0.024	0.020	0.022	0.155	0.168	0.146

Table A.22: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.02 mm on the camera translation and 0.03 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.015	**0.014**	**0.015**	**0.104**	**0.112**	0.098
Top Right	0.012	0.011	0.012	0.082	0.088	0.078
Center	0.014	0.013	0.014	0.096	0.102	0.090
Bottom Left	**0.016**	0.013	0.014	0.102	**0.112**	**0.094**
Bottom Right	0.016	0.013	0.014	0.100	0.110	0.092
Mean	0.015	0.013	0.014	0.097	0.105	0.090

Orientation simulations

Table A.23: Orientation results. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	**0.020**	0.018	0.021	**0.028**	**0.030**	**0.026**
Camera 2	**0.020**	0.018	**0.021**	**0.028**	**0.030**	**0.026**
Camera 3	0.016	**0.018**	0.017	0.026	0.028	0.024
Camera 4	0.015	0.017	0.016	0.024	0.028	0.024
Mean	0.018	0.018	0.019	0.027	0.029	0.025

Table A.24: Orientation results. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	**0.007**	0.006	**0.007**	**0.010**	**0.010**	**0.008**
Camera 2	**0.007**	0.006	**0.007**	**0.010**	**0.010**	**0.008**
Camera 3	0.005	**0.006**	0.006	0.008	**0.010**	**0.008**
Camera 4	0.005	0.006	0.005	0.008	**0.010**	**0.008**
Mean	0.006	0.006	0.006	0.009	0.010	0.008

Comprehensive simulations

Table A.25: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.03 mm on the camera translation and 0.04 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.024	**0.021**	**0.023**	**0.160**	**0.172**	**0.152**
Top Right	0.022	0.019	0.021	0.146	0.156	0.138
Center	0.018	0.016	0.017	0.122	0.132	0.116
Bottom Left	**0.024**	0.021	0.022	0.156	**0.172**	0.146
Bottom Right	0.019	0.016	0.017	0.120	0.132	0.112
Mean	0.021	0.019	0.020	0.141	0.153	0.133

Table A.26: Tracking a 0.5 m × 0.5 m × 0.5 m target frame represented by 56 points. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.01 mm on the camera translation and 0.02 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.012	0.011	0.011	0.080	0.086	0.076
Top Right	0.012	0.011	0.012	0.082	0.088	0.076
Center	**0.014**	**0.012**	**0.013**	**0.092**	**0.100**	**0.088**
Bottom Left	**0.014**	0.012	0.012	0.088	0.098	0.082
Bottom Right	0.013	0.011	0.012	0.086	0.094	0.080
Mean	0.013	0.011	0.012	0.086	0.093	0.080

A.2 Painting simulations

The tracking cameras used in these simulations match the AVT Stingray 5 Mpx cameras used with 8 mm lens. The tracked object is the 0.5 m × 0.5 m × 0.5 m target frame covered defined by fifty six points. An overview of all painting simulation results is given table A.27.

Table A.27: Overview of the painting simulation results.

	Input noise				Output accuracy		
	Picture coord. (µm)	Object coord. (mm)	Orientation translation (mm)	Orientation rotation (mrad)	Spatial (mm)	Angular (mrad)	See table
Bounding box surveyed by four cameras							
Frame*	0.345	0	0	0	0.028	0.196	A.28
	0.345	0	0	0	0.026	0.186	A.29
	0.345	0	0	0	0.029	0.214	A.30
Partial bounding box surveyed by three cameras							
Frame*	0.345	0	0	0	0.025	0.174	A.31
	0.345	0	0	0	0.026	0.170	A.32
	0.345	0	0	0	0.023	0.162	A.33
	0.345	0	0	0	0.025	0.172	A.34
	0.345	0	0	0	0.024	0.160	A.35
	0.345	0	0	0	0.026	0.172	A.36
Bounding box surveyed by six cameras							
Frame	0.345	0	0	0	0.016	0.116	A.37
	0.115	0	0	0	0.005	0.038	A.38
orientation	0.345	0	0	0	0.052	0.050	A.39
	0.115	0	0	0	0.017	0.016	A.40
Frame	0.345	0.10	0.05	0.04	0.029	0.198	A.41
	0.345	0.05	0.05	0.04	0.027	0.192	A.42
	0.345	0.05	0.06	0.06	0.032	0.216	A.43
	0.115	0.10	0.02	0.02	0.023	0.156	A.44
	0.115	0.05	0.02	0.02	0.015	0.106	A.45

* the tracked object and noise are fixed but the camera positions vary for these simulations.

A.2.1 Bounding box surveyed by four cameras

Table A.28: Variation (a) on the arrangement of the four tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.023	0.024	0.027	0.188	0.178	0.152
Top Right	0.024	**0.025**	**0.028**	**0.196**	0.186	0.158
Center	**0.025**	0.024	0.026	0.184	**0.190**	**0.166**
Bottom Left	0.024	0.021	0.024	0.166	0.176	0.150
Bottom Right	0.023	0.022	0.025	0.172	0.176	0.146
Mean	0.024	0.023	0.026	0.181	0.181	0.154

Table A.29: Variation (b) on the arrangement of the four tracking cameras. Gaussian noise of standard deviation $1/10^{\text{th}}$ of a pixel is added to the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.024	0.022	**0.024**	0.164	0.178	0.154
Top Right	0.021	0.021	0.024	0.162	0.164	0.142
Center	**0.026**	**0.023**	**0.024**	**0.172**	**0.186**	**0.164**
Bottom Left	0.024	0.021	0.024	0.160	0.176	0.150
Bottom Right	0.021	0.020	0.023	0.158	0.162	0.140
Mean	0.023	0.021	0.024	0.163	0.173	0.150

Table A.30: Variation (c) on the arrangement of the four tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.024	0.021	0.025	0.164	0.182	0.150
Top Right	0.024	0.021	0.026	0.170	0.184	0.150
Center	**0.029**	**0.024**	0.028	**0.188**	**0.214**	**0.178**
Bottom Left	0.026	0.023	**0.029**	0.184	0.200	0.164
Bottom Right	0.025	0.021	0.026	0.166	0.186	0.152
Mean	0.026	0.022	0.027	0.174	0.193	0.159

A.2.2 Partial bounding box surveyed by three cameras

Table A.31: Variation (a) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.018	0.018	0.019	0.134	0.132	0.122
Top Center	0.022	0.017	0.019	0.130	0.150	0.126
Center Left	**0.025**	**0.022**	**0.022**	**0.162**	**0.174**	**0.156**
Bottom Left	0.020	0.017	0.019	0.130	0.138	0.122
Bottom Center	0.023	0.017	0.019	0.130	0.158	0.132
Mean	0.022	0.018	0.020	0.137	0.150	0.132

Table A.32: Variation (b) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.019	0.016	0.018	0.124	0.130	0.118
Top Center	**0.026**	0.018	**0.020**	0.138	**0.170**	**0.142**
Center Left	0.024	**0.019**	0.020	**0.144**	0.160	**0.142**
Bottom Left	0.021	0.017	0.020	0.134	0.142	0.128
Bottom Center	0.023	0.016	0.018	0.126	0.154	0.128
Mean	0.022	0.017	0.019	0.133	0.151	0.132

Table A.33: Variation (c) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.018	0.018	0.019	0.136	0.132	0.120
Top Center	0.023	0.018	0.020	0.140	0.160	0.120
Center Left	0.021	**0.020**	0.020	**0.148**	0.150	**0.136**
Bottom Left	0.019	0.019	**0.020**	0.144	0.142	0.128
Bottom Center	**0.023**	0.018	**0.020**	0.142	**0.162**	0.134
Mean	0.021	0.019	0.020	0.142	0.149	0.128

Table A.34: Variation (d) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.018	0.017	0.018	0.128	0.130	0.118
Top Center	0.024	0.018	0.020	0.140	0.166	0.140
Center Left	**0.025**	**0.022**	**0.022**	**0.160**	**0.172**	**0.156**
Bottom Left	0.018	0.017	0.019	0.128	0.132	0.118
Bottom Center	0.024	0.017	0.020	0.134	0.162	0.132
Mean	0.022	0.018	0.020	0.138	0.152	0.133

Table A.35: Variation (e) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.017	0.015	0.017	0.116	0.124	0.110
Top Center	0.021	0.016	0.019	0.126	0.146	0.124
Center Left	0.021	0.020	0.018	0.140	0.144	0.140
Bottom Left	0.018	0.017	0.018	0.130	0.124	0.118
Bottom Center	**0.024**	**0.022**	**0.020**	**0.154**	**0.158**	**0.160**
Mean	0.020	0.018	0.018	0.133	0.139	0.130

Table A.36: Variation (f) on the arrangement of the three tracking cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.018	0.019	0.019	0.138	0.130	0.126
Top Center	**0.026**	**0.024**	**0.022**	**0.166**	**0.168**	**0.172**
Center Left	0.022	0.021	0.019	0.146	0.150	0.146
Bottom Left	0.019	0.016	0.018	0.124	0.136	0.120
Bottom Center	0.021	0.017	0.019	0.126	0.148	0.126
Mean	0.021	0.019	0.019	0.140	0.146	0.138

A.2.3 Bounding box surveyed by six cameras

Preliminary simulations

Table A.37: Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	**0.016**	0.014	0.015	**0.108**	**0.116**	**0.104**
Top Center	0.014	0.012	0.013	0.090	0.098	0.088
Top Right	0.016	0.015	**0.015**	0.110	**0.116**	0.102
Center Left	0.014	0.015	0.013	0.100	0.102	0.096
Center Right	0.015	**0.015**	0.013	0.102	0.102	0.098
Bottom Left	0.015	0.014	0.015	0.104	0.110	0.096
Bottom Center	0.014	0.012	0.012	0.088	0.096	0.086
Bottom Right	0.015	0.013	0.014	0.100	0.108	0.096
Mean	0.015	0.014	0.014	0.100	0.106	0.096

Table A.38: Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.005	0.004	0.005	0.032	0.036	0.032
Top Center	0.005	0.004	0.004	0.032	0.034	0.030
Top Right	**0.005**	**0.005**	**0.005**	**0.036**	**0.038**	**0.034**
Center Left	0.005	0.005	0.004	0.032	0.032	0.030
Center Right	0.005	**0.005**	0.004	0.034	0.034	0.032
Bottom Left	**0.005**	**0.005**	**0.005**	**0.036**	**0.038**	**0.034**
Bottom Center	0.005	0.004	0.004	0.030	0.034	0.030
Bottom Right	**0.005**	**0.005**	**0.005**	**0.036**	**0.038**	**0.034**
Mean	0.005	0.005	0.005	0.034	0.036	0.032

Orientation

Table A.39: Orientation results with $1/10^{\text{th}}$ of a pixel noise on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	0.034	0.033	0.046	0.042	0.038	0.028
Camera 2	**0.036**	**0.045**	**0.052**	**0.050**	0.038	**0.034**
Camera 3	0.032	0.030	0.042	0.042	0.040	0.030
Camera 4	0.033	0.034	0.049	0.048	**0.042**	0.032
Camera 5	0.034	0.034	0.048	0.046	0.038	0.028
Camera 6	0.034	0.038	0.049	0.048	0.036	0.030
Mean	0.034	0.036	0.048	0.046	0.039	0.030

Table A.40: Orientation results with $1/30^{\text{th}}$ of a pixel noise on the picture coordinates.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Camera 1	0.011	0.011	0.015	0.014	0.012	0.010
Camera 2	**0.012**	**0.015**	**0.017**	**0.016**	0.012	**0.012**
Camera 3	0.011	0.010	0.014	0.014	**0.014**	0.010
Camera 4	0.011	0.011	0.016	**0.016**	**0.014**	0.010
Camera 5	0.011	0.011	0.016	0.014	0.012	0.010
Camera 6	0.011	0.013	0.016	**0.016**	0.012	0.010
Mean	0.011	0.012	0.016	0.015	0.013	0.010

Comprehensive simulations

Table A.41: Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.1 mm on the object coordinates, 0.05 mm on the camera translation and 0.04 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.028	0.024	0.026	0.182	0.196	0.176
Top Center	0.027	0.023	0.023	0.168	0.182	0.164
Top Right	0.026	0.024	0.025	0.184	0.192	0.170
Center Left	0.025	0.025	0.022	0.174	0.176	0.168
Center Right	0.025	**0.026**	0.023	0.178	0.178	0.170
Bottom Left	0.028	0.025	**0.026**	**0.192**	0.200	0.176
Bottom Center	**0.029**	0.025	0.025	0.182	**0.198**	**0.178**
Bottom Right	0.026	0.023	0.024	0.172	0.184	0.164
Mean	0.027	0.024	0.024	0.179	0.188	0.171

Table A.42: Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.05 mm on the camera translation and 0.04 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.023	0.020	0.022	0.154	0.166	0.148
Top Center	0.021	0.018	0.019	0.132	0.144	0.130
Top Right	**0.027**	**0.024**	**0.025**	**0.182**	**0.192**	**0.168**
Center Left	0.024	0.024	0.022	0.168	0.168	0.160
Center Right	0.020	0.020	0.018	0.138	0.140	0.132
Bottom Left	0.021	0.019	0.020	0.144	0.150	0.132
Bottom Center	0.023	0.019	0.020	0.142	0.154	0.138
Bottom Right	0.024	0.021	0.023	0.160	0.174	0.154
Mean	0.023	0.021	0.021	0.153	0.161	0.145

Table A.43: Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.06 mm on the camera translation and 0.06 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.030	0.026	**0.028**	**0.198**	0.214	0.192
Top Center	**0.032**	**0.027**	0.028	**0.198**	**0.216**	**0.194**
Top Right	0.028	0.025	0.026	0.190	0.198	0.176
Center Left	0.023	0.023	0.020	0.158	0.160	0.152
Center Right	0.026	**0.027**	0.024	0.182	0.184	0.176
Bottom Left	0.023	0.020	0.021	0.154	0.162	0.142
Bottom Center	0.022	0.019	0.019	0.140	0.152	0.136
Bottom Right	0.023	0.020	0.021	0.154	0.162	0.142
Mean	0.024	0.022	0.022	0.163	0.170	0.154

Table A.44: Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.1 mm on the object coordinates, 0.02 mm on the camera translation and 0.02 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.022	0.019	0.020	0.144	**0.156**	**0.140**
Top Center	0.021	0.018	0.018	0.130	0.142	0.128
Top Right	0.022	0.020	**0.021**	**0.148**	0.154	0.136
Center Left	0.020	**0.021**	0.018	0.142	0.144	0.136
Center Right	0.018	0.018	0.016	0.124	0.126	0.120
Bottom Left	0.021	0.019	0.020	0.144	0.150	0.134
Bottom Center	**0.023**	0.019	0.020	0.144	**0.156**	**0.140**
Bottom Right	0.022	0.019	0.020	0.144	0.154	0.138
Mean	0.021	0.019	0.019	0.140	0.148	0.134

Table A.45: Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.02 mm on the camera translation and 0.02 mrad on the camera rotation.

	Spatial			Angular		
	x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
Top Left	0.013	0.011	0.012	0.086	0.092	0.082
Top Center	0.012	0.010	0.010	0.074	0.080	0.072
Top Right	0.011	0.010	0.010	0.074	0.076	0.068
Center Left	0.012	0.012	0.011	0.082	0.082	0.078
Center Right	0.011	0.011	0.010	0.078	0.078	0.074
Bottom Left	0.014	0.012	0.013	0.092	0.096	0.086
Bottom Center	0.012	0.010	0.011	0.078	0.084	0.076
Bottom Right	**0.015**	**0.013**	**0.014**	**0.100**	**0.106**	**0.096**
Mean	0.013	0.011	0.011	0.083	0.087	0.079

A.3 Statue simulations

As for the painting simulations the tracking cameras used in the simulations match the AVT Stingray 5 Mpx cameras used with 8 mm lens. The tracked object is the 0.5 m × 0.5 m × 0.5 m target frame covered defined by fifty six points. An overview of all statue simulation results is given table A.46.

Table A.46: Overview of the painting simulation results.

	Input noise				Output accuracy		
Cameras total / used	Picture coord. (µm)	Object coord. (mm)	Orientation translation (mm)	Orientation rotation (mrad)	Spatial (mm)	Angular (mrad)	See table
3/4	0.345	0	0	0	0.056	0.308	A.47
6/8*	0.345	0	0	0	0.031	0.196	A.48
6/8	0.345	0	0	0	0.031	0.210	A.49
	0.345	0.05	0.05	0.04	0.036	0.242	A.50
6/8	0.345	0.05	0.06	0.06	0.057	0.362	A.51
	0.115	0.05	0.02	0.02	0.019	0.122	A.52

* cameras are staggered instead of vertically aligned two by two.

Preliminary simulations

Table A.47: Partial bounding box surveyed by three cameras. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	**0.047**	0.037	0.031	0.226	**0.260**	**0.308**
4	3	0.041	0.030	0.027	0.196	0.212	0.256
2	2	0.037	0.041	0.030	0.222	0.224	0.278
2	3	0.036	0.048	0.034	0.250	0.244	0.290
0	2	0.035	0.053	0.037	0.274	0.232	0.284
0	3	0.036	**0.056**	**0.038**	**0.278**	0.244	0.290
Mean		0.039	0.044	0.033	0.241	0.236	0.284

Table A.48: Partial bounding box surveyed by six cameras. Top and bottom tracking cameras are staggered. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	**0.031**	0.026	**0.029**	**0.196**	**0.196**	0.200
4	3	0.027	0.021	0.024	0.156	0.174	0.166
2	2	0.023	0.025	0.022	0.164	0.158	0.170
2	3	0.025	0.024	0.020	0.156	0.158	0.174
0	2	0.025	**0.029**	0.024	0.186	0.166	0.180
0	3	0.024	0.026	0.020	0.160	0.150	**0.184**
Mean		0.026	0.025	0.023	0.170	0.167	0.179

Table A.49: Partial bounding box surveyed by six cameras. Top and bottom tracking cameras are vertically aligned. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	**0.027**	0.023	**0.023**	0.162	**0.174**	0.174
4	3	0.026	0.022	0.022	0.158	0.170	0.166
3	2	0.022	0.023	0.019	0.150	0.144	0.162
3	3	0.023	0.024	0.020	0.156	0.150	0.168
2	2	0.023	0.029	0.021	0.172	0.158	0.184
2	3	0.023	0.028	0.021	0.166	0.154	0.180
1	2	0.022	0.029	0.020	0.168	0.150	0.180
1	3	0.023	0.029	0.020	0.164	0.150	0.178
0	2	0.023	0.029	0.021	0.176	0.150	0.186
0	3	0.026	**0.031**	0.023	**0.182**	0.162	**0.210**
Mean		0.024	0.027	0.021	0.165	0.156	0.179

Comprehensive simulations

Table A.50: Partial bounding box surveyed by six cameras. Top and bottom tracking cameras are vertically aligned. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.05 mm on the camera translation and 0.04 mrad on the camera rotation.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	0.035	0.029	0.029	0.208	0.222	0.224
4	3	**0.035**	0.029	**0.030**	0.212	**0.228**	0.222
2	2	0.027	0.033	0.024	0.196	0.180	0.212
2	3	0.031	**0.036**	0.027	**0.218**	0.202	0.236
0	2	0.028	**0.036**	0.026	**0.218**	0.186	0.232
0	3	0.029	0.036	0.026	0.210	0.186	**0.242**
Mean		0.031	0.033	0.027	0.210	0.201	0.228

Table A.51: Partial bounding box surveyed by six cameras. Top and bottom tracking cameras are vertically aligned. Standard deviation of Gaussian noise is 0.345 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.06 mm on the camera translation and 0.06 mrad on the camera rotation.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	**0.048**	0.040	**0.041**	0.288	**0.308**	0.308
4	3	0.037	0.031	0.031	0.224	0.240	0.234
3	2	0.039	0.042	0.036	0.274	0.262	0.294
3	3	0.033	0.035	0.029	0.228	0.220	0.244
2	2	0.031	0.038	0.028	0.226	0.208	0.244
2	3	0.033	0.040	0.029	0.236	0.220	0.256
1	2	0.028	0.037	0.026	0.212	0.190	0.228
1	3	0.042	0.053	0.037	0.300	0.274	0.326
0	2	0.044	**0.057**	0.041	**0.340**	0.290	**0.362**
0	3	0.027	0.034	0.024	0.196	0.174	0.228
Mean		0.039	0.044	0.035	0.272	0.256	0.294

Table A.52: Partial bounding box surveyed by six cameras. Top and bottom tracking cameras are vertically aligned. Standard deviation of Gaussian noise is 0.115 µm on the picture coordinates, 0.05 mm on the object coordinates, 0.02 mm on the camera translation and 0.02 mrad on the camera rotation.

Row	Col.	Spatial accuracy			Angular accuracy		
		x (mm)	y (mm)	z (mm)	Ω (mrad)	Φ (mrad)	K (mrad)
4	2	**0.017**	0.014	**0.014**	0.100	**0.108**	0.108
4	3	0.016	0.013	0.014	0.098	0.104	0.102
2	2	0.014	0.017	0.013	0.104	0.096	0.112
2	3	0.016	**0.019**	0.014	**0.110**	0.104	0.120
0	2	0.012	0.015	0.011	0.092	0.078	0.098
0	3	0.015	0.018	0.013	0.106	0.094	**0.122**
Mean		0.015	0.016	0.013	0.102	0.097	0.110

Appendix B

Tracking Accuracy Complementary Data

B.1 Practical considerations

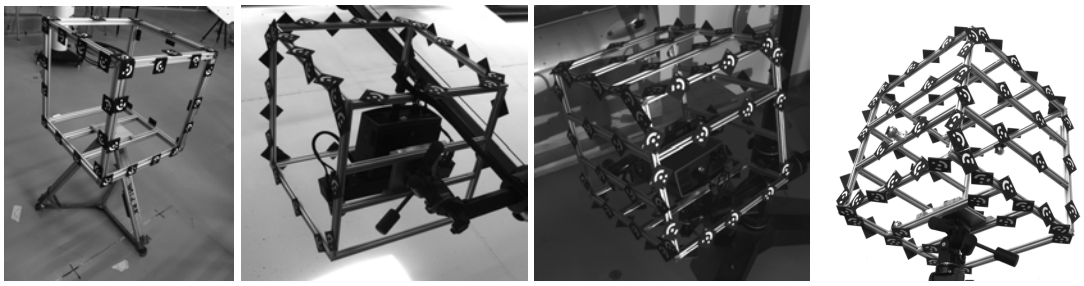
B.1.1 Target frame design

It is important for the final tracking and registration accuracy to design a target frame that is stable. The targets should not move with respect to one another, even of a few tenths of a millimeter. The target frame must also be light so that the frame and acquisition system can be supported by a tripod. Carbon is a light material with a low thermal expansion coefficient, but it is expensive and not adapted to creating several prototypes. Instead, we use aluminum profiles for flexible and rapid prototyping. These aluminum profiles were bought from item [*item*], which has distributors both in France and in Germany.

In the upcoming descriptions, we differentiate the following parts of the target frame:

- The **skeleton**, which is the cubic target frame structure with the additional sustaining bars.
- The **fixtures** used to attach the target frame to a tripod and to fix the acquisition systems to the target frame.
- The **targets**, including the target supports and screws. The targets themselves are created by Tritop and printed on adhesive paper.

The tripod head we use to sustain the target frame and acquisition system can bear up to 16 kg. This is the best compromise we found in terms of sustainable weight and flexibility (the tripod head must be able to rotate in all directions for the 3D digitizations). The heaviest acquisition system we use, the Gom Atos III, weighs 7.5 kg (the weigh of all three acquisition systems we consider is given in table 3.2). As a consequence, the weight of the target frame should not exceed 8.5 kg. A light target frame is also easier to manipulate. In practice it is difficult to design an aluminum target frame that weights less than 8.5 kg and matches our requirements.



(a) A (b) B1 (c) B2 (d) B3

Figure B.1: Target frame variations.

Table B.1: Weight of each target frame variation.

Frame	Number of targets	Weight (kg) Total	Skeleton	Fixtures	Targets
A	40	12.215	6.700	2.172	3.632
B1	48	6.654	3.213	2.188	1.253
B2	78	9.341	5.117	2.188	2.036
B3	80	9.665	5.389	2.188	2.088

Two different target frames were created. They are labeled A and B. There are a few variations on the number of targets covering target frame B. We quickly present these successive target frames, which are illustrated figure B.1 and described in table B.1.

Target frame iteration A

This target frame is made of profiles with a large cross section (30 mm × 30 mm) to maintain its stability. We use the lightest profiles available for this cross section. The target supports are composed of two rectangular brackets that are integrated with the system and can be easily fastened to it. These supports are made of steel and are very heavy. No photogrammetric targets are placed on the side of the target frame that will face the object under study, since it is completely hidden from the cameras. In an attempt to limit the weight of the target frame we do not place any targets on its bottom face. This side can hardly be seen from the tracking cameras when the target frame sits close to the ground to image the lower portion of the sarcophagus surface. We thus place photogrammetric targets on four of the six sides of the target frame.

This target frame was used for the first laser tracker tests. The empty frame was supported by a theodolite tripod, as shown figure B.1a. Such tripods can bear greater mass than our tripod head, which is why we tolerate the overweight target frame, but they are not flexible at all . To secure the target frame to the tripod, a metal plate is fixed between two bars on the bottom of the cube and screwed to the theodolite tripod. This metal plate is made of solid aluminum and measures 160 mm × 160 mm × 16 mm. It ensures a stable base on which the acquisition system, cube skeleton and tripod can be attached, but it weighs over 1 kg.

Target frame iteration B

The first target frame was suitable to be used empty for tracking tests performed with a theodolite tripod, but not to perform acquisitions with the Atos III. We designed a lighter target frame made of aluminum profiles with a smaller cross section (20 mm × 20 mm), even though this smaller cross section makes for a less stable cube.

In the previous configuration, a large proportion of the cube weight came from the targets. To decrease this weight we replaced the steel target holders by plastic target holders. We drilled two holes in opposite corners of these PVC squares of side 6 cm to fasten them to the target frame.

In the first iteration (target frame B1), the target frame was covered with 48 targets and was lighter than 8.5 kg. A few tracking tests with this target frame taught us that more targets were necessary. Increasing the number of targets to 78 or 80 (respectively target frame B2 and B3) makes the target frame weigh over 9 kg. This is in part due to the weight of the target holders, but mostly because of the necessary aluminum bars to support them. However, these additional aluminum bars also improve the stability of the target frame.

We used the same solid aluminum plate used in target frame A to attach the fixtures to the bottom of the target frame. We screwed a hexagonal plate holder to the metal plate to attach the acquisition systems to the target frame. Similarly, we screwed a hexagonal plate to the bottom of the target frame to ensure an easy and steady fastening to the tripod.

B.1.2 Fixing the tracking cameras

We also had to decide how to fix the tracking cameras. We needed a very stable support to keep the cameras from moving during the several hours that each experiment lasted. We also needed to position the cameras almost vertically on top of one another, some very close to the floor and the others almost 2 m high.

Though tripods are the natural solution for fixing cameras, they did not seem adapted to our specific problem: it is not easy to find stable tripods that can sit so low or so high. It would also be difficult to position the tripods close enough to position our cameras one on top of the other. Another concern was that it is fairly easy to trip over the feet of a tripod.

Instead, we used aluminum profiles to create a stand for the tracking cameras. We selected profiles with a cross section of 40 mm × 40 mm, larger than those used for the target frame. The first tests were performed with profiles available in the laboratory. The tracking cameras were fixed to the profiles with a hinge, which provided some flexibility for their orientation.

For the other tests, we designed a trestle-shaped frame that served as a camera support. By adding or removing a few pieces from this modular support we were able to adapt it both to the four camera configuration and to the six camera configuration. The cameras were fixed to the support using a small aluminum ball-joint, also from item, which was easily integrated to the stand.

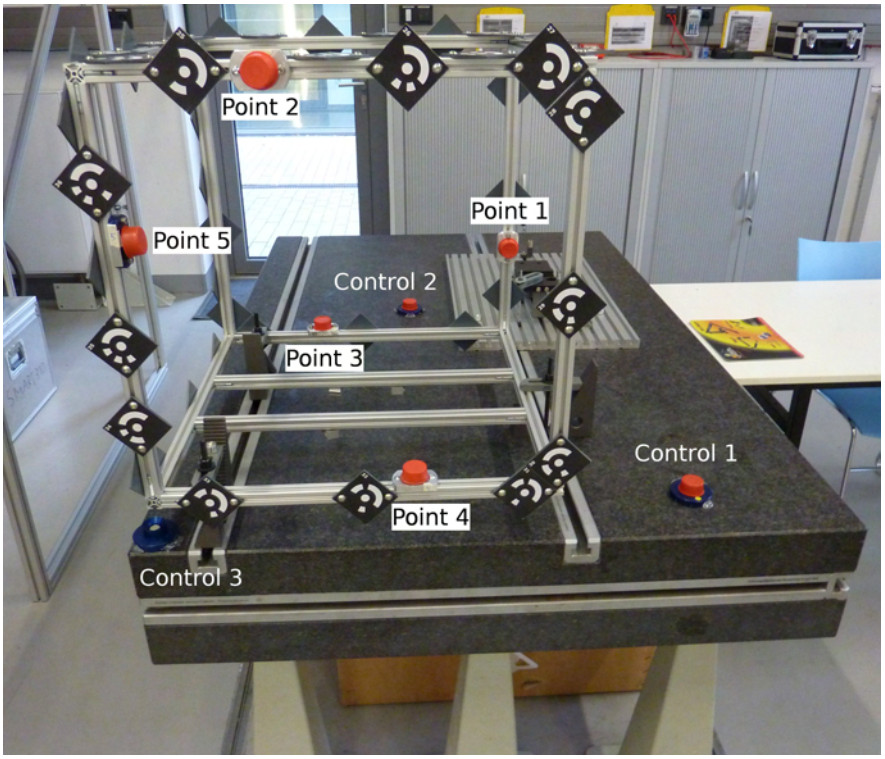


Figure B.2: Setup used to evaluate the laser tracker accuracy.

B.2 Laser tracker measurements

B.2.1 Evaluating the laser tracker accuracy

Before performing our measurements using the laser tracker, we wanted to evaluate its accuracy in our laboratory setup. Five CCR target holders were fixed to the target frame and three were glued to a marble table. The target frame was clamped to the marble table to fix it for the duration of the acquisitions (see figure B.2).

We acquire the points four times and the control points three times. The order of the acquisitions is the following:

- Points 1 – 5
- Control 1 – 3
- Points 1 – 5
- Control 1 – 3
- Points 1 – 5
- Control 1 – 3
- Points 1 – 5

The mean distance between each point and the laser tracker, as well as the standard deviation σ of each of these distances is given table B.2. The standard deviation if lower than 0.003 mm(1 σ) for all points. The points on the target frame do not have a larger deviation than the control points, which means that the target frame is securely fastened and there are no vibrations.

Table B.2: Repeatability testing of the laser tracker.

	Average (mm)	σ (mm)
Point 1	1922.967	0.0027
Point 2	1353.744	0.0006
Point 3	1971.182	0.0019
Point 4	1474.050	0.0015
Point 5	1418.963	0.0020
Control 1	1649.977	0.0026
Control 2	2293.404	0.0007
Control 3	1527.912	0.0010

B.2.2 Characteristics of photogrammetric tracking measurements

The target frame is clamped to the marble table. To compare the five successive measurements table B.3 shows the number of targets that are seen by one, two, three or for targets while table B.4 shows the average residuals of the targets in the images.

Table B.3: Number of targets used for each measurement.

Number of cameras	Measurement M1	M2	M3	M4	M5
1	1	0	1	2	1
2	23	29	26	21	38
3	0	0	3	1	1
4	10	10	11	10	13
Total	34	39	41	34	53

Table B.4: Average of residuals of targets in pictures for each measurement, in micrometers.

M1	M2	M3	M4	M5
1.264	1.280	0.443	0.547	0.821

B.2.3 Point to point distance

Point to point distances between the CCR targets are given for all acquisitions of the three laser tracker measurements in tables B.5 through B.7. Average and standard deviation σ of the distances over the acquisitions are also provided. All values are given in millimeters.

Table B.5: Point to point distances between the CCR targets during the laser tracker measurements with the target frame fixed to a theodolite tripod. Position of target 5 was not available for measurement M1.

Points	M1	M2	M3	M4	Average	σ
1 to 2	367.926	367.926	367.935	367.921	367.927	0.005
1 to 3	564.652	564.592	565.125	565.151	564.880	0.259
1 to 4	574.438	574.479	574.405	574.440	574.440	0.026
1 to 5	—	374.531	374.453	374.478	374.487	0.032
2 to 3	287.813	287.779	288.279	288.325	288.049	0.254
2 to 4	563.280	563.329	563.250	563.268	563.282	0.029
2 to 5	—	528.490	528.453	528.437	528.460	0.022
3 to 4	416.167	416.165	416.174	416.157	416.166	0.006
3 to 5	—	523.809	524.141	524.111	524.020	0.150
4 to 5	—	263.195	263.214	263.205	263.205	0.008

Table B.6: Point to point distances between the CCR targets during the laser tracker measurements with the target frame clamped to a marble table.

Points	M1	M2	M3	M4	M5	Average	σ
1 to 2	366.008	366.305	366.820	366.942	366.846	366.799	0.460
1 to 3	260.635	260.634	260.633	260.649	260.621	260.644	0.018
1 to 4	335.900	335.761	335.753	335.667	335.822	335.689	0.162
1 to 5	676.478	676.730	675.727	676.150	675.771	676.153	0.339
2 to 3	363.412	363.640	364.234	364.254	364.350	364.128	0.398
2 to 4	495.308	495.122	495.803	495.640	496.135	495.569	0.314
2 to 5	497.213	497.153	497.203	497.079	497.218	497.094	0.139
3 to 4	238.517	238.320	238.289	238.185	238.372	238.229	0.197
3 to 5	505.724	505.856	504.803	505.193	504.939	505.209	0.387
4 to 5	491.948	492.007	490.625	491.182	490.997	491.217	0.509

Table B.7: Point to point distances between the CCR targets during the laser tracker measurements with fixed CCR targets.

Points	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	Average	σ
1 to 2	358.366	358.371	358.348	358.341	358.359	358.359	358.365	358.360	358.349	358.352	358.357	0.009
1 to 3	552.808	552.775	552.753	552.739	552.787	552.775	552.785	552.762	552.770	552.757	552.771	0.019
1 to 4	649.941	649.953	649.973	649.975	649.955	649.947	649.938	649.932	649.945	649.942	649.950	0.014
1 to 5	293.347	293.477	293.370	293.408	293.352	293.376	293.369	293.370	293.387	293.387	293.384	0.035
2 to 3	310.651	310.678	310.660	310.661	310.667	310.653	310.665	310.638	310.672	310.642	310.659	0.012
2 to 4	487.921	487.928	487.948	487.951	487.923	487.922	487.948	487.922	487.954	487.932	487.935	0.013
2 to 5	576.401	576.492	576.374	576.398	576.396	576.433	576.414	576.406	576.410	576.413	576.414	0.030
3 to 4	397.986	397.891	397.881	397.870	397.949	397.944	397.958	397.937	397.961	397.960	397.934	0.037
3 to 5	666.834	666.776	666.690	666.697	666.795	666.802	666.780	666.766	666.771	666.785	666.770	0.042
4 to 5	650.068	650.070	650.051	650.054	650.080	650.120	650.040	650.045	650.041	650.048	650.062	0.023

B.2.4 Influence of calibrations

Table B.8 shows the varying tracking accuracy depending of the target frame calibration and exterior camera orientation used for the processing. The values correspond to the maximum spatial and angular standard deviation. The most accurate calibrations (B for the target frame and B* for the camera orientation) globally result in the most accurate tracking.

Table B.8: Influence of the target frame calibration and exterior camera orientation on the tracking accuracy when tracking with six cameras.

	frame B orient A*		frame B orient B*		frame C orient A*		frame C orient B*	
	(mm)	(mrad)	(mm)	(mrad)	(mm)	(mrad)	(mm)	(mrad)
M1	0.033	0.162	0.031	0.152	0.030	0.148	0.032	0.156
M2	0.024	0.124	0.026	0.130	0.022	0.110	0.022	0.114
M3	0.020	0.100	0.019	0.098	0.020	0.100	0.019	0.094
M4	0.026	0.122	0.027	0.124	0.025	0.118	0.026	0.120
M5	0.021	0.114	0.019	0.108	0.022	0.124	0.021	0.118
M6	0.021	0.108	0.019	0.098	0.024	0.122	0.021	0.110
M7	0.018	0.100	0.016	0.092	0.019	0.110	0.018	0.100
M8	0.019	0.100	0.017	0.092	0.021	0.112	0.020	0.106
M9	0.022	0.106	0.020	0.096	0.025	0.118	0.024	0.114
M10	0.018	0.096	0.017	0.090	0.019	0.104	0.019	0.102
M11	0.016	0.090	0.014	0.080	0.018	0.100	0.017	0.094
M12	0.016	0.080	0.015	0.076	0.018	0.092	0.017	0.090
M13	0.018	0.092	0.017	0.086	0.019	0.100	0.018	0.092
M14	0.024	0.114	0.023	0.108	0.025	0.118	0.024	0.112
M15	0.021	0.102	0.020	0.096	0.023	0.112	0.021	0.100
M16	0.014	0.092	0.013	0.086	0.015	0.100	0.015	0.096
Min.	0.014	0.080	0.013	0.076	0.015	0.092	0.015	0.090
Max.	0.033	0.162	0.031	0.152	0.030	0.148	0.032	0.156
Avg.	0.021	0.106	0.020	0.101	0.022	0.112	0.021	0.107

Appendix C

Improved Reflectance Estimation Complementary Data

C.1 Estimating the position of the light sources

There are several techniques that can be used to estimate the position of the light sources. The three techniques we considered are the following: using photogrammetry by sticking targets to the lights, detecting the specular highlight of a reflective sphere or fixing the lights to the target frame. We review the advantages and constraints of each of these methods.

Using photogrammetry

If we fix a few targets to the lights it is possible to estimate their position using photogrammetry. This technique presents the advantage of being easily integrated in our current acquisition protocol. The lights and their targets must simply visible from at least two of the tracking cameras. There is an inaccuracy in the depth estimation of the lights since all cameras are on a single side, but this disadvantage is mitigated by the dimensions in play and can be corrected by calibrating the position of the targets on the light.

The lights must be positioned to avoid casting shadows of the target frame on the surface under study. The possible position of the lights in the sarcophagus configuration is illustrated figure C.1. In this setup, the light will necessarily be close to raking the surface. We thus minimize the specular reflection but risk casting shadows from the object on itself.

Using a reflective sphere

Recognizing the specular highlights in a photograph of a reflective sphere enable us to evaluate the position of the light sources in a scene with respect to the position the imaging system. This technique is often used to detect the light position to create PTMs [Malzbender *et al.*, 2001]. The setup simply requires a reflective sphere, a camera and a software to process the images.

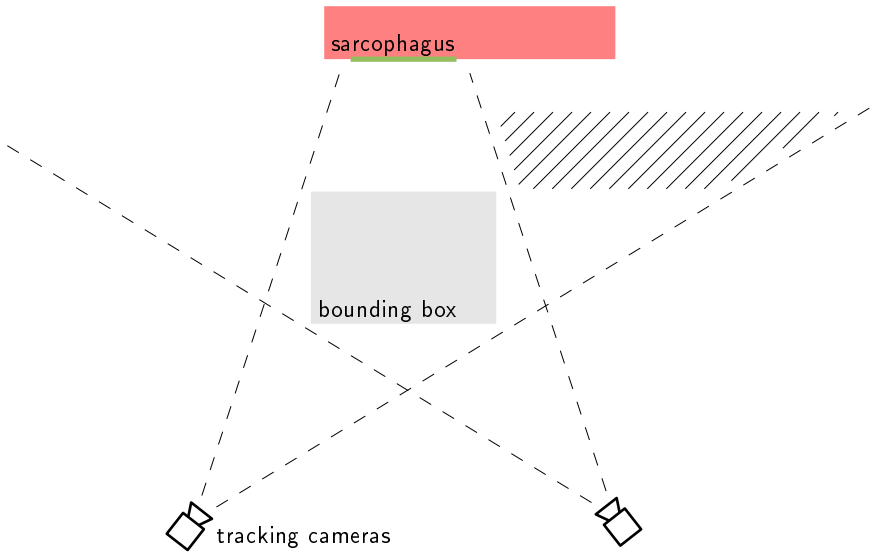


Figure C.1: Positioning of the lights if their position is estimated using photogrammety represented by hatching. We assume the light is at least 20 cm from the sarcophagus.

Open source software to create PTMs is available online at [*LPtracker*]. Such software is able to detect a sphere and its specular highlight in a set of images and to calculate the corresponding light position. The acquisitions can be performed by any of the many cameras available to us: the tracking cameras or the multispectral camera. When using the tracking cameras, the position of each image plane can be known with better accuracy than when using the mutlispectral camera.

A billiard ball is often used as the reflective sphere when acquiring PTMs but such sphere would be too small if the tracking cameras were used for the acquisitions. It is best that the sphere diameter be represented by 100 pixels in the image, which corresponds to a sphere over close to 10 cm in diameter in our setup with a 2 m distance between the tracking camera and sphere.

The possible positions of the lights in this setup are still constrained by the need not to cast shadows from the target frame on the surface under study. However, the lights need not necessarily be in the field of view of the tracking cameras, though they must illuminate this area (see figure C.2). Though the use of a reflective sphere offers a bit more flexibility in the positioning of the light sources than the photogrammetric setup, it is also a sensibly more complicated.

Fixing the lights to the target frame

The lights, for example a set of LED lights, can also be simply fixed to the target frame itself. A pre-calibration (for example using a reflective sphere) can provide their position in the target frame coordinate system. This has numerous advantages: the lights are not in the way of the user, and their is no risk of casting shadows from the frame on the object under study. The on site acquisitions could be further sped if such LED lights were controlled by the same software that controls the multispectral camera. Furthermore, the lights being fixed with respect to the

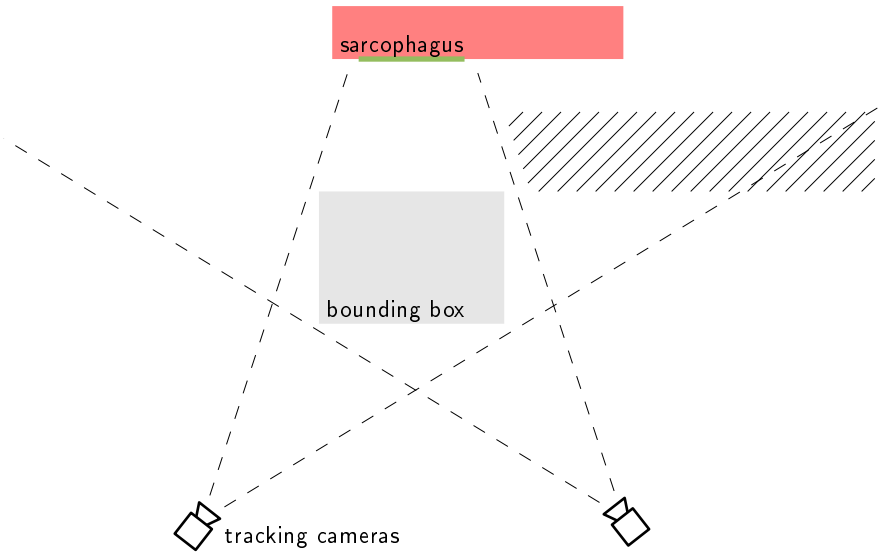


Figure C.2: Possible positioning of the lights if their position is estimated using a reflective sphere represented by hatching. We assume the light is at least 20 cm from the sarcophagus.

multispectral camera, the multispectral calibrations need only be performed once for any set of acquisitions. This drastically reduces the number of necessary acquisitions, greatly reducing the acquisition time. This however, requires a fair amount of work and preparation, that seems disproportionate for a first feasibility test.

Conclusion

The advantages and disadvantages of each technique are summarized in table C.1. The use of lights fixed to the target frame seems to be the most adapted method in the long run. For the first tests presented in chapter 6 however, using our current photogrammetry setup seems the best compromise. The use of a reflective sphere is more complicated and does not provide any significant improvement.

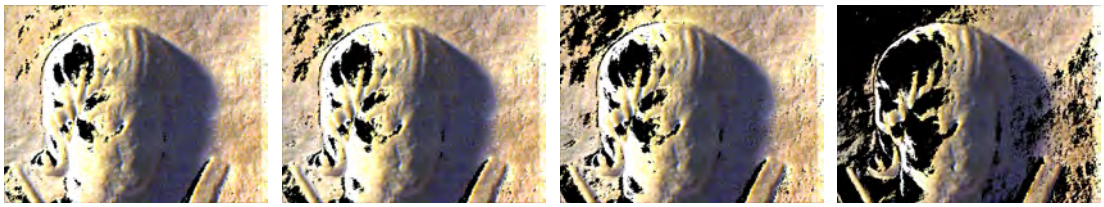
Table C.1: Comparison of possible methods to detect the position of the light sources.

Photogrammetry	Reflective Sphere	Fixed to Frame
Light positioning constraints		
In the field of view of at least two tracking cameras. Should not cast shadows from the target frame on the surface under study.	Should not cast shadows from the target frame on the surface under study.	Fixed to the target frame, no risk of casting shadows from the frame.
Light orientation		
Raking: Minimizes specular highlights but increases risk of shadows from surface itself.		**Direct:** Possible specular highlights but no shadows from surface.
Type of light		
Any available light, can be fixed on a tripod.		Light and soft since directly shines on object: LED.
Difficulty of first setup		
Easy: Setup is that already in place for tracking of the acquisition systems.	**Medium**: Must find a reflective sphere.	**Medium to hard**: Must buy, fix and calibrate lights. Must develop software that controls lights and multispectral camera together.
Difficulty of in situ use		
Easy: An extra calibration step not unlike the others. All light positions acquired in a single shot.	**Medium**: An extra calibration step with different software and procedure. Multiple acquisitions necessary to acquire the position of each light source.	**Very easy**: Light positions known through previous calibrations. Single software controls light and multispectral cameras. Same illumination is cast on each acquisition.

C.2 Detecting the specular highlights

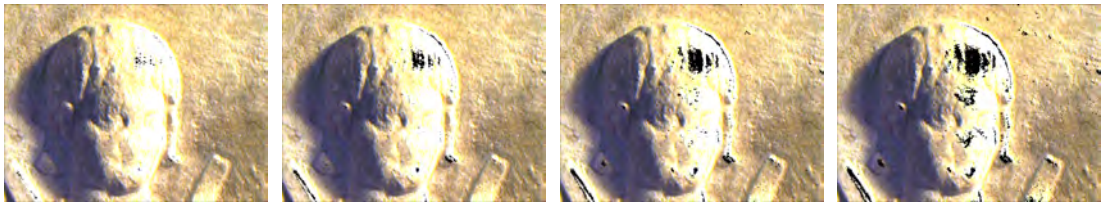
Figure C.3 through C.10 present the specular highlights detected for each multispectral acquisition position and for both the individual lights, for varying values of threshold T. The percentage p of specular pixels of each image is given in the legend. These results highlight the need to adapt the threshold to each image, even though the material is the same.

(a) T = 0.99, p = 0.14 % (b) T = 0.95, p = 0.86 % (c) T = 0.90, p = 1.95 % (d) T = 0.85, p = 3.32 %

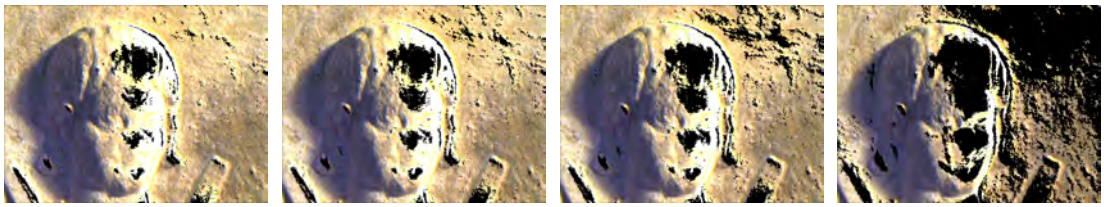


(e) T = 0.80, p = 5.59 % (f) T = 0.75, p = 9.71 % (g) T = 0.70, p = 14.90 % (h) T = 0.50, p = 40.06 %

Figure C.3: Specular highlights detected for the first acquisition with the left light.

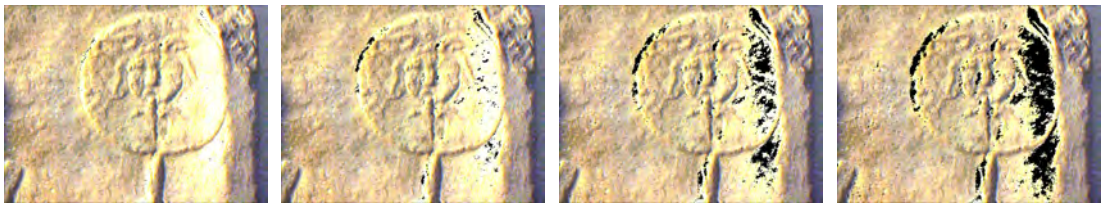


(a) T = 0.99, p = 0.15 % (b) T = 0.95, p = 0.78 % (c) T = 0.90, p = 1.74 % (d) T = 0.85, p = 3.26 %



(e) T = 0.80, p = 5.60 % (f) T = 0.75, p = 8.74 % (g) T = 0.70, p = 13.47 % (h) T = 0.50, p = 39.92 %

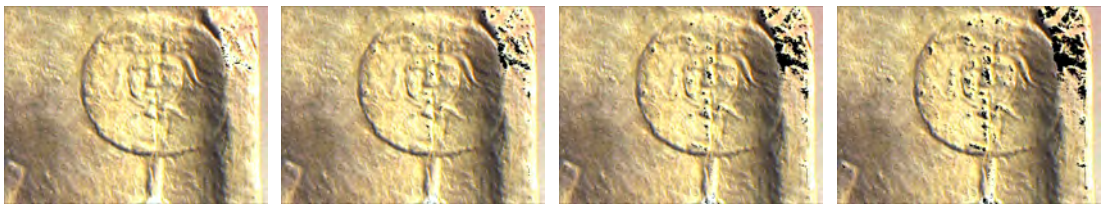
Figure C.4: Specular highlights detected for the first acquisition with the right light.



(a) T = 0.99, p = 0.09 % (b) T = 0.95, p = 1.00 % (c) T = 0.90, p = 4.31 % (d) T = 0.85, p = 8.48 %

(e) T = 0.80, p = 12.27 % (f) T = 0.75, p = 16.29 % (g) T = 0.70, p = 21.15 % (h) T = 0.50, p = 53.49 %

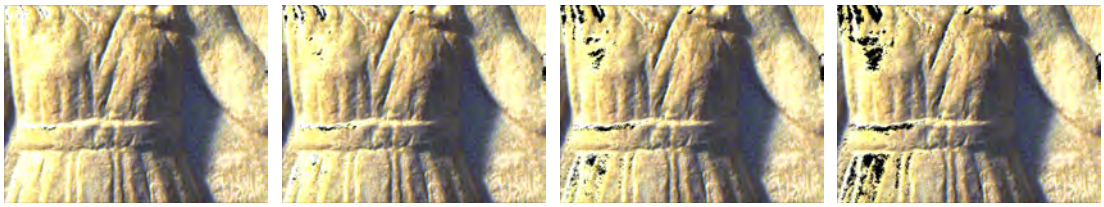
Figure C.5: Specular highlights detected for the second acquisition with the left light.



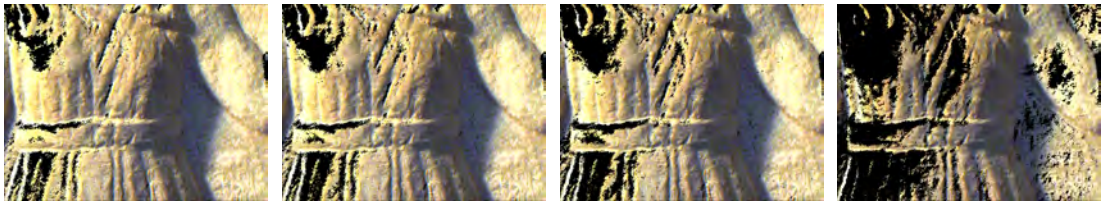
(a) T = 0.99, p = 0.15 % (b) T = 0.95, p = 1.07 % (c) T = 0.90, p = 2.31 % (d) T = 0.85, p = 3.92 %

(e) T = 0.80, p = 6.39 % (f) T = 0.75, p = 10.06 % (g) T = 0.70, p = 14.90 % (h) T = 0.50, p = 41.44 %

Figure C.6: Specular highlights detected for the second acquisition with the right light.



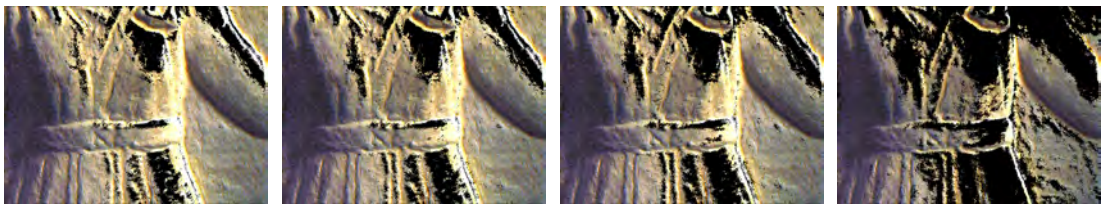
(a) T = 0.99, p = 0.10 % (b) T = 0.95, p = 0.67 % (c) T = 0.90, p = 2.06 % (d) T = 0.85, p = 4.53 %



(e) T = 0.80, p = 7.89 % (f) T = 0.75, p = 11.28 % (g) T = 0.70, p = 15.04 % (h) T = 0.50, p = 35.49 %

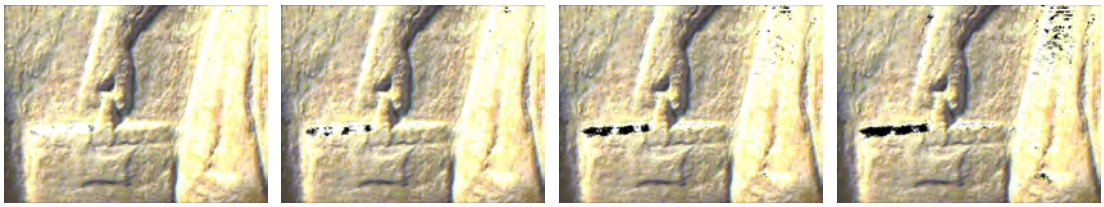
Figure C.7: Specular highlights detected for the third acquisition with the left light.

(a) T = 0.99, p = 0.32 % (b) T = 0.95, p = 2.05 % (c) T = 0.90, p = 4.80 % (d) T = 0.85, p = 8.03 %



(e) T = 0.80, p = 11.66 % (f) T = 0.75, p = 15.44 % (g) T = 0.70, p = 19.48 % (h) T = 0.50, p = 39.40 %

Figure C.8: Specular highlights detected for the third acquisition with the right light.



(a) T = 0.99, p = 0.04 % (b) T = 0.95, p = 0.35 % (c) T = 0.90, p = 0.98 % (d) T = 0.85, p = 1.88 %

(e) T = 0.80, p = 3.43 % (f) T = 0.75, p = 5.52 % (g) T = 0.70, p = 8.02 % (h) T = 0.50, p = 24.45 %

Figure C.9: Specular highlights detected for the fourth acquisition with the left light.

(a) T = 0.99, p = 0.02 % (b) T = 0.95, p = 0.11 % (c) T = 0.90, p = 0.26 % (d) T = 0.85, p = 0.46 %

(e) T = 0.80, p = 0.77 % (f) T = 0.75, p = 1.18 % (g) T = 0.70, p = 1.80 % (h) T = 0.50, p = 8.21 %

Figure C.10: Specular highlights detected for the fourth acquisition with the right light.

C.3 Evaluating the 3D – multispectral registration

To help complete the visual assessment of the accuracy of the 3D – multispectral registration, we create the same kind of checkered images described in chapter 6, for the previous acquisitions. These checkered images are composed of a visual representation of the light / normal dot product, and a greyscale image of the multispectral acquisition. In the two previous cases (cross stitch and bas-relief digitization), we did not detect the position of the light source, we thus calculate a light / normal dot product using the origin of the system as the position of the light source. Though this light position most certainly does not correspond to the real position of the light during the acquisitions, it is sufficient to create a 2D representation of the 3D data, to compare with the multispectral acquisitions.

The resulting checkered images are given figure C.11 and C.12 for the cross-stitch canvas and the bas-relief respectively. In the case of the cross-stitch canvas, we do not denote any other misalignment than the horizontal misalignment previously described. The bas-relief data also seems satisfyingly registered.

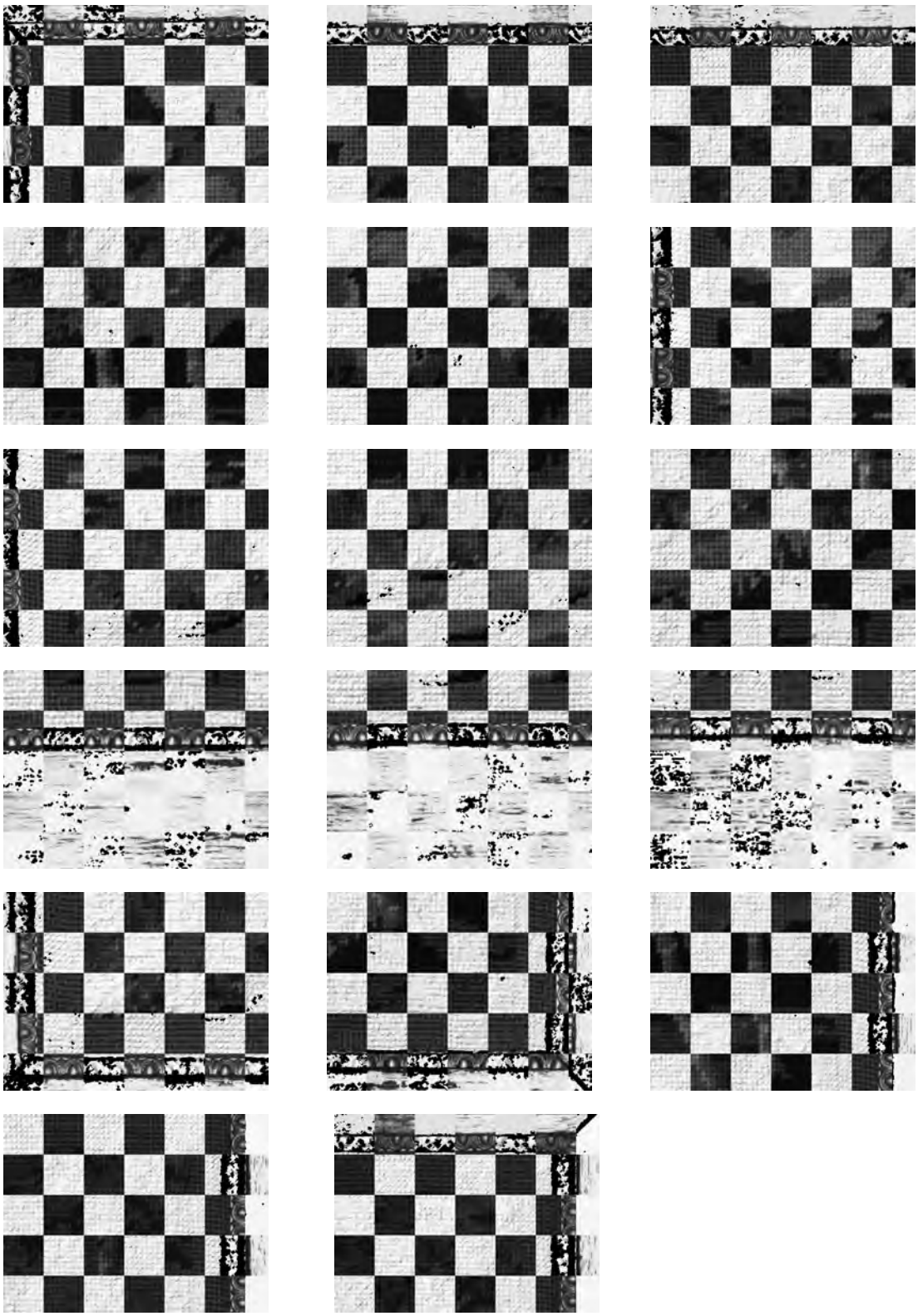


Figure C.11: Light / normal dot product, compared to the cross-stitch multispectral acquisitions.

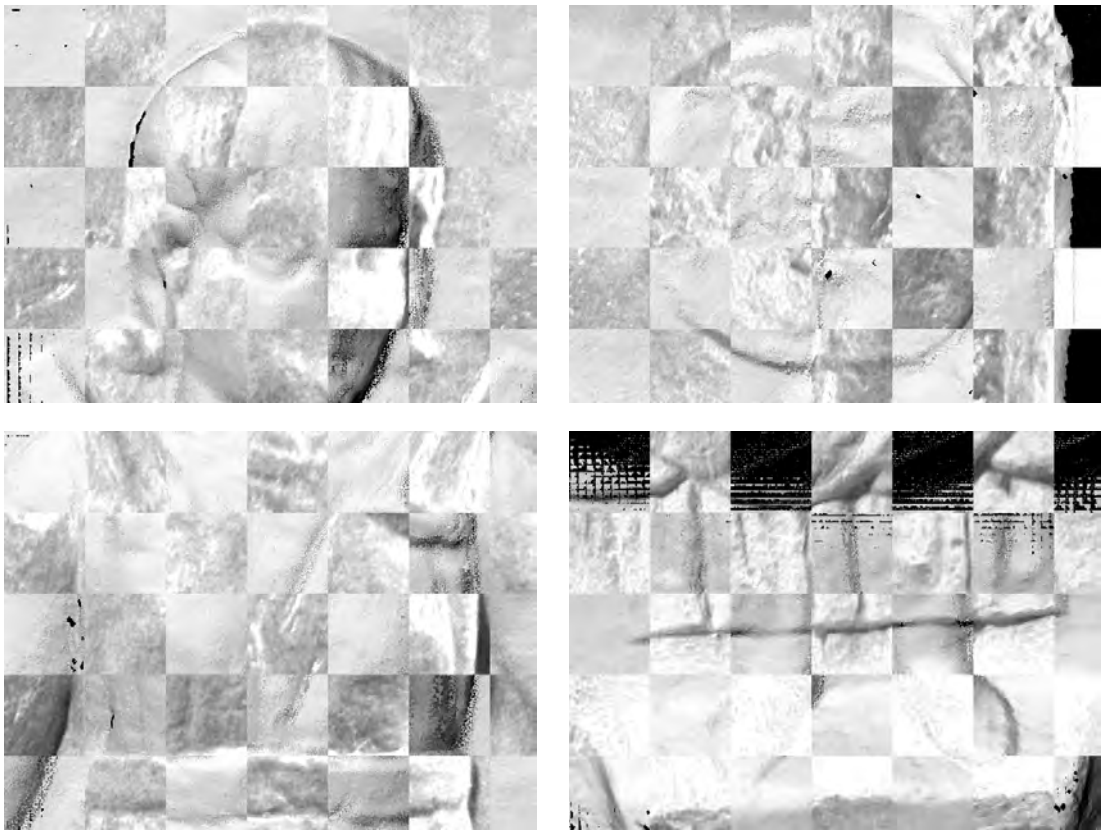


Figure C.12: Light / normal dot product, compared to the bas-relief multispectral acquisitions.

Publications

Journal articles

Simon Chane, C., Mansouri, A., Marzani, F. S., and Boochs, F. (2013a). "Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives". In: *Image and Vision Computing* 31 (1), pp. 91–102.

Simon Chane, C., Schütze, R., Boochs, F., and Marzani, F. (2013b). "Registration of 3D and Multispectral Data for the Study of Cultural Heritage Surfaces". In: *Sensors* 13, pp. 1004–1020.

Simon Chane, C., Schütze, R., Boochs, F., and Marzani, F. S. "Fusion of Arbitrary Multi-view 3D Acquisitions". In: *Computers in Industry, Special Issue on 3D Imaging in Industry.* Revision submitted March 11th 2012.

International conferences

Simon, C., Schütze, R., Boochs, F., and Marzani, F. S. (2012). "Asserting the Precise Position of 3D and Multispectral Acquisition Systems for Multisensor Registration Applied to Cultural Heritage Analysis". In: *Advances in Multimedia Modeling.* Ed. by K. Schoeffmann, B. Merialdo, A. G. Hauptmann, C.-W. Ngo, Y. Andreopoulos, and C. Breiteneder. Vol. 7131. Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg, pp. 597–608.

Simon, C., Huxhagen, U., Mansouri, A., Heritage, A., Boochs, F., and Marzani, F. S. (2010). "Integration of high-resolution spatial and spectral data acquisition systems to provide complementary datasets for cultural heritage applications". In: *Proceedings of SPIE Vol. 7531, Computer Vision and Image Analysis of Art.* Ed. by D. G. Stork, J. Coddington, and A. Bentkowska-Kafel, 75310L:1–9.

Boochs, F., Schütze, R., Simon, C., Marzani, F., Wirth, H., and Meier, J. (2010). "Increasing the accuracy of untaught robot positions by means of a multi-camera system". In: *International Conference on Indoor Positioning and Indoor Navigation.* Zurich, Switzerland: IEEE.

National conference

Schütze, R., Simon, C., Boochs, F., and Marzani, F. (2010). "Fusion von Sensordaten mittels eines photogrammetrischen Mehrkamerasystems". In: *Forum Bildverarbeitung*. Ed. by F. Puente León and M. Heizmann. Regensburg, Germany: KIT Scientific Publishing, pp. 337–348.

Workshops

Simon Chane, C., Schütze, R., Marzani, F. S., and Boochs, F. "Projection et recalage de données multispectrales sur des modèles 3D". In: *Journée thématique GDR ISIS : Imagerie multi/hyperspectrale : problématique et nouveaux domaines d'application*. (July 4, 2012). Paris, France.

Simon, C., Schütze, R., Boochs, F., and Marzani, F. "Repérage précis de caméras multispectrales et de scanners 3d pour le recalage de données multicapteurs appliqué à l'étude du patrimoine". In: *RFIA — Vision 3D et Patrimoine Culturel*. (Jan. 24, 2012). Lyon, France.

Boochs, F., Schütze, R., Simon, C., and Marzani, F. "Flexible and precise determination of the exterior orientation of optical sensors based on a multi-camera set up". In: *Large Volume Metrology Conference*. (Nov. 15–16, 2011). Manchester, UK.

Boochs, F., Huxhagen, U., and Simon, C. (2010). "Einsatz moderner Messentechnik für präzise dimensionelle Überwachung von Gesteinsoberflachen". In: *Naturwissenschaften in der Denkmalpflege — Festschrift zum 20 jährigen Bestehen des Instituts für Steinkonservierung e.V. IFS-Bericht*. Vol. 36.

Invited seminar

Simon, C. "Integration of high-quality multispectral and 3D datasets for the study of cultural heritage objects". In: *COST IE0601 Focused meeting on Non-destructive techniques to study Wooden Cultural Heritage Objects*. (May 6–7, 2011). Paris, France.

Bibliography

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Seitz, S., and Szeliski, R. (2009). "Building Rome in a Day". In: *International Conference on Computer Vision*. IEEE, pp. 72–79.

Andreetto, M., Brusco, N., and Cortelazzo, G. M. (2004). "Automatic 3-d modeling of textured cultural heritage objects". In: *IEEE Transactions on Image Processing* 13.3, pp. 354–369.

Autodesk. *3ds Max, 3D Modeling, Animation, and Rendering Software*. Last access October 2012. URL: http://www.usa.autodesk.com/3ds-max.

Axios3D. *Axori photogrammetric bundle block adjustment, AXOriLib 1.10.8*. URL: http://www.axios3d.de.

Bajard, A., Aubreton, O., Eren, G., Sallamand, P., and Truchetet, F. (2011). "3D digitization of metallic specular surfaces using scanning from heating approach". In: *Proceedings of SPIE Vol. 7864, Three-Dimensional Imaging, Interaction and Measurement*. Ed. by A. M. Baskurt, I. E. McDowall, J.-A. Beraldin, G. S. Cheok, M. Dolinsky, M. B. McCarthy, and U. Neuschaefer-Rube, 786413:1–7.

Bajcsy, R., Lee, S. W., and Leonardis, A. (1996). "Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation." In: *International Journal of Computer Vision* 17.3, pp. 241–272.

Barni, M., Pelagotti, A., and Piva, A. (2005). "Image Processing for the Analysis and Conservation of Paintings: Opportunities and Challenges". In: *IEEE Signal Processing Magazine* 22.5, pp. 141–144.

Barone, S., Paoli, A., and Razionale, A. V. (2012a). "3D Reconstruction and Restoration Monitoring of Sculptural Artworks by a Multi-Sensor Framework". In: *Sensors* 12.12, pp. 16785–16801.

— (2012b). "Shape measurement by a multi-view methodology based on the remote tracking of a 3D optical scanner". In: *Optics and Lasers in Engineering* 50.3, pp. 380–390.

Ben Hmida, H., Boochs, F., Cruz, C., and Nicolle, C. (2012). "Knowledge Base Approach for 3D Objects Detection in Point Clouds Using 3D Processing and Specialists Knowledge". In: *International Journal on Advances in Intelligent Systems* 5.1 and 2.

Beraldin, J.-A., Blais, F., Boulanger, P., Cournoyer, L., Domey, J., El-Hakim, S., Godin, G., Rioux, M., and Taylor, J. (2000). "Real world modelling through high resolution digital

3D imaging of objects and structures". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 55.4, pp. 230–250.

Bernardini, F., Martin, I., and Rushmeier, H. (2001). "High-quality texture reconstruction from multiple scans". In: *IEEE Transactions on Visualization and Computer Graphics* 7.4, pp. 318–332.

Bernardini, F. and Rushmeier, H. (2002). "The 3D Model Acquisition Pipeline". In: *Computer Graphics Forum* 21.2, pp. 149–172.

Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., and Taubin, G. (2002). "Building a Digital Model of Michelangelo's Florentine Pietà". In: *IEEE Computer Graphics and Applications* 22.1, pp. 59–67.

Berns, R. S. (2001). "The Science of Digitizing Paintings for Color-Accurate Image Archives: A Review". In: *Journal of Imaging Science and Technology* 45.4, pp. 305–325.

— (2005a). "Color-Accurate Image Archives Using Spectral Imaging". In: *Proceedings of the National Academy of Science - Scientific Examination of Art: Modern Techniques in Conservation And Analysis*, pp. 105–119.

— (2005b). "Rejuvenating the Appearance of Cultural Heritage Using Color and Imaging Science Techniques". In: *Congress of the International Colour Association*, pp. 369–374.

Berns, R. S., Taplin, L. A., Nezamabadi, M., Mohammadi, M., and Zhao, Y. (2005). "Spectral imaging using a commercial color-filter array digital camera". In: *Proceedings of the 14th Triennial ICOM-CC Meeting*. Vol. 2, pp. 743–750.

Berns, R. S., Taplin, L. A., Urban, P., and Zhao, Y. (2008). "Spectral Color Reproduction of Paintings". In: *Color in Graphics, Imaging and Vision*. Vol. 4. IS&T, pp. 484–488.

Besl, P. J. and McKay, N. D. (1992). "A Method for Registration of 3-D Shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2, pp. 239–256.

Bianco, S., Gasparini, F., Schettini, R., and Vanneschi, L. (2008). "Polynomial modeling and optimization for colorimetric characterization of scanners". In: *Journal of Electronic Imaging* 17.4, pp. 0430021–13.

Blais, F., Taylor, J., Cournoyer, L., Picard, M., Borgeat, L., Godin, G., Beraldin, J.-A., Rioux, M., and Lahanier, C. (2007). "Ultra High-Resolution 3D Laser Color Imaging of Paintings: The Mona Lisa by Leonardo Da Vinci". In: *Lasers in the Conservation of Artworks*. Ed. by M. Castillejo and P. Moreno, pp. 435–440.

Blais, F. and Beraldin, J. A. (2006). "Recent developments in 3D multi-modal laser imaging applied to cultural heritage". In: *Machine Vision and Applications* 17.6, pp. 395–409.

Blais, F., Taylor, J., Beraldin, J. A., Godin, G., Cournoyer, L., Picard, M., Borgeat, L., Dicaire, L., Rioux, M., Lahanier, C., and Aitken, G. (2005). "Ultra-High Resolution Imaging at 50 μm using a Portable XYZ-RGB Color Laser Scanner". In: *International Workshop on Recording, Modeling and Visualization of Cultural Heritage*, pp. 101–114.

Blinn, J. F. (1977). "Models of light reflection for computer synthesized pictures". In: *ACM SIGGRAPH Computer Graphics*. Vol. 11. 2, pp. 192–198.

Bloechl, K., Hamlin, H., and Easton, R. L. J. (2010). "Text recovery from the ultraviolet-fluorescence spectrum for a treatise in the Archimedes palimpsest". In: *Proceedings of SPIE Vol. 7531, Computer Vision and Image Analysis of Art*. Ed. by D. G. Stork, J. Coddington, and A. Bentkowska-Kafel, 753108:1–9.

Boochs, F., Schütze, R., Raab, C., Wirth, H., and Meier, J. (2009). "A flexible multi-camera system for precise tracking of moving effectors". In: *Robotics and Applications*.

Bouchard, M. B., Chen, B. R., Burgess, S. A., and Hillman, E. M. C. (2009). "Ultra-fast multispectral optical imaging of cortical oxygenation, blood flow, and intracellular calcium dynamics." In: *Optics Express* 17.18, pp. 15670–15678.

Brauers, J. and Aach, T. (2006). "A Color Filter Array Based Multispectral Camera". In: *12. Workshop Farbbildverarbeitung*.

— (2008a). "Longitudinal aberrations caused by optical filters and their compensation in multispectral imaging". In: *International Conference on Image Processing*. IEEE, pp. 525–528.

— (2008b). "Modeling and Compensation of Ghosting in Multispectral Filter Wheel Cameras". In: *IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 85–88.

Brauers, J., Helling, S., and Aach, T. (2009). "Multispectral Image Acquisition with Flash Light Sources". In: *Journal of Imaging Science and Technology* 53.3, pp. 0311031–10.

Brauers, J., Schulte, N., and Aach, T. (2008). "Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms". In: *IEEE Transactions on Image Processing* 17.12, pp. 2368–2380.

Breuckmann. *NaviSCAN — Inspection and Scanning without Limits*. Last access February 2011. URL: http://www.breuckmann.com/en/arts-culture/products/naviscan.html.

Brusco, N., Cortelazzo, G. M., Capeleto, S., Fedel, M., Paviotti, a., Poletto, L., and Tondello, G. (2006). "A System for 3D Modeling Frescoed Historical Buildings with Multispectral Texture Information". In: *Machine Vision and Applications* 17.6, pp. 373–393.

Callieri, M., Cignoni, P., Ganovelli, F., Impoco, G., Montani, C., Pingi, P., Ponchio, F., and Scopigno, R. (2004). "Visualization and 3D data processing in the David restoration". In: *IEEE Computer Graphics and Applications* 24.2, pp. 16–21.

Chan, Y., Delmas, P., Gimel'farb, G., and Valkenburg, R. (2010). "Fusing Large Volumes of Range and Image Data for Accurate Description of Realistic 3D Scenes". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by J. Blanc-Talon, D. Bone, W. Philips, D. Popescu, and P. Scheunders. Vol. 6474. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 332–343.

Chang, P.-R. and Hsieh, T.-H. (1995). "Constrained Nonlinear Optimization Approaches to Color-Signal Separation". In: *IEEE Transactions on Image Processing* 4.1, pp. 81–94.

Cheung, V., Westland, S., Li, C., Hardeberg, J., and Connah, D. (2005). "Characterization of trichromatic color cameras by using a new multispectral imaging technique". In: *Journal of the Optical Society of America A* 22.7, pp. 1231–1240.

Clarke, T. and Wang, X. (2000). "The control of a robot end-effector using photogrammetry". In: *International Archives of Photogrammetry and Remote Sensing* XXXIII.Part B5, pp. 137–142.

Colantoni, P., Pillay, R., Lahanier, C., and Pitzalis, D. (2006). "Analysis of multispectral images of paintings". In: *European Signal Processing Conference*.

Connah, D. R. and Hardeberg, J. Y. (2005). "Spectral recovery using polynomial models". In: *Proceedings of SPIE Vol. 5667, Color Imaging X: Processing, Hardcopy, and Applications*. Ed. by R. Eschbach and G. G. Marcu, pp. 65–75.

Cook, R. L. and Torrance, K. E. (1982). "A Reflectance Model for Computer Graphics". In: *ACM Transactions on Graphics* 1 (1), pp. 7–24.

Corsini, M., Dellepiane, M., Ponchio, F., and Scopigno, R. (2009). "Image-to-Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties". In: *Computer Graphics Forum* 28.7, pp. 1755–1764.

Cotte, P. and Dupouy, M. (2003). "Crisatel High Resolution Multispectral System". In: *Image Processing, Image Quality, Image Capture Systems Conference*. Vol. 6. IS&T, pp. 161–165.

Cotte, P. and Dupraz, D. (2006a). "Multispectral Photography of the Famous Mona Lisa Painting". In: *Color in Graphics, Imaging and Vision*. Vol. 3. IS&T, pp. 311–317.

— (2006b). "Spectral Imaging of Leonardo Da Vinci's Mona Lisa: An Authentic Smile at 1523 dpi with Additional Infrared Data". In: *Archiving 2006*, pp. 228–235.

Creaform. *MetraScan 70 and 120*. Last access September 2012. URL: http://www.creaform3d.com/en/metrology-solutions/optical-3d-scanner-metrascan.

Cucci, C., Picollo, M., and Vervat, M. (2012). "Trans-illumination and trans-irradiation with digital cameras: Potentials and limits of two imaging techniques used for the diagnostic investigation of paintings". In: *Journal of Cultural Heritage* 13.1, pp. 83–88.

Dana, K. J., Ginneken, B. van, Nayar, S. K., and Koenderink, J. J. (1999). "Reflectance and texture of real-world surfaces". In: *ACM Transactions on Graphics* 18.1, pp. 1–34.

Day, E. A. (2003). "The Effects of Multi-channel Visible Spectrum Imaging on Perceived Spatial Image Quality and Color Reproduction Accuracy". M.S. Thesis. Rochester Institute of Technology.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. (2000). "Acquiring the reflectance field of a human face". In: *SIGGRAPH Conference Proceedings*, pp. 145–156.

Delaney, J. K., Walmsey, E., Berrie, B. H., and Fletcher, C. F. (2005). "Multispectral Imaging of Paintings in the Infared to Detect and Map Blue Pigments". In: *Proceedings of the National Academy of Sciences*, pp. 120–136.

Drew, M. S. and Finlayson, G. D. (2007). "Analytic solution for separating spectra into illumination and surface reflectance components". In: *Journal of the Optical Society of America A* 24.2, pp. 294–303.

Du, H., Tong, X., Cao, X., and Lin, S. (2009). "A prism-based system for multispectral video acquisition." In: *International Conference on Computer Vision*. IEEE, pp. 175–182.

Easton Jr., R. L., Knox, K. T., and Christens-Barry, W. A. (2003). "Multispectral imaging of the Archimedes palimpsest". In: *Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop*. IEEE, pp. 111–116.

Eren, G., Aubreton, O., Meriaudeau, F., Sanchez Secades, L. A., Fofi, D., Naskali, A. T., Truchetet, F., and Ercil, A. (2009). "Scanning from heating : 3D shape estimation of transparent objects from local surface heating". In: *Optics Express* 17.14, pp. 370–375.

Everdell, N. L., Styles, I. B., Claridge, E., Hebden, J. C., and Calcagni, A. S. (2009). "Multispectral imaging of the ocular fundus using LED illumination". In: *Proceedings of SPIE Vol. 7371, Novel Optical Instrumentation for Biomedical Applications IV*, 73711C:1–6.

Faro. *Competitive Edge with laser line scanning*. Last access September 2012. URL: http://www.faro.com/edge/us/scanner.

Ferraton, M., Stolz, C., and Mériaudeau, F. (2009). "Optimization of a polarization imaging system for 3D measurements of transparent objects." In: *Optics Express* 17.23, pp. 21077–82.

FluxData. *The FD-1665-MS 7 Channel Camera*. Last access October 2012. URL: http://www.fluxdata.com/products/fd-1665-ms7.

Forbes, K., Nicolls, F., Jager, G. de, and Voigt, A. (2006). "Shape-from-Silhouette with Two Mirrors and an Uncalibrated Camera". In: *Computer Vision — ECCV 2006*. Ed. by A. Leonardis, H. Bischof, and A. Pinz. Vol. 3952. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 165–178.

Franken, T., Dellepiane, M., Ganovelli, F., Cignoni, P., Montani, C., and Scopigno, R. (2005). "Minimizing user intervention in registering 2D images to 3D models". In: *The Visual Computer* 21.8-10, pp. 619–628.

Fraunhofer. *Handheld optical 3D-scanner "kolibri Cordless"*. Last access September 2012. URL: http://www.iof.fraunhofer.de/en/business-fields/photonic-sensors-and-measuring-systems/3d-messsysteme/kolibri-cordless-handheld-3d.html.

Fukunaga, K. and Picollo, M. (2010). "Terahertz spectroscopy applied to the analysis of artists' materials". In: *Applied Physics A* 100.3, pp. 591–597.

Furukawa, Y. and Ponce, J. (2010). "Accurate, dense, and robust multi-view stereopsis". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.8, pp. 1362–1376.

Gom. *ATOS — Industrial 3D Scanning Technology*. Last access October 2012. URL: http://www.gom.com/metrology-systems/3d-scanner.html.

Gom. *Tritop Deformation Software*. Last access October 2012. URL: http://www.gom.com/3d-software/tritop-deformation-software.html.

Gooch, A., Gooch, B., Shirley, P., and Cohen, E. (1998). "A non-photorealistic lighting model for automatic technical illustration". In: *SIGGRAPH Conference Proceedings*, pp. 447–452.

Guarnieri, A., Guidi, G., Tucci, G., and Vettore, A. (2003). "Towards automatic modeling for cultural heritage applications". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXIV.Part 5/W12, pp. 176–181.

Guidi, G., Frischer, B., Russo, M., Spinetti, A., Carosso, L., and Micoli, L. L. (2006). "Three-dimensional acquisition of large and detailed cultural heritage objects". In: *Machine Vision and Applications* 17.6, pp. 349–360.

Guidi, G., Remondino, F., Russo, M., Menna, F., and Rizzi, A. (2008). "3D Modeling of Large and Complex Site Using Multi-sensor Integration and Multi-resolution Data". In: *International Symposium on Virtual Reality, Archeology and Cultural Heritage*. Ed. by M. Ashley, S. Hermon, A. Proenca, and K. Rodriguez-Echavarria. Eurographics Association, pp. 85–92.

Haneishi, H., Hasegawa, T., Hosoi, A., Yokoyama, Y., Tsumura, N., and Miyake, Y. (2000). "System Design for Accurately Estimating the Spectral Reflectance of Art Paintings". In: *Applied Optics* 39.35, pp. 6621–6632.

Hardeberg, J. Y., Schmitt, F. J., and Brettel, H. (2002). "Multispectral color image capture using a liquid crystal tunable filter". In: *Optical Engineering* 41.10, pp. 2532–2548.

Hardeberg, J. Y. (1999). "Acquisition and reproduction of colour images: colorimetric and multispectra approaches". PhD thesis. École Nationale Supérieure des Télécommunications de Paris.

Hawkins, T., Cohen, J., and Debevec, P. (2001). "A Photometric Approach to Digitizing Cultural Artifacts". In: *Conference on Virtual Reality, Archeology, and Cultural Heritage*. ACM, pp. 333–342.

He, X. D., Torrance, K. E., Sillion, F. X., and Greenberg, D. P. (1991). "A comprehensive physical model for light reflection". In: *Computer Graphics* 25 (4), pp. 175–186.

Hefele, J. (2002). "Real-time photogrammetric algorithms for robot calibration". In: *International Archives of Photogrammetry and Remote Sensing* XXXIV.Part 5, pp. 33–38.

Hefele, J. and Brenner, C. (2001). "Robot pose correction using photogrammetric tracking". In: *Proceedings of SPIE Vol. 4189, Machine Vision and Three-Dimensional Imaging Systems for Inspection and Metrology*. Ed. by K. G. Harding, J. W. V. Miller, and B. G. Batchelor, pp. 170–178.

Hernández-Andrés, J., Romero, J., and Lee, R. L. (2001). "Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain." In: *Journal of the Optical Society of America. A* 18.2, pp. 412–20.

Ho, J., Funt, B. V., and Drew, M. S. (1990). "Separating a Color Signal into Illumination and Surface Reflectance Components: Theory and Applications". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.10, pp. 966–977.

Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M., and Okamoto, Y. (2007). "The Great Buddha Project: Digitally Archiving, Restoring, and Analyzing Cultural Heritage Objects". In: *International Journal of Computer Vision* 75.1, pp. 189–208.

Imai, F. H. and Berns, R. S. (1999). "Spectral estimation using trichromatic digital cameras". In: *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, pp. 42–49.

Imai, F. H. and Berns, R. S. (2002). "Spectral estimation of artist oil paints using multi-filter trichromatic imaging". In: *Proceedings of SPIE Vol. 4421, 9th Congress of the International Color Association*. Ed. by R. Chung and A. Rodrigues, pp. 504–507.

Imai, F. H., Berns, R. S., and Tzeng, D.-T. (2000). "A Comparative Analysis of Spectral Reflectance Estimated in Various Spaces Using a Trichromatic Camera System". In: *Journal of Imaging Science and Technology* 44.4, pp. 280–287.

Imai, F. H., Rosen, M. R., and Berns, R. S. (2001). "Multi-spectral imaging of Van Gogh's self-portrait at the national gallery of art Washington D.C." In: *Image Processing, Image Quality, Image Capture Systems Conference*. Vol. 4. IS&T, pp. 185–189.

Imai, F. H., Taplin, L. A., and Day, E. A. (2003). *Comparative study of spectral reflectance estimation based on broad-band imaging systems*. Tech. rep. Rochester Institute of Technology, College of Science, Center for Imaging Science, Munsell Color Science Laboratory.

International Commission on Illumination. *Selected Colorimetric Tables*. Last access January 2013. URL: http://www.cie.co.at/index.php/LEFTMENUE/index.php?i_ca_id=298.

item. *Building Kit System for Industrial Applicatinos*. Last access October 2012. URL: http://www.item24.com/.

Jolivot, R., Vabres, P., and Marzani, F. (2011). "Reconstruction of hyperspectral cutaneous data from an artificial neural network-based multispectral imaging system". In: *Computerized Medical Imaging and Graphics* 35.2, pp. 85–88.

Jordan, J. and Angelopoulou, E. (2010). "Gerbil-A Novel Software Framework for Visualization and Analysis in the Multispectral Domain". In: *Vision, Modelling & Visualization*. Ed. by R. Koch, A. Kolb, and C. Rezk-Salama, pp. 259–266.

Karmacharya, A., Cruz, C., Boochs, F., and Marzani, F. (2010a). "Spatial Rules through Spatial Rule built-ins in SWRL". In: *Journal of Global Research in Computer Science* 1.2.

— (2010b). "Use of Geospatial Analyses for Semantic Reasoning". In: *Knowledge-Based and Intelligent Information and Engineering Systems*. Ed. by R. Setchi, I. Jordanov, R. Howlett, and L. Jain. Vol. 6276. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 576–586.

Kawakami, R., Wright, J., Tai, Y.-W., Matsushita, Y., Ben-Ezra, M., and Ikeuchi, K. (2011). "High-resolution Hyperspectral Imaging via Matrix Factorization". In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2329–2336.

Kawasaki, H. and Furukawa, R. (2004). "Entire model acquisition system using handheld 3D digitizer". In: *Proceedings 2nd International Symposium on 3D Data Processing, Visualization and Transmission*, pp. 478–485.

Kim, S. J., Zhao Shaojieand Deng, F., Fu, C.-W., and Brown, M. S. (2010). "Interactive Visualization of Hyperspectral Images of Historical Documents". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6, pp. 1441–1448.

Koirala, P., Hauta-Kasari, M., and Parkkinen, J. (2009). "Highlight Removal from Single Image". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders. Vol. 5807. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 176–187.

Koller, D., Trimble, J., Najbjerg, T., Gelfand, N., and Levoy, M. (2006). "Fragments of the city: Stanford's digital forma urbis romae project". In: *Proceedings of the Third Williams Symposium on Classical Architecture, Journal of Roman Archaeology Supplement*. Vol. 61, pp. 237–252.

Lafortune, E. P. F., Foo, S.-C., Torrance, K. E., and Greenberg, D. P. (1997). "Non-Linear Approximation of Reflectance Functions". In: *SIGGRAPH Conference Proceedings*. ACM, pp. 117–126.

Lahanier, C., Aitken, G., Pillay, R., Beraldin, J.-A., Blais, F., Borgeat, L., Cournoyer, L., Picard, M., Rioux, M., Taylor, J., Breuckmann, B., Colantoni, P., and Deyne, C. de (2005). "Two-dimensional multi-spectral digitization and three-dimensional modelling of easel paintings". In: *Proccedings of the 14th Triennal ICOM-CC Meeting*. The Hague, Netherlands, pp. 30–42.

Leica. *T-Scan*. Last access September 2012. URL: http://metrology.leica-geosystems.com/en/Leica-T-Scan_1836.htm.

Levoy, M. (2002). *An interactive kiosk for the Tribune del David*. Last access September 2011. URL: http://graphics.stanford.edu/projects/mich/kiosk/kiosk.html.

Levoy, M., Rusinkiewicz, S., Ginzton, M., Ginsberg, J., Pulli, K., Koller, D., Anderson, S., Shade, J., Curless, B., Pereira, L., Davis, J., and Fulk, D. (2000). "The Digital Michelangelo Project: 3D Scanning of Large Statues". In: *SIGGRAPH Conference Proceedings*. ACM, pp. 131–144.

Lewis, R. R. (1994). "Making Shaders More Physically Plausible". In: *Computer Graphics Forum* 13.2, pp. 109–120.

Li, R., Luo, T., and Zha, H. (2010). "3D digitization and its Applications in Cultural Heritage". In: *Digital Heritage*. Vol. 6436. Lecture Notes in Computer Science, pp. 381–388.

Li, W., Böhler, M., Schütze, R., Marzani, F. S., Voisin, Y., and Boochs, F. (2008). "Use of Local Surface Curvature Estimation for Adaptive Vision System Based on Active Light Projection". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders. Vol. 5259. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 530–541.

Li, W., Schütze, R., Böhler, M., Boochs, F., Marzani, F. S., and Voisin, Y. (2009). "Preprocessing of region of interest localization based on local surface curvature analysis for three-dimensional reconstruction with multiresolution". In: *Optical Engineering* 48.6.

Lin, S. and Shum, H. Y. (2001). "Separation of Diffuse and Specular Reflection in Color Images". In: *Conference on Computer Vision and Pattern Recognition*. Vol. 1, pp. 341–346.

LPtracker. Last access November 2012. URL: http://lptracker.sourceforge.net/.

Lu, Z., Tai, Y.-W., Ben-Ezra, M., and Brown, M. S. (2010). "A Framework for Ultra High Resolution 3D Imaging". In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1205–1212.

Luhmann, T. (2010a). "Close range photogrammetry for industrial applications". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.6, pp. 558–569.

— (2010b). *Nahbereichsphotogrammetrie, Grundlagen – Methoden – Anwendungen*. Wichmann Verlag.

Maas, H.-G. (1997a). "Concepts of real-time photogrammetry". In: *Human Movement Science* 16.2-3, pp. 189–199.

— (1997b). "Dynamic photogrammetric calibration of industrial robots". In: *Proceedings of SPIE Vol. 3174, Videometrics V*, pp. 106–112.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). "Multimodality image registration by maximization of mutual information". In: *IEEE Transactions on Medical Imaging* 16.2, pp. 187–98.

Maloney, L. T. and Wandell, B. A. (1986). "Color constancy: a method for recovering surface spectral reflectance". In: *Journal of the Optical Society of America A* 3.1, pp. 29–33.

Malzbender, T., Gelb, D., and Wolters, H. (2001). "Polynomial Texture Maps". In: *SIGGRAPH Conference Proceedings*. ACM, pp. 519–528.

Mansouri, A. (2005). "Étude, conception et réalisation d'un système multispectral de vision pour des applications de proximité, et développement d'algorithmes de reconstruction de la réflectance". PhD thesis. Université de Bourgogne.

Mansouri, A., Lathuiliere, A., Marzani, F., Voisin, Y., and Gouton, P. (2007). "Toward a 3D Multispectral Scanner: An Application to Multimedia". In: *IEEE MultiMedia* 14.1, pp. 40–47.

Mansouri, A., Marzani, F. S., and Gouton, P. (2005a). "Development of a Protocol for CCD Calibration: Application to a Multispectral Imaging System". In: *International Journal of Robotics and Automation* 20.2, pp. 94–100.

— (2005b). "Neural networks in two cascade algorithms for spectral reflectance reconstruction". In: *IEEE International Conference on Image Processing*. Vol. 2. Genoa, Italy, pp. 718–721.

Mansouri, A., Marzani, F. S., Hardeberg, J. Y., and Gouton, P. (2005c). "Optical calibration of a multispectral imaging system based on interference filters". In: *Optical Engineering* 44.2, pages.

Martinez, K., Cupitt, J., Saunders, D., and Pillay, R. (2002). "Ten Years of Art Imaging Research". In: *Proceedings of the IEEE* 90.1, pp. 28–41.

Miao, L., Qi, H., and Snyder, W. E. (2004). "A generic method for generating multispectral filter arrays". In: *IEEE International Conference on Image Processing*. Vol. 5, pp. 3343–3346.

Miyazawa, K., Hauta-Kasari, M., and Toyooka, S. (2001). "Rewritable Broad-Band Color Filters for Spectral Image Analysis". In: *Optical Review* 8.2, pp. 112–119.

Morvan, Y. (2009). "Acquisition, compression and rendering of depth and texture for multi-view video". PhD thesis. Technische Universiteit Eindhoven.

Müller, G., Meseth, J., and Klein, R. (2003). "Compression and real-time rendering of measured BTFs using local PCA". In: *Vision, Modeling and Visualisation*. Ed. by T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel, E. Steinbach, and R. Westermann. Akademische Verlagsgesellschaft Aka GmbH, Berlin, pp. 271–280.

Müller, G., Meseth, J., Sattler, M., Sarlette, R., and Klein, R. (2005). "Acquisition , Synthesis and Rendering of Bidirectional Texture Functions". In: *Computer Graphics Forum* 24.1, pp. 83–109.

NDI. *ScanTrak*. Last access September 2012. URL: http://www.ndigital.com/industrial/optotrakproseries-models.php.

Nehab, D., Rusinkiewicz, S., Davis, J., and Ramamoorthi, R. (2005). "Efficiently Combining Positions and Normals for Precise 3D Geometry". In: *ACM Transactions on Graphics* 24.3, pp. 536–543.

Neverova, N., Muselet, D., and Trémeau, A. (2012). "Lighting Estimation in Indoor Environments from Low-Quality Images". In: *Computer Vision — ECCV 2012. Workshops and Demonstrations*. Ed. by A. Fusiello, V. Murino, and R. Cucchiara. Vol. 7584. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 380–389.

Nicodemus, F., Richmond, J., Hsia, J., Ginsberg, I., and Limperis, T. (1977). *Geometrical Considerations and Nomenclature for Reflectance*. Tech. rep. U.S. Dept. of Commerce, National Bureau of Standards.

Novati, G., Pellegri, P., and Schettini, R. (2005). "An affordable multispectral imaging system for the digital museum". In: *International Journal on Digital Libraries* 5.3, pp. 167–178.

Ohta, Y. and Hayashi, Y. (1994). "Recovery of illuminant and surface colors from images based on the CIE daylight". In: *Proceedings of the Third European Conference on Computer Vision*. Ed. by J. Eklundh. Vol. 801. Lecture Notes in Computer Science. Spring Berlin Heidelberg, pp. 234–246.

Palma, G., Corsini, M., Cignoni, P., Scopigno, R., and Mudge, M. (2010a). "Dynamic shading enhancement for reflectance transformation imaging". In: *Journal on Computing and Cultural Heritage* 3.2, 6:1–6:20.

Palma, G., Corsini, M., Dellepiane, M., and Scopigno, R. (2010b). "Improving 2D-3D registration by mutual information using gradient maps". In: *Eurographics Italian Chapter Conference*, pp. 89–94.

Paperno, E., Sasada, I., and Leonovich, E. (2001). "A New Method for Magnetic Position and Orientation Tracking". In: *IEEE Transactions on Magnetics* 37.4, pp. 1938–1940.

Park, J.-I., Lee, M.-H., Grossberg, M. D., and Nayar, S. K. (2007). "Multispectral Imaging Using Multiplexed Illumination". In: *IEEE International Conference on Computer Vision*.

Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., and Tsioukas, V. (2007). "Methods for 3D digitization of Cultural Heritage". In: *Journal of Cultural Heritage* 8, pp. 93–98.

Pelagotti, A., Del Mastio, A., De Rosa, A., and Piva, A. (2008). "Multispectral imaging of paintings". In: *IEEE Signal Processing Magazine* 25.4, pp. 27–36.

Phong, B. T. (1975). "Illumination for computer generated pictures". In: *Communications of the ACM* 18.6, pp. 311–317.

Poger, S. and Angelopoulou, E. (2001). "Selecting Components for Building Multispectral Sensors". In: *IEEE Conference on Computer Vision and Pattern Recognition, Technical Sketches*.

Pollefeys, M., Van Gool, L., Vergauwen, M., Cornelis, K., Verbiest, F., and Toops, J. (2003). "3D Recording for Archaeological Fieldwork". In: *Computer Graphics and Applications* 23.3, pp. 20–27.

Pulli, K. (1997). "Surface Reconstruction and Display from Range and Color Data". PhD thesis. University of Washington.

— (1999). "Multiview registration for large data sets". In: *International Conference on 3-D Digital Imaging and Modeling*, pp. 160–168.

Raab, F. H., Blood, E. B., Steiner, T. O., and Jones, H. R. (1979). "Magnetic Position and Orientation Tracking System". In: *IEEE Transactions on Aerospace and Electronic Systems* 15.5, pp. 709–718.

Rapantzikos, K. and Balas, C. (2005). "Hyperspectral imaging: potential in non-destructive analysis of palimpsests". In: *IEEE International Conference on Image Processing*. Vol. 2, pp. 618–621.

Remondino, F. and El-Hakim, S. (2006). "Image-based 3D modelling: a review". In: *The Photogrammetric Record* 21.115, pp. 269–291.

Remondino, F., El-Hakim, S., Gruen, A., and Zhang, L. (2008). "Turning images into 3-D models". In: *IEEE Signal Processing Magazine* 25.4, pp. 55–64.

Remondino, F., Pelagotti, A., Del Mastio, A., and Uccheddu, F. (2009). "Novel data registration techniques for art diagnostics and 3d heritage visualization". In: *9th Conference on Optical 3D Measurement Techniques*.

Ribés, A., Schmitt, F., Pillay, R., and Lahanier, C. (2005). "Calibration and Spectral Reconstruction for CRISTATEL: an Art Painting Multispectral Acquisition System". In: *Journal of Imaging Science and Technology* 49.6, pp. 563–573.

Rocchini, C., Cignoni, P., Montani, C., Pingi, P., and Scopigno, R. (2001a). "A low cost 3D scanner based on structured light". In: *Computer Graphics Forum* 20.3, pp. 299–308.

Rocchini, C., Cignoni, P., Montani, C., Pingi, P., Scopigno, R., Fontana, R., Pezzati, L., Cygielman, M., Giachetti, R., and Gori, G. (2001b). "3D scanning the Minerva of Arezzo". In: *International Cultural Heritage Informatics Meeting*, pp. 266–272.

Rocchini, C., Cignoni, P., Montani, C., and Scopigno, R. (2002). "Acquiring, stitching and blending diffuse appearance attributes on 3D models". In: *The Visual Computer* 18.3, pp. 186–204.

Rushmeier, H. and Bernardini, F. (1999). "Computing Consistent Normals and Colors from Photometric Data". In: *International Conference on 3-D Digital Imaging and Modeling*, pp. 98–108.

Rusinkiewicz, S., Burns, M., and DeCarlo, D. (2006). "Exaggerated Shading for Depicting Shape and Detail". In: *ACM Transactions on Graphics* 25.3, pp. 1199–1205.

Rusinkiewicz, S. and Levoy, M. (2001). "Efficient variants of the ICP algorithm". In: *International Conference on 3-D Digital Imaging and Modeling*, pp. 145–152.

Sanchez, M., Mansouri, A., Marzani, F. S., and Gouton, P. (2005). "Spectral reflectance estimation from multispectral images using neural networks". In: *Physics in Signal and Image Processing*.

Schechner, Y. and Nayar, S. (2002). "Generalized mosaicing: wide field of view multispectral imaging". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.10, pp. 1334–1348.

Schütze, R., Raab, C., Boochs, F., Wirth, H., and Meier, J. (2009). "Optopose - a multi-camera system for fast and precise determination of position and orientation for moving effector". In: *9th Conference on Optical 3D Measurement Techniques*.

Seulin, R., Stolz, C., Fofi, D., Millon, G., and Nicolier, F. (2006). "Three-dimensional tools for analysis and conservation of ancient wooden stamps". In: *The Imaging Science Journal* 54.2, pp. 111–121.

Shen, H.-L., Cai, P.-Q., Shao, S.-J., and Xin, J. H. (2007). "Reflectance reconstruction for multispectra imaging by adaptive Wiener estimation". In: *Optics Express* 15.23, pp. 15545–15554.

Sherman, J. T., Lubkert, J. K., Popovic, R. S., and DiSilvestro, M. R. (2007). "Characterization of a Novel Magnetic Tracking System". In: *IEEE Transactions on Magnetics* 43.6, pp. 2725–2728.

Shi, M. and Healey, G. (2002). "Using reflectance models for color scanner calibration". In: *Journal of the Optical Society of America A* 19.4, pp. 645–656.

Simon Chane, C., Mansouri, A., Marzani, F. S., and Boochs, F. (2013a). "Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives". In: *Image and Vision Computing* 31 (1), pp. 91–102.

Simon Chane, C., Schütze, R., Boochs, F., and Marzani, F. (2013b). "Registration of 3D and Multispectral Data for the Study of Cultural Heritage Surfaces". In: *Sensors* 13, pp. 1004–1020.

Simon, C., Huxhagen, U., Mansouri, A., Heritage, A., Boochs, F., and Marzani, F. S. (2010). "Integration of high-resolution spatial and spectral data acquisition systems to provide complementary datasets for cultural heritage applications". In: *Proceedings of SPIE Vol. 7531, Computer Vision and Image Analysis of Art*. Ed. by D. G. Stork, J. Coddington, and A. Bentkowska-Kafel, 75310L:1–9.

Simon, C., Schütze, R., Boochs, F., and Marzani, F. S. (2012). "Asserting the Precise Position of 3D and Multispectral Acquisition Systems for Multisensor Registration Applied to Cultural Heritage Analysis". In: *Advances in Multimedia Modeling*. Ed. by K. Schoeffmann, B. Merialdo, A. G. Hauptmann, C.-W. Ngo, Y. Andreopoulos, and C. Breiteneder. Vol. 7131. Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg, pp. 597–608.

Sitnik, R., Krzeslowski, J., and Maczkowski, G. (2012). "Archiving shape and appearance of cultural heritage objects using structured light projection and multispectral imaging". In: *Optical Engineering* 51.2, 021115:1–7.

Sitnik, R., Maczkowski, G., and Krzeslowski, J. (2010a). "Calculation Methods for Digital Model Creation Based on Integrated Shape, Color and Angular Reflectivity Measurement". In: *Digital Heritage*. Ed. by M. Ioannides, D. Fellner, A. Georgopoulos, and D. G. Hadjimitsis. Vol. 6432. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 13–27.

— (2010b). "Integrated shape, color, and reflectivity measurement method for 3D digitization of cultural heritage objects". In: *Proceedings of SPIE Vol. 7526, 3D Image Processing and Applications*. Ed. by A. M. Baskurt, 75260Q:1–10.

Steinbichler. *T-Scan 3*. Last access September 2012. URL: http://www.steinbichler.com/products/surface-scanning/3d-digitizing/t-scan-3.html.

Tan, R. T. and Ikeuchi, K. (2005). "Separating Reflection Components of Textured Surfaces Using a Single Image". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.2, pp. 178–93.

Taylor, J., Beraldin, J.-A., Godin, G., Cournoyer, L., Baribeau, R., Blais, F., Rioux, M., and Domey, J. (2003). "NRC 3D Technology for Museum and Heritage Applications". In: *Journal of Visualization and Computer Animation* 14.3, pp. 121–138.

Toler-Franklin, C., Finkelstein, A., and Rusinkiewicz, S. (2007). "Illustration of complex real-world objects using images with normals". In: *International Symposium on Non-Photorealistic Animation and Rendering*. Ed. by B. Gooch, M. Agrawala, and O. Deussen. ACM, pp. 111–119.

Tominaga, S. (1996). "Multichannel vision system for estimating surface and illumination functions". In: *Journal of the Optical Society of America A* 13.11, pp. 2163–2173.

Tominaga, S. and Tanaka, N. (2008). "Spectral image acquisition, analysis, and rendering for art paintings". In: *Journal of Electronic Imaging* 17.4, 043022:1–13.

Tominaga, S. and Wandell, B. A. (1989). "Standard surface-reflectance model estimation and illuminant estimation". In: *Journal of the Optical Society of America A* 6.4, pp. 576–584.

Tonsho, K., Akao, Y., Tsumura, N., and Miyake, Y. (2001). "Development of goniophotometric imaging system for recording reflectance spectra of 3D objects". In: *Proceedings of SPIE Vol. 4663, Color Imaging: Device-Independent Color, Color Hardcopy, and Applications VII*. Ed. by G. G. M. Reiner Eschbach, pp. 370–378.

Torrance, K. E. and Sparrow, E. M. (1967). "Theory for Off-Specular Reflection From Roughened Surfaces". In: *Journal of the Optical Society of America* 57.9, pp. 1105–1114.

Tsai, R. (1987). "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". In: *IEEE Journal on Robotics and Automation* 3.4, pp. 323–344.

Valkenburg, R., Penman, D., Schoonees, J., Alwesh, N., and Palmer, G. (2006). "Interactive hand-held 3d scanning". In: *Image and Vision Computing New Zealand*. Ed. by P. Delmas, J. James, and J. Morris, pp. 245–250.

Vergauwen, M. and Van Gool, L. (2006). "Web-based 3D Reconstruction Service". In: *Machine Vision and Applications* 17.6, pp. 411–426.

Vergne, R., Pacanowski, R., Barla, P., Granier, X., and Schlick, C. (2009). "Light warping for enhanced surface depiction". In: *ACM Transactions on Graphics* 28.3.

Viola, P. A. and Wells, W. M. (1997). "Alignment by Maximization of Mutual Information". In: *International Journal of Computer Vision* 24, pp. 137–154.

Whitted, T. (1980). "An improved illumination model for shaded display". In: *Communications of the ACM* 23.6, pp. 343–349.

Wolff, L. B. (1990). "A Polarization-Based Matrerial Classification from Spectular Reflection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.11, pp. 1059–1071.

Yamaguchi, M., Haneishi, H., and Ohyama, N. (2008). "Beyond Red–Green–Blue (RGB): Spectrum-Based Color Imaging Technology". In: *Journal of Imaging Science and Technology* 52.1, 010201:1–15.

Yilmaz, H., Yakar, M., and Yildiz, F. (2008). "Documentation of historical caravansaries by digital close range photogrammetry". In: *Automation in Construction* 17, pp. 489–498.

Zhang, R., Tsai, P.-S., Cryer, J. E., and Shah, M. (1999). "Shape-from-shading: a survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21.8, pp. 690–706.

Zhang, X. and Xu, H. (2008). "Reconstructing spectral reflectance by dividing spectral space and extending the principal components in principal component analysis". In: *Journal of the Optical Society of America A* 25.2, pp. 371–378.

Zhao, Y. and Berns, R. S. (2007). "Image-based spectral reflectance reconstruction using the matrix R method". In: *Color Research & Application* 32.5, pp. 343–351.

Zheng, J. Y., Zhang, Z. L., and Abe, N. (1998). "Virtual recovery of excavated archaeological finds". In: *Proceedings International Conference on Multimedia Computing and Systems*, pp. 348–357.

Zhu, J., Wang, L., Yang, R., and Davis, J. (2008). "Fusion of time-of-flight depth and stereo for high accuracy depth maps". In: *Conference on Computer Vision and Pattern Recognition*. IEEE.

Zuzak, K. J., Francis, R. P., Wehner, E. F., Smith, J., Litorja, M., Allen, D. W., Tracy, C., Cadeddu, J., and Livingston, E. (2009). "DLP Hyperspectral Imaging for Surgical and Clinical Utility". In: *Proceedings of SPIE Vol. 7210, Emerging Digital Micromirror Device Based Systems and Applications*. Ed. by L. J. Hornbeck and M. R. Douglass, 721006:1–9.

Abstract:

The concern and interest of this PhD thesis is the registration of featureless 3D and multispectral datasets describing cultural heritage objects. In this context, there are few natural salient features between the complementary datasets, and the use of targets is generally proscribed. We thus develop a technique based on the photogrammetric tracking of the acquisition systems in use.
A series of simulations was performed to evaluate the accuracy of our method in three configurations chosen to represent a variety of cultural heritage objects. These simulations show that we can achieve a spatial tracking accuracy of 0.020 mm and an angular accuracy of 0.100 mrad using four 5 Mpx cameras when digitizing an area of 400 mm $\times$ 700 mm. The accuracy of the final registration relies on the success of a series of optical and geometrical calibrations and their stability for the duration of the full acquisition process.
The accuracy of the tracking and registration was extensively tested in laboratory settings. We first evaluated the potential for multiview 3D registration. Then, the method was used for to project of multispectral images on 3D models. Finally, we used the registered data to improve the reflectance estimation from the multispectral datasets.

Keywords: 2D – 3D registration, close range photogrammetry, optical calibration, 3D digitization, multispectral imaging, cultural heritage

Résumé :

Cette thèse s'intéresse au recalage de données issues de capteurs 3D et multispectraux pour l'étude du patrimoine. Lorsque l'on étudie ce type d'objet, il y a souvent peu de points saillants naturels entre ces jeux de données complémentaires. Par ailleurs, l'utilisation de mires optiques est proscrite. Notre problème est donc de recaler des données multimodales sans points caractéristiques. Nous avons développé une méthode de recalage basé sur le suivi des systèmes d'acquisition en utilisant des techniques issues de la photogrammétrie.
Des simulations nous ont permis d'évaluer la précision de la méthode dans trois configurations qui représentent des cas typiques dans l'étude d'objets du patrimoine. Ces simulations ont montré que l'on peut atteindre une précision du suivi de 0,020 mm spatialement et 0,100 mrad angulairement en utilisant quatre caméras 5 Mpx lorsque l'on numérise une zone de 400 mm $\times$ 700 mm. La précision finale du recalage repose sur le succès d'une série de calibrations optiques et géométriques, ainsi que sur leur stabilité pour la durée du processus d'acquisition.
Plusieurs tests ont permis d'évaluer la précision du suivi et du recalage de plusieurs jeux de données indépendants ; d'abord seulement 3D, puis 3D et multispecrales. Enfin, nous avons étendu notre méthode d'estimation de la réflectance à partir des données multispectrales lorsque celles-ci sont recalées sur un modèle 3D.

Mots-clés : recalage 2D – 3D, photogrammétrie, calibrations optiques, numérisation 3D, imagerie multispectrale, étude du patrimoine

SPIM

■ École doctorale SPIM - Université de Bourgogne/UFR ST BP 47870 F - 21078 Dijon cedex
■ tél. +33 (0)3 80 39 59 10 ■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr