Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives
Camille Simon, Alamin Mansouri, Marzani Franck, Boochs Frank

To cite this version:

HAL Id: hal-00783985
https://hal.archives-ouvertes.fr/hal-00783985
Submitted on 6 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Integration of 3D and multispectral data for cultural heritage applications: survey and perspectives

Camille Simon Chanea,b,∗, Alamin Mansouric, Franck S. Marzania, Frank Boochsb

aLe2i, Université de Bourgogne, Bâtiment Mirande - UFR Sciences & Techniques, B.P. 47870 - 21078 Dijon Cedex, France.
bIM2mainz, Institute for Spatial Information and Surveying Technology, University of Applied Sciences Mainz, Germany.
cLe2i, Université de Bourgogne, B.P. 16, Route des plaines de l’Yonne, 89010 Auxerre Cedex, France.

Abstract
Cultural heritage is increasingly put through imaging systems such as multispectral cameras and 3D scanners. Though these acquisition systems are often used independently, they collect complementary information (spectral vs. spatial) used for the study, archiving and visualization of cultural heritage. Recording 3D and multispectral data in a single coordinate system enhances the potential insights in data analysis.

We present the state of the art of such acquisition systems and their applications for the study of cultural heritage. We also describe existing registration techniques that can be used to obtain 3D models with multispectral texture and explore the idea of optically tracking acquisition systems to ensure an easy and precise registration.

Keywords: cultural heritage, multispectral imaging, 3D digitization, registration, data fusion, photogrammetry

1. Introduction
Cultural heritage is a favorite application 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 mostly two dimensional objects such as paintings, multispectral imaging is the path of predilection. For statues, we favor 3D digitization.

∗Corresponding author
Email addresses: camille.simon@u-bourgogne.fr (Camille Simon Chane), alamin.mansouri@u-bourgogne.fr (Alamin Mansouri), franck.marzani@u-bourgogne.fr (Franck S. Marzani), frank.boochs@geoinform.fh-mainz.de (Frank Boochs)
Of course it is not un-common 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. Most importantly, they 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. Thus, creating multispectral 3D models is not a straightforward task.

Our goal is to review and centralize the techniques (acquisition systems and algorithms) that can be used to obtain multispectral 3D models of cultural heritage objects. An overview of the existing applications of each of these techniques in the field of cultural heritage is meant to inform the reader of the possible uses of such data. This article is voluntarily high-level and meant as a resource survey.

Our scope is limited to the study of cultural heritage objects as opposed to full buildings, archaeological sites or cities. We use the term "object" to describe anything from the smallest archaeological fragment to 5m 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.

We only describe non-contact acquisition systems since contact acquisition systems may damage the object and are thus rarely used for the study of cultural heritage. We foreground portable systems that can be used directly in museums, monuments or any other non-laboratory site. 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 review is organized as follows: Section 2 presents the state-of-the-art in 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. Section 3 presents 3D digitization techniques, great 3D digitization campaigns and the various uses of 3D models for conservation. 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 describe how photogrammetry-based optical tracking can be used to ease multimodal registration.
2. Multispectral imaging

2.1. 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, at the least. 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. Theoretically, 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 [1]. However, systems that acquire more than three spectral components permit a better estimation of the reflectance. CIE recommends that multispectral acquisition systems acquire 31 bands between 400 nm and 700 nm (that is, one channel every 10 nm) but Berns et al. [2] 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.

2.2. 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 [3].

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 [4-7]. Various algorithms can be used to compensate geometric distortion [4, 5], longitudinal aberrations [8] and ghosting [9] 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 [10] though this is not adapted to the study of cultural heritage objects.

**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 [11-14]. Devices based on such filters are also very flexible since the number and width of bands can be changed programmatically.

Pogger and Angelopoulou [15] have compared the characteristics of these three types of electronically tunable filters. Most importantly, AOTF have a high transmission rate (98%) but they require collimated light while LCTF and Fabry-Perot devices suffer from a low transmission rate (lower than 50%).

**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 [2, 16, 17]. The practical use of this type of multispectral system is still being investigated.

**Tunable Light Sources** It is possible to successively project light at different wavelengths on the object (excitation filtering) instead of filtering the light reflected from the object (emission filtering). A common strategy is to use multiplex multicolor LED illumination as a tunable light source [18-20]. This solution is gaining popularity with the increasing availability of LEDs in an increasing variety of colors. It is also possible to use a slide projector and color filters [21] or DLPs [22] as 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 [23] or prisms [24] 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. [25] 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 object.

2.3. 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 1 summarizes the technical characteristics of existing acquisition systems developed for the study of cultural heritage.

In the early 1990's the VASARI project [26] focused on the development of a 12 band multispectral acquisition system based on a filter-wheel tunable light source. Using a 3000 × 2000 pixels CCD and a gantry the system captured 20000 × 20000 pixel acquisitions. At the time, such high spectral resolution came at a high cost: it took about 3 hours to digitize a 1 m × 1 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 [27] also aimed at developing a multispectral acquisition system for the analysis of paintings. The device is based on a linear CCD (12000 pixels) which is mechanically displaced to acquire 20000 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 cm × 82 cm) in an hour and a half [29]. 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 [6, 14, 21, 28]. 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. [6] is particularly impressive as it acquires visible reflectance, infrared reflectography and UV-fluorescence, a combination of data most adapted for the study of paintings.

Multispectral data alone lacks the spectral resolution necessary to distinguish artist pigments [26] but it can be used as a complement to spot techniques for material identification [6, 30] or to distinguish metamers [30, 31]. Multispectral imaging has been used to increase the readability of palimpsests.
Table 1: Multispectral systems developed for cultural heritage applications given in rough chronological order.

<table>
<thead>
<tr>
<th>References</th>
<th>Filtering type</th>
<th>Spectral bands</th>
<th>Spectral range (nm)</th>
<th>Image size (pixels)</th>
<th>Dynamic range (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VASARI project [26]</td>
<td>Tunable light</td>
<td>12</td>
<td>450 – 1000</td>
<td>20000 × 20000</td>
<td>8</td>
</tr>
<tr>
<td>CRISATEL project [27]</td>
<td>Interferential</td>
<td>13</td>
<td>400 – 1000</td>
<td>12000 × 30000</td>
<td>12</td>
</tr>
<tr>
<td>Novati et al. [14]</td>
<td>LCTF</td>
<td>31</td>
<td>400 – 700</td>
<td>1392 × 1040</td>
<td>12</td>
</tr>
<tr>
<td>Berns et al. [28]</td>
<td>CFA</td>
<td>6</td>
<td>380 – 730</td>
<td>22 megapixels</td>
<td>16</td>
</tr>
<tr>
<td>Pelagatti et al. [6]</td>
<td>Interferential</td>
<td>15</td>
<td>380 – 1130</td>
<td>2048 × 2048</td>
<td>14</td>
</tr>
</tbody>
</table>
Figure 1: Spectral model of the acquisition process in 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.

using pseudocoloring [32–34]. In the same vein, virtually restoring the original colors of paintings [35] can increase their readability. Multispectral images can be used for the analysis of cracks and the study of varnish aging [29, 36]. This is both important to understand the history of the painting and to guide a virtual restoration. The use of multispectral data for virtual archives is also a common application [21, 26].

The possibility to view the digitized surface under arbitrary illuminants [21, 27] 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 [28, 37].

2.4. 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 1.

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

$$d_k = \int_{\lambda_{\min}}^{\lambda_{\max}} I(\lambda) r(\lambda) o(\lambda) c(\lambda) t_k(\lambda) d\lambda + \eta_k.$$  (1)
If we assume that the noise is removed by preprocessing [4, 38], 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 1 thus becomes:

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} S_k(\lambda) r(\lambda) \, d\lambda$$  \hfill (2)

By regular sampling of the spectral range at $N$ wavelengths, equation 2 can be written in matrix notation as follows:

$$d_k = S_k^T(\lambda)r(\lambda),$$  \hfill (3)

where $S_k^T(\lambda) = [s_k(\lambda_1) s_k(\lambda_2) \ldots s_k(\lambda_N)]^T$ is the vector which describes the spectral sensitivity of the $k^{th}$ channel of the acquisition system, $r(\lambda) = [r(\lambda_1) \ r(\lambda_2) \ldots r(\lambda_N)]^T$ 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 3 can be written as:

$$d = S^T r,$$  \hfill (4)

where $d$ is the vector containing all $d_k$ camera outputs and $S = [s_1 \ s_2 \ldots s_K]^T$ is the matrix containing the channels’ spectral sensitivities $S_k$. The final goal is to recover $r$ from the camera output given by equation 4. This is obtained by finding an operator $Q$ that satisfies $\tilde{r} = Q d$. Many methods permit this estimation [39-45], most of which can be categorized depending on how the matrix $S$ is determined:

- If $S$ is obtained by a direct physical system characterization [46-49], the operator $Q$ is the inverse of $S$. However, since $S$ 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 $Q = S^+$ where $+$ is the pseudo-inverse operator.

- $S$ can also be obtained indirectly by matching a set of $M$ color patches for which we know the theoretical reflectance with an image of these patches captured by the camera [50, 51]. We then have a set of corresponding pairs $(d_m, r_m)$, for $m = 1, \ldots M$, 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 $R$ and the camera outputs for the $M$ patches are gathered in matrix $D$. The operator $Q$ is obtained directly by calculating this match. Any optimization method can fulfill this aim (neural networks, least squares regression, etc.). The operator $Q$ is obtained from $R = Q D$ through the pseudo-inverse of $D$ : $Q = R D^+$.

- The third paradigm for spectral reflectance estimation consists of directly interpolating the camera outputs $d_k$ [52, 53]. No knowledge about the matrix $S$ 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, it is often necessary to separate the diffuse and specular components of the images \([54-57]\). The first and third paradigms also require the separation of the illumination and surface reflectance \([58-63]\).

2.5. 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 used by conservators; improving their 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.

However, when the purpose of multispectral imaging is solely for visualization, reflectance imaging \([64, 65]\) can be more appropriate than multispectral imaging. The process consists in capturing a set of images from a single view, under varying illumination conditions. It is then possible to render the subject under user-defined illumination. Though there is a very heavy setup for a single view, with even greater constraints for multiple views, the results are very realistic.

For the widespread adoption of multispectral imaging by conservators it is also necessary to improve existing visualization and analysis tools \([66, 67]\). 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 \([68]\) must be adapted to multispectral imaging. Specific algorithms that exploit the wealth of data acquired by multispectral systems must also be developed. Such tools have to
be developed hand in hand with the conservation community to provide them with methods they need.

3. 3D digitization

3.1. 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.

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 various 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.

3.2. Principles and devices

3D digitization techniques are generally sorted between 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. It consists in digitizing cultural heritage using images from calibrated cameras [69]. Lately however, several algorithms have emerged which automatically create 3D models from un-calibrated images [70], for example using Multi-View Stereopsis [71]. This type of techniques is well adapted to the documentation of archaeological sites [72].

Other passive digitization techniques can usually be classified under "shape from-X", where X can be shading, silhouette, edges, texture, focus, etc. Though generally easy to setup, these techniques suffer from low accuracy [73].
Active 3D digitization techniques. The most used active digitization techniques are time of flight 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 objects — even full archaeological sites [74] — 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 [75] 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 [76], and many commercial systems exist such as those developed by GOM and Breuckmann, the Comet5 from Steiblicher, and the compact, cordless handless Kolibri device by Fraunhofer [77].

3.3. From 3D acquisitions to a 3D model

When digitizing but the smallest objects, multiple acquisitions must be performed and stitched together. Even if we acquire only 3D data, a registration phase is often necessary to create a complete 3D model.

In the case of highly spatially-structured objects such as statues, the registration may profit from the spatial characteristics of the object itself. Local shape information expressed in grouped points can are then transformed by algorithms like the Iterative Closest Point algorithm (ICP) [78]. 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 registering views of flat objects such as painting or wall paintings, though some variants tackle this issue [79].

There is great focus on automating the registration process to boost the use of 3D models by the conservation community [80, 81]. [80] describes 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. 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, though another technique must be used to register top and bottom views of the object.

The registration process highly influences the accuracy of the end model and is thus a critical step in digitizing cultural heritage artifacts. The interested reader can turn to [82] for more details.


3.4. 3D digitization of cultural heritage

We now present a few remarkable projects where these techniques were applied.

In 1998-99 Levoy et al. [83] 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 [84]. 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 [85]. A similar project of smaller scope consisted in monitoring the restoration of a statue of Minerva by digitizing it before and after restoration [86].

The focus of the Great Buddha Project [87] 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 [88], restoring the object’s original color, reconstructing missing information [89]. In the same vein, it is also possible to interact with archaeological fragments to reconstruct broken objects, either automatically or based on algorithms [90, 91] while digitized wooden stamps [75] can be used for virtual printing.

3D models also provide very rich communication tools. For example, visitors can digitally approach a 5m high statue with dedicated interactive kiosks in a museum [92], or immerse themselves in larger settings through virtual displays [89].

3.5. Perspectives

Passive techniques are generally easy to set up but do not provide very accurate results. They can be sufficient for visualization purposes, but active sensors are better suited for applications which include surface analysis and metrology. 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 when there is insufficient contrast.

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, they require some training to be used at the best of their capacity to use them.

Active sensors are thus difficult to use for the digitization of archaeological sites. They are also 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.

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 [93, 94]. Recently, comparable results have been achieved using a simpler system based on a projector, a low resolution camera and a high resolution camera [95].

Generally speaking, the digitization of cultural heritage often involves large objects that must be scanned with high precision [83, 96]. The acquisition systems must be portable so it can be used in situ. 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.

4. 2D - 3D registration

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 [101] where the authors successively perform multispectral acquisitions of "La dame en prière" and two 3D digitizations of it’s surface but they rely on independent viewers to explore each dataset.

4.1. Integrated acquisition systems

Since multispectral imaging is seldom used in combination with 3D digitization, we also present the integration of color and 3D models, using algorithms or unique sensors. Multispectral imaging is an extension of color imaging and the methods used in this context can be adapted from one to the other.

*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 [102–106]. Equivalent devices include a fringe projection system with a color sensor [76] and an integrated laser stripe projection scanner with and a color camera [91].

*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
Table 2: Selected characteristics of integrated 3D digitization and multispectral acquisition systems.

<table>
<thead>
<tr>
<th>References</th>
<th>Number of sensors</th>
<th>Filtering type</th>
<th>Spectral bands</th>
<th>Spectral range (nm)</th>
<th>3D digitization system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mansouri et al. [97]</td>
<td>single</td>
<td>interference filters</td>
<td>7</td>
<td>400 – 1000</td>
<td>projection of structured light</td>
</tr>
<tr>
<td>Sitnik et al. [98]</td>
<td>single</td>
<td>interference filters</td>
<td>20</td>
<td>420 – 740</td>
<td>projection of structured light</td>
</tr>
<tr>
<td>Tonsho et al. [99]</td>
<td>multiple</td>
<td>optical filters</td>
<td>5</td>
<td>380 – 780</td>
<td>triangulation laser scanner</td>
</tr>
<tr>
<td>Brusco et al. [100]</td>
<td>multiple</td>
<td>spectrograph</td>
<td>60</td>
<td>400 – 1000</td>
<td>time of flight laser scanner</td>
</tr>
</tbody>
</table>
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 setup were developed to acquire 3D models of cultural heritage [83, 87, 107].

Integrated MS - 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. Mansouri et al. [97] and Sitnik et al. [98] 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. [99] 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.

Although [97-99] all cite cultural heritage as a possible application, to our knowledge only Brusco et al. [100] 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.

4.2. 2D-3D registration algorithms

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 [108-110]. 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. Though targets may be used to produce a more accurate registration, they are usually not adapted to cultural heritage applications where we want to minimize the disturbance to the object.
**Mutual Information methods.** Though there have been efforts to minimize user intervention [111] in this type of registration setup, the need to find corresponding points is altogether eliminated if one uses fully automated maximization of mutual information methods [112-116] 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 [114], gradient map [115], silhouette map, reflection map and other illumination-based renderings [116]).

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 [116]. This method provides good results for visualization purposes, even using low quality images [114].

Ikeuchi et al. [87] present a color/3D registration method suitable when the 3D data is obtained by time of flight. They extract edges and discontinuities from the reflectance images created as a side product of the range data. These images are naturally registered to the 3D data and are used as a first alignment with the color images. Then the 3D reflectance edge points are registered to the 2D color edge points through an iterative process.

### 4.3. Registration based on known sensor position

When using separate sensors, 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 fulfill this purpose, but they are too cumbersome to be used in situ. Another possibility is to use magnetic tracking [117, 118] to derive the position and orientation of the sensor in use. However, even recent sensors [119] 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 [88]. Optical devices can be used to determine the position and orientation of the sensor in use, this approach will be detailed in the next section.

### 4.4. Perspectives

Integrated multispectral/3D digitization systems can acquire an impressive amount of data. The ease of the eventual registration procedure is the strongest advantages of integrated color/multispectral and 3D acquisition systems, be they single sensor or multisensor. However, these systems lack flexibility. They are often built with a specific application in mind and, especially in the case of single sensor integrated devices, it is impossible or difficult to adapt the acquisition system 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 sensor, the accuracy of the registration greatly depends on the quality of the data. 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.

Post processing is greatly simplified when the position of the sensor at the time of the acquisition is known. This technique is flexible to the extent that it works with any optical sensor. However, the use of mechanical structures is cumbersome and magnetic tracking does not yet reach the necessary accuracy.

There is no unique and best technique to register 3D and multispectral data with sub-pixel accuracy. We believe the use of optical tracking is a good compromise in terms of flexibility, portability and end precision, as described in the next section.

5. Optical tracking for 2D-3D registration

5.1. Optical tracking of acquisition systems

Optical devices can be used to extract the position and orientation of any acquisition system. For example Blais et al. [104] 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.

There are few examples of using optical tracking for registration purposes in cultural heritage settings. However, photogrammetry is widely used in industrial settings for the precise positioning of robot end effectors. It is a flexible technique that has long been 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 different 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 [120-122] or cameras survey the workspace to obtain the position and orientation of the effector, which is enhanced with a target object [121, 123, 124]. 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 [125].

A new generation of hand-held laser scanners also relies on optical tracking either using photogrammetry [126-128] or using a laser tracker [129]. Breuckmann and Metronor have developed a multisensor digitization setup, the NaviSCAN [130], based on two photogrammetric cameras, a structured light projection scanner, a frame to fix on the digitization system and a probing unit. 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 coordinate system. We intend to extend
Figure 2: Global setup of a photogrammetry-based tracking system. The tracking cameras record the position and orientation of the acquisition devices while they successively capture the position of the surface of the object under study. A target-frame fixed to the acquisition system ensures precise tracking.

this type of setup to the tracking of multiple optical acquisition systems in cultural heritage settings to permit a precise integration of multispectral and 3D data.

5.2. Calibration

Figure 2 represents the in-situ acquisition setup: a group of photogrammetric/tracking cameras observe the acquisition devices as they successively digitize the surface under study from various positions. Several common photogrammetric calibration steps are necessary before, during and after the acquisition process to obtain a precise registration.

Pre-processing. These steps can be performed in a laboratory setting, either before or after the on-site acquisitions.

Frame calibration: A scale bar and additional targets are placed around the frame which is then photographed from multiple directions. A bundle adjustment provides us with the relative position of all targets fixed on the frame.
Optical calibration of the photogrammetric cameras and of the acquisition systems: A known target field is photographed from various angles. A bundle adjustment provides us with the internal camera parameters. A high number of images (30 to 80 per sensor) with a varied view of the target field provide precise calibration results.

Frame to acquisition system calibration: We also need the relative position of the targets fixed on the frame and the optical axis and optical center of the acquisition system when they are fastened together. A known target field is placed in the field of view of the acquisition system and thus captured. Several acquisitions containing the frame and the target field are acquired by another sensor. Since the frame, target field and acquisition system have been calibrated previously we obtain the required information via triangulation.

On-site acquisitions. These steps must be performed after the photogrammetric cameras have been set up to observe the acquisition devices in all their planned positions:

Calibration of the tracking cameras: The same procedure as the optical calibration of the acquisition system provides us with the exterior parameters of the tracking cameras, but less acquisitions are necessary (about 10) and the frame can be used as a calibration object.

Data acquisitions: The tracking cameras simultaneously grab an image for each acquisition by each sensor.

5.3. Data Processing

Once the calibration and acquisition steps have been performed, we can project all the acquired data in the coordinate system defined by the tracking cameras $C_0 (O_0, x_0, y_0, z_0)$. $C_F (O_F, x_F, y_F, z_F)$ and $C_S (O_S, x_S, y_S, z_S)$ are the coordinate systems associated respectively to the frame and to the acquisition system sensor (see figure 3). $O_S$ is the center of the sensor, and coordinate system $C_S$ is oriented in such way that $z_S$ is collinear to the optical axis of the sensor.

For every acquisition, the tracking cameras provide us with the position and orientation of the frame in $C_0$. This position is that of the coordinates of $O_F$ in coordinate system $C_0$, which we write $O_F|_{C_0}$. The orientation is given by the rotation matrix $P_{0,F}$. This matrix is defined so that, for any point $A$ for which we know the coordinates $A|_{C_F}$ in coordinate system $C_F$, then we can calculate the coordinates $A|_{C_0}$ in $C_0$ using:

$$A|_{C_0} = O_F|_{C_0} + P_{0,F} A|_{C_F}. \quad (5)$$

On the other hand, the calibration between the frame and the acquisition system provides us with the position and orientation of the sensor in $C_F$, namely
Figure 3: Coordinate systems associated to the tracking cameras, the frame and the sensor in use.

$O_S|_{C_F}$ and $P_{F,S}$. For any point $A$, the relation between $A|_{C_S}$ and $A|_{C_F}$ is, as previously,

$$A|_{C_F} = O_S|_{C_F} + P_{F,S} A|_{C_S}. \quad (6)$$

Combining equations 5 and 6 we obtain the coordinates of $A$ in coordinate system $C_0$ given its coordinates in $C_S$:

$$A|_{C_0} = O_F|_{C_0} + P_{0,F} (O_S|_{C_F} + P_{F,S} A|_{C_S}). \quad (7)$$

This equation can be re-written as:

$$A|_{C_0} = O_S|_{C_0} + P_{0,S} A|_{C_S}, \quad (8)$$

since $P_{0,F} O_S|_{C_F} = (O_S - O_F)|_{C_0}$ and $P_{0,S} = P_{0,F} P_{F,S}$.

If $A$ is a point of the surface under study captured by the acquisition system sensor, then $A'$ image of $A$ on the sensor has only two coordinates $x_{a'}$ and $y_{a'}$ in $C_S$, given how the coordinate system $C_S$ is defined. Using a pinhole camera model that takes into account the focal length $f$, we obtain the coordinates of point $A$ given $x_{a'}$ and $y_{a'}$ using:

$$A|_{C_S} = \lambda \begin{pmatrix} f & \tau & 0 \\ 0 & \eta f & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_{a'} \\ y_{a'} \\ 1 \end{pmatrix}, \quad (9)$$

where $\tau$ represents the slant of the pixels ($\tau = 0$ for rectangular pixels), $\eta$ is the pixel height/length ratio ($\eta = 1$ for square pixels) and $\lambda$ is a scaling factor. If the distance $Z$ between the sensor and the object is known then $\lambda = Z$. 

20
In our case, $Z$ is only known when the acquisition system acquires 3D data (we then have at least two sensors). If the sensor is that of a multispectral camera, we can infer $Z$ from the intersection between the ray defined by equation 9 and the 3D model built with another acquisition system. The camera model can be extended to account for lens distortion. The distortion parameters, as well as the precise values of $f$, $\tau$ and $\eta$ are obtained from the optical calibration of the acquisition system.

For each pixel $(x'_a, y'_a)^T$, we can thus calculate the possible positions of point $A$ in coordinate system $C_0$ using equations 8 and 9:

$$A|_{C_0} = O_{S}|_{C_0} + \lambda P_{0,S} K \begin{pmatrix} x'_a \\ y'_a \\ 1 \end{pmatrix}.$$  

(10)

5.4. Achievable tracking accuracy

Optimizing each calibration step ensures that the error accumulation caused by the cascade of coordinate systems stays within reasonable margins. The precision of the final registration depends on many parameters such as the number, focal length, position and sensor of the photogrammetric cameras; the dimension of acquisition area; the dimensions of the frame and the number of targets attached to it, etc. Simulations enable us to test many configurations and evaluate how precisely we can detect the position and orientation of each acquisition system.

We have performed simulations involving four cameras with a 5Mpixel sensor and a $0.5 \times 0.5 \times 0.5$ m frame covered with 56 targets. The resulting point accuracy over an area of $75 \times 100$ m is 0.007 mm ($1\sigma$). Comparable results are achieved using six cameras over an area of $2 \times 1.5$ m.

Preliminary tests have been performed using a $0.5 \times 0.5 \times 0.5$ m frame with 40 targets spread over four faces (see figure 4). The achieved positional point accuracy over an area of $75 \times 100$ m is 0.018 mm. This is not as good as the simulation results, but we are confident that an improved control of the illumination setup as well as increasing the number of targets on the frame from 40 to 60 will enable us to reach the accuracy predicted by the simulations. As a comparison, the NaviSCAN can measure a 1 m stick 6 m away from the tracking cameras with a 0.03 mm ($1\sigma$) accuracy.

5.5. Perspectives

Optically tracking the acquisition systems in use can enable seamless integration of multisensor data in a single coordinate system. The ensuing registration precision is independent from the content of the acquired data. This technique is thus suitable to register data with no salient features and with no artificial markers such as targets. This method is adapted to tracking any optical sensor, though the post-processing and registration strategies do depend on the type of data acquired. Many parameters can be tweaked to fit different application needs in working volume, final precision and setup cost.
Though the necessary equipment (frame, cameras and tripods) is cumbersome, the method is transportable and adapted to acquisitions in museums, monuments and archaeological sites. In the context of cultural heritage we believe this method can be particularly useful for multisensor digitization of paintings, wall-paintings or bas-reliefs. For objects with a high 3D component such as statues, the number of tracking cameras necessary may outbalance the practicality of the system.

6. Conclusion

There is no doubt that multispectral imaging and 3D models have a wide variety of possible applications in the field of cultural heritage. They can be used to create improved communication tools and to provide greater insight to
conservators regarding the history and condition of the object. On the acquisition side, both types of sensors must be improved to capture dark, transparent, shiny materials.

The integration of 3D and multispectral data is a difficult task. A few integrated devices have been developed, but most of these seem too bulky to be used for the in-situ study of cultural heritage. Systems developed specifically for cultural heritage are application-driven and not easily adapted to different types of objects. We believe the use of optical tracking for the multi-sensor integration of data on a 3D model offers a good compromise between the required materials and efforts, and the achieved precision. This technique fairs especially well for the study of mainly two-dimensional objects such as paintings, wall-paintings and bas-reliefs. When analyzing highly three-dimensional objects such as statues, the use of registration algorithms may be better adapted.

Optical tracking can be used to project the variety of data acquired when studying cultural heritage (color images, multispectral acquisitions, infra-red reflectography data, fluorescence images, etc.) on a single 3D model. Data from other optical sensors such as polarimetry imaging, thermal imaging, shearography, etc. can also be integrated on the same model.

As the quantity of data acquired increases, there is a growing need for new data analysis techniques to help make sense of the data. Segmentation, recognition and classification algorithms must now be based on the whole integrated dataset, and not a single aspect (3D or multispectral). Post-processing must be developed hand in hand with the conservation community to provide them with valuable information. New visualization and interaction methods must be developed so that specialists and non-specialists can interact intuitively with these augmented models. As multi-modal models gain widespread use, there is a great need of standards to share data between different entities.

Acknowledgments

The authors are grateful for the financial support provided by the Conseil Régional de Bourgogne (France) and i3mainz, Institute for Spatial Information and Surveying Technology (Germany).

References


