UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Dottorato di ricerca in HIGH MECHANICS AND AUTOMOTIVE DESIGN & TECHNOLOGY / MECCANICA AVANZATA E TECNICA DEL VEICOLO

(Nell'ambito della Scuola di Dottorato in HIGH MECHANICS AND AUTOMOTIVE DESIGN & TECHNOLOGY / MECCANICA AVANZATA E TECNICA DEL VEICOLO)

Ciclo XXVI

VALIDATION TESTS AND BEST PRACTICES SUPPORTING AUTOMATED PROCEDURES IN IMAGE-BASED 3D MODELLING

Candidato: Isabella Toschi

Relatore (Tutor): Prof. Alessandro Capra

Direttore della Scuola di Dottorato: Prof. Paolo Tartarini

Messen ist Wissen

(Measurement is knowledge)

Georg Simon Ohm (1789-1854)

CONTENT

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Three-dimensional (3D) digital imaging systems consist of optical sensors that extract information about the geometry of visible surfaces, both natural and man-made, without contact. The term "3D" refers to the type of acquired (sampled) information, i.e. a set of 3D coordinates (triplets) on a surface, organized as a 2D array ("imaging" systems) that covers an area of the surface itself. The 3D information gathered by these systems is processed by a series of transformation steps, that convert raw data into meaningful measurements, such as features and models (i.e. surface digital representations), that can be further visualized and analysed. Many market sectors benefit from these 3D imaging systems, that have recently gained large diffusion in several application fields as diverse as manufacturing, automotive, geomatics, cultural heritage, space exploration, robot guidance, biomedical, quality control, reverse engineering, architecture, power generation/transmission and gaming. Of course, such different applications require also different needs to be met: in fact, if surface resolution, completeness and appearance (texture and reflectance) are paramount for visual communication purposes, the possibility of extracting accurate geometrical features, surfaces and edges from acquired 3D data is the basis of inspection activities. 3D imaging systems should deal with all these requirements if they want to reach always more diverse niche markets: this is a complex task, especially during a global recession time. Nevertheless, they seem to have successfully faced this challenge, as a recent market research[1] proves.

Many 3D imaging companies are, in fact, showing +20% annual growth since 2006. In particular, the global optical digitizer and scanner market reached $358.3 million in revenue in 2011, growing by 6.6 percent over 2010. Furthermore, revenue is expected to reach $467.1 million in 2016, with new product introductions and software development driving demand. This market will grow normally at a compound annual growth rate of about 5.4 percent in the forecast period (2011 to 2016). The strength of this sector is mainly due to its competitiveness: the market remains consolidated, with major participants having significant market share, whereas small- and medium-size participants have region-specific expertise. In particular, the top two companies (i.e. Hexagon Metrology and FARO Technologies) held together the 50% of market share, as a percent of sales.

Many key market drivers can potentially lead this expected growth, such as: the automation and inline inspection, that create need for high-end scanners; a consistent product introduction by leading vendors, that provides end-users with more purchasing options; the increase in R&D spend, that helps manufacturers to cope with evolving end-user technology. On the other hand, some market restraints should be identified as well, mainly consisting of the end-user industry maturity, compelling the identification of new growth market, and the price pressure, affecting the overall market growth. In particular, this latter factor will be paramount

[1] Source: "Analysis of the Global Optical Digitizer and Scanner Market", by Frost & Sullivan, NA89-30, September 2012. Forecast period: 2011 to 2016.

during the next few years, especially if one considers the mean pricing values for 3D laser scanner systems, that can be categorized by machine type, as follows:

- From $15,000 to more than $65,000 for 3D laser scanners mounted on portable arms (excluded the cost of the arm);
- From $20,000 to more than $50,000 for scanners mounted on fixed CMMs (excluding the cost of the Coordinate Measuring Machines);
- From $20,000 to more than $70,000, for stand-alone 3D laser scanners (including both hardware and software components);
- From $10,000 to more than $200,000 for withe-light scanners;
- From $75,000 to more than $150,000 for laser trackers.

All the above mentioned data give an immediate idea of the market potentialities and desirability; at the same time, they point out its drawbacks too, especially in terms of costs.

In this context, mainly dominated by laser scanners, several solutions for the automatic generation of textured dense 3D surface models from 2D images have recently appeared. These so-called "passive 3D imaging" systems, based on optical sensors recovering surface information without any artificial source of light, offer nowadays both low-cost software packages and open-source solutions, that provide the user with the possibility of automatically retrieving dense or sparse point clouds from a set of un-oriented digital images. The topic of image-based 3D modelling, though introduced well before the appearance of laser scanners, has recently become a very active research field in both the photogrammetric and computer vision communities: this was, in particular, due to two significant improvement processes, that produced remarkable developments in the field of passive 3D imaging systems. First of all, the emergence of cheap and high quality (in term of resolution and signal-to-noise ratio) digital cameras, together with recently advances in both hardware and software technologies, have significantly increased the power of images, that have always played an important role, by transmitting a huge amount of information in a compact, intuitive and visually appealing format. Moreover, the last four decades have also seen the rapid evolution of computational power in personnel computers and the development of a GPU-based approach. All these enhancements helped photogrammetry and computer vision-based algorithms to become very attractive for 3D modelling, leading to the development of automated procedures that provide the end-users with low-cost and versatile image-only methods. As a consequence, many new users are now exploiting these novel techniques, although if they are not experts or have only few ideas of photogrammetric principles. A number of metrological concerns have thus been voiced by both researchers[2] and non-experts, who are worried about the quality of the results delivered by these recent methods.

Even if 3D image-based modelling techniques are now widely available, in fact, neither international standards, nor best practices and comparative data have appeared yet. This lack

[2] Many scientific and technical communities have understood the importance to benchmark algorithms and systems. CIPA (International Committee for Documentation of Cultural Heritage, http://cipa.icomos.org) and ISPRS (International Society of Remote Sensing, http://www.isprs.org) Commission V are examples of international scientific communities that are actively supporting this process.

is critical for the end-user confidence and may prevent, or simply delete, the market growth itself. There is, thus, the need for the development of accuracy assessment tools, i.e. systematic methodologies specifically designed in order to evaluate the overall performance of these new methods, in terms of both resolution, uncertainty and accuracy of the thereby recovered 3D information. Key factors and critical configurations affecting 3D model accuracy should be analysed too, especially in terms of the adopted image acquisition protocols. Some efforts have already been made by the research community, that has looked at the actual accuracy of image-derived 3D models through comparison tests performed by using reference data acquired with laser scanners. Nevertheless, no internationally recognized standards can be found in the field of laser scanners too, whose measurements are still performed on the basis of datasheets provided by manufacturers and linked only to internal companies guidelines. Thus, the context of inter-comparisons between different 3D acquisition technologies (e.g. 3D data acquired from a time-of-flight laser scanner versus image-derived 3D data) gives rise to many concerns and issues that should be dealt with.

To answer some of these questions, many accuracy tests have been performed during the three years of PhD research, using reference data acquired with time-of-flight laser scanners, triangulation laser scanners, one total station, one laser tracker, some contrast targets and GPS-measured points. The application fields range from small sculptural elements and architectural objects within the cultural heritage sector, to a larger spatial scale case study covering an extensive portion of territory. A custom-made scene located in a ISO 1 environmentally controlled laboratory was adopted too. Many different platforms for image data acquisition were tested, by processing both terrestrial, UAV- (Unmanned Aerial Vehicle) based and spaceborne imagery. Among the open-source image-based software solutions, the attention was especially focused on the public domain Apero and MicMac tools, created by the French mapping agency. On the other hand, within the huge amount of commercially-available software packages, the ERDAS Imagine Suite 2011 by Intergraph was employed. All experiments were carried out with a metrological approach, trying thereby to fill the void created by the lack of standards in 3D imaging by reviewing the metrological aspects of the problem and by proposing an avenue for solution for inter-comparisons between dissimilar technologies. In particular, the main goals of this study can be summarized as follows:

- The identification and discussion of a traceable methodology to evaluate the metric quality and limitations of automated image-based 3D modelling techniques in a metrological context. Since the experiments rely on metrological inter-comparisons among results achieved with dissimilar instruments, the study aims at estimating the most important uncertainty components of each measurement within a proper error budget calculation.
- The provision of comparative data and accuracy evidences, in order to identify a set of the most significant parameters influencing an image-based method. Though the employed test environments are limited in volume, these experiments aim at highlighting many aspects that can be critical even when considering a wider scope project.

- The study of some key factors affecting 3D model accuracy, such as the image acquisition protocol and the digital camera calibration approach. In particular, the goal is to identify (and, thus, avoid) weak geometric configurations, low redundancy network design and incorrect calibration.
- The testing of different environment conditions, by performing the experiments in both outdoor, quite "hostile" environments and indoor environmentally controlled facilities. The former allow the study of the algorithm performances in dealing with unfavourable and uncontrollable boundary conditions; the latter offer a privileged context for traceable measurements where the accuracy of a measurement in terms of the uncertainty can be evaluated without worrying about the effect of temperature, humidity and pressure.

Within this context and general purposes, the thesis work will be structured as follows. Chapter 1 provides a general overview on available techniques in the field of 3D modelling from reality and reviews current literature on inter-comparisons and best practices in 3D documentation. A deepen description of passive, multi-view 3D imaging systems is presented in Chapter 2, whereas Chapter 3 covers the topic of available algorithmic and software solutions addressing the problem of 3D modelling from images. A glance at active 3D imaging systems is casted in Chapter 4, in order to describe how reference data, and associated uncertainties, can be measured. Chapter 5 provides a discussion on the issue of digital camera calibration, by presenting some tests performed in order to evaluate self-calibration and test-range calibration approaches. 3D modelling from terrestrial imagery and its metrological assessment are the main subjects of Chapter 6, where three applications are described, i.e. the creation of detailed 3D surface models of sculptural elements (Cathedral of Modena, Italy), an experiment in an outdoor environment within the architectural field (Cathédrale de la Majore, Marseille, France), and some tests performed with a custom-made scene in a ISO1 environmentally controlled facility for 3D imaging metrology (ISO1 Laboratory, Ottawa, Canada). Chapter 7, then, discusses the possibility of extracting 3D dense point clouds from images acquired with a UAV system, in order to reconstruct the complete geometry of a tower (Basilica Santo Stefano, Bologna, Italy); in particular, the topic of multi-sensor, multi-platform data fusion and integration is here addressed. Finally, the main steps that constitute the procedure of Digital Terrain Model and Digital Surface Model extraction form spaceborne imagery are metrically evaluated in Chapter 8. General conclusions are drawn at the end.

1. MODELLING FROM REALITY: A GENERAL OVERVIEW

1.1 Optical sensors for three-dimensional measurements

Three-dimensional (3D) modelling of an object can be defined as the complete process that starts from data acquisition and ends with a virtual model visually interactive on a computer (Remondino and El-Hakim, 2006). This research work will focus on **modelling from reality** (Ikeuchi and Sato, 2001), i.e. the 3D reconstruction, digital recording and representation of the shape and appearance of an existing object or scene. This reality-based 3D modelling and as-built documentation should be clearly distinguished from the computer graphics creation of artificial world models using graphics and animation software. Computer graphics is mainly applied in the fields of movie production, games, web-based applications and architectural or object design: these kind of applications fall outside of the general aim of this research thesis.

The most general classification of 3D object reconstruction techniques is summarized in Figure 1.1, where the fundamental distinction between contact and non-contact methods is graphically explained.

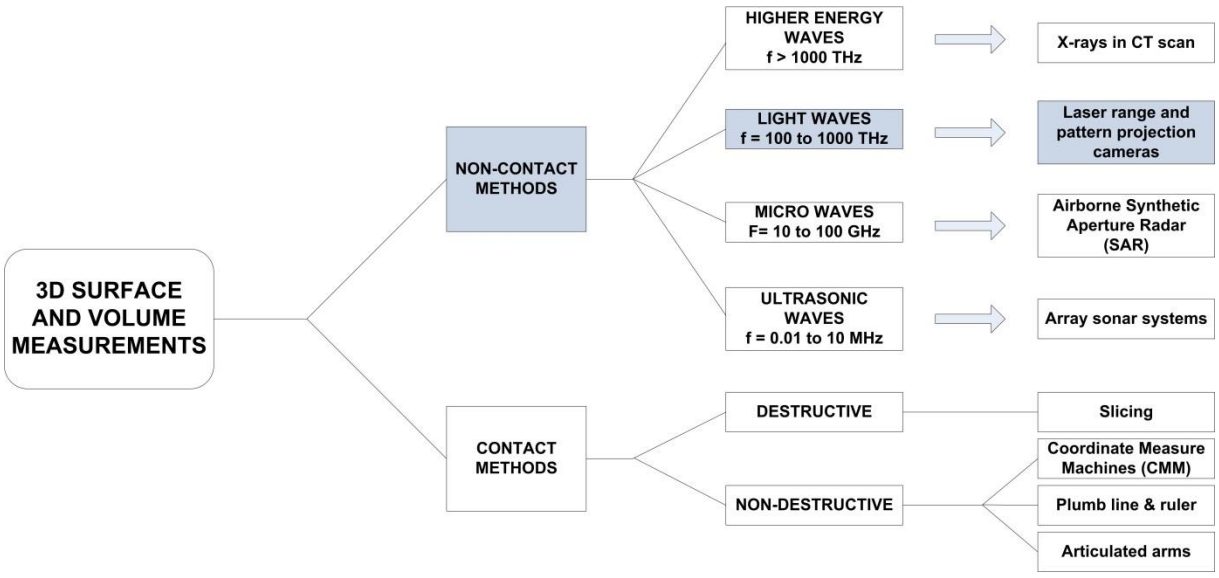


Figure 1.1 General overview of measuring 3D shape systems
(Jähne et al., 1999)

Nowadays the generation of a 3D model is mainly performed using non-contact three-dimensional measurement techniques based on light waves, highlighted in blue in Figure 1.1. These methods are generally based on the use of active or passive optical sensors (Figure 1.2); in some applications, other information derived from CAD (Computer Aided Design) models (Yin et al., 2009) or classical surveying, e.g. Total Station (TS) or Global Positioning System (GPS), may also be integrated with the sensor data.

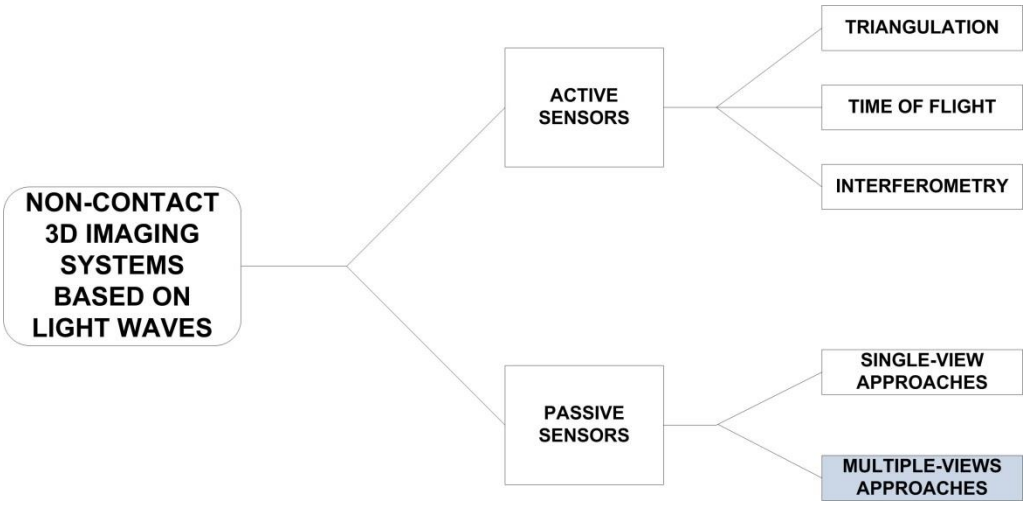


Figure 1.2 General overview of non-contact 3D imaging systems based on light waves

Active 3D imaging systems use an artificial illumination, usually either a spatially coherent light source (e.g. laser) or an incoherent one (e.g. halogen lamp), to directly capture the 3D geometric information of an object. Optical active sensors (Blais 2004; Vosselman and Maas, 2010) acquire dense range maps and point clouds from a visible surface, producing a quantifiable 3D digital representation of it in a specified finite volume of interest and with a particular measurement uncertainty. Although this sensors can be based on different measurement principles, commercially available systems generally use time-of-flight, triangulation and interferometry principles for the measurement of objects (Drouin and Beraldin, 2012). Though being popularized more than 30 years ago (Bandiera et al. 2011), it is only in the last 10 years that active sensors and range data have received greater attention and hence becoming a very common tool for both the scientific community and non-expert end-users, such as professionals working in the Cultural Heritage (CH) field. This diffusion is mainly due to two advantages shown by these systems: first of all, they can directly provide the user with dense 3D information, e.g. point clouds, even if this huge amount of data usually require a time-consuming post-processing phase to be done, in order to obtain a geometrically detailed and textured 3D polygonal model. Secondly, the use of an artificial light source makes it possible for active 3D imaging systems to deal also with featureless surfaces, requiring minimal operator assistance; furthermore, 3D delivered information is quite insensitive to ambient illumination and surface colour. On the other hand, these sensors are strongly affected by the reflective characteristics of the surface and do not respond adequately with translucent materials, such as some types of marble (Godin et al., 2001; Guidi et al., 2010). Moreover, they require a good knowledge of the capability, operative distance camera-to-object and measurement uncertainty of each different technology for the desired application. Other problems may arise in the case of very large data-sets and in the modelling of sharp edges, resulting in blunders or smoothing effects, especially for long-range sensors (Remondino and El-Hakim, 2006). On the other hand, the range-based modelling production process is nowadays quite straightforward (Cignoni and Scopigno, 2008) and active sensors can provide accurate and detailed 3D models of small and medium-size objects, with a high

degree of automation (Beraldin et al., 1999). As regards the qualitative aspect of the delivered 3D information, most of the active systems provide only a monochrome intensity value for each range value, focusing only on the acquisition of the 3D geometry. Even if some sensors have a digital colour camera connected with the instrument in order to register texture together with geometry, this approach does not usually provide the best results: the practical external scanning conditions may not in fact coincide with the ideal ones required by the digital image acquisition. Therefore, an high resolution photo-realistic texture should be added to the 3D model in the post-processing phase, but this is still nowadays a complex and long process (Beraldin et al., 2002). Finally, active sensors usually show some practical drawbacks: high costs for the acquisition of the instrument and the software for the data-treatment phase, time consuming data acquisition phase, limited flexibility and cumbersome instrumentation. A general overview of active 3D imaging systems will be given in Chapter 4.

Passive 3D imaging systems use natural illumination, thus they can recover 3D scene information without the use of an artificial source of light or other electromagnetic radiations. Optical passive sensors (Remondino and El-Hakim, 2006) are usually digital camera provided with a Charge Coupled Device (CCD) or a Complementary Metal Oxide Semiconductor (CMOS) sensors, both of which are able to capture light and convert it into electrical signals. These systems can't directly provide 3D data, since they require a mathematical formulation to transform 2D image measurements (correspondences) into 3D information. This transformation can be based on projective geometry (Pollefeys et al., 2004) or use a prospective camera model. Passive 3D imaging systems originate from the mature field of photogrammetry and, more recently, from the younger field of Computer Vision (CV), that, in contrast with photogrammetry, is mainly focused on the development of fast and automatic techniques, sometimes at the expensive of accuracy (Se and Pears, 2012). Image-based modelling techniques are nowadays receiving great attention from both the surveying community and non-experts, who benefit from the advantages of these systems, e.g. low cost, short data collection time, flexibility, space-saving and light instrumentation and direct extraction of photo-textured point clouds. In particular, the wide diffusion of high quality consumer-grade digital cameras, along with an increasing focus on photo-realistic 3D modelling in many application fields, e.g. CH, have recently forced both the photogrammetric and the CV communities to work together, in order to develop adequate solutions to the problem of extracting point clouds from a set of un-oriented images. This effort has led to the development of different image-based software packages, a general description of which can be found in Chapter 3.

In order to recover 3D information from 2D camera images, at least two images are generally required, even if 3D shape can also be inferred from a single viewpoint using other information sources: according to this distinction, 3D vision techniques can be categorized as Multiple-View and Single-View approaches. **Single-View approaches** can record the 3D geometry of an object using a single image together with other cues, such as shading, texture and focus. Shape-from-shading techniques (Horn and Brooks, 1989; Zhang et al., 1999), for example, use the shades in a greyscale image to infer the shape of the surfaces: after the

surface normal have been computed for each pixel, they can be converted into a depth map using a regularized surface fitting. This approach require some assumptions to be made, such as uniform albedo, reflectance and known light source directions; the computations involved are considerably complicated and there are open issues with convergence to a solution. Shape-from-texture approaches (Kender, 1981; Garding, 1992) recover the shape of the observed object using the distortion of the surface texture caused by the image process. In this case, the required assumptions are those of having a texture surface and the presence of a regular pattern. The degree of blur is a strong cue for object depth estimation and is used by Shape-from-focus techniques (Pentland, 1987; Nayar and Nakagawa, 1994) in order to retrieve 3D information from two input images captured from the same viewpoint but at different camera depths of field. The depth of the scene is in this case derived from the amount of defocus, that is estimated by averaging the square gradient in a region and increases as the object moves away from the camera focusing distance. Other single-view approaches use methods such as Shape–from-specularity (Healey and Binford, 1987), Shape-from-contour for medical applications (Ulupinar and Nevatia, 1995) and Shape-from-2D-edge-gradients (Winkelbach and Wahl, 2001). All such methods, although capable to recover 3D information from a single viewpoint, are often not practical, in terms of both robustness and speed. Thus, the most commonly applied approaches are those based on multiple views. **Multiple-View approaches** are able to infer information on the 3D structure of a scene, observing it from two or more viewpoints. In particular, the multiple images can be taken simultaneously using two or more cameras (Stereo) or in a sequence captured over a period of time by a single moving camera (Structure from Motion). Stereo methods are based on the triangulation process, that intersects 3D corresponding rays to determine the 3D position of the scene point. This operation, despite its apparent simplicity, requires the solution of many difficult problems that may be summarized as: camera calibration, image correspondences and dense 3D reconstruction. In contrast to stereo systems, Structure from Motion (SfM) refers to variable viewpoints with a sequential image capture. These methods, typical of the CV community, can recover at the same time the 3D motion of the camera and the 3D structure of the observed scene.
This research work will mainly focus on image-based, multiple-view approaches, highlighted in blue in Figure 1.2; in particular, Chapters 2 and 3 will be devoted to present in detail the various aspects of both stereo and structure from motion 3D imaging.

1.2 Comparison and integration between active and passive optical sensors: related work

After a general overview on active and passive 3D imaging systems, this section is devoted to present some research works, practical projects and publications where the topic of comparison and integration between these digital reconstruction techniques has been deepen dealt with. Since their introduction in the early 90s, active sensors like laser scanners should

directly compete with photogrammetry, analytical and digital, in the field of 3D surface and object measurements: this has led to many publications where **comparisons** between the two acquisition systems are presented (Georgantas et al., 2012). (Baltsavias, 1999) introduces a general comparison between airborne laser scanning and traditional manual photogrammetry for the extraction of Digital Terrain Models (DTM) and Digital Surface Models (DSM). Many advantages of laser scanning techniques are shown, such as: density of measurements, automation and rapidity. Nevertheless, the author concludes that the two technologies are complementary to each other since the one can compensate for the individual weaknesses of the other. Other authors compare and discuss practicality issues of laser scanning and digital close range photogrammetry (CIPA&ISPRS, 2002; Velios and Harrison, 2002). (Georgantas et al., 2012) presents a comparison between automatic photogrammetric technique and terrestrial laser scanning for 3D modelling of complex interior spaces. The authors conclude that, even if the results of the image-based approach may be less accurate than the ones delivered by the range-based method, photogrammetry can be considered an interesting solution to laser scanning thanks to its scalability, low cost and on the field rapidity.
Since laser scanning techniques have been widely adopted for the CH documentation, many papers have studied and presented the advantages and disadvantages of these methods if compared with manual close range photogrammetry (Alshawabkeh and Haala, 2004; Böhler and Marbs, 2004; Kadobayashi et al., 2004; Böhler, 2005; Remondino, 2005; Grussenmeyer et al., 2008). The general conclusion of all these works is that the choice of the method is heavily correlated to the characteristics of the scene that has to be modelled; a combination of the two approaches may be very useful in many cases and therefore desirable.

Starting from these evidences, range-based and image-based 3D vision systems have recently begun to be part of the "**multi-sensor data fusion**" research topic, i.e. the issue concerning those techniques that combine data from multiple sensors and related information from associated databases, in order to achieve improved accuracies and more specific inferences than could be achieved by the use of a single sensor alone (Hall and Llinas, 1997). Generally speaking, the benefits that should derive from the application of these multi-sensor data fusion techniques may be summarized as follow (Hong, 1999):

- Robust operational performance;
- Extended spatial/temporal coverage;
- Reduced ambiguity;
- Increased confidence;
- Improved detection performance;
- Enhanced resolution (spatial/temporal);
- Increased dimensionality.

This topic usually cover two categories of applications, i.e. information augmentation and uncertainty management (Beraldin, 2004). (Hong, 1999) defines information augmentation as referred to a situation where each sensor provides a unique piece of information to an application: fusion can therefore extend, for example, the system's spatial/temporal coverage. On the other hand, uncertainty management covers those situations where the same 3D scene

is acquired by different sensors, from different locations or times or even users. In these cases, data fusion should get the most out of each employed sensor, minimizing the impact of the uncertainties related to different variables (e.g. sensing devices, external environment and user skills): the lowest global uncertainty, that justifies the use of an expensive multi-sensor solution, is reached if one can manage all these uncertainties.
Different investigations on sensor integration have been performed in (El-Hakim and Beraldin, 1994, 1995). Active and passive optical sensors have been integrated especially for complex or large architectural objects, where the use of a single technique cannot provide a complete and detailed 3D model. In these cases, elementary shapes (e.g. planar surfaces) are recovered by image-based methods, while fine details (e.g. detailed reliefs) are described by range sensors (Flack et al., 2001; Sequeira et al., 2001; Bernardini et al., 2002; Borg and Cannataci, 2002; El-Hakim et al., 2004; Beraldin et al., 2005). In (Beraldin, 2004) the complementary between laser scanning and photogrammetry is examined. The author also presents a short review of the basic theory, measurement uncertainties and best practices associated to laser range scanners and digital photogrammetry. The integration of image-based modelling and laser scanning techniques is also examined in (El-Hakim et al., 2008), where the challenges presented to both approaches by the problematic texture of marble surfaces are underlined. Finally, one can find many other practical projects where a combination of these two modelling approaches has been used, coupled with survey information and maps for external geo-referencing and scaling, in (Stumpfel et al., 2003; Rönnholm et al., 2007; Stamos et al., 2008; Guidi et al., 2009; Remondino et al., 2009).

In this research thesis both issues of comparison and integration between image-based and range-based 3D imaging techniques are presented. In Chapter 6, the metric potentiality and associated measurement uncertainty of dense image-matching algorithms in dealing with different 3D objects and scenes will be assessed through comparisons with high resolution data acquired with active optical sensors of known uncertainty. The integration of the two different approaches is tested as well: Chapter 7 presents a case study where the integration of range and image data is necessary in order to deliver a complete model of a complex architectural structure.

1.3 3D modelling applications

The creation of photo-realistic 3D models of an observed scene has been an active research topic for many years, since these three-dimensional representations may be very useful in many applications for both visualization and metric measurement purposes. Some recent applications will be briefly presented in this section, where the focus will be mainly from the photogrammetric and CV-based perspective. A more general overview of both 3D active and passive vision system applications can be found in (Sansoni et al., 2009).

Many academic and commercial projects have efficiently applied 3D passive vision systems in order to reconstruct large-scale urban scenes. These **city modelling** applications include, for example, the 4D Cities Project (Schindler et al., 2007), which aims at creating a spatial-

temporal model of Atlanta city from historical photographs; the Stanford CityBlock Project (Román et al., 2004), that uses video of city blocks to generate multi-perspective strip images, and the UrbanScape project of (Akbarzadeh et al., 2006; Mordohai et al., 2007), are other two examples of this kind of applications. The objective of the latter is to develop a real-time data collection and processing system for the automatic geo-registered 3D reconstruction of urban scenes from video data. GPS and Inertial Navigation System (INS) measurements are integrated as well, in order to geo-reference the reconstructed photo-realistic 3D models.

For autonomous vehicles and planetary rovers, the generation of 3D terrain models of the environment is applied for visualization and path planning purposes (Barfoot et al., 2006). **Planetary Rover Navigation** requires in fact the ability to sense the nearby 3D terrain: the use of stereo cameras turns out to be suitable for planetary exploration, since they exhibit low power, low mass requirements and no moving parts. In this application field, the NASA Mars Exploration Rovers, named Opportunity and Spirit, may be mentioned: both of them use passive stereo image processing in order to measure geometric information about the nearby environment (Maimone et al., 2006). In particular, they use a pair of rectified stereo images to generate a 3D point cloud by matching and triangulating their pixels.

Mobile robot localization and mapping represents a third kind of application: in this case, the simultaneous tracking of the position of a mobile robot relative to its environment and the building of a map of the environment itself, are the required tasks. To achieve a simultaneous localization and mapping capability, high resolution passive vision systems can be efficiently applied, since they can capture images in milliseconds, being thus suitable for moving platforms such as mobile robots. The NASA Mars Exploration Rovers are, for example, equipped with this kind of capability (Maimone et al., 2007): the rover's position is updated by tracking the motion of selected terrain features between two pairs of stereo images. Apart from localization and mapping, passive 3D imaging techniques can also be applied for obstacle and hazard detection in mobile robotics: image matching approach is here used to derive 3D dense point clouds and cluster of them that are above the ground plane are considered to be hazards.

Airborne mapping and surveillance represents essential tasks for military missions: in this application field, Unmanned Aerial Vehicles (UAV) are becoming the favourable platforms for such surveillance operations, using on-board video cameras. Photo-realistic 3D models can be thereby generated in order to provide situational awareness and an easier understanding of the scene (Se et al., 2010).

Accident reconstruction and forensic applications represent other examples of fields where the use of contactless, accurate and fast acquisition systems would be of great advantage. With passive 3D imaging systems, 3D models of the crime scene can be created quickly without the generation of changes or disturbances to the crime scene itself (Little et al., 1999; Bruschweiler et al., 2003; Se et al., 2010). The extracted 3D model can then be used by the police to perform additional measurements and be shown in court so that the judge and the jury can understand the crime scene better.

Photo-realistic 3D models are useful also for survey and geology in **underground mining**: consecutive 3D models of a mine can be created as it advances, in order to update the mine

map after each drill/blast/ore removal cycle (Se et al., 2010). In this way, any deviation from the plan is minimized and the mining companies can monitor how much material is taken at each blast.

The possibility to create, display, manipulate, archive and share a digital representation of the shape and appearance of an existing object finds a most challenging class of applications in high resolution recording of heritage-related objects and sites (Beraldin et al., 2002). **Cultural Heritage** applications can benefit from 3D model representations thanks to their wealth of information that can be analysed and enhanced. Small objects or features can be interactively examined, even if they are only visible from a distance; experts and tourists can still study and visit those sites that must be closed for conservation reasons; once a 3D model has been created, computer-based visual enhancement and analysis techniques can be efficiently applied to it, in order, for example, to perform a virtual restoration of an historical site/object. For instance, virtual restoration allows the experts to deepen explore textual and artistic informative data, without operating directly on the original copy; furthermore, the site can be re-viewed in his correct historical context after having virtually removed architectural elements that have been added over the years. Moreover, tourism applications can be enhanced by the introduction of virtual 3D visits and tours, that guide the visitor through the discover of the historical site and of its developments both in space and over the centuries. Other CH applications of 3D modelling techniques can be found in the field of accurate archiving, 3D catalogue generation and rapid prototyping, that allows the creation of physical copies of the object without a direct contact with it. All these CH documentation applications (Godin et al., 2002; Ikeuchi and Miyazaki, 2007) find in passive vision systems a good response to their basic needs, i.e.: accuracy, portability, low cost requirements, on-the-field rapidity, flexibility, scalability and dense 3D data recording. For all these reasons, a significant number of 3D recording and modelling projects, mainly led by research groups, have been performed in the last decade. Among them, (Levoy et al., 2000; Bernardini et al., 2002; Grün et al., 2004; El-hakim et al., 2008; Guidi et al., 2009) have realized very good quality and complete digital models also exploiting the integration between active and passive 3D imaging systems. In (Beraldin et al., 2002) the virtualization of a Byzantine Crypt is presented, as an application of an effective approach for the photo-realistic 3D model building from the integration between photogrammetry and range data. In (Beraldin et al., 2011) the authors address the problem of the 3D photo-realistic reconstruction of a Neolithic cave, showing best practices and the processing pipeline related to this work. Finally, photogrammetric and remote sensing technologies and methodologies for CH 3D documentation and modelling are presented in (Remondino, 2011).

Although this section has briefly showed how many different applications of 3D imaging systems are today efficiently applied, the projects presented in this research work will mainly focus on the Cultural Heritage field: 3D data are, in fact, a critical component in the preservation activities and efforts demanded by our heritages, both natural and historical. Permanently recording the form of objects and sites that suffer from on-going attrition may thus allow their passing down to future generations.

1.4 Best practices for 3D imaging systems

Best practices and guidelines are a fundamental metrological topic, especially when a given technology matures enough that many users, not really familiar with it, decide to approach that technology and to make it mainstream. This is, for example, the case of 3D modelling methodologies: many users, also non-experts, have tried to approach them, especially after the emergence and diffusion on the market of high quality, non-metric and relatively cheap digital camera, together with many software solutions for the 3D image-based modelling. So, clear statements and information about these optical 3D measurement systems should be established, in order to show not only their advantages, but also their metric limitations. If many technical standards have already been adopted for the traditional surveying and dimensional contact metrology fields, it is only in the last few years that best-practice-related projects and information have appeared in the field of 3D Cultural Heritage. Since this research work aims at producing comparative data and metric evidences that may support the best practice and standard production, a general overview of some initiatives regarding best practices for 3D will be presented in this section, with a focus on Cultural Heritage. A more deep explanation of these projects can be found in (Beraldin et al. 2011).

A standard is "*A document established by consensus and approved by a recognized body that provides for common and repeated use, rule, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context*" ("ISO/IEC GUIDE 2", 2004). On the other hand, best practices and guidelines are able to ensure, or at least increase, the chance of performing quality data acquisition and subsequent use in a given field: therefore, as clearly stated in (Beraldin et al. 2011), best practices help to increase the chances of success of a given project.
Starting from these assumptions and definitions, some fundamental principles that should allow metrological experts to make accurate measurements can be found in (Flack and Hannaford, 2005). This **Good Practice Guide** set up six guiding principles to perform good measurements, i.e.:

- The Right Measurement;
- The Right Tools;
- The Right People;
- Regular Review;
- Demonstrable Consistency;
- The Right Procedures.

The American Society for Testing and Materials, Committee E57, **ASTM E57**, was formed in 2006 with a focus on 3D imaging systems ("Committee E57", 2008; Cheok et al., 2008). The Committee is divided into four subcommittees: Terminology, Test Methods, Best Practices and Data Interoperability. In particular, the third mentioned subcommittee defines a best practice as a process or method that, when executed effectively, leads to enhanced project performance. The general aim of this subcommittee is to develop, validate and communicate best practices in the field of 3D imaging technology; its primary focus was concentrated on

safety requirements and led to the developments of the "E2641, Practice for Best Practices for safe application of 3D imaging technology".

Another initiative regarding best practices, the **Heritage3D project** ("Heritage3D"), was undertaken by the School of Civil Engineering and Geosciences at Newcastle University. This two-year project developed best practices in laser scanning for archaeology and architecture; in particular, useful information on defining a typical project workflow is derived.

The **E-Curator project** ("E-curator research project") is focused on the application of 3D scanning and e-Science technologies to museum work and artefact analysis. Generally speaking, the project contributes to the CH field by reducing some of the practical barriers to the movement of people and objects, enhancing international scholarship, facilitating the safe moving of artifacts and a better understanding of best practices for small to medium size museum artifacts ("E-curator research project").

The **London Charter** is an initiative concerned with the use of 3D visualization in the research and communications of Cultural Heritage. As stated in ("The London Charter for the Computer-Based Visualisation of Cultural Heritage"), the project aims at establishing what 3D visualization requires to be as intellectually rigorous and robust as any other research methods.

Many others best-practice-related projects have been undertaken by **academic research groups**. In (Benedetti et al., 2010), the authors summarize the methodology and pipeline followed to model a large and complex historical site, like the Forum of Pompeii, Italy. In particular, the best practices adopted in this case and the state of the art described in current literature are presented as well. More recently, another publication addresses the problem of 3D documenting and modelling a complex and large object, like the Grotta dei Cervi in Apulia, Italy (Beraldin et al. 2011). The authors present the project that led to the generation of a textured 3D model of the Neolithic cave; the best practices emerged from the work together with the processing pipeline are described as well. (Pavlidis et al., 2006) provides a discussion on the whole life cycle of the general cultural content, addressing the main issues affecting it. In particular, the authors identify five main processes in digital recording that require new advances, i.e.:

- Digitization in 3D;
- Processing and storage of 3D data;
- Archiving and management of 3D data;
- Visualization and dissemination of 3D data;
- Replication and reproduction of 3D data.

On the other hand, (Renisson et al., 2009) deals with the lack of standards that affects the 3D modelling field for archaeological applications. The research team decided to compare different datasets and to evaluate results of other people's work in the same field. For these purposes, a collaborative scanning project with four other institutions, both academic and cultural, was undertaken. As a result of this pilot project, the team concludes that the establishment of methodological guidelines for providing best practices still represents a difficult target to be achieved.

2. PASSIVE 3D IMAGING

2.1 Introduction

In this chapter a general description of passive, multiple-view 3D imaging systems is presented. The aim of the dissertation is to introduce the fundamental principles of passive 3D vision systems; a more complete and detailed review of passive 3D reconstruction methods can be found in these recent publications (Ma et al., 2003; Moons et al., 2008; Se and Pears, 2012). The first topic addressed is the definition of a **mathematical model** that can describe the image formation process. Generally speaking, image formation can be seen as the inverse problem of vision: the former studies how objects give raise to images, while the latter tries to use images in order to recover a description of objects in 3D space. Therefore, the description of vision algorithms requires first the development of a suitable image formation model. The term suitable should suggest that the level of abstraction and complexity in modelling image formation must trade off physical constraints and mathematical simplicity: in other words, the formulation should result in a manageable mathematical model. Both the classical photogrammetric collinearity model and the computer vision-based one will be deepen described. The estimation of the parameters in the developed camera models is then discussed, starting from the mathematical parameterization of the perturbation effects affecting real camera systems and caused by different physical sources, such as lens distortions and atmospheric refraction. This issue is addressed within the photogrammetric pipeline by performing a proper camera calibration procedure: two main calibration approaches will be detailed, together with a description of the general rules required to avoid system instability and parameter coupling. Once the image formation process and its mathematical model are provided, their "reverse" process can be undertaken in order to infer the geometry of the acquired 3D scene starting from its imaged representation. In multiple-view geometry this requires the solution of the **correspondence problem**, i.e. the search for homologous points in different images of the same 3D scene. In order to efficiently solve this problem, a good understanding of two-view geometry (i.e. the relationship between two camera views) is necessary: the epipolar geometry will be discussed therefor, together with its algebraically formulation. Following this, two main classes of correspondence algorithms will be described, after having presented a useful shortcut (the rectification process) that simplifies the correspondence search to be across the same horizontal scanlines in each image. Finally, a second question will be addressed, i.e.: given two corresponding points, how can the 3D position of the object point be computed? This is the **3D reconstruction problem**, that will be dealt with by describing the process of generating a 3D point cloud from a set of image correspondences. The details of both stereo and structure from motion, i.e. the two essential forms of multiple-view 3D reconstruction technique, will be thus presented.
Figure 2.1 summarizes the procedural workflow followed in this chapter.

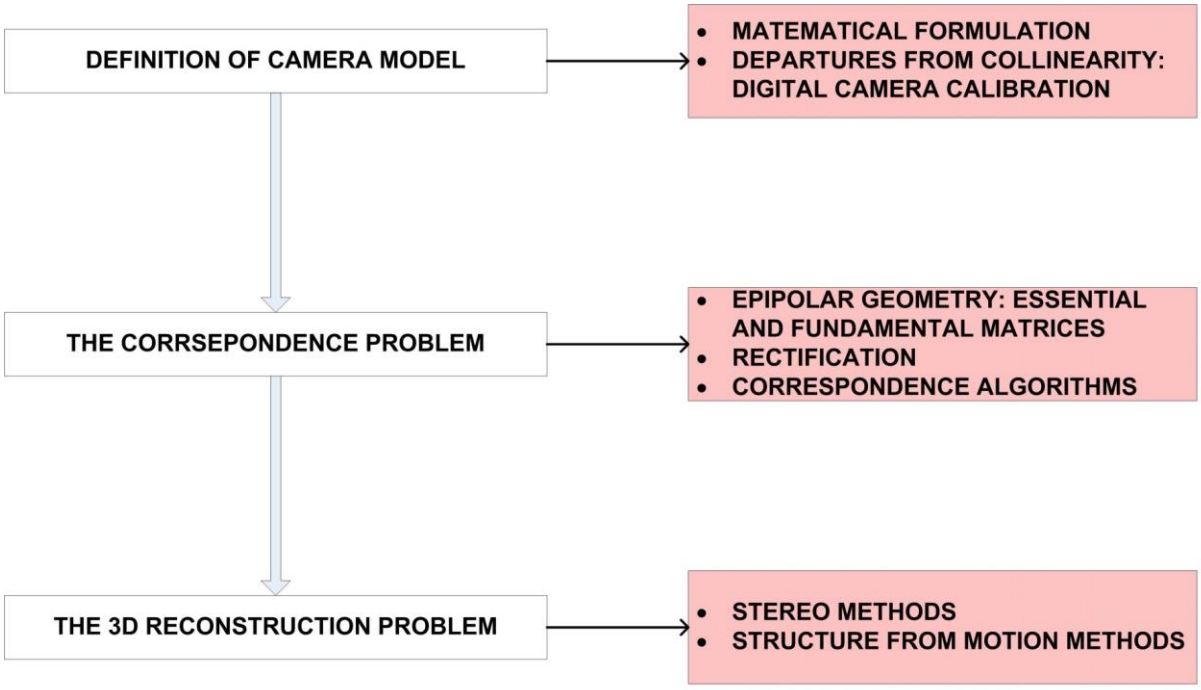


Figure 2.1 Procedural workflow (Passive 3D imaging)

2.2 Definition of camera model

An **image** may be defined as a two-dimensional brightness array; in mathematical words, it is a map I, defined on a compact region Ω of a two-dimensional surface, taking values in the positive reals. In particular, in the case of a camera, Ω is a planar, rectangular region occupied by the photographic medium (i.e. by a CCD/CMOS sensor for a digital camera), so that:

$$I: \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}_+; \quad (x, y) \rightarrow I(x, y) \qquad [2.1]$$

In the case of a digital image, both the domain Ω and the range $\mathbb{R}_+$are discretized: the image can then be represented as an array of numbers, whose values depend upon physical properties of the scene being viewed.

The image formation process takes place within a camera, where the 3D scene is back-projected down on the 2D image itself. The most commonly used projection model is the ideal **3D perspective projection**, that is geometrically described in Figure 2.2. The illustrated perspective projection is based on the ideal pinhole camera model, where **C** is the position of the pinhole, defined as the camera centre or the centre of projection. A first conventional point to note is that, although the real image plane (image sensor) is behind the camera centre (centre of the lens system), it is common practice to use a virtual image plane in front of the camera: in this way, in fact, the image is conveniently at the same orientation as the scene, i.e. not inverted top to bottom and left to right. The path of the imaged light is modelled by a ray that passes from a 3D object point **X** through the camera centre. The position of the corresponding 2D image point $\mathbf{x_c}$ lies at the intersection of this ray with the image plane.

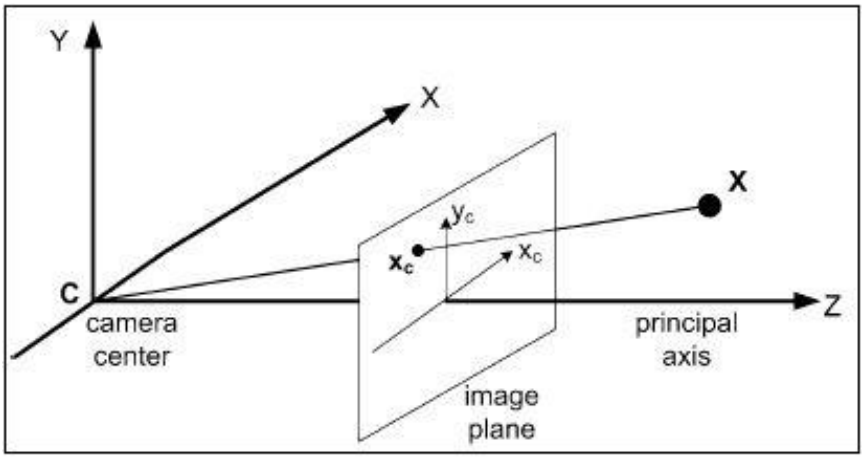


Figure 2.2 3D perspective projection
(Se and Pears, 2012)

It's now possible to geometrically introduce the process of back-projecting an image point to an infinite ray that extends out into the 3D scene: for each point on the image plane, its corresponding object point must in fact lie somewhere along the ray connecting the centre of projection **C** and that imaged point $\mathbf{x_c}$. In other words, when the scene is imaged with a "perfect" camera system, the perspective centre, the object point and the corresponding image one should lie along the same straight line. This is referred to as "**collinearity**" condition and forms the basis of the classical analytical photogrammetry. The process of analytical photogrammetric restitution requires, in fact, a perspective transformation between image and object space. The fundamental problem results in an optical triangulation process between two or more digital images, i.e. the computation of the (X,Y,Z) object space coordinates as intersection of the optical rays back-projected from the camera centres through the corresponding (x, y) image coordinates. To solve this problem, the following geometric relations should be established:

- The spatial direction of each of the intersecting rays and the position of the perspective centre for each camera system, with respect to the XYZ object coordinate system. This involves the computation of six parameters (three orientation angles and three camera station coordinates) per image, describing the so-called **exterior orientation** of the image itself.
- The relationship between the perspective centre and the x_cy_c image coordinate system, defined by the camera's **interior orientation**. In particular, this requires the knowledge of three parameters, i.e. the camera principal distance f (the distance between the image plane and the camera centre) and the coordinates (x_0, y_0) of the principal point (intersection between the optical axis and the image plane).

The above defined collinearity condition is thus implicit in the perspective transformation between image and object space, and can be mathematically expressed through the well-known collinearity equations (Kraus, 1994):

$$x - x_0 = -f\frac{r_{11}(X - X_0) + r_{12}(Y - Y_0) + r_{13}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)}$$
$$y - y_0 = -f\frac{r_{21}(X - X_0) + r_{22}(Y - Y_0) + r_{23}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)} \quad [2.2]$$

where $r_{i,j}$ are the elements of a unitary-orthogonal matrix describing the relative orientation between the xyz image space and XYZ object space coordinate systems. X_0, Y_0, Z_0 are the object space coordinates of the camera station.

Equations [2.2] show that each object point always corresponds to an image point; however, an infinite amount of possible object points exist for each image point. The resulting problem is concerned with the fact that we do not know how far along the ray the 3D scene point lies: in other words, explicit depth information is lost in the imaging process. This represents the main source of geometric ambiguity in a single image and clears up the requirement of a stereo image system (or of other cues in single-view approaches) in order to recover the depth information.

The above discussed collinearity equation-based model represents the perspective geometrical formulation used in close-range photogrammetry in order to perform sensor orientation and calibration; due to its non-linear nature, it requires approximate parameter inputs for a least-squares bundle adjustment (Brown, 1971) computation. A 3D perspective projection, based on the pinhole camera model, is also the most commonly used projection model in computer vision. This mathematical camera model will be deepen described in the following subsections, starting from a brief introduction to the concept of homogeneous coordinates that represent the natural coordinate system of analytic computer vision-based projective geometry.

2.2.1 Homogeneous coordinates

In analytical projective geometry, points and lines are typically described by homogeneous coordinates (also called projective coordinates), where (n+1) coordinates are employed to describe points in an n-dimensional space. For example, the representations of a general point and a general line in homogeneous coordinates are described as follows:

$$\mathbf{x} = [x_1, x_2, x_3]^T \quad [2.3]$$
$$\mathbf{l} = [l_1, l_2, l_3]^T \quad [2.4]$$

And the general equation of a line is given by:

$$\mathbf{l}^T\mathbf{x} = 0 \quad [2.5]$$

[2.5] is an homogeneous equation, since its right hand side is zero: this means that any non-zero multiple of the point $\lambda[x_1, x_2, x_3]^T$ is the same point and, similarly, any non-zero multiple of the line's coordinates is the same line. This symmetry is directly related to the dual theories of projective geometry, in which points and lines can be exchanged due to the symmetry itself. For example, in homogeneous coordinate space the cross product of two

lines yields their intersecting point and the cross product of a pair of points gives the line between them. Homogenous coordinates allow the relevant transformations in the image formation process to be represented as **linear mappings**, expressed as matrix/vector equations: in other words, the mapping between homogeneous object coordinates and homogeneous image coordinates is linear. Nevertheless, the mapping from homogeneous to inhomogeneous coordinates is still non-linear, because this conversion requires the division by the third element (i.e., $[x_1, x_2, x_3]^T$ maps to $[x_1/x_3, x_2/x_3]^T$).

Homogeneous coordinates are usually employed in analytic projective geometry, since their use fits well with the relationship between image points and their associated back-projected rays into the object space. In fact, if a virtual image plane is placed at a distance of one metric unit in front of the centre of projection, the 3D scene ray is defined as $[\lambda x, \lambda y, \lambda]^T$, where ($\lambda$>0) is the unknown distance along the ray. Thus, there is an intuitive link between the previously mentioned depth ambiguity and the definition of homogeneous coordinates up to an arbitrary non-zero scale factor. Finally, the definition of point at infinity should be stated: it is referred to any point with zero as its third homogeneous element. There is an infinite set of points $[x_1, x_2, 0]^T$, that define a ray parallel to the image plane and hence meet it at infinity. This line $[0,0,1]^T$, where the points at infinity lie, is called line at infinity. In the following sections, a tilde symbol will be used to differentiate n-tuple inhomogeneous coordinates from (n+1)-tuple homogeneous coordinates.
A more detailed reading on homogeneous coordinates and projective geometry can be found in (Coxeter, 2003; Hartley and Zisserman, 2004).

2.2.2 Perspective projection camera model

The perspective projection camera model maps 3D object points (in standard metric units) into 2D image points (expressed in pixel coordinates of an image sensor). To model this kind of process, it is convenient to divide it into three successive steps:

1. A **coordinate transformation**, i.e. a rigid transformation that maps points expressed in object coordinates (world reference system) to corresponding points expressed in camera centred coordinates (camera reference frame). This 6 Degree-Of-Freedom (DOF) transformation consists in a rotation R (3 DOF) and a translation **t** (3 DOF).
   Referring to Figure 2.1, the camera reference frame has its (X,Y) plane parallel to the image plane and Z is the direction of the principal axis of the lens, termed the optical axis. Let's suppose that, in the world reference system, $\tilde{\mathbf{C}}$ is the camera centre inhomogeneous position and R_c is the rotation of the camera reference system relative to the world one. So, a generic object point can be expressed in the inhomogeneous camera reference system (subscript c symbol) as:

$$\tilde{\mathbf{X}}_\mathbf{c} = R_c^T\left(\tilde{\mathbf{X}} - \tilde{\mathbf{C}}\right) = R\tilde{\mathbf{X}} + \mathbf{t} \qquad [2.6]$$

   Where:

- $R = R_c^T$ is the rigid rotation;
- $\mathbf{t} = -R_c^T \tilde{\mathbf{C}}$ is the rigid translation.

Equation [2.6] can be expressed as a projective mapping, that is linear in homogeneous coordinates:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = P_r \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad [2.7]$$

Where:

- $P_r = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ is the homogeneous matrix representing the rigid 6 DOF coordinate transformation.

2. A **perspective projection**, that maps object points (expressed in camera reference frame) to the corresponding image points (expressed in metric image coordinates). Starting from the similarity of triangles in the geometry of perspective imaging (Figure 2.1), one can derive:

$$\frac{x_c}{f} = \frac{X_c}{Z_c} \quad ; \quad \frac{y_c}{f} = \frac{Y_c}{Z_c} \quad [2.8]$$

Where:

- (x_c, y_c) is the position of a point in the camera reference frame (metric units);
- (f) is the distance between the image plane and the camera centre, termed principal distance (metric units). This parameter is usually set to the focal length of the camera lens.

The equations [2.7] can be written in linear form as:

$$Z_c \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = P_p \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad [2.9]$$

Where:

- $P_p = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ is the perspective projection matrix, defined by the value of f.

If the distance between the camera centre and the virtual image plane is fixed at one metric unit (f=1), then points on this plane are characterized by normalized image coordinates, given by:

$$x_n = \frac{X_c}{Z_c} \quad ; \quad y_n = \frac{Y_c}{Z_c} \quad [2.10]$$

The subscript n will be used to indicate image coordinate normalizations.

3. An **image sampling**, necessary to map from metric image coordinates to pixel coordinates. The image sensor, a CCD or CMOS device, usually samples the image produced on the image plane at the locations identified by an array of pixels. A final step in the image formation modelling is so necessary, in order to define how pixel coordinates can be generated. In the more general case, pixels in an image sensor are not square and the different scaling may be defined as follows:
   - (m_x) is the number of pixels per unit distance in the x_c direction;
   - (m_y) is the number of pixels per unit distance in the y_c direction.

   Pixel positions usually have their origin at the upper left corner of the image sensor, so that the intersection between the principal axis and the image plane, termed principal point, may be localized by the pixel coordinates $[x_0, y_0]^T$. Finally, a general camera model should also consider a skew parameter, s, that cater for any lack of orthogonality between the two directions of the image sensor sampling. Starting from all these assumptions and definitions, the mapping into pixel coordinates is given by:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P_c \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \quad [2.11]$$

   Where:

   - $P_c = \begin{bmatrix} m_x & s & x_0 \\ 0 & m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$ is the projective matrix, defined by the five parameters m_x, m_y, x_0, y_0 and s.

After having described the three successive steps of the image formation process, it is now possible to concatenate them, giving:

$$\lambda \mathbf{x} = P_c P_p P_r \mathbf{X} \quad \rightarrow \quad \lambda \mathbf{x} = P\mathbf{X} \quad [2.12]$$

Where

- λ is a non-zero and positive value;
- P is the final projection matrix.

It is clear that any non-zero scaling of the projection matrix ($\lambda_p P$) performs the same projection since in equation [2.12] any non-zero scaling of homogeneous image coordinates is mathematically equivalent. As a consequence of this propriety, any camera with projection matrix P (termed camera P) is defined only up to scale.

Let's now focus on the structure of projection matrix P; it can be expressed by:

$$P = K[R|\mathbf{t}] \quad [2.13]$$

Where:

- $K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$ is the matrix defining the camera's **intrinsic parameters**[3] (5 DOF), in which $(\alpha_x = fm_x)$ and $(\alpha_y = fm_y)$ represent the focal length in pixels in the x and y directions;
- R and **t** define the camera rigid rotation and translation, that are termed the camera's **extrinsic parameters**[3] (6 DOF).

Thus, dealing with homogeneous coordinates, a camera projection matrix has only 11 DOF: in fact, as previously noticed, the overall scale of P does not matter.
Equation [2.12] may be expanded into:

$$\lambda \overbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}^{\text{homogeneous image coordinates}} = \overbrace{\begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}}^{\text{intrinsic camera parameters}} \overbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}}^{\text{extrinsic camera parameters}} \overbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}^{\text{homogeneous object coordinates}} \quad [2.14]$$

Two important consequences may be deduced from the above equation:

- Both the intrinsic and extrinsic camera parameters are necessary in order to metrically define a ray in the 3D space and to make absolute measurements in multiple-view 3D reconstruction;
- Any non-zero scaling of homogeneous object coordinates produces the same homogeneous image coordinates up to scale.

2.2.3 Digital camera calibration

The above derived linear projective model is based on the ideal pinhole camera assumption: this is usually not suitable to accurately describe real physical cameras, especially if a low-cost lens or a short focal length lens (such as a fisheye) are employed. Typical lenses are, in fact, characterized by different kind of distortions, whose principal effects act along the radial direction and are shown in Figure 2.3.

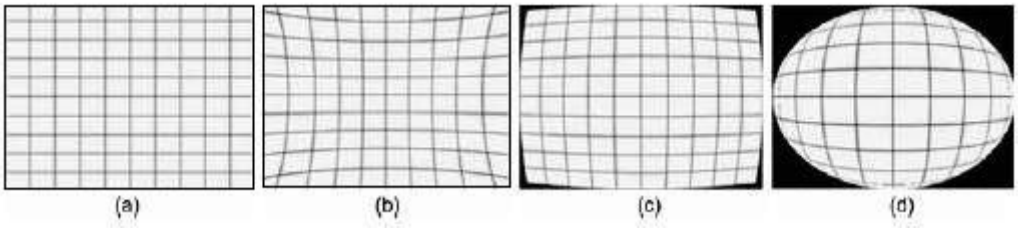


Figure 2.3 Example of radial distortion effects: (a) Ideal lens (no distortion); (b) Pincushion distortion; (c) Barrel distortion; (d) Fisheye distortion
(Se and Pears, 2012)

[3] The terms "intrinsic" and "extrinsic" parameters are used in the machine vision literature, and correspond to the "interior" and "exterior" orientations, as defined by the photogrammetric community.

The effect is non-linear and, referring to the previously described three-stage development of the projective camera model, it affects the second step, that is the 3D to 2D projection. Of course, distortion is then sampled by the image sensor too.
Many different distortion models have been formulated over the years, containing a huge number of parameters that model both radial and tangential distortion effects (Brown, 1966). One of the most noteworthy parameterization is formulated in (Fraser, 2001), and it will be detailed described in Chapter 3. Here a simple general mathematical expression is derived, in order to take into account the radial distortion effect, that usually represents the dominant factor. Using a low-ordered polynomial, the radial distortion can be modelled as:

$$\begin{bmatrix} x_{nd} \\ y_{nd} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} x_n \\ y_n \end{bmatrix} (k_1 r^2 + k_2 r^4) \qquad [2.15]$$

Where:

- $[x_n, y_n]^T$ are the normalized coordinates of the undistorted image position;
- $[x_{nd}, y_{nd}]^T$ are the normalized coordinates of the distorted image position;
- (k_1, k_2) are the unknown radial distortion parameters;
- $r = \sqrt{x_n^2 + y_n^2}$ is the radial distance from the principal point.

Finally, assuming zero skew, the distorted position can be expressed in pixel coordinates, as follow:

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} (x - x_0) \\ (y - y_0) \end{bmatrix} (k_1 r^2 + k_2 r^4) \qquad [2.16]$$

Where:

- $[x_d, y_d]^T$ are the pixel coordinates of the distorted image position;
- $[x, y]^T$ are the pixel coordinates of the undistorted image position;
- r is still defined in normalized image coordinates.

Both Equation [2.16] and Figure 2.1 show that the distortion effect increases away from the centre of the image and requires the image points to be moved towards the centre of the image (barrel distortion) or in the opposite direction (pincushion distortion). It is now necessary to understand how to compute the parameters of this more complete camera model.

Digital camera calibration (Brown, 1971) is the process of recovering the parameters of the camera that produced a given image of a 3D scene. Both extrinsic and intrinsic parameters should be thereby computed, comprising those modelling the camera distortion effects. Once all these parameters are known, the camera projection matrix P is totally defined and it's possible to back-project any image pixel to a 3D ray in space.
Generally, the end-user cannot get the required calibration information to the required accuracy directly from camera manufacturer's specifications and from external measurements of the camera poses in some reference frame. Thus, a camera calibration procedure should be performed and there is an extensive body of literature on this subject, whose main topics are: overall reviews (Fryer, 1996; Fraser, 2001; Remondino and Fraser, 2006); general investigations (Fraser and Shortis, 1995; Jantos et al., 2002); low-cost digital cameras (Läbe

and Förstner, 2004; Cronk et al., 2006); stability of parameters (Peipe and Stephani, 2003) and accuracy aspects (Salvi et al., 2002; Fraser and Al-Ajlouni, 2006). In order to give a general overview of camera calibration issues, common models and classifications are briefly summarized below.

Camera calibration techniques may adopt two different basic functional models, i.e.:

- A perspective projection camera model, that is based on the collineaity equations [2.2][4], specifically adapted in order to include perturbations to collinearity. This model requires five or more point correspondences within a multi-image network; due to its no-linear nature, it needs approximations for parameter values within the least-squares bundle adjustment in which the calibration parameters are computed.
- A projective camera model, characterized by the Essential and the Fundamental matrix approaches (see Section 2.2). It needs a minimum of 6-8 point correspondences to facilitate a linear solution.

A second classification can be made according to the parameter estimation and optimization technique employed (Remondino and Fraser, 2006):

- Linear techniques, that have the advantages of simplicity and rapidity, but generally cannot handle lens distortion and need a control point array of known coordinates. An example of this approach is the well-known Direct Linear Transformation (DLT) algorithm (Abdel-Aziz and Karara, 1971), that is equivalent to an Essential matrix model.
- Non-linear techniques, which are based on the perspective projection camera model and are typical of the photogrammetric approach. They require an iterative least-squares estimation process in order to provide a rigorous and complete modelling of all camera calibration parameters.
- A combination of linear and non-linear techniques, which is based on a two-stage approach. A linear method is initially employed for the parameter initialization process; then, orientation and calibration are iteratively refined (Heikkilä and Silven, 1997).

The most common classification is based on the approach employed in order to recover the unknowns of Equations [2.2][4]:

- A **Test-range Calibration** approach (Fraser, 2001), where only the interior and exterior camera parameters are computed, through a spatial resection approach. Any subsequent triangulation of object points would then be achieved through ray intersection, based on the recovered calibration information. This method requires the provision of a suitable control field comprising pre-surveyed targets of known XYZ coordinates. The computational effort involved with the iterative lest-squares estimation process of test-range calibration is reasonably modest, but to avoid system instability and coupling between parameters, some general rules should be followed:

[4] The collinearity equations are modified by introducing the terms Δx and Δy, that model the perturbations to collineairity (Fraser, 2001).

i. The control point field should be well distributed in three dimensions in order to facilitate the computation of intrinsic parameters.
ii. The use of multiple photo stations, with varying camera-object distances (but the same focal setting) and camera roll angles, greatly enhances the recovery of intrinsic parameters. This configuration, in fact, reduces the possibility of having a projective coupling between the parameters, such as for example between intrinsic and extrinsic parameters and between distortion parameters and principal point coordinates.
iii. The greater is the image point density, the better will be the expected recovery of the calibration parameters.

- A **Self-Calibration approach** (Fraser, 2001; Grün and Beyer, 2001), that implies the simultaneous recovery of all unknowns, i.e. interior and exterior parameters, distortion coefficients and object point coordinates. This method does not require any a-priori knowledge of object point coordinates or scale information: the correspondences across the images of the same scene provide enough constraints to perform a generalized relative orientation of all bundles of rays. Self-calibration is more rigorous and flexible than Test-range calibration, since it fully employs all geometric relationships modelled in the collineaity equations [2.2]: with this kind of approach, also the relative orientation between bundles of rays from different images is taken into account. Of course, a significant computational effort is required and critical to the accuracy of the process are the overall network geometry and, especially, the camera station configuration. The rules listed for the Test-range calibration approach are still valid for the Self-calibration one; furthermore, different experimental tests (Grün and Beyer, 2001; El-Hakim et al., 2003) have proved that:

  i. The accuracy of the process increases with increasing convergence angles for the imagery. Of course, this evidence implicitly means to increase the base-to-depth (B/D) ratio and may conflict with the requirements of a successful image matching process: thus, a compromise should be set up and a reasonable base-to-depth ratio should be kept.
  ii. The accuracy of the process is enhanced by increasing the number of rays to a given object point.
  iii. The accuracy of the process increases with the number of measured points per image, but the incremental improvement is small beyond a few tens of points.
  iv. Self-calibration does not require the object space to be well distributed in three dimensions and a planar object point array can be efficiently employed: in this case, however, the images should be acquired within a good network geometry, i.e. orthogonal roll angles, high degree of convergence and varying camera-object distances. Anyway, a 3D object should be preferred over a 2D distribution of object points.

A more general classification refers to the different calibration approaches adopted by the photogrammetric and CV communities, starting from their different accuracy requirements:

- Camera **calibration in Computer Vision** needs to position object points to an accuracy of only, say, 5% of the camera-to-object distance (Remondino and Fraser, 2006). The calibration models adopted have traditionally determined the calibration matrix K using images of a known object point array (usually, a checkerboard pattern). All the commonly adopted methods (Tsai, 1987; Heikkilä and Silven, 1997; Zhang 2000) are based on the pinhole camera model and include terms for modelling radial distortion. Self-calibration procedure in CV is generally adopted to upgrade a projective 3D reconstruction to one that is metric (see Section 2.3); in general, three kind of constraints are applied to perform it: scene constraints, camera motion constraints or constraints on the camera intrinsic parameters. Furthermore, only the focal length is usually determined, while lens distortion are neglected.
- **Photogrammetric camera calibration** is usually designed to support a subsequent object space measurements requiring 1:20000 accuracy (Remondino and Fraser, 2006). Different camera models have been formulated, but generally a perspective geometrical model based on the bundle adjustment procedure (Brown, 1971) is the most commonly adopted approach for sensor orientation and calibration. Non-linear techniques are employed and a favourable network geometry is required. The basic mathematical model requires to extend Equations [2.2][4] adding correction terms, usually expressed by additional parameters (Fraser, 2001). Within the bundle adjustment, these additional parameters can be treated as:
  i. Camera-invariant or focal setting-invariant, if one set of parameters is used for all images acquired by the same camera or from the same focus setting in a multi-camera calibration which may also include different zoom settings from the same camera.
  ii. Image-variant, if a different set of additional parameters is used for each image, if the acquisition has been performed using different cameras or different zoom settings. Although the adoption of image-specific additional parameters may be very problematic, a recently developed process termed zoom-dependent camera calibration (Fraser et al., 2012) presents a possible solution to the problem and describes the practical implementation of zoom-dependent calibration within software.

Finally, further criteria may also be adopted to classify camera calibration techniques, as:

- Implicit or Explicit models. The photogrammetric approach, based on its explicit physically interpretable calibration model, is usually opposed against implicit models used in CV techniques.
- Methods using a 3D or a planar point array.
- Point-based or line-based methods (Fryer and Brown, 1986). Point-based approaches are more popular in photogrammetry, where the only noteworthy line-based method is the one of plumb-line calibration.

2.3 The correspondence problem

The previous section described how images are formed and formulated a mathematical model of this process. In this section a first step along the reverse process will be described: in other words, it is now necessary to understand how to "undo" the image formation process, in order to infer the geometry of the acquired 3D scene. In multiple-view geometry, the first gap to fill is to establish how points "move" from one image to the next one. Or, more in general, it is first necessary to establish "which points correspond to which" in different images of the same 3D scene: this is called the **correspondence problem** and stands at the core of the process of converting 2D measurements of light (i.e. images) into 3D measurements of geometry (i.e. acquired scene). Following subsections will be mainly devoted to address this search problem in the field of two-view geometry, starting from a detailed description of this geometry, that establishes the relationships between two camera views.

2.3.1 Epipolar geometry

Starting from the assumption that two images of the same 3D scene have been acquired from two distinct vantage points, epipolar geometry establishes the relationship between these two camera views. Once this geometry is known, a useful constraint can be then employed, i.e. for any image point in one image, its corresponding point in the second image must lie on a line (termed epipolar line) associated with the original point. Consequently, this epipolar constraint greatly reduces the correspondence problem from a 2D-search over the whole image to a 1D-search along the epipolar line only: thereby, computational cost and ambiguities are reduced too. The discussion below will be focused only on two-view geometry; a more general constraint applicable to n-view geometries can be found in (Hartley and Zisserman, 2004).

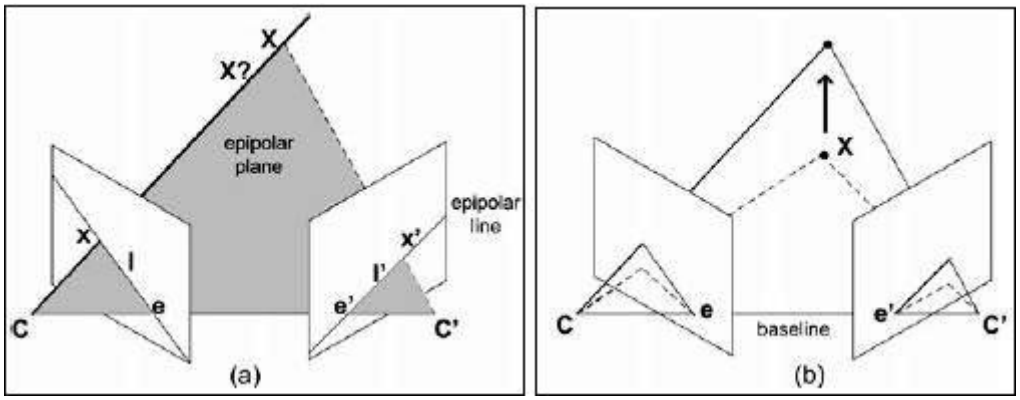


Figure 2.4 The epipolar geometry: relationship between two camera views (a); pencil of epipolar planes having the baseline as the pencil axis (b) (Se and Pears, 2012)

For any point **x** in the left image, the point **x'** in the right image is termed "corresponding point" if **x** and **x'** are images of the same physical scene point **X.** As Figure 2.4(a) shows, the image points **x** and **x'**, the object point **X** and the two camera centres **C** and **C'** are co-planar

and this plane, shaded in the figure, is termed the **epipolar plane**. The line **l'** is the intersection of the epipolar plane with the second image plane: this line, called epipolar line, is the image in the second view of the ray back-projected from **x**. This geometry immediately explained the previously mentioned epipolar constraint: if any point on epipolar line **l** has a corresponding point in the second image, it must lie on epipolar line **l'**, and vice-versa: thus, the correspondence search does not need to cover the entire image, but can be restricted only to the epipolar lines and **l** and **l'** are called conjugate **epipolar lines**.
The **epipole** is the point of intersection between the baseline (i.e. the line joining the two camera centres, whose length is the magnitude of the extrinsic translation vector **t**) and the image plane: in particular, the epipole **e** is the projection of the right camera centre on the left image, while the epipole **e'** is the projection of the left camera centre on the right image.
More in general, there is a pencil of planes having the baseline as the pencil axis, and a pencil of lines in each image given by all epipolar lines that intersect at the epipole of the respective image, as described in Figure 2.4(b). If this is the general two-view epipolar geometry, some degenerations can occur. When, for example, the cameras are oriented in the same direction and are separated by a translation parallel to both image planes, the epipoles are at infinity and the epipolar lines are parallel. Furthermore, if the translation is only along the X direction and the cameras have the same intrinsic parameters, the conjugate epipolar lines lie on the same image rows. The latter represents an ideal set up when the correspondence search should be performed, but it needs an image acquisition configuration that may conflict with the correspondence finding (Subsection 2.2.5) requirements: some camera convergence is in fact advisable in order to improve the field-of-view overlap between the two cameras. As Subsection 2.2.4 will discuss, image rectification will break into this loop: convergent images can be, in fact, warped so that the epipolar lines become horizontal again.

2.3.2 Essential and Fundamental matrices

The previously described epipolar constraint can also be represented algebraically by a 3x3 matrix, that is called **Fundamental Matrix** (F) in the more general case, i.e. when we are dealing with raw pixel coordinates. This matrix mathematically imposes a constraint on matching features in that they must lie on each other's epipolar line. When the internal calibration of the cameras is known, it is possible to upgrade this concept to that of an **Essential Matrix** (E), that deals with metrically expressed coordinates in the image plane.
Both the essential and fundamental matrices can be derived from a simple co-planarity constraint. In this section, the epipolar relation will be formulated using the essential matrix first; then, a relation involving pixel-based image coordinates will be derived, that will introduce the fundamental matrix.

In Figure 2.5, the object point **X** projects to image points $\mathbf{x_c}$ and $\mathbf{x_c'}$ and these two image plane points are metrically expressed by their homogeneous image coordinates in their own camera frame (subscript c). From the epipolar constraint, the three vectors $\mathbf{Cx_c}$, $\mathbf{C'x_c'}$ and **t** should be co-planar. Choosing the right frame as the reference one, this co-planarity can be expressed using the scalar triple product as follows:

$$\mathbf{x}_c'^{\mathrm{T}}(\mathbf{t} \times \mathrm{R}\mathbf{x}_c) = 0 \quad \rightarrow \quad \mathbf{x}_c'[\mathbf{t}]_x \mathrm{R}\mathbf{x}_c = 0 \qquad [2.17]$$

Where:

- $[\mathbf{t}]_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$ is the skew-symmetric matrix.

Thus, the essential matrix can be defined mathematically as:

$$\mathrm{E} = [\mathbf{t}]_x \mathrm{R} \qquad [2.18]$$

The relation [2.16] can so be rewritten as:

$$\mathbf{x}_c' \mathrm{E} \mathbf{x}_c = 0 \qquad [2.19]$$

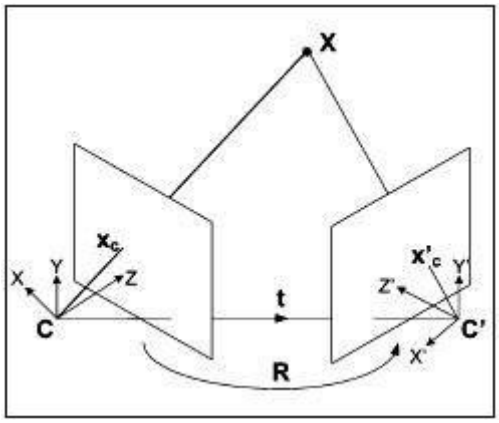


Figure 2.5 The epipolar geometry encoded by the essential and fundamental matrices (Se and Pears, 2012)

From equation [2.18], it is clear that the essential matrix is constructed only from the extrinsic parameters of the cameras, i.e. it encapsulates only the rotation and translation associated with the relative pose of the two cameras. Thus, in applications where R and **t** have not been computed by a calibration procedure, the may be directly derived from an estimation of E (Subsection 2.3.2).

In a more general case, it is necessary to deal with raw pixel coordinates, since the two cameras may generally be un-calibrated and, thus, their intrinsic parameters are unknown. The mapping between metric image coordinates and raw pixel values requires shifting and scaling operations, that can be expressed through the matrices K and K', containing the intrinsic parameters of the two cameras:

$$\mathbf{x} = \mathrm{K}\mathbf{x}_c \quad ; \quad \mathbf{x}' = \mathrm{K}'\mathbf{x}_c' \qquad [2.20]$$

If these relations are inserted into Equation [2.18], the following expressions can be derived:

$$\mathbf{x}'^{\mathbf{T}} \mathrm{K}'^{-\mathbf{T}} \mathrm{E} \mathrm{K}^{-1} \mathbf{x} = 0 \quad \rightarrow \quad \mathbf{x}'^{\mathbf{T}} \mathrm{F} \mathbf{x} = 0 \qquad [2.21]$$

where the following definition of F has been employed:

$$F = \mathrm{K}'^{-\mathbf{T}}\mathrm{E}\mathrm{K}^{-1} = \mathrm{K}'^{-\mathbf{T}}[\mathbf{t}]_{\mathrm{x}}\mathrm{R}\mathrm{K}^{-1} \quad [2.22]$$

Thus, the fundamental matrix encapsulates both intrinsic and extrinsic camera parameters.

Starting from the previously deduced mathematical expressions, it is possible to summarize some key properties of the fundamental matrix, such as:

- If F is the fundamental matrix between camera P and camera P', then F^{T} is the fundamental matrix between camera P' and camera P.
- The epipolar relation $\mathbf{x}'^{\mathbf{T}}\mathrm{F}\mathbf{x} = 0$ expresses the geometrical epipolar constraint, i.e. the observation that the x' in the second image, which correspond to x in the first image, should lie on the epipolar line $\mathbf{l}' = \mathrm{F}\mathbf{x}$. Symmetrically, $\mathbf{l} = \mathrm{F}^{T}\mathbf{x}'$ is the epipolar line in the first image corresponding to **x'** in the second image.
- F is a projective mapping that maps a point to a line. In fact, if **l** and **l'** are conjugate epipolar lines, then any image point **x** on **l** maps to the same line **l'**. Hence, there is no inverse mapping (F has zero determinant, i.e. it is rank 2).
- F has seven degree of freedom, even if a 3x3 homogeneous matrix has normally eight independent ratios. F, in fact, should satisfy an additional constraint, i.e. its determinant is zero (F is rank 2).
- The fundamental matrix F can be computed, up to a non-zero scalar factor, from the image data alone. Subsection 2.2.3 will explained different method for the computation of F.
- The epipoles are determined as the left and right null-spaces of the fundamental matrix. Thus, given F, the epipole e' in the second image is the unique 3-vector with third term equal to one, that should satisfy $\mathbf{e}'^{\mathrm{T}}\mathrm{F} = 0$. Similarly, the epipole e of the second camera in the first image is the unique 3-vector with third term equal to one, that should satisfy $\mathrm{F}\mathbf{e} = 0$.
- If the two cameras are identical (i.e. K=K') and they are separated by a pure translation (i.e. R=I), the fundamental matrix has a simpler form (Hartley and Zisserman, 2004):

$$\mathrm{F} = [\mathrm{K}\mathbf{t}]_{\mathrm{x}} = [\mathbf{e}']_{\mathrm{x}} = \begin{bmatrix} 0 & -\mathrm{e}'_{\mathrm{z}} & \mathrm{e}'_{\mathrm{y}} \\ \mathrm{e}'_{\mathrm{z}} & 0 & -\mathrm{e}'_{\mathrm{x}} \\ -\mathrm{e}'_{\mathrm{y}} & \mathrm{e}'_{\mathrm{x}} & 0 \end{bmatrix} \quad [2.23]$$

  and the epipoles are at the same location in both images. If the translation is parallel to the image plane, the epipoles are at infinity (i.e. $\mathrm{e}_{\mathrm{z}} = \mathrm{e}'_{\mathrm{z}} = 0$) and the epipolar lines are parallel in both images. If the translation is parallel to the camera X direction, the epipolar lines are parallel and horizontal, thus corresponding to the same image row. In this case, $\mathrm{e}_{\mathrm{z}} = \mathrm{e}'_{\mathrm{z}} = \mathrm{e}_{\mathrm{y}} = \mathrm{e}'_{\mathrm{y}} = 0$ and $\mathrm{e}_{\mathrm{x}} = \mathrm{e}'_{\mathrm{x}} = 1$, thus the fundamental matrix is:

$$\mathrm{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad [2.24]$$

The relationship between corresponding image points **x** and **x'** can be thus reduced to $y = y'$. These simplifications, already introduced at the end of Subsection 2.2.1, will be further discussed in Subsection 2.2.4.

2.3.3 Computation of the Fundamental Matrix

As observed at the end of the previous subsection, F can be computed from image correspondences alone: no camera calibration is needed and pixel coordinates are directly employed. Starting from Equation [2.20] and expanding it using

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad ; \quad \mathbf{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad ; \quad F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \qquad [2.25]$$

we obtain:

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0 \qquad [2.26]$$

Every corresponding point pair (**x**, **x'**) delivers one constraint on the nine components of the 3x3 matrix F; so, for n correspondences, the following set of linear equations can be derived:

$$\begin{bmatrix} x'_1x & x'_1y & x'_1 & y'_1x & y'_1y & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx & x'_ny & x'_n & y'_nx & y'_ny & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \mathbf{0} \qquad [2.27]$$

Equation [2.26] can be compactly expressed as:

$$A\mathbf{f} = \mathbf{0} \qquad [2.28]$$

where A is called the data matrix and **f** is the vector containing the unknown elements of F.

A very popular algorithm to solve Equation [2.18] is the **linear eight-point algorithm** (Longuet-Higgins, 1981). This method allows the linear solution of F using eight correspondences, that may be obtained by human interaction or using an automated seed point matching (Subsection 2.2.5). With eight correspondences, in fact, Equation [2.28] can be solved by linear methods and the solution is the null-space of A. More precisely, the system of equations is easily solved by Singular Value Decomposition (SVD), whose principles are detailed in (Golub and Van Loan, 1996). However, the matrix we obtain from **f** in this way may be inaccurate when the employed correspondences do not provide enough strong constraints, e.g. they are not well spread over the images or some of them are near-collinear or

co-planar. Thus, in the presence of noise, the solution for F may not satisfy the rank-2 constraint: this means that the epipolar lines will not intersect at a single point.

A possible solution to this problem is to exploit the fact that F has rank 2: using this constraint, the fundamental matrix F can even be computed, up to a non-zero scalar factor, from seven point correspondences between the images. However, since the 2-rank condition involves a relation between products of three entries on F, the algorithm is **non-linear** and the computation ultimately boils down to solving a non-linear equation. This approach requires to characterize the right null-space of the system of linear equations originating from the seven point correspondences and to impose the 2-rank condition (i.e. the determinant of F should be set equal to zero). A detailed explanation of this **seven-point algorithm** can be found in (Moons et al., 2008).

In general, it is always preferable to use more than the minimum needed amount of correspondences. In this case, it is possible to employ all these matches to compute F, by writing down the equation for each match and stacking all these equations in the matrix A. Then, the SVD of this matrix A allows the computation of F, that is the best solution in a least-squares approach. This method is based on a minimization of an algebraic error; however, the error that should be minimized is a geometric one, i.e. the distance between the points and the corresponding epipolar lines. Algebraically, the distance d between **x** (Equation [2.24]) and the line **l** (Equation [2.3]) is given by:

$$d^2 = \frac{(xl_1 + yl_2 + l_3)^2}{l_1^2 + l_2^2 + l_3^2} \qquad [2.29]$$

In order to minimize this geometric error, the so-called **Reweighted Least Squares** method can be efficiently employed: this approach is based on an iterative computation of F. The SVD of the matrix A computes a first estimate of F. Then, the residual distance for each correspondence is derived with Equation [2.29] and every row of A is divided by the residual computed in this way for its corresponding point match. Finally, F is re-computed by solving the SVD of the reweighted matrix A. This leads to a better estimation of F, since it doesn't give equal weight to every match: in fact, correspondences with small residuals will have more impact on the computation, whereas the worst correspondences producing higher distance errors will have a smaller impact.

Another approach that may be adopted to take into account multiple correspondences into the computation of F is a **non-linear refinement technique**. For example, the linear SVD approach can be used in order to produce the initial solution of a Levenberg-Marquardt algorithm (Hiebert, 1981), that uses a non-linear least squares minimization. It is an iterative technique, in which every iteration computes a new solution, that minimizes the error function. Some non-linear cost function may be minimized: a geometric cost function can, for example, be formulated as the sum of the squared distances computed from Equation [2.28]. This is averaged over both points in a correspondence and over all point matches. A more

detailed discussion of this non-linear refinement approach can be found in (Hartley and Zisserman, 2004).

All the above mentioned approaches are based on a SVD computation of A, that can only work robustly if the different columns of the matrix A have approximately the same magnitude. The third coordinate of each image point is always 1 and it is clearly typically much smaller the other two coordinates, that are usually in the order of 1000 or more. This means that the solution may be heavily biased. In order to solve this problem, it is essential to **normalize** the pixel coordinates of each image before applying SVD (Hartley, 1997). For example, a simple scaling of the 2D coordinates of the image points to the range [-1,1] can yield satisfactory results. Of course, the estimate of normalized F will map points to the epipolar lines in the normalized image space; thus, at the end a de-normalization of the fundamental matrix should be performed.

Both linear and non-linear methods above described are only effective if all input data are reliable. In general, however, the correspondence set between a pair of images includes not only inliers, but also outliers. In other words, incorrect matches, termed outliers, may pollute the set of equations building A, giving rise to a biased fundamental matrix. In order to deal with these outliers, a well-known algorithm can be used: given a correspondence data-set polluted by outliers, this algorithm can compute the epipolar geometry and at, the same time, it identifies the inliers and outliers in the data-set. This robust statistic technique is called Random Sample Consensus (RANSAC) and was proposed by (Fishler and Bolles, 1981). It can be summarized as follows:

1. Extract features in both images (Subsection 2.2.5);
2. Perform feature matching between images in order to compute a set of potential correspondences;
3. Repeat the following steps N times:
    a. select eight (or seven) putative correspondences randomly;
    b. compute F using these eight (or seven) points as above described;
    c. determine which of the remaining matches in the data-set are inliers for this F, i.e. the numbers of point matches that lie within some threshold of its expected position predicted by F.
4. Find the matrix F with the highest number of inliers among the N trials.

In other words, RANSAC uses an iterative scheme: in every iteration, it assumes that all seven or eight selected correspondences are inliers and searches for outliers among the remaining matches in the data-set. This iterative procedure is repeated until a stop-criterion has been reached. This solution with the largest number of inliers is finally selected as the true solution. The stop-criterion is expressed by N, i.e. the number of trials needed to get at least one set of correct matches (whose number is s) with a high probability (e.g. 99%). If s matches are selected from the data-set, the probability of having p correct matches is ε^p, where ε is the percentage of inliers in the date-set. Thus, if the process is performed N times, the probability of getting a good sample is given by:

$$P = 1 - (1 - \varepsilon^p)^N \quad [2.30]$$

The above equation can then be rewritten in order to isolate N, as follows:

$$N = \frac{\log(1 - P)}{\log\left(1 - (1 - \varepsilon^p)\right)} \quad [2.31]$$

where P stands for the certainty we wont to achieve (e.g. 99%).
This method can also be improved, to make it more stable and faster. For example, an adaptive RANSAC method can be adopted, where the numbers of outliers at each iteration is used to re-compute the total number N of iterations required. In addition to RANSAC, other alternative techniques that allow the detection of possible outliers in the observations are: Least Median of Squares (Rousseeuw and Leroy, 1987) and Maximum A Posteriori SAmple Consensus (MAPSAC) (Torr, 2002).

2.3.4 Rectification

As introduced in Sections 2.2.1 and 2.2.2, the epipolar constraint reduces the dense matching computational efforts tremendously, since it requires that each pixel in the first image needs to be compared only with pixels on the corresponding epipolar line in the second image. However, in general epipolar lines can lie in all directions, complicating the search for correspondences. To overcome this limit, it is possible to take the advantages given by standard rectilinear stereo rig, i.e. those configurations where the two cameras have parallel principal axes (no vergence) and identical intrinsic parameters. In this case, in fact, corresponding epipolar lines would lie along the same horizontal scanline in each image. In order to retain the favourable verged configuration (that improves the stereo viewing volume) and yet take the advantages associated with rectilinear rig, a pre-processing step can be applied. This operation is called **rectification** and is able to warp the raw images acquired by the verged system so that conjugate epipolar lines become collinear and lie on the same scanline. Afterwards, correspondences must only be searched along scanlines with the same y coordinates and hence they only differ in horizontal displacement, called **disparity** (Section 2.4.1).

Under the hypothesis of a **calibrated stereo rig**, both the intrinsic and extrinsic parameters are known. This information makes the rectification approach simpler, reducing it to finding an image mapping that generates, from the original images, a pair of images that would have been obtained from a rectilinear stereo. Thus, assuming some vergence, the required operation should map the image points onto a pair of virtual image planes that are parallel to the baseline and co-planar. An homographic structure can be employed to perform this task, i.e. to warp images between a pair of rotated views. The first step is the determination of the rotation matrices associated with the rectification of the left and right views. Assuming the optical centre of the left camera as the origin of the stereo system and (R,**t)** as the known

calibration information defining the rigid position of the right camera relative to the left one, the rectifying rotation matrix can be defined as follows:

$$R_{rect} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \quad [2.32]$$

Where the three mutually orthogonally unit vectors are defined as:

- $\mathbf{r}_1 = \frac{\mathbf{t}}{\|\mathbf{t}\|}$, that lies in the direction of the translation to the right camera, **t**
- $\mathbf{r}_2 = \frac{1}{\sqrt{t_x^2 + t_y^2}} \begin{bmatrix} -t_y \\ t_x \\ 0 \end{bmatrix}$, that is orthogonal to the first vector;
- $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$, that is mutually orthogonal to the first two vectors.

Finally, since the real right camera is rotated relative to the left camera, the rotation RR_{rect} should be applied to the image points of the right camera. Of course, as the rectified coordinates thereby obtained are in general not integer, a final resampling using some form of interpolation should be added.

Without calibration information, rectification can be performed using an estimate of the fundamental matrix, that is computed from correspondences within the raw image data. A common approach is detailed described in (Hartley, 1999; Mallon and Whelan, 2005) and is based on a pair of rectifying homographies computed for the left and the right images. This technique fails when the epipol lies within the image, that is a very common situation in Structure from Motion problems (Section 2.3.2), when the camera is moved along its Z direction. Several authors have tried to solve this problem, and examples of these solutions can be found in (Pollefeys et al., 1999; Pollefeys et al., 2004).

2.3.5 Finding correspondences

In the previous subsections, the epipolar constraint and its associated algebraic expression encapsulated into the fundamental matrix, have been proved to facilitate the search for correspondences: in fact, once F is known, the search for the corresponding point in one of the images can be restricted to the epipolar line in the other image. However, a reduced amount of correspondences is also needed to determine the epipolar geometry itself. In other words, there is something of a circular dependency here; luckily, the RANSAC sampling approach described earlier, represents a way to break into this loop. In fact, the search for correspondences can proceed in two subsequent stages. At first, a number of salient features are brought into correspondence: these seed correspondences are then employed to derive the fundamental matrix. Once F has been computed, a search for dense, i.e. per pixel, correspondences can be started based on the epipolar constraint, that reduces the search space from a 2D search to the epipolar line only.

In this subsection, some methods for finding correspondences in image pairs will be briefly described. Most of these approaches are based on the following two assumptions, that hold

when the distance between the object point and the two cameras is much larger that the baseline:

- Most scene points are visible from both viewpoints.
- Corresponding image regions are similar.

Starting from these hypotheses, two main classes of correspondence algorithms can be described: they represent two different solutions to the main involved issues, i.e. which image element is suitable to match and what good similarity measure can be adopted. The following general overview aims at presenting the fundamental principles of these two approaches.

- **Correlation-Based Methods** represent the more traditional approach. They can recover dense correspondences on the base of the continuity assumption, which asserts that, at the resolution level where image matching is performed, most of the image window depicts a portion of a continuous and planar surface element (Remondino et al., 2008). Therefore, adjacent pixels in the image window will generally represent contiguous points in the object space. With this approach, the element to be matched is the centre of a small window of pixels in a reference image, that is statistically compared with equally sized windows of pixels in another image. The similarity criterion is a measure of the correlation between the two windows: in particular, a correspondence is given by the window that maximizes a similarity criterion or minimizes a dissimilarity criterion within a search range. Afterwards, once a match is found, the offset between the two corresponding windows is computed and called disparity.
  The window function, that moves on the rectified images and has a size equal to m (odd integer), can be mathematically expressed as follows:

$$W_m(x,y) = \left\{(u,v) \middle| \, x - \frac{(m-1)}{2} \leq u \leq x + \frac{(m-1)}{2}, y - \frac{(m-1)}{2} \leq v \leq y + \frac{(m-1)}{2}\right\} \quad [2.33]$$

  A dissimilarity criterion can be, for example, formulated by the Sum of Squared Differences (**SSD**) cost, that computes the intensity difference as a function of the disparity d:

$$SSD(x,y,d) = \sum_{(u,v)\in W_m(x,y)} [I_l(u,v) - I_r(u-d,v)]^2 \quad [2.34]$$

  where I_l and I_r are the intensities of the left and right images respectively. Thus, for each pixel in the left image, correlation-based methods would compare the SSD measure for pixels that lie within a search range along the corresponding epipolar line in the right image. The best match will be the one indicated by the disparity value giving the lowest SSD.
  A slight different dissimilarity criterion is represented by the Sum of Absolute Difference (**SAD**), where the cost measure to be minimized is expressed by:

$$SAD(x,y,d) = \sum_{(u,v)\in W_m(x,y)} |I_l(u,v) - I_r(u-d,v)| \quad [2.35]$$

In this case, the required computational efforts are less expensive; however, the first approach penalizes more the large intensity difference, thanks to the squaring operation.
Both the two previously mentioned approaches are based on intensity measures, that can present huge variations caused by illumination changes and non-Lambertian reflection. In other words, SSD and SAD may not give a low (i.e. correct) value, when non-Lambertian reflection or differences in the gain and sensitivity cause intensity variations. In order to solve this problem, a normalization process of pixels within each window can be performed. For example, the intensities in each window can be forced to be zero-mean; then, these zero-mean intensities can be scaled so that they either have the same range or, preferably, unit variance. This normalization requires that, after the zero-mean operation, each pixel intensity is divided by the standard deviation of all window pixel intensities. Thus, if $\bar{I}$ is the mean intensity and σ_I is the standard deviation of window intensities, normalized pixel intensity is given by:

$$I_n = \frac{I - \bar{I}}{\sigma_I} \qquad [2.36]$$

A different approach to the problem of correlation computation is offered by the Normalized Cross-Correlation (**NCC**) method, that uses a similarity criterion and not a dissimilarity one. The first step is, again, the pixel normalization, that is performed by subtracting the average intensity of the window so that only the relative variation would be considered in the correlation procedure. Then, the NCC measure is computed as follows:

$$NCC(x, y, d) = \frac{\sum_{(u,v)\in W_m(x,y)} (I_l(u,v) - \overline{I_l}) \times (I_r(u,v) - \overline{I_r})}{\sqrt{\sum_{(u,v)\in W_m(x,y)} (I_l(u,v) - \overline{I_l})^2 \times (I_r(u,v) - \overline{I_r})^2}} \qquad [2.37]$$

Where:

- $\overline{I_l} = \frac{1}{m^2} \sum_{(u,v)\in W_m(x,y)} I_l(u,v)$;
- $\overline{I_r} = \frac{1}{m^2} \sum_{(u,v)\in W_m(x,y)} I_r(u,v)$.

All the above mentioned methods have a very high accuracy potential if well-textured image windows are used. Disadvantages of correlation-based approaches are the need for small searching range for successful matching and the huge amount of data that should be handled. In particular, problems occur in area with occlusions, area with a lack of texture or with a repetitive one, and if the surface does not respect the continuity assumption (Remondino et al., 2008).

- **Feature-Based Methods** usually can recover only sparse image correspondences, also termed seed correspondences or interest points. In this case, the element to be matched is an image feature and the similarity measure is the distance between descriptors of these image features. In other words, feature-based methods require a two-stage procedure: at first, interesting features are detected and associated with feature descriptors for their

representation; then, the corresponding features are determined using similarity measures involving the feature descriptors (for example, the Euclidian distance between the descriptor elements can be computed). Of course, the two steps are connected to each other, since the feature extraction and descriptor computation should be such that the final correspondence determination is easy, accurate and robust.
Starting from the first stage, three different types of features can be detected, i.e.:

- Interest Points (Schmid et al., 2000; Mikolajczyk and Schmid, 2005); these features can be extracted applying contour-based methods, signal-based methods or methods based on template fitting. Point-based features have been usually preferred by researchers thanks to their advantages, e.g. accuracy, stability, sensitivity, controllability and speed. An example of these features, well-localized in two mutually orthogonal direction, are represented by corners; thus, many corner detectors have been presented in literature, such as the Harris corner detector (Harris and Stephens, 1988) and the corner detector SUSAN (Smith and Brady, 1997).
- Edges (Schmid and Zisserman, 2000); in this case, the property used for feature extraction is the intensity change, which is represented via the gradient of the image. Thus, edge detectors usually require the same steps, i.e.: smoothing, applying edge enhancement filters applying a threshold and edge tracing (Remondino et al., 2008). A problem connected with linear features is that the matching can be poorly localized along the length of a line, especially if the linear feature is fragmented.
- Regions (Mikolajczyk et al., 2005); these features are represented by homogeneous areas of the images with intensity variations below a certain threshold.

Feature-based detectors should deal with wide baseline configurations, that, as previously mentioned, improve the stereo viewing volume, but, on the other hand, produce large geometric and photometric variations between the images. In order to operate successfully over these difficult conditions, many interest point detection algorithms have been proposed in recent years, that are not sensitive to these variations. These methods can be used even when epipolar geometry is not yet known, since their descriptors allow correspondences to be efficiently searched over the whole image. An example of these approaches is the Scale Invariant Feature Transform (**SIFT**), that provides highly distinctive features that are invariant to image translation, scaling and rotation and partially invariant to illumination changes and affine or 3D projections. The operator acts in four stages, i.e.: scale-space extreme detection, key point localization, orientation assignment and key point descriptor creation. The last step assigns to the extracted features a local image vector with 128 elements. A detailed discussion on this approach can be found in (Lowe, 1999 and 2004). The second most popular feature type developed for image feature generation is the Speeded-Up Robust Feature (**SURF**), that gives comparable results to SIFT while requiring a lower computational time. This method uses a Hessian matrix-based measure for the detector and the distribution of the first-order Haar wavelet responses for the descriptor. (Bay et al., 2008) includes a comprehensive explanation of this topic.

2.4 The 3D reconstruction problem

In the previous section, the correspondence problem has been addressed using fundamental notions such as the epipolar relation and the fundamental matrix. Now it is necessary to answer to a second question, i.e.: given two corresponding points **x** and **x'**, how can the 3D position of the object point **X** be computed? This is called the **3D reconstruction** problem and aims at recovering the geometric structure of a scene from two or more of its images.
Depending on what is still known about the camera setup, the geometric uncertainty of 3D reconstruction can be very different; thus, different types of 3D reconstruction can be obtained based on the amount of *a-priori* knowledge available. Figure 2.5 and Table 2.1 summarize the different possible approaches, that are here briefly discussed. The simple case of two-view configuration will be treated; a more detailed discussion can be found in (Moons et al., 2008).

- **Euclidean 3D reconstruction** is the simplest method to recover 3D information from images and requires that the intrinsic and extrinsic parameters of the cameras are known. In other words, it refers to the situation where the position and orientation of the second camera relative to first one are known and both cameras are internally calibrated. On the other hand, camera position and orientation relative to an absolute world coordinate frame are generally unknown. In this case, from projective equations [2.14] is it possible to recover the 3D structure of the scene up to a 3D Euclidian transformation of the scene itself; thus, the actual metric dimensions of the 3D object can, for example, be determined. Of course, it is impossible to recover absolute information about the camera external parameters in the absolute world coordinate frame.

- **Metric 3D reconstruction** refers to the same cameras setup described in the previous point, but in this case let's suppose that the distance between the cameras is unknown. So, this approach is based on the knowledge of camera relative orientation and of the direction along which the second camera is translated with respect to the first one. Again, both cameras are supposed to be internally calibrated. In this case, the 3D reconstruction of the object point **X** will then be recovered up to un unknown scalar factor, that does not depend on the scene point **X** and is a constant value. This approach brings the geometry uncertainty about the 3D scene up to un unknown 3D similarity transformation, that can also be explained intuitively: if a 3D scene is scaled together with the cameras in it, then no impact on the images will follow. In other terms, this would only change the distance between the cameras, and not their relative orientation, shift direction and internal calibration. Finally, the issue of the unknown overall scale of the 3D scene is not worrisome in practice: the knowledge of a single distance or length in the scene, in fact, is enough to lift this uncertainty about scale.

- **Affine 3D reconstruction** abandons also the hypothesis of known internal camera calibration: from a geometric point of view, this implies that the calibration matrices are unknown and cannot be used to convert the digital images into pure perspective

projections of the scene. On the other hand, it is possible to prove (Moons et al., 2008) that if the vanishing points of three independent directions of the scene can be identify in the images, then a system of affine reconstruction equations for the scene can be computed. This statement is based on the definition of vanishing points in the images, that represents the perspective projections onto the image planes of points on the plane at infinity in the scene. For example, the vanishing points of the line connecting the two camera centres (i.e. the baseline) are the intersection of the line with each image plane, that is just the epipoles.

- **Projective 3D reconstruction** represents the final case, in which no knowledge about the camera configuration or about the scene is available. Of course, the possibility of extracting images correspondences is still valid, in order to compute the fundamental matrix. In this case, typical of Structure from Motion applications (Section 2.3.2), only a projective 3D reconstruction of the scene can be achieved. This means that the 3D structure is known only up to an arbitrary projective transformation: thus, it is, for example, possible to know how many planar faces the object has and which point features are collinear, however nothing about the scene dimensions and angular measurements can be said.

***A-priori* knowledge**	**3D reconstruction type**
• Intrinsic and extrinsic parameters	Euclidian 3D reconstruction
• Intrinsic parameters • Cameras relative orientation and shift direction (no shift magnitude)	Metric 3D reconstruction
• Vanishing points of three independent directions of the scene	Affine 3D reconstruction
• No available information (only correspondences)	Projective 3D reconstruction

Table 2.1 Different types of 3D reconstruction

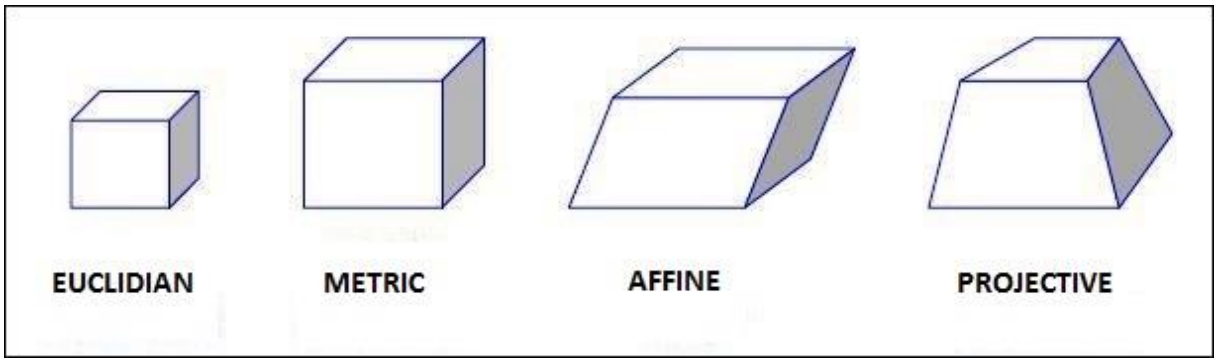


Figure 2.6 Different types of 3D reconstruction
(Moons et al., 2008)

Summing up, Figure 2.6 shows the effects of the different 3D reconstruction methods in dealing with a cube. In an Euclidian reconstruction, the original object is found, whereas in a metric reconstruction the actual size of the recovered cube is undetermined. The affine reconstruction is able to produce a parallelepiped, since affine transformations preserve parallelism, but they do not preserve metric relations such as lengths and angles. Finally, in a projective transformation the scene is reconstructed as an irregular hexahedron: these kind of transformations, in fact, only preserve incidence relations such as collinearity and coplanarity. The above described system of mutual relations between the different kind of 3D reconstruction is usually termed **stratification** of the geometries: in this scheme, in fact, the transformations higher in the list can be treated as special types of the transformation lower down. Of course, the uncertainty related to the 3D reconstruction goes up when going down the list, but it is not "fixed": each kind of reconstruction can be, in fact, upgraded to an upper level. For example, a projective reconstruction can be upgraded into an affine one if parallel lines in three independent directions are known. Moreover, a projective reconstruction can also become a metric one with a sufficient amount of scene metric information (e.g. known lengths or distances, known angles or orthogonality relations between scene structures).

As stated in Chapter 1, in multiple-view approaches the 3D scene is observed from two or more viewpoints, by either multiple cameras at the same time (stereo vision) or a single moving camera at different time (structure from motion). The following two subsections will address the problem of 3D reconstruction in these two different cases.

2.4.1 Stereo

Stereo vision refers to the process of recovering information on the 3D structure of a scene from two or more images taken from different viewpoints. Generally speaking, a disparity map can be computed by the disparities of all the image points; then, if the stereo system is calibrated, this disparity map can also be converted into a 3D point cloud. The discussion here will be focused only on two-view systems, whereas multiple-view stereo methods will be discussed in Chapter 3.

Image matching is, in general, defined as the "establishment of correspondences between two or more images to reconstruct surface in 3D" (Remondino et al., 2008). Starting from this generic definition, it is simple to derive the aim of **dense stereo matching**, i.e. the computation of as many pixel matches as possible between the two images of a stereo pair. In other words, in order to achieve a dense 3D point cloud, the disparity values for all the image points should be computed. Since some interest points are not sufficient, correlation-based methods should be used here, employing different criteria for matching, such as SSD, SAD and NCC. However, this approach can have some difficulties in finding a unique match (multiple possible correct matches) or no match at all and has unpredictable results if different pixels are similar. Fortunately, these ill-posed problems can be converted into well-posed ones by introducing constraints. Typical **constraints** that can make the matching more robust, are:

- Epipolar constraint, that reduces the search from 2D to the epipolar line only.
- Uniqueness constraint, based on the assumption that a 3D point cannot be projected to two 2D points in the same image.
- Ordering constraint, meaning that, if the 3D scene consists of piecewise planar surfaces, the ordering of pixels along the epipolar lines should be preserved.
- Disparity range constraint, that limits the disparity search range according to the prior information of the expected scene.
- Smoothness constraint, that reduces the problem of low-textured surfaces in the image. This constraint enforces the disparity to evolve smoothly along the epipolar line, except at sudden changes (e.g. depth discontinuities and edges).

Dense image matching methodologies have been generally categorized into local and global (Brown et al., 2003). **Local methods** require a suitable choice of window size and are very sensitive to locally ambiguous areas in the images (e.g. occluded area or poorly textured area). Concerning the first issue, the window size m of Equation [2.33] should be chosen as a trade-off between two opposing effects: while using a larger window size provides more intensity variation and more context for matching, on the other hand this can cause problems around the occluded area and at object boundaries. Thus, a small window size will produce a more detailed disparity map, but it will be also more noisy. On the other hand, increasing the window size will infer cleaner disparity maps, but not well-defined around the boundaries.

In order to improve the stereo matching quality by using information outside the local window regions, **global methods** have been introduced. These approaches perform a global search which searches for the optimal global solution to the matching problem. In particular, they are usually formulated as an energy minimization problem and the dynamic programming algorithm (Van Meerbergen et al., 2002) is a common example of this approach. After rectification, all pixels of one scanline are compared to the pixels on the corresponding scanline, using NCC. The goal of the stereo algorithm is to search for the optimal path inside the extracted matching matrix: in order to determine this optimal solution, a cost function is assigned to every path. The stereo algorithm will then select the cheapest path as the optimal solution. Other examples of global methods are graph cuts, max flow (Roy and Cox, 1998) and belief propagation (Sun et al., 2003): all these approaches produce better disparity maps than local methods, but they are very computationally expensive.

The complexity of the above described algorithm, in fact, is $\mathcal{O}(\mathrm{WH}\delta^2)$, where W and H are the width and the height of the image, whereas δ is the disparity range. Thus, large disparity ranges can have a significant bad effect on the execution time. To reduce this limit, stereo algorithms are generally executed in a hierarchical way. After rectification, images are down-sampled a number of times into a pyramid (Bergen et al., 1992). The optimal path is searched for at the smallest level, and the solution is used as an initialization for the next level. By executing the algorithm for all levels, the optimal path at the last level can be computed with a complexity of $\mathcal{O}(\mathrm{WH})$. Another advantage of this coarse-to-fine hierarchical approach is the possibility to easily deal with large disparity range, since a narrower disparity range can be used for the original image.

Recently, **semi-global methods** (Hirschmüller, 2005) have also be introduced: these approaches use an approximation of the global model to provide an efficient solution. Thus, optimisation is only approximated (hence, "semi"), and the general procedure can be divided into four steps: pixel-wise cost computation, cost aggregation, disparity calculation and disparity refinement.

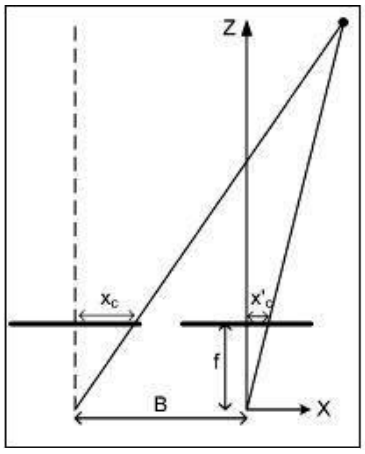


Figure 2.7 The stereo geometry after image rectification (Se and Pears, 2012)

Moving to a more geometrical point of view, it is now possible to understand how 3D world points can be computed once disparity values have been calculated. Generally speaking, after the extraction of the corresponding left and right image points, the **triangulation** process allows the back-projection of the two rays from the camera centres through those left and right image points. Considering two rectified views (Figure 2.7) and assuming intrinsic and extrinsic parameters as known (e.g. computed via a calibration procedure), the following two relations can be derived:

$$x'_c = f\frac{X}{Z} \quad ; \quad x_c = f\frac{X+B}{Z} \qquad [2.38]$$

where:

- (x'_c, x_c) are the corresponding horizontal image coordinates (expressed in metric units);
- f is the focal length;
- B is the baseline.

Since the disparity d is defined as the difference between the horizontal image coordinates of corresponding left and right image points,

$$d = x_c - x'_c = \frac{fB}{Z} \qquad [2.39]$$

the 3D coordinates of the object point can be computed as follows:

$$Z = \frac{fB}{d} \quad ; \quad X = \frac{Zx'_c}{f} \quad ; \quad Y = \frac{Zy'_c}{f} \qquad [2.40]$$

where y'_c is the vertical image coordinate in the right image.

Thus, through Equations [2.39], disparity maps can be converted into depth maps to generate a 3D point cloud.

Finally, by computing the derivatives of Equations [2.40], the standard deviation of the depth measurement can be expressed as follows:

$$\Delta Z = \frac{Z^2}{Bf}\Delta d \qquad [2.41]$$

where Δd is the standard deviation of the disparity. This equation shows which parameters can have a direct influence on the depth uncertainty. In particular, depth uncertainty increases quadratically with the camera-object distance: thus, stereo systems are usually employed within a limited depth range. Furthermore, Equation [2.41] proves that the depth error can be reduced by increasing the baseline, the focal length or image resolution. Unfortunately, each of these actions has also a negative effect. Increasing the baseline, for example, makes the stereo matching harder and causes more occluded area; increasing the focal length reduces the lens depth of field and, finally, processing time will grow up if image resolution is higher. Thus, concluding, the design of stereo cameras should involve a range of performance trade-offs, that must be selected in accordance with the application requirements.

2.4.2 Structure from motion

Structure from Motion (SfM) refers to the situation where images are captured by a single moving camera. The process requires the simultaneous recovery of both the camera relative poses and 3D reconstruction of the scene. Assuming a static scene, i.e. without any moving objects, a SfM application can be divided into two sub-problems:

- The correspondence problem, that refers to the possibility of recovering which elements of an image frame correspond to which elements of the next one;
- The camera motion and 3D reconstruction problem, that refers to the determination of camera motion (sometimes termed ego-motion) and the 3D structure of the observed scene.

If the hypothesis of a static scene is not valid and some objects in the scene may have moved between frames, a third problem should be added, i.e. the problem of segmentation. This is a relatively recent problem in SfM applications and requires the extraction of regions corresponding to one or more moving objects: in other words, features belonging to moving objects should be identified and removed as outliers. In the following discussion dynamic scenes are not dealt with and the assumption of a static scene will be made.

The two above mentioned issues can be addressed with two different approaches. The first one requires initially the computation of the fundamental matrix, that can be recovered by matching at least eight corresponding features between frames. F can then be directly use to achieve a projective reconstruction of the scene, without any knowledge about camera calibration parameters. In most cases, however, a metric reconstruction is at least required, in order to reconstruct a 3D scaled version of the real scene. In order to upgrade a projective transformation into a metric one, camera intrinsic parameters should be estimated: a self-calibration technique can be efficiently used to perform this task. Once the intrinsic camera parameters are known, the essential matrix can be derived from the fundamental one and, then, Equation [2.18] can be used to recover the camera motion. In particular, SVD can be applied to extract $\mathbf{t}$ and R from E, following the approach explained in (Hartley and Zisserman, 2004), where $\mathbf{t}$ is determined only up to a scale factor. Finally, once the camera extrinsic parameters have been extracted, a sparse scene structure can be recovered by

computing the intersection between the back-projected rays. Of course, in this way the scene structure will be determined only up to a scale factor, but it can be upgraded to an Euclidean transformation if some measurement is known in the scene.

A different approach to the SfM problems is offered by the Bundle Adjustment (Triggs et al., 1999) method, that appeared several decades ago in the photogrammetric literature and is now widely used also in the computer vision community. This technique is illustrated in Figure 2.8 and offers a more accurate method that simultaneously optimizes the 3D structure and the extrinsic camera parameter set for each frame in the sequence.

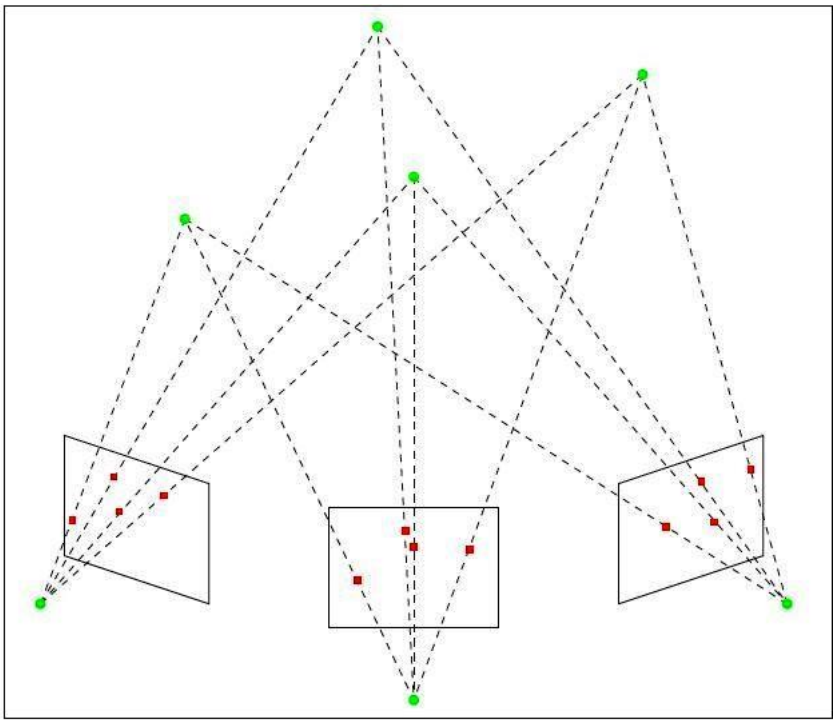


Figure 2.8 The process of Bundle Adjustment
(Moons et al., 2008)

If a simple setup of three images and four object points is considered, let's suppose of having already recovered both the camera and the 3D point positions. In an ideal case, the re-projection of the computed 3D points into the images should perfectly coincide with the extracted feature points. In general, due to measurement noise, deviations from this ideal solution will occur. In other words, a re-projection error should be considered and defined as the Euclidean distance between an image feature and its re-projection into the image plane starting from its 3D computed position and the extracted camera poses. Bundle Adjustment aims at minimizing this re-projection error by iteratively adapting the position of the 3D points and the extrinsic parameters of the cameras. In other words, it is a batch process, that iteratively refines the camera parameters and the 3D structure (hence, more in general, the bundles of projection rays) in order to minimize the sum of the re-projection errors. Since a specific re-projection error is only dependent on its own scene point and viewpoint, the structure of the equations is sparse and sparse linear algebra algorithms can be applied to solve it. That is way this approach is usually called Sparse Bundle Adjustment.

3. 3D MODELLING FROM IMAGES: ALGORITHMIC AND SOFTWARE SOLUTIONS

3.1 State of the art, algorithmic solutions

Recently, two significant improvement processes have produced remarkable developments in the field of passive 3D vision systems. First of all, the tremendous emergence of cheap and high quality digital cameras has amplified the phenomenon of contemporary multi-media society, that is based on multi-source information including text, audio, images, video and interactive content. Among these different forms, images have always played a hugely important role, since they include and transmit a huge amount of information in a compact, intuitive and visually appealing format. Thus, recently advances in both hardware and software technologies have increased the power of images: for example, digital cameras are now embedded in many devices, such as smartphones, and the recent explosion in social networking allows more and more people to share images. Moreover, the last four decades have also seen the rapid evolution of computational power in personnel computers and the development of a GPU-based approach: many algorithms are now highly suitable to run on Graphics Processing Units (GPUs), in order to free up the Central Processing Unit (CPU) for other tasks. GPUs, in fact, have now a parallel throughput architecture that supports executing many concurrent processes, proving an immense speed-up for parallelized algorithms.
These enhancements have led, in the past decade, to a very active community of research in both photogrammetry and computer vision: the dream of modelling the world in 3D at scale one using only images seems now, in fact, almost affordable, at least from an algorithmic point of view.

However, until the 2000 year, these two communities, i.e. the photogrammetric and CV ones, have always worked almost independently on the problem of 3D reconstruction starting from a set of un-oriented images (Förstner, 2009). This division and dual-approach is mainly due to the different goals that constitute the core of these two disciplines (Hartley and Mundy, 1993).
The field of **computer vision** has evolved under the central theme of achieving a human-level capability in the extraction of information from image data. Generally speaking, the goal of computer vision is to automate the analysis of image through the use of computers; this general principle is then applied to three main issues: object recognition (i.e. the process of arriving at the same class for an object as that defined by the human conceptual framework), navigation (i.e. the function to provide guidance to an autonomous vehicle) and object 3D modelling (i.e. the recovery of 3D structure of a scene).
On the other hand, the central theme of **photogrammetry** is geometric accuracy. Photogrammetry is a much older discipline, whose development started almost at the same time as the discovery of photography itself. The photogrammetric approach is based on the collinearity equations, that model the perspective transformation between image and object spaces (Kraus, 1994 and 1997); two main application fields are generally covered: the

production of topographic maps and close-range applications (e.g. architecture, anthropometrics, industrial metrology and archaeological surveying).

Thus, despite the conceptual differences, an intersection of interests for these two fields can be found; in particular, both research communities are dealing with problems related to camera calibration, pose determination and model 3D reconstruction. On these issues and until the end of last century, the photogrammetrists were mainly working on well calibrated images for metrological applications, using bundle adjustment in Euclidean space and processes requiring a lot of human interaction (Kasser and Egels, 2002). During the same time, the computer vision community was working on low resolution images for real-times automated applications, focusing on relative pose computation in a projective space (Faugeras, 1993). Since the begun of 2000 years, the advantages coming from a possible integration between the two approaches have started to be contemplated: by taking the best of both photogrammetry and computer vision, in fact, fully automatic and accurate tools could be designed, allowing 3D modelling of complex environments from images acquired by cheap amateur cameras.

This innovative idea was also enforced through several algorithmic improvements that have been efficiently produced during last decade.
First of all, given two partially overlapping images, the problem of automatically detecting a reliable and stable set of tie points should be addressed: the introduction of SIFT and SURF tools (Subsection 2.2.5) has introduced this possibility for the first time. Secondly, a solution to the problem of automatic orienting large blocks of images using only tie points should be found: in this case, the development of software solutions, such as the one in (Snavely et al., 2008), represented the breakthrough, since they merged for the first time the already known necessary algorithmic components. The last step that had to be improved was the dense automatic matching of oriented images: multi correlation techniques and optimization tools (Section 2.3) represented the necessary means to efficiently perform this task.

All these enhancements have recently led to an impressive development of image-based automated procedures and software for detailed 3D modelling; a general overview of these operational solutions will be given in Section 3.2. Since their final step of dense image matching still represents a very hot topic, a brief description of multiple-view stereo reconstruction algorithms will be discussed earlier, in the next subsection.

3.1.1 Multiple-view stereo reconstruction algorithms

In Chapter 2 (Subsection 2.3.1) dense stereo matching through two-view systems has been discussed; here, on the contrary, a general survey of multiple-view stereo methods will be briefly discussed, focusing on algorithms that reconstruct dense object models from calibrated views. More detailed studies on this topic can be found in (Koch et al., 1998; Dyer, 2001; Slabaugh et al., 2001; Seitz et al., 2006).

The aim of multi-view stereo is to recover the 3D complete structure of a scene, starting from a collection of images taken from multiple and known camera viewpoints. Available techniques try to perform this task with different assumptions, operating ranges and approaches. Before giving an overall description of several recently developed algorithms, six general fundamental properties of them will be described below. The taxonomy presented in (Seitz et al., 2006) will be used, in order to differentiate these key properties.

- **Scene representation**. Multi-view algorithms can use different kinds of representation to model the geometry of an object; some techniques can also employ different representations for various steps in the procedural pipeline. Most approaches use a regularly sampled 3D grid, that represents a very common geometry thanks to its advantages, i.e. simplicity, uniformity and ability to approximate almost any surface. Other techniques represent the surface as a set of connected planar facets, that are called polygon meshes: they represent a very efficient means of storing and rendering, thus they are a very popular output format for many multi-view algorithms. Finally, some methods describe the geometry of the scene as a set of depth maps, one for each point of view; these 2D representations are especially useful for small datasets and avoid the geometry resampling on a volumetric domain.
- **Photo-consistency measure**. Different measures can be used in order to evaluate the visual compatibility of a reconstruction with a set of input images. Among them, photo-consistency measures (Kutulakos and Seits, 2000) compare pixels in one image to pixels in other images to see how well they correlate. These measures can be defined in scene (i.e. object) space or in image space. In the former case, a point, patch or volume of the reconstructed geometry is back-projected into the input images: the amount of mutual agreement between these projections is then evaluated. For example, the variance of the projected pixels in the input images is used as a mutual agreement measure. Image space approaches, on the other hand, measure the photo-consistency through a prediction error. An image is first warped from a viewpoint to predict a different view, using an estimate of scene geometry; then, these two images, i.e. the measured image and the predicted one, are compared.
- **Visibility model**. Modern algorithms should model visibility, considering also occlusions. In other terms, visibility models specify which views should be considered when photo-consistency measures are computed. One possible approach, termed geometric technique, tries to model the image formation process and the scene shape in order to determine which scene structures are visible in which images. Quasi-geometric techniques use an approximate geometric approach to determine visibility relationships: for example, they limit the photo-consistency analysis to clusters of nearby cameras. Finally, a third approach avoids explicit geometric reasoning and treats occlusions as outliers: in this case, simple outlier rejection techniques (Stewart, 1999) can be employed in order to select the good views.

- **Shape prior**. Some methods impose shape priors that bias the reconstruction to have desired geometric characteristics. Some techniques, for example, seek minimal surfaces with a small overall surface area; other methods, on the contrary, prefer maximal surfaces. Finally, rather than impose global priors on the overall size of the surface, other approaches use shape priors that encourage local smoothness.
- **Reconstruction algorithm**. Four main classes of reconstruction algorithms can be identified. The first one uses a two-stage approach: it first computes a cost function on a 3D volume, then a surface is extracted from this volume. A second kind of technique employs an iterative approach, in which the surface is step by step adjusted, in order to minimize a cost function. A third possibility is represented by those methods that reconstruct the geometry of the scene by computing a set of depth maps. Finally, a fourth class of algorithms includes those techniques that initially extract and match a set of feature points and then fit a surface to them.
- **Initialization requirements**. In order to eliminate trivial shapes, almost all multi-view reconstruction algorithms require some input information about the geometric extent of the acquired scene. Some techniques only require a rough bounding box, whereas others need a foreground/background segmentation for each input image. Finally, a third kind of constraint can be represented by an allowable range of disparity or depth values: in this case, typical of image-space algorithms, scene geometry should lie within a near and a far depth plane for each camera viewpoint.

After having presented the main characteristics according to which multi-view stereo reconstruction algorithms can be categorized, six of these algorithmic solutions are here briefly described. A qualitative comparison and a quantitative evaluation of them can be found in (Seitz et al., 2006). (Ahmadabadian et al., 2013) presents a comparison of dense image matching algorithms for scaled surface reconstruction using stereo camera rigs. Furthermore, the European Spatial Data Research Organization (EuroSDR) started a benchmark on image based DSM generation using aerial images in February 2013. Some of the results provided from different research and commercial groups are presented in (Haala, 2013), where they are used to evaluate the potential of these image-based procedures. Finally, the Middlebury Stereo Vision Page (Scharstein and Szeliski, 2002) is another well-known example of benchmark, providing useful datasets and evaluation tests on the performance of state-of-the-art matching algorithms.

Among the recently published algorithms, the following ones can be mentioned (Seitz et al., 2006):

- (Furukawa and Ponce, 2006) uses wide baseline stereo matching in order to recover the 3D coordinates of seed feature points. Then it shrinks a visual hull model so that the recovered points lie on its surface; the final result is so refined using an energy function minimization.

- In (Goesele et al., 2006) a depth map is computed from each camera viewpoint and finally results are merged.
- (Hernandez and Schmitt, 2004) first extracts a depth map from each viewpoint and merges the results into a cost volume. The computed mesh is then iteratively deformed to find a minimum cost surface in a visual hull volume.
- (Kolmogorow and Zabih, 2002) uses graph cut techniques to compute a set of depth maps from multi-baseline stereo; then the results are merged into a voxel volume by computing the intersections of the occluded volumes from each viewpoint.
- (Pons et al., 2005) extracts a minimum cost surface by adjusting a surface using a prediction-error measure.
- (Vogiatzis et al., 2005) computes a cost volume in the neighbourhood of the visual hull; then a minimum cost surface is extracted using volumetric min-cut techniques.

3.2 State of the art, software solutions

After having presented a general review of the state of the art from an algorithmic point of view, the most common software solutions are described below. Software tools for image-based 3D point cloud generation are currently developed by a number of both research institutes and photogrammetric software vendors. Thus, the field of image data processing is changing at a great rate and it is difficult to give a complete and exhaustive overview of all the possible existing approaches. So, in order to cover the current state of the art, the most recently published software solutions are here presented; among them, two suites of tools will be described in detail in the following part of this research thesis, since they have been efficiently evaluated from a metrical point of view. A list of websites providing more details of all the following software solutions is reported at the end of the Reference Chapter.

1. **Commercial software**. In the following list, some of the most common tools distributed by software vendors are presented. Since they are nor open-source nor free of charge, the most practical drawback of these solutions is the necessity of paying a license to install and use them; moreover, the user has no control on them and it may be an hard task to adapting their processes to specific requirements and data-sets. On the other hand, they are usually very user-friendly and detailed user guides and manuals are provided together with them.
    - Agisoft PhotoScan is a multi-view 3D reconstruction software that allows for precise textured mesh model reconstruction. Even though the user can set a large number of input parameters, the reconstruction itself is a painless four-stage procedure. First of all, the digital camera calibration should be performed and the possibility of using a provided planar chessboard pattern is given to the user. PhotoScan also enables the use of available metric camera information, since the input of custom calibration

parameters is supported. Self-calibration through CV-based techniques is provided as well. Secondly, camera poses and orientations should be computed, through a feature detection and matching phase, followed by a bundle adjustment refinement. A dense 3D reconstruction phase is then applied and several dense stereo-matching algorithms are supported. The surface of the reconstructed object is described through a mesh geometry, on which a photo-texture is finally projected. Even if this software can handle different kinds of dataset, it is especially designed for aerial imagery processing. A more detailed description of the implemented procedure and input parameters is given in (Verhoeven, 2011).

- CLORAMA (CLOse RAnge MAtching) is a software offered by the ETH spin-off company 4DiXplorer; although it is not a traditional commercial software, it is nor open-source nor free-of-charge, thus it has been included in this first category of software solutions. CLORAMA is a software package for DSM generation, starting from oriented terrestrial images. It is based on a coarse-to-fine hierarchical approach with an effective combination of several image matching techniques and automatic quality control tools. The software performs three consecutive steps: first of all, images are optimised through the application of several filters, such as an adaptive smoothing filter and an enhancement filter; images pyramids are generated as well. Secondly, multiple matching primitives (interest points, edges and grid points) are extracted, matched and combined. Finally, at the original image resolution level a refined matching is performed, in order to identify and eliminate some possible mismatches. A more detailed description of the software is discussed in (Barazzetti et al., 2010).
- iWitness (Fraser and Hanley, 2004) is a photogrammetric data processing system, which has been mainly developed for the non-specialist user. It is configured as a flexible, robust and ease-to-use software package for integrated image measurement, orientation, subsequent photogrammetric triangulation and post-orientation processes. Its primary application domain is accident reconstruction and forensic measurements; however, the software can also deal with many other measurement tasks in engineering, architectural, photogrammetry and cultural heritage application fields.
- IMAGINE Photogrammetry (formerly Leica Photogrammetric Suite, LPS) is a complete suite of photogrammetric production tools. It is constituted by different and efficiently connected modules, that provide the users with all necessary tools required to transform raw images into reliable geospatial data. The module LPS eATE, for an accurate and automatic DTM/DSM extraction from multiple images, will be extensively described in Chapter 8, where an application of the tool will be presented.
- ImageStation by Intergraph is a suite of different modules that handle several photogrammetric workflows, such as orientation and triangulation, 3D feature collection and editing, interactive digital terrain model extraction and editing, and

orthophoto production using aerial and satellite imagery. In particular, it can generate dense point clouds from stereo imagery using a semi-global matching technique.

- Match-T DSM is one of the modules implemented into Trimble INPHO software in order to generate DSM and DTM with an image-based approach. It applies a sequential multi-image matching procedure in several scales combining feature-based and least squares matching. The main advantages of this module are: the generation of very dense point clouds, a significant reduction of noise in final point clouds, short processing times, the support of large datasets and the distributed processing (Heuchel et al., 2011).
- PhotoModeler Scanner by EOS Systems (Alby et al., 2009) is specifically designed to create 3D models and measurements from images. It adopts an area-based algorithm and, in the case of stereo-matching, it needs images taken with parallel axes. The image matching is controlled by several different parameters, such as reference surface, correlation area and texture type. A single surface model is extracted by each stereo-pair correlation then all models generated are registered and merged together. Finally, a triangulated mesh can be produced.
- Shape Capture by ShapeQuest Inc. is a tool for accurate 3D measurement and modelling from single or multiple photos. The software (Beraldin et al., 2002) allows to calculate camera calibration with simple and well defined steps: the user should only print a calibration grid of proper dimensions, taking into account the extension of the scene to be acquired in order to measure the actual subject. After calibration, a set of convergent images can be oriented by first selecting homologous points over them and then launching a bundle adjustment process. Finally, accurate measurements on the 3D coordinates of specific points can be performed; moreover, highly accurate 3D models can be created from the registered images. A second software by the same company is ShapeScan: it allows the use, also at the same time, of both structured white light scanning and image-based dense stereo matching.
- SOCET SET (module Next-generation Automatic Terrain Extraction, NGATE) by BAE Systems is an image matching tool that uses a hybrid matching process to create precise elevation data for DTM and DSM generation. If compared to the previous ATE (Automatic Terrain Extraction) version, it shows many advantages, such as less smooth surfaces, less blunders and capability of using also a feature-based matching (Remondino and Menna, 2008). The main features and characteristics of this package are summarized in (Zhang et al., 2006).
- 3DF Zephyr Pro by 3DFLOW is a new fully automatic image-based 3D reconstruction tool. Its completely automatic procedure requires nor coded target nor manual editing and is able to estimate both intrinsic and extrinsic camera parameters automatically. Geo-referencing process can be performed as well, by providing the system with Ground Control Points and know distance information. The final 3D

model is generated with a fully unwrapped texture, generated using the images at their original resolution.

2. **Open-source / free of charge software**. These solutions offer free and, generally, more flexible tools, even if these advantages come sometimes with the price of requiring a less user-friendly pipeline.

   - Apero/MicMac is a set of photogrammetric tools developed by Marc Pierrot-Deseilligny at the MATIS Laboratory of the French Mapping Agency (*Institut Géographique National*, IGN) and delivered as open-source starting from 2007. The two main tools currently distributed are Apero, a software for computing orientation of images, and MicMac, a software for computing depth maps from oriented images. Both tools will be discussed in detail in Section 3.3, where the main features and characteristics of this software suite will be extensively described.

   - Bundler is a well-known open-source software that computes camera interior and exterior parameters and produces sparse 3D point representations (Snavely et al., 2006). The tool is based on Structure from Motion techniques, that are performed in a multi-step procedure. As the first step, SIFT detector and descriptor are used to find key-point locations and associate local descriptor to each key-point. Then, the approximate nearest neighbour's package (Arya et al., 1998) is applied to match key-point descriptors. Furthermore, the incremental reconstruction step is started: at the beginning, an image pair with optimal overlap and intersection geometry is selected and its relative orientation is determined. Subsequently, new images are added by resecting the available 3D points using a DLT technique inside a RANSAC procedure. A bundle adjustment refinement is performed after the orientation of each image. This procedure is repeated until an orientation is available for all images.

   - PMVS (Patch-based Multi-view Stereo Software) – Version 2 is the most common additional step to the Bundler workflow, in order to extract a dense point cloud starting from the output of Bundler itself (Ducke et al., 2010; Furukawa and Ponce 2010). The software takes as input data the undistorted images, their orientation parameters, a sparse point cloud and the projection matrices: this information is then converted into a dense and accurate set of rectangular patches. The procedure implemented in PMVS includes three subsequent steps. First, the matching phase is performed by finding corner and blob features: they are then matched across multiple images by taking into account local photometric consistency with NCC, and a sparse set of patches is thereby generated. Then an expansion step is added, where initial patches are extended to nearby pixels in order to generate a dense set of patches. Finally, within a filtering phase, incorrect matches are deleted using visibility constraints. A known object space distance can be used in order to accurately solve the scale ambiguity of the final model.

   - Insight3d is an open-source image-based 3D modelling software that automatically matches a series of un-ordered photos and then computes the internal and external

camera parameters. A 3D point cloud of the scene is finally extracted; furthermore the user can apply some provided modelling tools to create textured polygonal models.

- SURE is a dense image matching software, which has been developed by the Institute for Photogrammetry at the University of Stuttgart (Rothermel et al., 2012). It uses a multi-view stereo approach, where at first stereo pairs are matched; this stereo matching step is based on a hierarchical semi-global matching approach that enables the determination of 3D information for almost each pixel. Then, the results of image matching are fused by triangulating rays for multiple stereo models at once: this process improves the precision of object points and enables the rejection of outliers as well as the determination of quality values for each 3D point. The whole process is parallelized and optimized for scalability.
- VisualSFM (Wu, 2013) is a GUI application of Structure from Motion that integrates three different projects developed by the same author: SIFT on GPU (SiftGPU), Multicore Bundle Adjustment and an incremental SfM system. In addition to the sparse 3D reconstruction, the software provides also an interface to run the PMVS2 tool. VisualSFM runs very fast by exploiting multi-core acceleration of the three main processes, i.e. feature detection, feature matching and bundle adjustment.

3. **Web services**. These solutions offer two main advantages, i.e. easiness for the end-user (no software should be installed) and cost synergy by sharing the CPU. On the other hand, they require the availability of an internet access and the possibility of sharing the images on the web. Moreover, their code is generally not accessible, so it's not possible to improve or adapt it to the user needs.

   - ARC3D (Automatic Reconstruction Cloud) is a group of free tools which allow users to upload their digital images to a server, where the 3D reconstruction is performed (Vergauwen and Van Gool, 2006); the output is finally sent back to the user in few hours. The necessary time depends on size, number and quality of the images that have been uploaded, but the process is very fast since the reconstruction is computed over a distribution network of PCs. Moreover, ARC3D provides a tool for visualizing the 3D scene reconstructed by the server. Of course, the success of the automated reconstruction procedure is strongly related to the input images, that should be acquired with small convergent angles and with fixed focal setting.
   - Photosynth by Microsoft is a software of Image Based Rendering, i.e. devoted to the problem of synthesizing new views of a scene from a set of input photographs. The software is free, but a Microsoft Live ID is anyway required. This photo visualization tool is largely based on (Snavely et al., 2006) and allows the user to capture the world in 3D through two different visualization styles: panoramas and synth. The latter is useful to reconstruct different sides and details of an object; furthermore, the software efficiently streams image data using a system called "Seadragon", that computes which parts of which images are visible on the screen and at what resolution each

photo is viewed (Snavely et al., 2008). This software is primarily designed for visualization and web-based touristic applications.

- Autodesk service, provided through the following main solutions (now merged into ReCap, a point cloud and image-based 3D modelling software for 3D documentation):
    i. Project PhotoFly is based on the technology of RealViz, acquired by Autodesk in 2008. The tool (Abate et al., 2011) is a web service technology that is able to give back to the user a complete and triangulated 3D model, with a texture-mapping built using the input photos. Like almost all web services, it is completely automatic and only the final resolution of the model may be selected, using a simple low-medium-high option.
    ii. 123D Catch by Autodesk is a free web-service that overcomes the previous Autodesk PhotoFly technology project. It is based on Computer Vision algorithms and through them it is able to reconstruct the internal parameters of the digital camera and the 3D position of homologous points, starting from a number of extracted image correspondences. The pictures should be acquired according to a path of continuity around the object and uploaded to the server following this order. The 3D reconstruction is a completely automated process. The tool allows the user to improve the final result through a manual stitching of homologous points on triplets of images: the scene can then be submit again to the server. The approach employed by this software technology is well described in (Hiep et al., 2009). Detailed studies on the service performances can be found in (Santagati and Inzerillo, 2013).

3.3 Apero/MicMac

After the general review presented in the previous section, it is clear that the landscape of the available image-based 3D modelling approaches is wide and still growing up. Certainly, the flexibility offered by the open-source tools, usually based on computer vision techniques, is a great advantage especially for those users, who are non-experts and hence have few ideas of photogrammetric constraints and principles. However, the price to pay for this peculiarity is usually a lack of reliable and accurate results, that can only be used for simple visualization, image-rendering and web-based applications. In other words, the low photogrammetric rigour of the most available solutions leads to unacceptable precision for metric and accurate documentation purposes. Some weaknesses can be summarized as follows:

- Over-parameterization of orientation algorithms, that are usually based on the fundamental matrix model; more camera parameters than the minimum necessary amount are thereby evaluated, e.g. each image is usually allowed to have its own focal length. This approach leads to system imprecision and perturbations.

- Weak calibration procedures, often based on simple and incomplete camera distortion models. Most of the open-source available solutions can only model radial distortion effects and cannot deal with images acquired by fish-eye lenses.
- Sparse image matching techniques, that usually work on a subset of well contrasted interest points. These approaches lead to sparse 3D reconstruction results, that may be inadequate for those applications requiring final dense point clouds.

All the points listed above suggested the need for reliable, accurate and flexible solutions, based on solid photogrammetric principles and able to deliver detailed and precise 3D reconstructions, useful for metric purposes. The French National Geographic Institute, IGN, tried to address these requests by making an open-source deposit of some internally developed photogrammetric tools. The history of this "effort" can be briefly summarized as follows.

In 2003 an image matching software was developed in order to produce DSM and DTM for satellite and aerial photogrammetry applications. Two years later, the software was called "MicMac" and combined with an XML (eXtensible Markup Language) interface. In 2007 the software was delivered as open-source code and further improved in order to take into account also different application fields, such as close-range terrestrial imagery. As a consequence, in 2008 MicMac was completed by Apero, a photogrammetric bundle adjustment software for automatic orientation of images. Many other tools were then added, in order to cover all the most common photogrammetric workflows; thereby a real suite of tools was completed. Starting from 2010, finally, many training sessions have been organized in order to encourage the use of these tools; some simplified interfaces (without XML) have also been developed.

The Apero and MicMac open-source software are released under the CeCILL-B licence (CeCILL, 2005): it is essentially an adaption to the French law of the L-GPL licence, being thus quite permissive from a legacy point of view. All the other simplified tools born from the extension and evolution of MicMac obey to the same licence. All the tools are written in C^{++} and mainly distributed in source code format that the user has to compile. Pre-compiled binaries are also provided; moreover, one can also prefer the versioning software Mercurial, that enables a more easy installation and update of the suite of tools. Some external tools are also needed as system's pre-requisites, such as make, ImageMagick, exiftool/exiv2 and proj4, whose related web sites are listed at the end of the Reference Chapter.

The suite is composed by simplified and complex tools. The former are command lines tools, whose parameters can be directly written on the command line; on the other hands, for more complex processes, requiring a huge amount of arguments and attributes, this command line format would not be manageable. For this reason, complex and parametrical tools need a XML file to be written by the user, specifying all the parameterization options. Anyway, all the tools are available via one single command, "mm3d", that represents the input to all processes: this has led to many advantages, such as more compactness of the binaries and the possibility to factorize the future developments. The number of available tools is very large

and still growing up; a short summary of the most common tools is presented in Table 3.1, with a brief description of the corresponding functionality. All tools are working on Linux, MacOs and Windows, with the exception of the set of *Saisie* tools, which can run only on Linux platforms.

Category	Tool	Functionality
XML tools	Pastis	Tie point detection
	Apero	Internal and external orientations
	MicMac	Image matching from oriented images
	Porto	Global orto-photo
Simplified versions of XML tools	Tapioca	Interface to Pastis
	Tapas	Interface to Apero (internal and external orientations)
	Campari	Interface to Apero (compensation of heterogeneous measures)
	Malt	Interface to MicMac
	Tarama	Image rectification
	Tawny	Interface to Porto
Coordinate transformations	GCPBascule	Absolute orientations using GCP
	CenterBascule	Absolute orientations using embedded GCP
Saisie tools	SaisieAppuisInit	Interactive tool for GCP capture
	SaisieMasq	Interactive tool for mask capture
Visualization/Export	AperiCloud	Sparse 3D representation of tie points and camera poses
	GrShade	Shading computation from depth map
	Nuage2Ply	Point cloud computation from depth map

Table 3.1 Some of the most common tools of the IGN's suite

Some of the functionalities of these tools are also available with an end-user GUI with dedicated context interfaces, such as (Pierrot-Deseilligny et al., 2011):

- A general interface for the Apero/MicMac pipeline, developed at the IGN.
- A specific interface for the entire 3D reconstruction, integrated into the Maya plug-in NUBES Forma (De Luca et al., 2011) and developed at the CNRS MAP-Gamsau Laboratory of Marseille.
- A web-viewer for image-based 3D navigation and point clouds visualization, developed at the CNRS MAP-Gamsau Laboratory of Marseille; this viewer allows to jump into the different image point of views, back-projecting the point clouds onto the images.

From an algorithmic point of view, this open-source suite of tools uses mathematical formulations derived from both the photogrammetric and the computer vision fields, focusing thereby on the accuracy and metric content of the final results. This more robust photogrammetric rigour makes these tools mainly addressed to experts (geomatic professionals, cartographers, engineers, architects, …) who have some basic knowledge on photogrammetric acquisition and processing principles. On the other hand, every single step of the 3D reconstruction procedure, even if complex, may be controlled by the user, who can adapt the parameterization to his/her personal requirements; furthermore, these tools offer the possibility of dealing with very big datasets. The main steps of the photogrammetric procedural pipeline (Figure 3.1) will be discussed in the following subsections, where some specific bibliographical references will also be mentioned. More detailed issues are discussed and addressed in the software documentation (Documentation MicMac) and on-line forum (Forum MicMac). The ortho-photo generation process, although possible, will not be here explained.

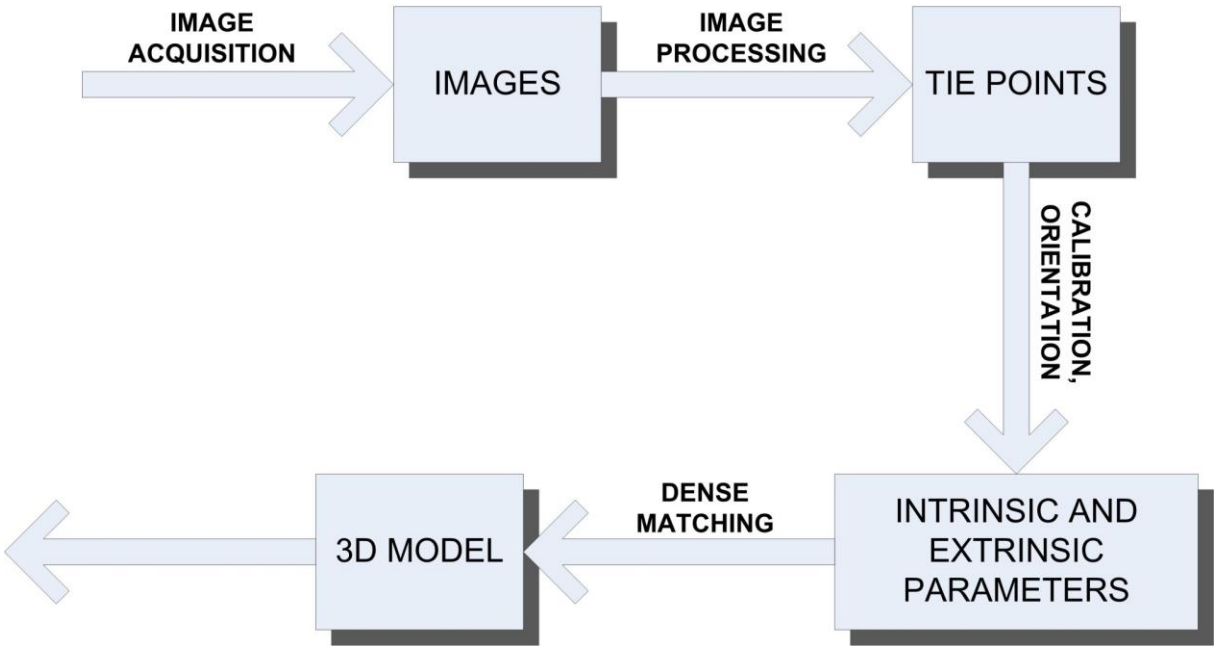


Figure 3.1 The main steps of the Apero/MicMac procedural pipeline

3.3.1 Image acquisition

The approach of Apero (Subsection 3.3.3) for the initialization of the orientation procedure, based on both photogrammetric and computer vision techniques, allows a certain flexibility in the data acquisition; however, this first phase, performed on the field, is still a very crucial step in the photogrammetric pipeline. Thus, the user should be aware of the two most important issues that should be considered during the acquisition phase:

- Images must be sufficiently connected by tie points. The SIFT algorithm, used in the phase of homologous point extraction and matching (Subsection 3.3.2), is sensitive to affine differences between the images. This weakness should be compensated by an appropriate strategy of image acquisition. In particular, small baselines produce more overlap between the images and a good similarity between two different views of the same scene: these factors guarantee more success in the automatic tie point extraction procedure. Furthermore, this configuration reduces the possibility of having occluded area.
- For dense stereo matching, every point of the object must be seen from different directions, with good viewing angles. Equation [2.40] in Chapter 2 computed the standard deviation of the depth measurement on the final reconstruction: according to it, the depth error can be reduced by increasing the baseline between the images.

Starting from these observations, it is clear that a compromise between the two opposite requirements should be reached. The shooting configuration should generally be convergent, in order to acquire more possible hidden details; the choice of an optimal angle of convergent images (α in figure 3.2) should also be searched for. If α is too strong the two images are too different, and geometric distortions as well as hidden parts will occur (Figure 3.3).

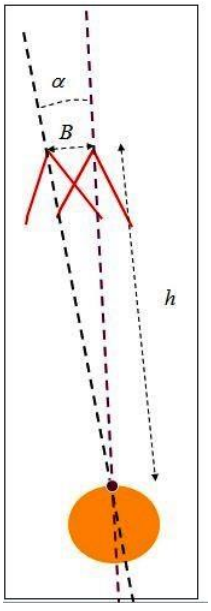


Figure 3.2 Convergent shooting acquisition

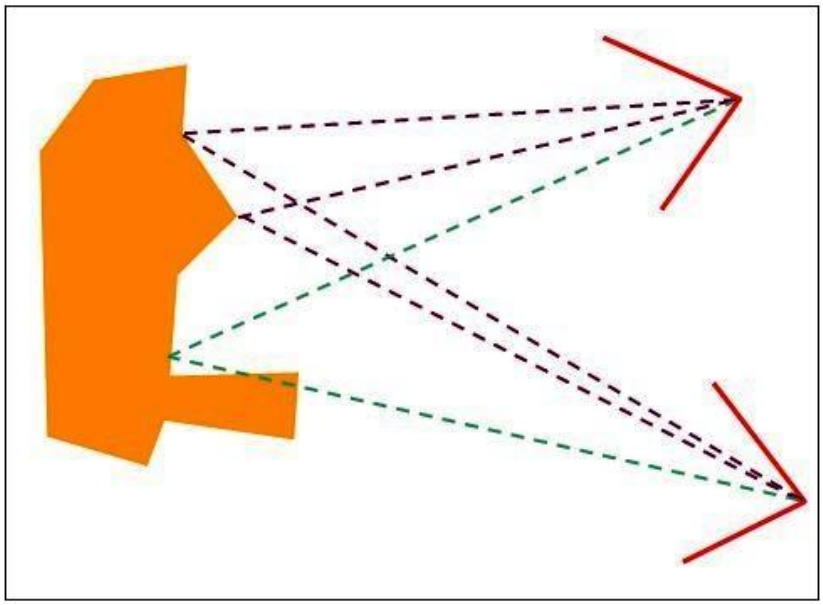


Figure 3.3 Geometric distortions (purple lines) and hidden parts (green lines) caused by too strong values of α

On the other hand, accuracy and robustness of image matching is inversely proportional to α. Of course, the optimal angle α depends also on the following factors: the relief of the scene (with its consequent amount of possible occluded area); the texture of the object (with good texture, matching can use small windows, thus stronger angles); the final purpose (orthophoto or 3D model); the accessibility of the area. So, a compromise choice can be formulated as:

- $5^{\circ} < \alpha < 20^{\circ}$ for terrestrial convergent multi-stereo images;
- $10^{\circ} < \alpha < 20^{\circ}$ for aerial photogrammetry on dense urban areas;
- $\alpha > 60^{\circ}$ for aerial photogrammetry on open rural areas.

An optimal acquisition configuration is the crosswise convergent image shooting (Figure 3.4), where a central master image (blue pyramid) is associated with four secondary images (white pyramids), which form a vertical and horizontal multi-stereo geometry.

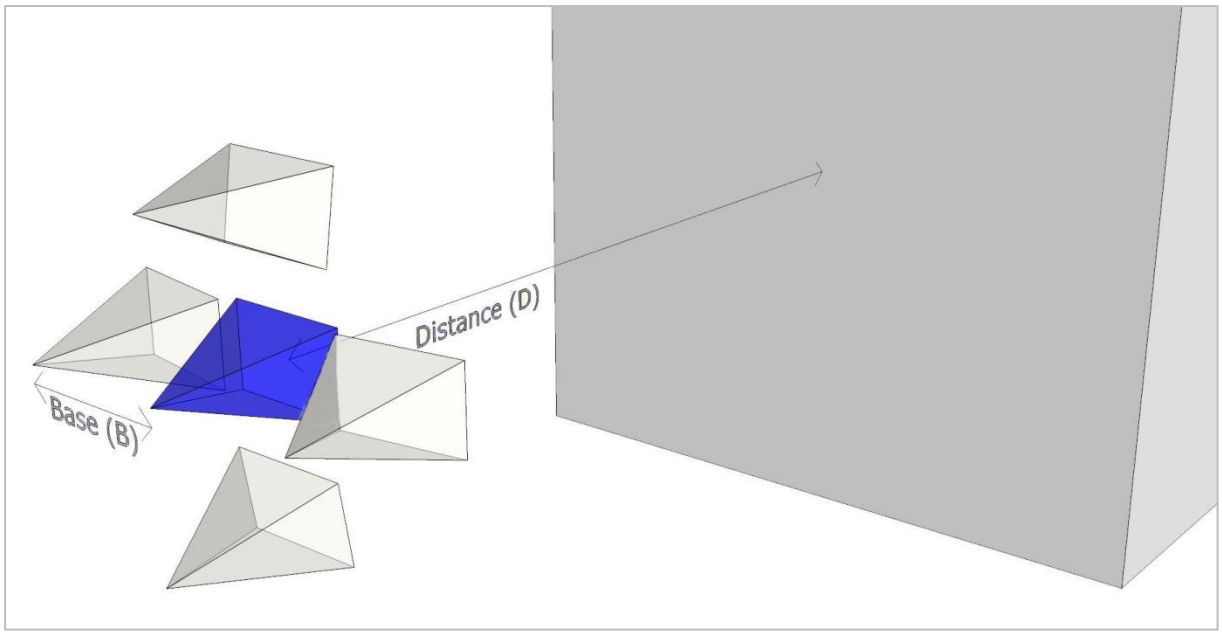


Figure 3.4 The crosswise convergent image configuration

Considering all the above mentioned issues, some rules have to be respected during the image acquisition phase, in order to support the convergence of the subsequent relative orientation procedure:

- For each desired point cloud (i.e., for each point of view), take a master image with four closed associated images (crosswise convergent image configuration);
- Between each master image, take a sufficient number of intermediary images, in order to assure the connection between the different points of view during the orientation step;
- In both the previous cases, keep a reasonable base-to-depth (B/D) ratio, according to the mentioned optimal ranges of α;

- Avoid to move a punctual light source during the photo acquisition; the scene should be lighted with a diffuse zenith light (prefer an overcast sky for outdoor acquisition and a studio flash with a big light box for indoor acquisition);
- Use a context with enough details spread on all the acquired background;
- Fix as many parameters of the cameras as possible (as an optimal choice, focal and focus setting should be fixed);
- Put a metrical reference in the scene (object of known distance or measured Ground Control Points, for example) in order to solve the scale ambiguity.

The number of images necessary to cover all the object depends essentially on the dimensions, shape and morphology of the scene and on the employed focal length. Thus, the acquisition layout is not unique, and many different protocols can be adopted according to the specific characteristics of the application. Two examples are presented below:

- A simple and small object (i.e. an artefact with a simple morphology situated in an easy external context) can be acquired with a close circle protocol, as it is represented in Figure 3.5, where master images (in blue), secondary images (in white) and intermediary images (in green) are depicted.

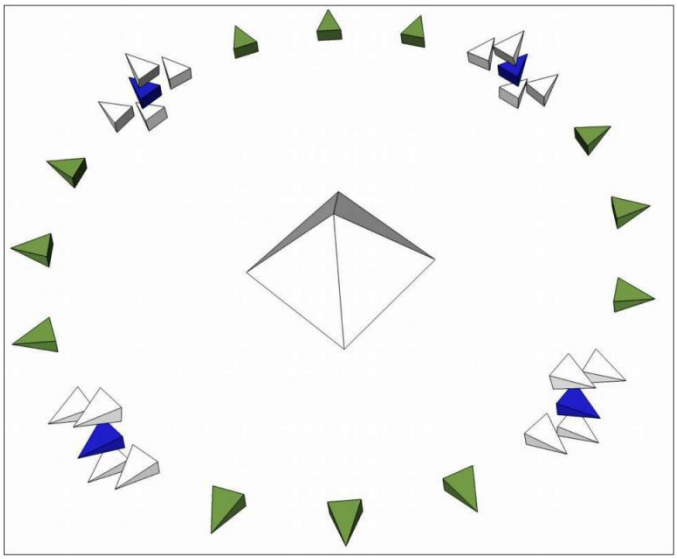


Figure 3.5 Acquisition protocol for an object with a simple morphology (Martin-Beaumont et al., 2013)

- A more complex and big object should be acquired with a dual-focal-length strategy: a lens with long focal length should be used to acquire the images of small portions with good resolution; a lens with short focal length should be used to acquire the images needed to orientate the data-set. In other words, short focal length images will be first oriented, since they offer a more field-of-view overlapping; then, the long focal length images will be oriented on this previously computed canvas.

A more detailed discussion on the over mentioned issues related to the image acquisition phase can be found in (Martin-Beaumont et al., 2013; Pierrot-Deseilligny et al., 2011; Pierrot-Deseilligny and Clery, 2011).

3.3.2 Tie point extraction

INPUT	**OUTPUT**	**TOOL**	
		XML	Simplified
A set of images	Tie points between images	Pastis	Tapioca

Table 3.2 Tie point extraction phase

Table 3.2 briefly describes the tie point extraction phase, whose purpose is the computation of homologous points between the input images. Thanks to its good invariance to scaling, rotation, translation and contrast, the SIFT algorithm is used. In particular, this first step is performed through the SIFT^{++} implementation of SIFT algorithm (Vedaldi, 2010), that has been modified in order to work with large images: SIFT^{++} is a lightweight C^{++} implementation of SIFT detector and descriptor, directly derived from the MATLAB/C implementation. From a legacy point of view, the SIFT code used as tie point generator for Apero is submitted to the SIFT patent: if the user cannot employ it, he/her could replace it with any other detector.

The SIFT algorithm is based on a three-stage process: at first characteristic points are computed (maximum of the Laplacian in scale and space); then, for each characteristic point the SIDT descriptor is estimated. Finally, Euclidian distances between the descriptors are used in order to match the homologous characteristic points. From a practical point of view, this process is completely automatic; however, the user can choose among some different options. First of all, the search mode should be selected between:

- All possible pairs of images;
- Multi-scale approach, that can save significant computation time with large datasets. In this mode, a first computation of tie points is performed for all possible pairs of images at very low resolution; then, the tie points are computed again only among the image pairs that, at low resolution, have reached a certain threshold. This second computation is performed at higher resolution, specified by the user.
- Line approach, suitable for linear image acquisition. If the photo canvas has a linear structure, the Kth image will be matched only with images in the interval [K-δ ; K+δ], where δ is specified by the user;
- File approach, where the user explicitly defines the image pairs to be considered through an XML file.

Furthermore, the parameter Size is used to shrink the images: it specifies the desired width for shrinking the images during the tie point extraction phase. A suitable Size factor is important in order to reduce the dimensions of the transmitted data, especially in case of large datasets.

3.3.3 Calibration and orientation

INPUT	OUTPUT	TOOL	
		XML	Simplified
• Tie points • Other known measures	• Camera calibration • Camera orientations compatible with all measurements	Apero	Tapas (+ other optional tools)

Table 3.3 Calibration and orientation phases

Table 3.3 describes the calibration and orientation phases within the Apero/MicMac 3D reconstruction pipeline. These processes make use of both computer vision techniques (in the initialization phase) and photogrammetric ones (in the refinement phase); more detailed information about the algorithmic aspects can be found in (Pierrot-Deseilligny and Clery, 2011). Generally speaking, Apero is a software that computes camera orientations, positions and calibrations compatible with a set of input, redundant observations, that are declared by the user with the associated confidence levels (weighting functions).
Basically, the input observations can be:

- A set of computed tie points between images; this is usually the predominant information (sometimes, the only one);
- Ground Control Points (GCP), i.e. object points with known 3D coordinates and visible in two or more images;
- Information about the position of camera projection centres through embedded-GPS observations;
- Results of previous orientation computations.

The output of the process is represented by the computed values of the unknowns, that are:

- Internal calibration parameters;
- Camera poses and orientations (in a local or absolute reference frame).

The general problem to be solved is the computation of unknown parameters, that should be compatible with all the input observations. In order to address this task, the main processes performed by Apero can be summarized as follows:

1. Computation of the initial solution. This step is the most difficult one, since no known algorithms exist that can compute directly a set of orientations compatible with tie points. In other words, there is still a lack of validated algorithms that can find a direct solution to a global relative orientation problem. On the other hands, one can find many algorithms that can solve the following elementary problem: given a set of already oriented images and only one additional image, use tie points to compute the orientation of the new image relative to the others. These algorithms can then be run many times to build, step by step, the global orientation of a large block of images. Thus, the adopted approach is the following one, that is repeated many times, step by step, until all the images are oriented:
    - Choose the first image and set it in an arbitrary position;

- Select the next best image to add. This choice is performed by computing a stability estimator: for each possible image, the cloud of tie points with the already oriented images (the first selected one in this case) is analysed, in order to determine which image offers the most numerous and homogeneously spread set of tie points. For this task, the smallest value of the cloud inertial matrix is used as stability estimator.
- Use elementary algorithms and tie points to compute the orientation of the second image relatively to the first one (or, to the already oriented ones). Apero tests essential matrix with RANSAC and, if there are sufficient tie points, tries also the space resection algorithm with RANSAC. Finally, the best solution is chosen.
- Since these direct elementary algorithms don't use all possible information, an error accumulation can occur, leading to system divergence; to avoid this problem, a second process, i.e. a solution refinement through bundle adjustment, is regularly performed on the already oriented images.

2. Bundle adjustment refinement. This approach follows the classical photogrammetric one presented in (Triggs, 1999). In particular, for each tie point:

    - An estimation of the corresponding object point is computed by bundle intersection of all images where it is seen and using the current values of internal and external orientations;
    - The re-projection error is computed and the sum of the re-projection errors is formulated using a weighting function and minimized;
    - The resulting system of equations is linearized and solve by least mean squares.

    At each step of the minimization process, the user can select which unknown are free to evolve in the computation, and which are temporarily frozen. Usually, the internal calibration parameters are kept frozen at the beginning of the refinement process, then they are released in this order: distortion coefficients, focal length, distortion centre, principal point. Of course, each time one parameter is released, a bundle adjustment computation is performed.

3. Camera calibration process. The camera self-calibration may be performed during the previously described bundle adjustment procedure. First of all, for each camera the tool has to handle, it should find an initial value. If the user does not provide a set of calibration parameters, previously computed via, for example, a test-range calibration, the initialization is performed according to the following rules:

    - The focal length is computed in pixel-values starting from the image EXIF (EXchangeable Image File) information and the width of the camera sensor (a global camera database is, in fact, provided with the suite and the user can eventually update it according to the employed digital cameras). If the EXIF file does not provide the expected information, it can be indicated dynamically by creating specific key in a given XML file (MicMac-LocalChantierDescripteur.xml).

- Distortion is initially set to zero (ideal pinhole camera model);
- The principal point is initially located at the centre of the image (no offsets).

Apero, then, proposes a large set of internal calibration models:

- RadialExtended model, which considers only radial distortion effects; it has 10 degrees of freedom (DoF), 1 for focal length, 2 for principal point, 2 for distortion centre and 5 for coefficients of the radial distortion polynomial (r^3, r^5, …, r^{11});
- RadialBasic model, that is just a subset of the previous one and suitable where the RadialExtended formulation may fail due to system divergence; this model keeps only 5 DoF, since the principle point and distortion centre are supposed to be the same point, and the radial distortion polynomial is truncated to degree 5.
- Fraser model, that is deduced from the mathematical formulation presented in (Fraser, 2001). According to it, the principal sources of departures from the ideal camera model are four and their cumulative influence produces the following image displacements:

$$\Delta x = \Delta x_r + \Delta x_d + \Delta x_u + \Delta x_f$$
$$\Delta y = \Delta y_r + \Delta y_d + \Delta y_u + \Delta y_f \quad [3.1]$$

  where the subscript r is for radial distortion, d for decentring distortion, u for out-of-plane unflatness and f for in-plane image distortion. The resulting camera model has 12 DoF: 1 for focal length, 2 for principal point, 2 for distortion centre, 3 for coefficients of radial distortion (r^3, r^5, r^7), 2 for coefficients of decentring distortion and 2 for affine parameters. Optionally, one can set to false the LibAff and LibDec options, if, respectively, decentring or affine parameters should be kept frozen.
- FraserBasic model, that immediately derives from the previous one by constraining the principal point and distortion centre to have the same value (10 DoF);
- FishEyeEqui model, that is made by a combination of a theoretical model and additional polynomial parameters, resulting in a 14-DoF formulation;
- HemiEqui model, that is a modified version of the previous one, adapted to hemispherique equilinear fisheyes;
- The AutoCal and Figee options, finally, don't define a calibration model, since they refer to the situation in which a calibration parameter set (and, eventually, also the camera orientation one) has already been computed by the user. In particular, with AutoCal the parameters are re-evaluated during the bundle adjustment refinement; with Figee they are, vice versa, kept frozen.

4. Geo-referencing of solution. If an absolute orientation is needed, Apero offers to the user different means of geo-referencing:

- If a GPS is embedded and synchronized with the camera (it is, for example, a very common case in aerial photogrammetry), direct information about the position of the camera projection centre can be provided to the software. This information is usually employed in two steps. At first, the relative orientation computed with tie points can be transformed to a first geo-referenced solution. This global transformation is then used in the compensation phase by adding a new observation equation to the global system.
- Ground control points can also be used if at least three of them are known and measured in at least two images; of course, also these observations are used during the compensation process and allows the computation of absolute orientation.
- Sometimes, no metric geo-referencing information is available, but Apero allows anyway the computation of an orientation that is coherent with some physical constraints. For example, an horizontal plane can be specified, or a line can be used to fix an orientation. Furthermore, an object of known size can be employed in order to set the scale of the model.

If Apero is a parametrical software that "reads" the parameters stored in a XML file, Tapas is a simplified command-line version of it, offering most of the possibilities of Apero for computing only relative orientation. In combination with Tapas, other simplified tools can be used in order to run all the functionalities of Apero, such as:

- GCPBascule or CenterBascule, that use respectively ground control points and embedded GPS information to make a global transformation from a relative orientation to an absolute one.
- Campari, that compensates together heterogeneous measures (tie points and ground control points);
- AperiCloud, that generates a visualization of the camera poses and a sparse 3D model.

3.3.4 Dense image matching

INPUT	**OUTPUT**	**TOOL**	
		XML	Simplified
• Tie points • Internal & External orientations	• Depth (or disparity or height) map	MicMac	Malt (+ other optional tools)

Table 3.4 Dense image matching phase

Table 3.4 describes the dense image matching phase within the Apero/MicMac 3D reconstruction pipeline. MicMac was initially developed to match aerial and satellite images, then it was adapted to deal also with convergent terrestrial images. It uses a NCC stereo matching algorithm, with a multi-scale, multi-resolution and pyramidal approach: a detailed

discussion on the algorithmic aspects can be found in (Pierrot-Deseilligny and Paparoditis, 2006). In order to give a general overview of the software functionalities, three main issues will be addressed below:

1. Algorithmic aspect.

MicMac adopts the Normalized Cross Correlation (NCC) coefficient as a similarity measurement among the images. If the formulation derived for the two-view geometry has been presented in Chapter 2 (Subsection 2.2.5), it should be here adapted to the case of multi-view stereo imagery. Given N images and (U_1,…,U_N) vectors to be compared, a first possible formulation is the following one:

$$\mathrm{Corr}(U_1)(U_2 \dots U_N) = \frac{1}{N-1}\sum_{k=2}^{N} \mathrm{Corr}(U_1, U_k) \qquad [3.2]$$

The above definition gives greater importance to image 1, thus it is suitable when an image plays a special role among the others. If all images play a symmetrical role, as it usually occurs in common acquisition geometry, a different formulation of multi-image NCC should be considered, computing the mean over all possible pairs of images:

$$\mathrm{Corr}(U_1 \dots U_N) = \frac{2}{N(N-1)} \sum_{1 \le i \le j \le N} \mathrm{Corr}(U_i, U_j) \qquad [3.3]$$

As evident, the complexity of formulation [3.3] is $\mathcal{O}(N^2)$, and this may represent a drawback from the point of view of the required computational time. Thus, MicMac uses an algorithm that, although derived from equation [3.3], enables the computation of multi-stereo NCC with a linear trend of computational time, i.e. with a complexity of $\mathcal{O}(N)$. This novel formulation will not be discussed here: a detailed mathematical explanation can be found in (Apero/MicMac documentation, 2013).

Furthermore, MicMac implements a multi-scale, multi-resolution image matching approach. This choice, besides further improving the computational time, reduces the possibility of having erroneous matching: this approach, in fact, implements the idea that real homologous points are similar at every scale, while false homologous points usually appear different in different scales. So, in order to make the computation more robust, MicMac follows the following strategy:

- It computes a pyramid of N images at the different resolution 2^k, $k \in [1, N]$;
- The image-matching procedure is first performed at the lowest resolution dataset, that is treated normally, i.e. the whole incertitude interval (see the parameter list) is explored;
- The subsequent resolution level, 2^k, is then computed using the previous one, of resolution 2^{k+1}. MicMac forces the solution computed at resolution 2^k to be as close as possible to the one derived from the previous step, at resolution 2^{k+1}, by limiting it

within the spatial interval defined by the parameters of planar and altitude dilatation (see the parameter list).

In order to reconstruct a geometric surface that satisfies two different conditions at the same time, i.e. the *a-priori* knowledge of the surface regularity and the coherence with the extracted homologous points, MicMac uses a regularization algorithm based on an energetic formulation. This approach defines a global function on the image field, that can be written as follows:

$$E(Z) = \sum A\big(x, y, Z(x, y)\big) + \alpha * F\big(\vec{G}(Z)\big) \qquad [3.4]$$

Where:

- Z is the unknown height (or depth) function;
- $A\big(x, y, Z(x, y)\big)$ is the image matching term, measuring the image consistency/similarity computed at the image projection 3D point $\big(x, y, Z(x, y)\big)$. It is derived from the NCC computation;
- $F\big(\vec{G}(Z)\big)$ is the regularization term, which expresses the *a priori* knowledge of the surface regularity;
- α is a parameter which allows to adjust the importance of the regularity term relatively to the image term.

In order to globally minimize the energetic function [3.4], MicMac offers several possible algorithms, that always include a parameter playing the role of α in the above formulation. In particular, the main available algorithmic approaches are the following ones:

- A Roy-Cox implementation of the classical minimal cut and maximal flow graph theory algorithms, that tries to compute the exact minimum of the energetic function. Even though polynomial, this approach requires significant efforts, both in terms of computational time and memory. Thus, MicMac implements a multi-resolution variant of the Roy-Cox algorithm, using the above explained pyramidal approach.
- A 2D generalization of classical dynamic programming approaches, that doesn't produce the exact minimum of the energetic function, but a pseudo-optimum, i.e. a solution generally very close to the optimal one. This solution is calculated by analysing the image line by line, and looking for the optimum within a line (or column). The drawback of this approach is the dissymmetry that it introduces in the processing of image lines and columns: as a consequence, some errors may be introduced in the final result. On the other hand, this algorithm is generally faster and more flexible.

Finally, the matching algorithm used by MicMac makes a quantification of the disparity (or height or depth according to the selected geometry), producing undesirable jumps in the final results. Thus, a de-quantification process is usually performed at the final higher resolution step: this post-processing phase eliminates the quantification artefacts, producing a floating point map.

2. Geometric aspect.

From the geometric point of view, MicMac offers three main types of input-output geometries:

- Epipolar geometry; if images have been warped in order to take the advantages of rectilinear stereo-rigs, corresponding epipolar lines are forced to lie along the same horizontal scan line in each image. Thus, given an image point (x, y), the aim of epipolar matching is to compute the disparity $d(x, y) = x' - x$, so that the point $(x + d(x, y), y)$ is the homologous of point (x, y). A disparity map is finally computed. This kind of geometry has, of course, the advantage of simplicity. However, it lacks of generality, since it cannot be used for multi-stereoscopy; furthermore, it requires an unnecessary re-sampling of the images.
- Ground geometry; generally speaking, the function that defines the orientation of an image k can be formulated as

$$\pi_k: \mathbb{R}^3 \to \mathbb{R}^2; \quad \pi_k(x, y, z) = (i, j) \qquad [3.5]$$

  Where:

  - (x, y, z) is an object point;
  - (i, j) is its projection in image k;
  - $\pi_k(x, y, z) = (i, j) = I\left(p_0\left(R_k\left((x, y, z) - C_k\right)\right)\right)$ is the mathematical model for the stenope projection, as it is calculated by Apero; in this formulation, I includes the intrinsic parameters and $p_0(x, y, z) = \frac{(x,y)}{z}$.

  Considering the stenope model stored in the format generated by Apero (although different generic models can be also defined), the matching in ground geometry aims at defining a height map $Z = Z(x, y)$, that must satisfy two conditions, i.e.: the windows centred on the $I_k\left(\pi_k\left(x, y, Z(x, y)\right)\right)$ should be similar; $Z(x, y)$ should respect the regularity formulation. In this approach, the output geometry is equal to the input one: the resulting image (height map) is, in fact, directly understood as a grid $Z = f(x, y)$. Working with this Euclidian geometry is suitable for the reconstruction of quasi-planar object, such as: the earth surface acquired with aerial imagery, building facades and small planar objects, like bas-reliefs.
- Image-Ground geometry; when the reconstruction of fully 3D objects is needed, a third kind of geometry offered by MicMac may be used, i.e. the image-ground geometry. With this approach, in fact, the user can select a specific geometry for each different point of view. Image-ground geometry is based on a master image I_m defined for each point of view, whose pixels (i, j) represents the geometry (x, y). Among the different image-ground geometry models, the 1D-depht-of-field is generally used, since it is optimized for well calibrated stenope cameras. In this case, the input geometry defines the correspondence between a point (x, y, z) of the Euclidian space and its projection in each image. The output geometry is, vice-versa, defined by the

correspondence between a point (A, B, C) of the result space and its homologous in Euclidian space (x, y, z), so that:

- (A, B) are a pixel of the master image I_m;
- C is the inverse of the depth of field;
- (x, y, z) is the point located on the ray emerging from (A, B) at depth equal to $1/C$.

This geometry, hence, computes a disparity map for each master image (i.e. for each point of view), which can be directly superimposed to the master image itself.

3. Parametrical aspect.

As Apero, also MicMac is a very parametrical software. It is quite impossible to mention all the different parameters that the user can specify into the XML file. The following list includes only the main parametrical choices that have a direct influence on the image matching phase:

- SzW is the size of the correlation window;
- AlgoRegul specifies which regularization algorithm should be used;
- Regul is the regularization factor, defining the weight of the regularization term within the energetic function;
- Px1Pas specifies the quantification step of the disparity map in order to have a matching precision better than the pixel; this quantification process requires, of course, an interpolation approach, since it needs to get values of images at non integer point;
- Px1DilatAlti and Px1DilatPlani are the coefficients of planar and altitude dilatation, defining a range around the solution at 2^{k+1}resolution, where the surface at resolution 2^k is forced to lie.
- ZIncCalc specifies the amplitude of the incertitude interval that should be explored in ground geometry matching; it is expressed as a proportion of the average height computed and sored by Apero;
- BoxTerrain specifies the ground surface on which the matching should be done in ground geometry;
- ZPas defines the altitude resolution; while the planar resolution is selected by MicMac equal to the average resolution of the image (stored in files generated by Apero), the altitude one is forced to be equal to the planar resolution multiplied by ZPas;
- MulZMin and MulZMax specifies, in image-ground geometry, the depth of field interval to be explored, i.e. $[\text{MulZMin} * D_0; \text{MulZMax} * D_0]$, where D_0 is the average depth computed and stored by Apero;
- ZoomF and ZoomI define the final and initial resolution for the pyramidal, multi-resolution image matching approach.

If MicMac is a parametrical software that "reads" the parameters stored in a XML file, Malt is a simplified command-line version of it. Currently, it can handle image matching in both ground geometry and image-ground geometry: the former is specified by the arguments Ortho (for a matching adapted to orthophoto generation) and UrbanMNE (for a matching adapted to

digital surface model generation); the latter is, vice-versa, specified with the argument GeomImage. The user can then specify the values that should be assigned to the parameters associated with the selected geometry; if no value is set, the default one will be applied by the software. In specific cases, the use of Malt should follow the one of other auxiliary tools: for example, in ground geometry a basic rectification of the images should be performed (tool Tarama), after having defined an appropriate photogrammetric reference system (tools SaisieBasc and RepLocBascule).

To run the image matching process (whether with MicMac or with Malt), the tool should decide which space is to be explored. If no auxiliary information is given, the software will adopt a default strategy, that can be summarized as follows: all points of the scene visible on at least N images (being N usually equal to 2 or 3) will be selected, making the assumption that the scene is globally flat. Usually, this creates a useless, too large area, where both the object and its background are considered and matched. To avoid this problem, the user can create a mask on the image (or on each master image), that specifies which terrain points should be matched. For this purpose, the tool SaisieMasq can be employed: it creates both a masked image and a meta-data file that geo-references the masked image on the original one. Of course, the user may prefer to employ other software for the mask definition phase.

The final output of MicMac/Malt computation is a depth map (or disparity map or height map according to the selected geometry). This information, together with the other outputs of all the so far performed processes (internal and external orientations, origin and step of depth quantification, image mask), can be used in order to generate a 3D point cloud. The tool Nuage2Ply of the suite can be used in order to perform this conversion: it back-projects in the object space each pixel of the master image, according to the image orientation parameters and the associated depth (or height or disparity) values. Then, the tool associates to each re-projected 3D point a RGB attribute, directly detected from the master image itself. Thereby, a photo-textured 3D point cloud for each point of view (in the general case of image-ground geometry) is finally achieved.

4. ACTIVE 3D IMAGING AT A GLANCE

4.1 General overview

This Chapter reviews the basic principles that underline active 3D imaging systems, i.e. those non-contact measurement instruments that use an artificial illumination to produce a quantifiable 3D digital representation of a surface in a selected volume of interest and with a particular measurement uncertainty. The title of the Chapter emphasizes that only a very general overview of these systems will be here given, since they will mostly be used only to define adequate reference[5] data of known uncertainty: the core of this research thesis, in fact, is represented by passive 3D vision systems and novel approaches dealing with them.
From an historical point of view, many advances have been made during the last 50 years in the field of solid-state electronics, photonics and computer vision: these active research fields have produced many important changes, leading the growth of active 3D imaging system technology. For example, the availability of affordable and fast digital computers and reliable light sources (such as lasers, halogen lamps and Light Emitting Diode, LED) has offered the possibility of capturing large amounts of 3D data with reliable, accurate and high-resolution 3D active optical sensors. Furthermore, the ability of post-processing the acquired dense point clouds in an efficient and cost-effective way has led the diffusion of these systems in a myriad of different application areas, ranging from military activities, up to medical, industrial, entertainment and commercial ones.
Active optical sensors that employ light waves for 3D measurements can be categorized according to their basic measurement principle. Many different taxonomies exist in the literature on this topic (Nitzan, 1988; Jähne et al, 1999; Drouin and Beraldin, 2012). Here, the schema of Figure 4.1 will be adopted and, according to it, the main instrument classes will be briefly described in the following subsections.

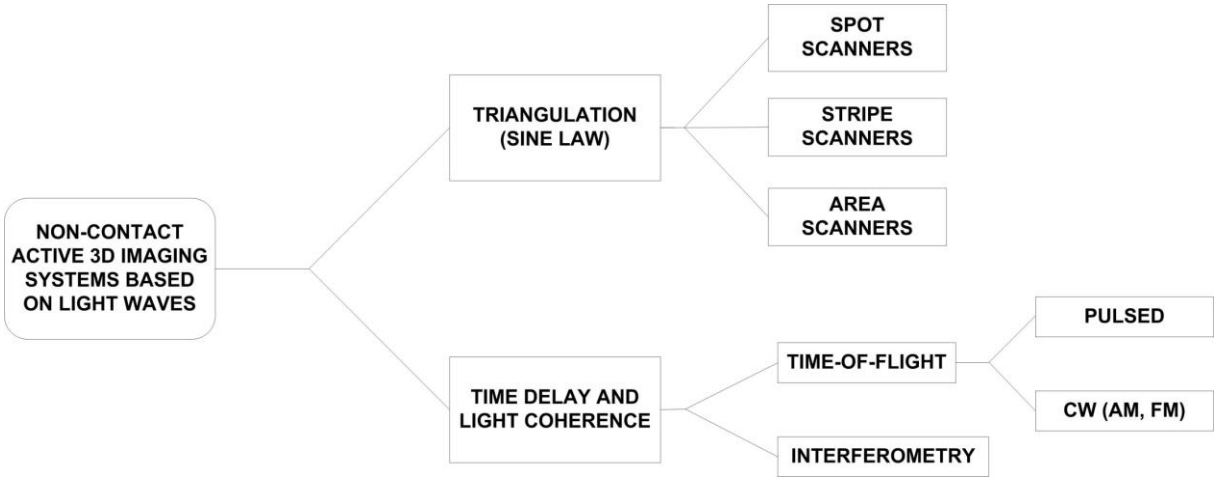


Figure 4.1 Classification of active 3D imaging systems based on light waves (λ from 400 nm to 1600 nm)

[5] Some authors use the expression "ground truth" to signify a reference dataset and the (VIM3) gives explanatory notes on the use of the word "true". Hereinafter, the term "reference dataset" will be however preferred.

4.1.1 Triangulation-based methods

According to (Seitz, 2007), triangulation systems can be described as methods based on geometry; in particular, they refer to the same geometric principle of passive 3D imaging systems, i.e. the intersection of light rays in the 3D space. Compared to passive methods, active triangulation-based systems replace one camera by a projection device, that can be a digital video projector, an analogue slide projector or a laser. The collection of a scattered laser light from the surface is done from a vantage point, that is distinct from the projected light source (Figure 4.2). The 3D measurement is then based on the solution of a triangle, by knowing the projection and collection angles (α and β) and the baseline (H).

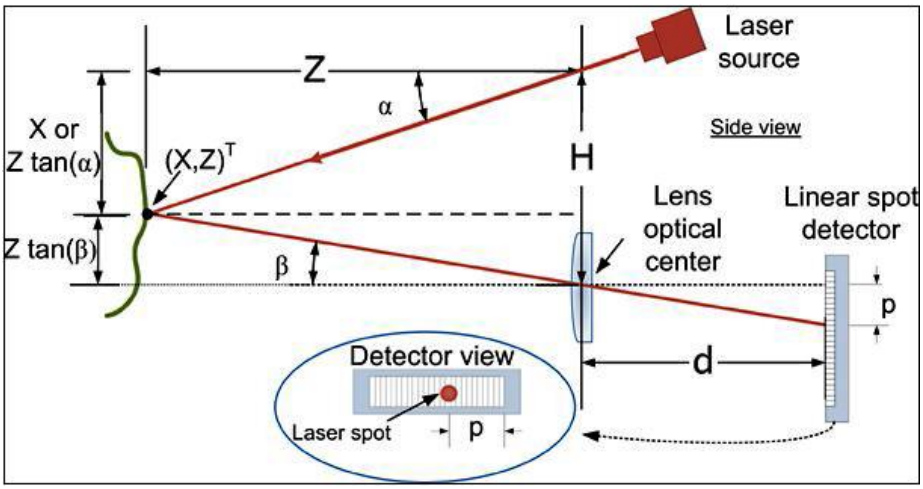


Figure 4.2 Triangulation-based active 3D imaging (single spot)
(Drouin and Beraldin, 2012)

Triangulation-based systems, that usually require a mean scanner-object distance of about 0.1 cm to 500 cm, can be classified according to different distinctive characteristics, such as their opto-mechanical components, construction and performance. Here, the classification defined in (Beraldin, 2004) will be adopted, based on the way in which the active 3D imaging system illuminates the scene. Accordingly, three main categories can be identified: spot scanners, stripe scanners and systems using structured light patterns (Figure 4.3).

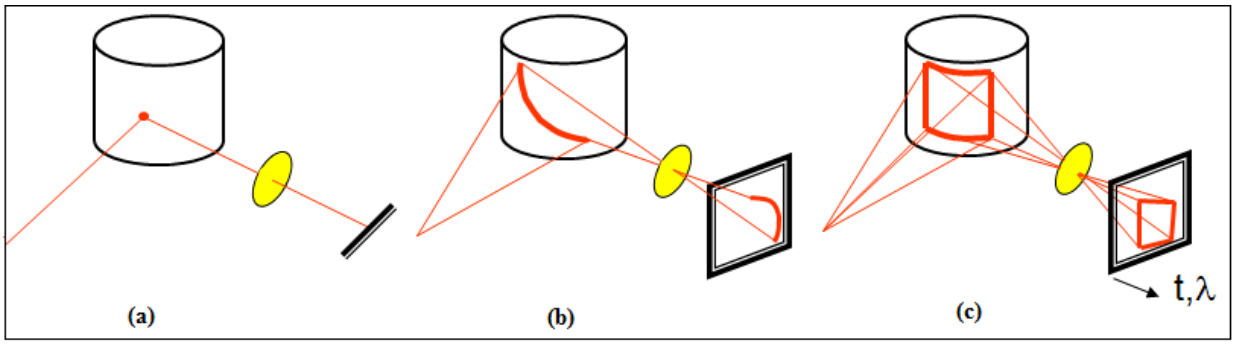


Figure 4.3 The three main categories of triangulation-based methods: spot scanners (a), stripe scanners (b) and area scanners (c)
(Beraldin, J.A., Blais, F., Godin, G., Tutorials from NRC Canada at 3DIM, 1997-2003)

Spot scanners (also called point-based scanners) use a collimated or focused laser beam to illuminate a very small circular or elliptical part of the scene for each measurement capture. This approach (Figure 4.3a) offers many advantages, such as the "elimination" of the correspondence problem: the illumination of the surface is, in fact, spread temporally, i.e. in the temporal dimension. Furthermore, these systems offer the possibility of controlling the spatial sampling of the surface measurements and their laser power can also be modulated according to this 3D sampling. Of course, this comes at the price of requiring an additional opto-mechanical complexity: the laser spot, in fact, should be scanned either by mounting the sensor on an translation bar or by using two galvanometer-mounted mirrors in order to orient the laser along two rotation axes. Limiting this discussion on laser-based spot scanners and referring to the geometrical configuration of Figure 4.2, the linear spot detector can be viewed as an angle sensor, providing signals that are measured as a position p: this position on the linear spot detector is computed using a peak detector. From the knowledge of p and d (i.e. the distance between the laser spot detector and the collection lens, also termed effective focus distance), it is immediately possible to define the collection angle β (in radians) as follows:

$$\beta = \tan^{-1}\left(\frac{p}{d}\right) \qquad [4.1]$$

Using trigonometry, one can then derive:

$$Z = \frac{H}{\tan\alpha + \tan\beta} \quad ; \quad X = Z\tan\alpha \qquad [4.2]$$

and thus

$$Z = \frac{Hd}{p + d\tan\alpha} \qquad [4.3]$$

As previously mentioned, in order to acquire a complete measurement profile, a 2D translation stage can be employed. Otherwise, the laser beam can be scanned around a specific $[X, Z]^T$ position, using a mirror mounted on a galvanometer drive. If this solution is adopted, the scan angle α is then varied according to a pre-defined field of view.

Stripe scanners, also termed profile scanners, illuminate the scene through a "sheet of light", that is usually generated by passing a collimated laser beam through a cylindrical lens. In other solutions, this cylindrical lens is replaced by a diffractive optical element (DOE) or a diffraction grating. The triangulation principle applied to stripe scanners (Figure 4.3b) refers to the intersection between a back projected 3D ray (generated from a pixel in a conventional camera) and a projected 3D plane of light (generated by a digital projector that can be viewed as an inverse camera). This intersection allows the definition of a point in the 3D scene: of course, in order to assemble a complete range image, the projected stripe should be scanned in one direction relative to the scene. This can be achieved through two different technological solutions: the scanner head may be translated or rotated relative to the object. Otherwise, a rotating single mirror may produce the rotation of the laser beam. Of course, the object may also be translated or rotated instead, but this set up is common almost only for industrial 3D

imaging applications. Considering a rotating plane of projected light, an image is acquired for each laser plane orientation. Then, the camera pixels that "view" the intersection between the laser plane and the 3D object are transformed into observation directions. The latters can be obtained in the camera by applying a peak detector on each raw or column of the image (Drouin and Beraldin, 2012).

Finally, **area based scanners** project a structured light pattern (i.e. many planes of light simultaneously) onto the 3D scene (Figure 4.3c). These devices (Jähne et al, 1999) are not really "scanners", since they do not scan the projected light over the object: the scene is, in fact, usually completely illuminated by the projected pattern. This configuration represents a great advantage, because it strongly reduces the acquisition time and, consequently, minimizes distortion effects typical of dynamic scenes. On the other hand, the correspondence problem is here more difficult than it was for the other triangulation-based methods: it is, in fact, an hard task to determine which part of the projected pattern corresponds to which part of the imaged pattern. In order to solve this problem, these systems use a coding strategy that structures the projected light either spatially, temporally or both. Thus, even if many coding strategies can be used to establish the correspondences (Salvi et al., 2004), the two main categories of coding are: spatial coding and temporal coding. Systems based on the latter principle project the patterns one after the other and capture an image for each pattern. In this case, thus, the matching to a specific projected stripe is performed using the time sequence of imaged intensity at a particular location in the scanner's camera. An example of temporal coding strategy is the time-multiplexing codification (Salvi et al., 2004), that is based on intensity measurements and includes, for example, the well-known Gray code. Another temporal coding strategy is the one based on phase measurement. On the other hand, spatial coding techniques project just a single pattern and use this grayscale or colour pattern within a local neighbourhood to perform the required correspondence matching. Also these spatial neighbourhood methods are extensively discussed in (Salvi et al., 2004).

Considering the simpler case of a spot scanner and starting from Equation [4.3], the **range uncertainty** on the Z measurement is, according to the propagation error equation, approximately given by:

$$\delta_Z^2 \approx \left(\frac{Z^2}{H \times d}\right)^2 \delta_p^2 + \left(\frac{Z}{\cos(\alpha)^2}\right)^2 \delta_\alpha^2 \quad \rightarrow \quad \delta_Z \approx \frac{Z^2}{H \times d}\delta_p \qquad [4.4]$$

where δ_p and δ_α are, respectively, the uncertainty in laser spot position and in light ray projection angle. From the equation above, one can find that the Z measurement uncertainty is inversely proportional to both the baseline and the effective focus distance, while it is directly proportional to the square of the distance itself. Thus, by decreasing the range distance Z, one will directly reduce also its uncertainty: unfortunately, this would increase the shadows effect (Figure 4.4), i.e. the presence of scene regions that cannot be measured due to the separation between the laser source and the detector. Furthermore, also H and d cannot be made as large as possible. Increasing the triangulation baseline H would in fact cause stronger self-occlusion

problems (shadows effects) and reduce the stability of the whole system: H is, in fact, mainly limited by the mechanical structure of the optical set up. The effective position of laser spot sensor is limited too: an increase of d will, in fact, reduce the field of view of the sensor (ϕ_{XZ} in Figure 4.5).

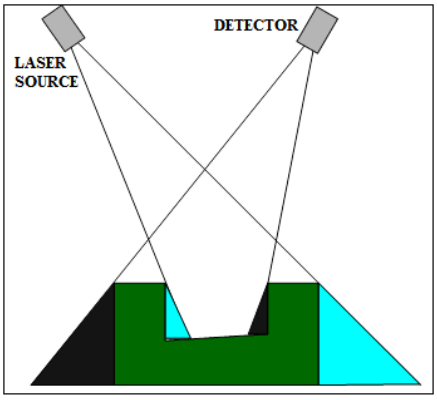


Figure 4.4 Shadow effects

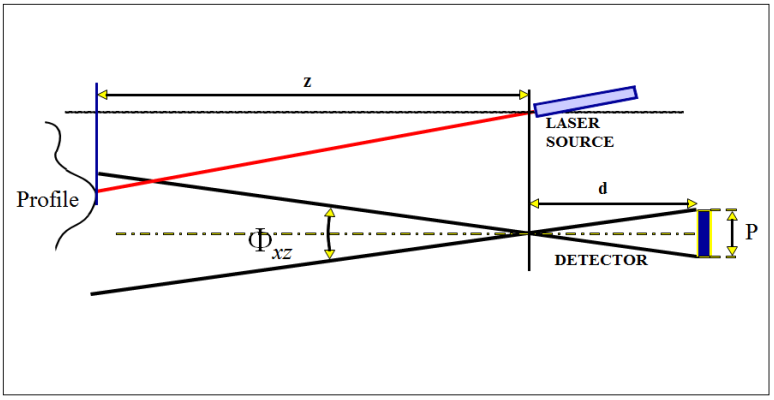


Figure 4.5 Lateral field of view

(Beraldin, J.A., Blais, F., Godin, G., Tutorials from NRC Canada at 3DIM, 1997-2003)

The uncertainty δ_p of the laser image position on the detector depends especially on the following factors (Beraldin. 2004): the type of laser spot sensor used, the peak detector algorithm, the signal-to-noise ratio (SNR) and the imaged laser beam shape. In particular, it is limited by the detector and other electrical noise, the detector inter-pixel gap, the ambient/background light and the laser speckle phenomenon. This latter aspect is the prevalent one in the case of discrete response laser spot sensors, sub-pixel laser spot position estimation and high SNR (Jähne et al, 1999; Amann et al, 2001). In particular, speckle is caused by the interference of many light waves having the same wavelength but different phases. The projection system, in fact, emits different waves that are then reflected by the object surface at slightly different positions: thus, the waves reach the detector with slightly different phases and are finally added together by the detector, that measures a varied intensity. As a consequence, speckle mostly depends on the surface micro-structure or roughness of the object that is acquired. The effect of this speckle noise on spot position uncertainty can be computed as (Baribeau and Rioux, 1991; Dorsch et al., 1994; Beraldin, 2004):

$$\delta_p = \frac{1}{\sqrt{2\pi}} \lambda f_n \qquad [4.5]$$

where λ is the laser wave length and f_n is the receiving lens f-number. This formulation, as mentioned above, is for high values of SNR; the SNR, however, deteriorates rapidly with distance. This deterioration is mainly due to the fact that the amount of light collected decreases with the square of the distance: unfortunately, most of these triangulation-based systems don't have a gain mechanism embedded in the optical sensor. For all these reasons, the maximum distance range of triangulation-based laser scanners, even with a baseline of 1 m, cannot exceed 10 m.

Besides the range uncertainty, also the lateral resolution of a laser scanner is a very important factor to be computed. It is defined as the size of the laser light spot projected on the object (Blais and Beraldin, 2006) and is limited by diffraction. The latter, in fact, causes light waves to spread transversely during their propagation, making it impossible to have a perfectly collimated beam. In other words, laser beam and collected ray cannot be assumed as infinitely thin, since they don't maintain focus with distance. The laser beam propagation can be expressed mathematically as a function of the distance Z, through the following equation (Blais and Beraldin, 2006), that is based on a beam propagation with a Gaussian shape transversal profile:

$$W(Z) = W_0 \sqrt{1 + \left[\frac{\lambda(Z - Z_0)^2}{\pi W_0^2}\right]^2} \qquad [4.6]$$

where Z_0 is the laser focusing distance (i.e. the distance from the lens to the point at which the beam radius is minimal), W_0 is the minimum beam radius at $1/e^2$ irradiance (e is the base of the natural algorithm) and λ is the wavelength of the laser source.
Considering diffraction effects, the lateral resolution can thus be intuitively seen as the capability of the scanner to discriminate two adjacent structures on the illuminated surface. Generally speaking, it depends on both structural and spatial resolutions. The former is directly related to the beam radium and, thus, to the laser beam propagation; the latter, on the contrary, is determined by the smallest possible variation of the scan angle α.

4.1.2 Methods based on time delay and light coherence

Systems based on time delay and light coherence are generally divided into Time-of-Flight (ToF) and interferometry methods: in particular, (Seitz, 2007) describes ToF as systems based on an accurate clock and interferometry as the method that uses accurate wavelengths. The remainder of this subsection will focus on ToF systems, while a description of interferometry-based methods is discussed in (Jähne et al, 1999). In order to give only a general glance at interferometry, however, one can say that it is based on the superposition of two beams of light: in fact, a laser beam is usually split into two paths. One path is characterized by a known length, while the length of the other one is unknown: thus, this difference between the path lengths creates a phase difference between the light beams. Finally, the two beams are combined together before reaching the detector. The interference pattern seen by the detector and resulting from the superposition of the two light beams can then be converted into a distance measurement, since it depends on the path difference. Other interferometry-based commercially available systems are, vice-versa, based on other principles, such as conoscopic holography. Interferometry systems, finally, are characterized by a very wide distance range.

Generally speaking, methods based on time delay (Figure 4.6) use a fundamental property of a light wave, i.e. its velocity of propagation: in a given medium, in fact, light waves travel with a finite and constant velocity. Thus, as their names suggest, these systems try to evaluate

the time delay created by light travelling in a specific medium from a source to a reflective surface and back to the source again: this measurement is, in fact, obviously related to the covered distance. If this is the common basic principle, different strategies have then been developed in order to exploit it: ToF systems can thereby be divided into pulsed sensor and systems based on Continuous Waves (CW), which in turn can be split into Amplitude Modulation (AM) and Frequency Modulation (FM) systems. Concerning the scanner-object mean distance, finally, ToF systems can operate between 100 cm up to several km (long-range sensors), depending on the implemented measurement principle.

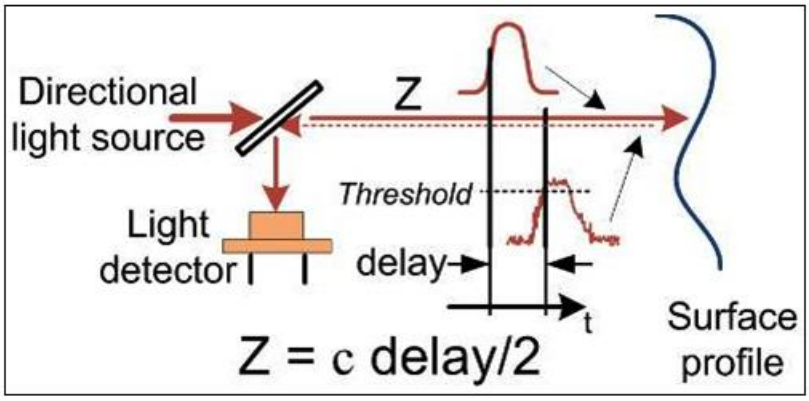


Figure 4.6 Methods based on time delay
(Beraldin, J.A., Blais, F., Godin, G., Tutorials from NRC Canada at 3DIM, 1997-2003)

Tim-of-Flight **pulsed sensors** measure the camera to object distance Z, by sending a relatively short impulse of light on a reflective surface and measuring the round trip transit time τ, deriving

$$Z = \frac{c}{n}\frac{\tau}{2} \qquad [4.7]$$

where c is the speed of light (equal to 299,792,458 m/s in a vacuum[6]) and n is a correction factor that depends on the properties of the medium in which the beam travels. If the light waves travel in air, n is equal to the air refractive index, that depends on the air temperature, pressure and humidity: it is generally fixed at 1.00025.
Thus, the range uncertainty for a single pulse, δ_{r-p}, is approximately given by the following equation:

$$\delta_{r-p} \approx \frac{c}{2}\frac{T_r}{\sqrt{SNR}} \qquad [4.8]$$

where T_r is the pulse rise time and SNR is the above mentioned signal-to-noise ratio. Furthermore, averaging N different measurements will reduce δ_{r-p} by a factor proportional to the square root of N, as follows:

[6] It is **exactly** 299,792,458 m/s, since the length of the meter is defined from this constant and the international standard for time.

$$\delta_{r-p} \approx \frac{c}{2}\frac{T_r}{\sqrt{SNR}}\frac{1}{\sqrt{N}} \quad [4.9]$$

Finally, for a single pulse and high SNR, the uncertainty in range estimation is given by:

$$\delta_{r-p} \approx \frac{c}{2}\frac{T_r}{\sqrt{SNR} \times BW} \quad [4.10]$$

where BW is the root mean square signal bandwidth.
From the above equations, one can derive that the range uncertainty may be decreased by increasing the SNR and/or the effective signal bandwidth. In particular, the increase in bandwidth corresponds to a signal pulse with sharp edges, thus offering a better discrimination against background noise (Beraldin, 2004). Anyway, a more accurate estimation of the range uncertainty may be derived by including walked errors caused by variations in pulse amplitude and shape (Amann et al., 2001). Finally, ToF systems are characterized by an ambiguity interval, that depends on the time spacing between consecutive pulses.

Other methods based on time delay get around the measurement of short pulses by modulating the power or the wavelength of emitted Continuous Wave (CW). Among them, **Amplitude Modulation (AM)** systems project the modulated signal onto a surface and collect the scattered light into a single photodiode. A circuit then measures the phase difference between the two waveforms, that represents in fact a time delay.
The measurement uncertainty of AM systems is approximately given by:

$$\delta_{r-AM} \approx \frac{1}{4\pi}\frac{\lambda_m}{\sqrt{SNR}} \quad [4.11]$$

where $\lambda_m = c/f_m$ is the wavelength of the amplitude modulation.
The above equation shows that a low frequency, f_m, corresponds to an higher range uncertainty, because it makes the phase detection less reliable. Furthermore, it is not possible to compute the absolute distance measurement from a simple AM approach: the collected wave, in fact, cannot be directly associated with a specific part of the original signal. This limiting factor is called range ambiguity interval, and is equal to $\lambda_m/2$. In order to solve this problem, AM systems usually employ multiple frequency waveforms: in a two-tone AM system, for example, a low frequency (10 MHz) and a high frequency (150 MHz) are both emitted into the scene.

Finally, CW systems can also be based on **Frequency Modulation** methods, that modulate the frequency of a laser radar with coherent detection. In particular, the frequency of the laser beam is linearly modulated either directly at the laser diode or using an acousto-optic modulator. In both cases, the linear modulation has generally a triangular or saw-tooth shape and is known as a chirp. Among the different key-aspects of this technology, one should cite:

- The optical detector performs a coherent detection;

- The resulting beat frequency "encodes" the time delay using a much smaller bandwidth compared to ToF systems (Schneider et al., 2001);
- An absolute distance measurement can be achieved.

A mathematical formulation of the range uncertainty associated with these systems is not here provided; however, as reference data, one can say that most of the commercially available systems usually have a measurement uncertainty of about 40 μm over a range of 2-10 m, with a data rate of 1000 points/second (Beraldin, 2004).

4.2 Experimental characterization

Since some tests on laser scanner performance will be presented in the following chapters, a brief overview on experimental characterization of 3D active imaging systems will be here provided. Without any international standard about these optical 3D measurement systems, both manufacturers and end-users are very interested in verifying that their scanner really performs within predetermined specifications. Generally speaking, the most common experimental practices use known objects as reference data and characterize the instrument through one of the following two approaches:

- An object of simple shape, like a plane or a sphere, is first manufactured with great attention; then, it is scanned with the 3D imaging system and these measurements are finally compared with the nominal values;
- A different instrument, such as a Coordinate Measuring Machine (CMM), is used to characterize the manufactured and scanned object; then, the measurements acquired with the scanner are compared with those achieved with the CMM. With this approach, the measurements acquired by the reference instruments should be an order of magnitude more accurate than those acquired with the scanner.

The most important factors that have a direct impact on the uncertainty of a 3D imaging system are summarized in (Beraldin, 2009), where the following list of uncertainty sources is provided: hardware means, software means, method, ambient, material and people. The latter, in particular, represents a significant source of error in the measurement chain. It is usually due to a lack of experience, training and understanding of the performance limitations of the instruments.

In order to characterize 3D imaging systems, four different types of experimental tests will be briefly described in this section, assuming that the range images generated by the scanner are composed by point clouds arranged in a grid format. A more detailed discussion on this topic can be found in (Drouin and Beraldin, 2012).

The first type of test provides a **low-level characterization** and examines the errors contained in a small area of the 3D point grid. For example, multiple scans of a planar surface at different positions in the reconstruction volume can be performed and analysed: the flatness measurement error is thereby estimated. These tests are not affected by instrumental miscalibration, especially if the area used to fit a plane is small with respect to the

reconstruction volume. Thus, these experimental approaches make it possible to identify systematic errors, that are independent of the calibration. In order to decorrelate the different error sources, the error analysis should be performed using the raw output of the scanner, without any further post-processing phase.

The second type of test provides a **system-level characterization** and looks at the errors introduced when examining the interaction among different small portions of the point grid. An example is represented by angular measurements performed between intersecting surfaces: at first, planes are fitted to the 3D points acquired by the scanner, then the angles between them are measured. An alternative approach refers to sphere-to-sphere measurement and, usually, requires two spheres mounted on a bar with a known centre-to-centre distance: this bar is then placed and acquired at different significant positions in the reconstruction volume. The errors of centre-to-centre distance are finally employed to characterize the scanner performance. This kind of analysis is significantly affected by miscalibration effects.

The third category of test provides a **characterization of errors caused by surface properties** and evaluates the impact of object surface properties on the reconstructed geometry. A very common approach is based on resolution charts, that are also used to assess the lateral resolution of cameras and scanners. In this case, the chart may be employed in order to evaluate the impact of sharp intensity variations on the metric performance of scanners: texture changes, in fact, may cause errors in the surface geometry. Furthermore, besides the object micro-structure surface, also the spectral distribution of the light source can greatly influence the accuracy of the recovered geometry. In order to evaluate these kinds of effects, an optical flat surface can be scanned with the same instrument, but using different light sources.

Finally, an **application-based characterization** can be performed in order to evaluate the fitness of a particular scanner to a specific task. This last family of test is typical of industrial applications: its objective is to identify defects that create an unacceptable variation on the surface of a product. For example, an object with known defects is scanned and the capability of the systems to localize these defects is assessed: this analysis characterizes the performance of the system in imaging small structural details.

5. DIGITAL CAMERA CALIBRATION PROCEDURES

5.1 Introduction

In order to gather accurate and metrically reliable information from digital images, perturbation effects affecting the real camera systems should be adequately formulated and parameterized. If within an ideal camera model, in fact, the perspective centre, the object point and the corresponding image one should lie along the same straight line, in real physical cameras, however, this collinearity condition cannot be perfectly achieved, especially if a low-cost lens or a short focal length lens (such as a fisheye) are employed. This problem should be solved within the photogrammetric pipeline through an adequate camera calibration procedure, that allows a mathematical parameterization of these departures from collinearity, by modelling the associated perturbations caused by different physical sources, such as lens distortions and atmospheric refraction. For this reason, the issue of camera calibration has always played an essential role in the field of photogrammetric measurements, especially with the rapid development and diffusion of consumer-grade digital cameras, that are increasingly employed for metric purposes thanks to their significant image quality (especially in terms of resolution and signal-to-noise ratio). A general overview of the topic is provided in Chapter 2 (Subsection 2.1.3), where a discussion of the two main calibration approaches, i.e. test-range calibration and self-calibrating bundle adjustment, is detailed together with a description of the general rules required to avoid system instability and parameter coupling.

These observations apply, of course, also to low-cost and open-source software solutions for image-based 3D modelling, that can achieve a metrically reliable 3D reconstruction of the object only if they adequately deal with the issue of camera calibration. In this context, the IGN's suite of tools provides the user with the possibility of performing a camera self-calibration within the bundle adjustment procedure, by adopting one of the several different calibration models proposed therefor (Section 3.3). In order to assess the accuracy of this calibration approach, many validation tests were performed during the three years of PhD research, starting from the results achieved in the Master's Thesis (Toschi, 2010). The attention is here paid to the study of the algorithmic performance in dealing with different image datasets and environmental conditions (measure reproducibility). The repeatability of such a method is analysed too. Since the digital camera employed in the experiments is not a metric-one, no manufacturer's specifications could be here used as reference values. A different approach was thus adopted, by performing a classical test-range calibration procedure with a pre-surveyed laboratory test-field and a commercial photogrammetric software. Resulting parameters and distortion profiles were then employed in order to evaluate the metric uncertainty of the algorithmic solution provided by the tested open-source tools: in particular, results achieved with the two software were analysed through different procedural strategies, based on both direct and *a-posteriori* validation tests. The general workflow followed in this case-study is discussed in Subsection 5.1.1, whereas Subsections 5.1.2 and 5.1.3 provide a description of the laboratory test-field and image acquisition

protocols adopted in the tests. A summary of the whole experiment can be found in (Toschi et al., 2013), together with a discussion on the results thereby achieved.

5.1.1 Procedural workflow

Figure 5.1 describes the synthetic procedural workflow, designed in order to perform a metric evaluation of the tested open-source procedure for digital camera calibration.

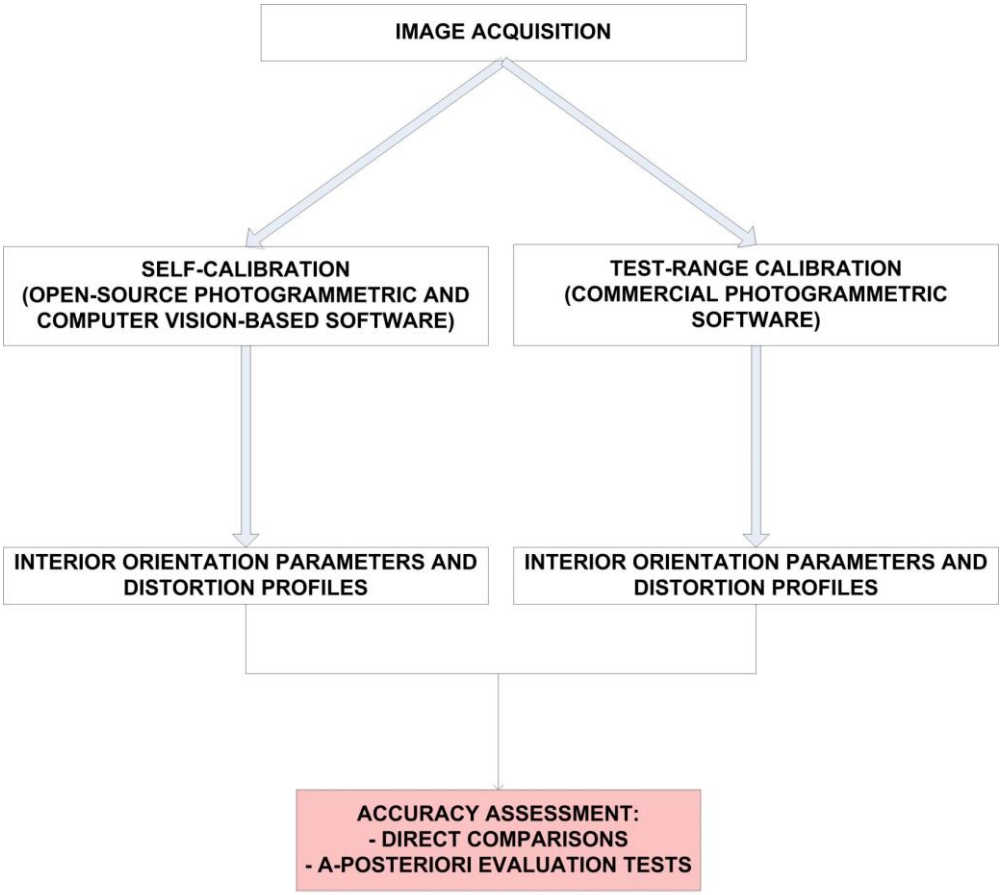


Figure 5.1 Procedural workflow (Accuracy assessment of camera calibration procedures)

Starting from the same image datasets, the camera interior parameters (i.e. focal length and coordinates of principal point) and radial distortion profiles were computed with the self-calibration procedure provided by the tool Tapas and through a separate stand-alone photogrammetric calibration performed with an *ad-hoc* commercial software. Many tests were carried out by adopting different combinations of images: computed calibration results were then statistically analysed in order to evaluate the dispersion associated with significant parameters within each tested approach. The self-calibration performance was also studied by including different objects within the acquired 3D scene. Afterwards, the accuracy assessment of the open-source procedure was carried out through direct comparisons, where both calibration results and derived statistical parameters were considered. In this studies, results achieved via the classical test-range calibration were assumed as reference values. The

analysis was finally completed by preforming further validation tests specifically designed in order to assess the influence of the computed calibration results on the image orientation accuracy, using some externally measured check points as reference observations.

5.1.2 Laboratory test-field

As mentioned, calibration parameters were firstly recovered with a test-range approach, that requires the use of an adequate object space control field of known XYZ coordinates. A laboratory test-field was thus specifically designed in order to provide a 3D object point array, homogeneously distributed along the three orthogonal directions. Figure 5.2 shows a view of this control field, with the identification codes of the measured points.

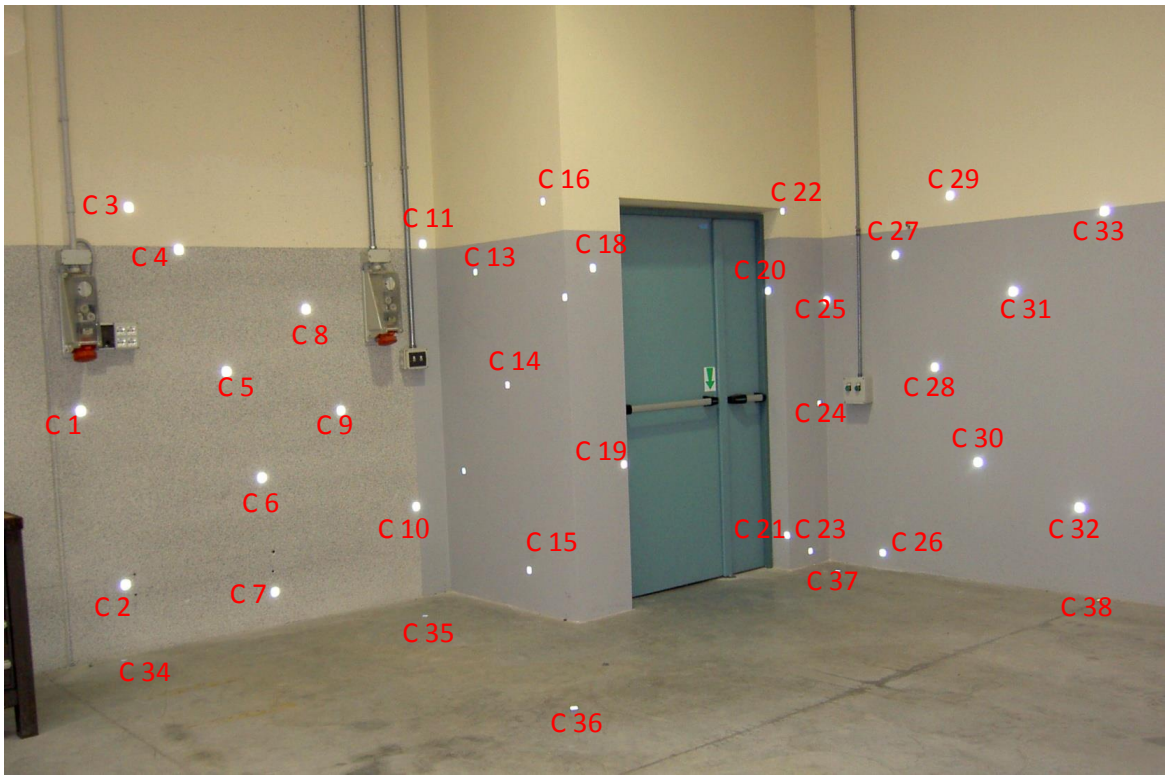


Figure 5.2 Calibration test-field

The test-field is constituted by 38 square targets (2 cm per side) glued on the walls of an indoor mechanical laboratory, built at the Department of Engineering "Enzo Ferrari" (Modena, Italy). In particular, a corner of a the building is specifically selected for the control field, in order to fulfil the following requirements:

- Presence of many depth variations through the consecutive succession of several mutually orthogonal planes. Some targets were placed also on the floor, in order to complete the 3D array configuration;
- High availability of open space in front of the test-field, in order to allow the calibration of several lenses with appropriate image acquisition configurations, that require different camera-object distances;

- Environmentally controlled boundary conditions, even if a specific system for temperature and humidity monitoring is not provided yet. The indoor space, limited by thick walls of reinforced concrete, is, in fact, characterized by stable environmental conditions; furthermore, the possibility of having a fixed and homogeneously diffused illumination is provided too.

The XYZ coordinates of the targets were measured with the Total Station Leica TPS1201+ by Leica Geosystems, whose main technical specifications are listed in Table 5.1. A redundant number of observations (distances, horizontal and vertical angles) were measured from three well-distributed survey stations and then statistically compensated with STAR*NET vs 6.0.36, a least square survey adjustment software developed by MicroSurvey, formerly STARPLUS SOFTWARE INC. (STAR*NET). The 3D positions of all measured targets were thereby computed with a sub-millimetre accuracy.

LEICA TPS1201+	
ANGLE MEASUREMENT	
PRECISION	1” (Hz) – 1” (V)
COMPENSATOR SETTING ACCURACY	0.5”
DISTANCE MEASUREMENT PRECISION	
WITH REFLECTOR	1 mm + 1.5 ppm
WITHOUT REFLECTOR	2 mm + 2 ppm

Table 5.1 Technical specifications of LEICA TPS1201+

5.1.3 Image acquisition

A Canon EOS 5D Mark II (5616 x 3744 pixel) mounting the zoom lens CANON EF 16-35mm f2.8L USM was employed for the tests. The focal length was fixed at maximum zoom (narrowest angle). The technical specifications of the photogrammetric equipment are listed in Table 5.2.

Canon EOS 5D Mark II	
BODY TYPE	Mid-Size SRL
SENSOR RESOLUTION	21 Mpixel
SENSOR SIZE	Full Frame
SENSOR TYPE	CMOS

ISO	Auto, 100-6400 in 1/3 stops, plus 50, 12800, 25600 as option
MIN SHUTTER SPEED	30 sec
MAX SHUTTER SPEED	1/8000 sec
Canon EF 16-35mm f2.8L USM	
FOCAL LENGHT	16-35 mm
MAX APERTURE	f2.8
MIN APERTURE	f22.0
MIN FOCUS	0.28 m

Table 5.2 Technical specifications of Canon EOS 5D Mark II and lens employed

Starting from results previously achieved with the same photogrammetric equipment (Toschi, 2010), images were always acquired at fixed zoom and focus settings (35 mm, infinity focus), in conformance with the following photogrammetric rules (Fraser, 2001): multiple photo stations with varying camera-object distances, different roll angles (horizontal, vertical, oblique), great image point density and covering the entire image format with measured grid points. During the acquisition phase, a photographic tripod was used and a constant homogeneous illumination was provided to the 3D scene.
A total of 40 convergent images were thereby acquired and divided into seven classes that correspond to seven different camera setups, i.e.:

- Horizontal camera, central test-field[7];
- Horizontal camera, bottom test-field[7];
- Horizontal camera, upper test-field[7];
- Vertical camera (+/- 90°), left test-field[7];
- Vertical camera (+/- 90°), right test-field[7];
- Partially rotated camera (+30°);
- Partially rotated camera (-30°).

Figure 5.3 summarizes the above mentioned camera setups, by showing one representative image for each class. The seven classes were finally employed in order to group the original 40 images into 20 different combinations, each containing the same number of shots extracted from each class. This criterion allowed the selection of 20 different image datasets that are comparable to each other's in terms of both adopted acquisition protocol and number of total shots (20).

[7] The first name refers to the camera orientation, whereas the second name refers to the position of the test-field within the image

Figure 5.3 The seven different camera setups

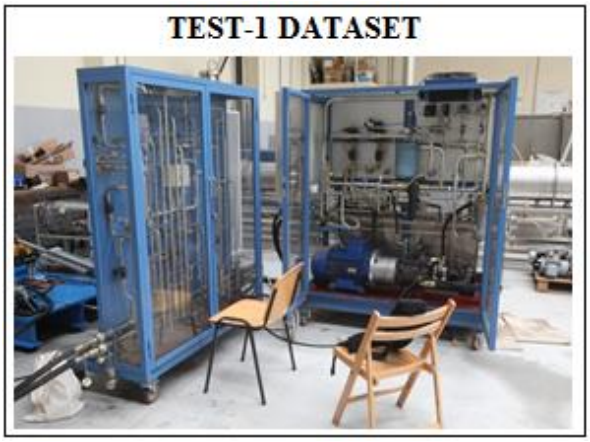


Figure 5.4 A representative image of the Test-1 dataset

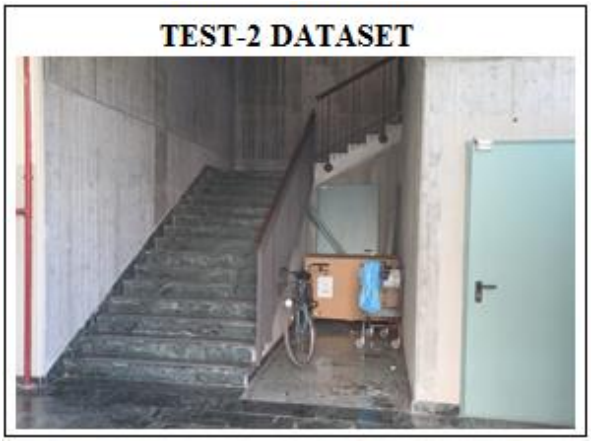


Figure 5.5 A representative image of the Test-2 dataset

Two additional image datasets were then acquired in order to analyse the performance of the tested self-calibration procedure in dealing with different 3D scene. The first dataset (referred to as "Test-1 dataset") is constituted by 11 images of a group of mechanical instruments and equipment (Figure 5.4): this 3D scene includes some reflective surfaces that may represent possible noise sources within the image-based processing pipeline. The second test-object (Figure 5.5) consists of a low-textured stairway and the corresponding image dataset (referred to as "Test-2 dataset") includes 6 images.

Both datasets were acquired by following the above mentioned basic photogrammetric rules under natural conditions of illumination.

5.2 Image processing

The 20 test-field datasets were initially processed using the commercial photogrammetric software MicroMap vs 2.0.0.143 (MicroMap), released by Geoin. This software consists of a suite of different modules (e.g. image orientation, stereo-model editing, orthophoto production, DTM generation and aerial triangulation), that cover the entire photogrammetric and cartographic production pipeline. For the recovery of the interior camera parameters, a test-range calibration procedure with a general spatial resection approach is provided by the software, that computes thereby three parameters of interior orientation (calibrated focal length, c, and principal point coordinates, x_{PP}, y_{PP}) and four parameters for optical distortion parameterization (image resolution and three coefficients of symmetric radial distortion). Only the radial term of lens distortion is modelled, since it represents the prevalent effect in common digital cameras (Fraser, 2001). In particular, the associated symmetrical radial distortion profile is computed through the following polynomial balanced formulation:

$$\Delta r = A_1 \times r^2 + A_2 \times r^4 + R_0 \times r^6 \qquad [5.1]$$

where r is the radial distance from the principal point (PP) and (A_1, A_2, R_0) are the provided distortion coefficients. In order to compute the above mentioned calibration parameters, the user should collimate at least three common points in the images and associate these 2D positions with their corresponding measured 3D coordinates: the algorithm is then able to refine the initial camera model assumption (here consisting of the ideal pinhole camera formulation) through a spatial resection approach. This procedure was carried out for each of the 20 image combinations and a statistical analysis was finally performed on the derived results. Table 5.3 lists the standard deviations delivered by the experiment.

TEST-RANGE CALIBRATION	
Parameter	Standard Deviation
c	4.570 pixels
x_{PP}	1.009 pixels
y_{PP}	1.049 pixels
A_1	0.000038116438427655
A_2	0.000000213853841987
R_0	0.000000000451779118

Table 5.3 Standard deviations computed by analysing the results achieved from the 20 test-field datasets with the test-range calibration approach (pixel size = 6.4 μm)

The IGN photogrammetric tools were then employed to process the same 20 test-field data-sets. The acquired 3D scene, i.e. the laboratory control field, doesn't actually represent a

favourable test-object for the image matching algorithm, since it is characterized by homogeneous and low-textured surfaces. However, it was selected for the tests in order to allow the recovery of calibration parameters with both procedures (test-range and self-calibration approaches) by starting from exactly the same image datasets. Furthermore, the possibility of computing stable and metric reliable calibration results using the tested open-source tool and a representative 3D scene for indoor applications, is thereby assessed too.
In order to extract sparse correspondences between all possible pairs of images, the SIFT^{++} implementation of SIFT algorithm provided by the tool Tapioca (Subsection 3.3.2) was initially employed. The original image resolution (5616 x 3744 pixel) was always adopted within the computations. Afterwards, starting from the detected tie points, the camera calibration parameters and relative poses were computed with the tool Tapas (Subsection 3.3.3). The automated calibration approach starts from a typical initialization phase, where input values are derived from the image EXIF (EXchangeable Image File) information and by adopting the ideal camera model (no distortion effects and no principal point offsets). These parameters are then refined within the bundle adjustment procedure by exploiting the redundant observations provided by homologous points. Among the different proposed internal calibration models, the FraserBasic formulation was chosen: as discussed in Chapter 3, this model considers four different physical sources of perturbations, resulting in a 10-DoF (Degree of Freedom) mathematical formulation. For each dataset, three parameters of interior orientation (calibrated focal length, c, and principal point coordinates, x_{PP}, y_{PP}), three coefficients of symmetric radial distortion (K_1, K_2, K_3), two coefficients of decentring distortion (P_1, P_2) and two coefficients of affine terms (b_1, b_2) were computed. A statistical analysis was finally performed on the derived results, delivering the standard deviations listed in Table 5.4.

SELF-CALIBRATION	
Parameter	Standard Deviation
c	11.461 pixels
x_{PP}	8.677 pixels
y_{PP}	6.588 pixels
K_1	1.17479874231822000E-10
K_2	1.94348998880236000E-17
K_3	1.25938249704315000E-24
P_1	7.27522828859090000E-08
P_2	5.05721138960506000E-08
b_1	9.98636033733927000E-05
b_2	1.12490338386486000E-04

Table 5.4 Standard deviations computed by analysing the results achieved from the 20 test-field datasets with the self-calibration approach (pixel size = 6.4 μm)

The statistical parameters delivered by the analyses offered the possibility of estimating the repeatability of the two tested calibration procedures (test-range and self-calibrating approaches), i.e. the property that describes the agreement within sets of measurements acquired under the same conditions in terms of equipment, place, time and human factors (NPL, 2010). The stand-alone camera calibration performed by employing the commercial software and the total station-derived information provides an higher level of stability, if the attention is focused on the standard deviations associated with the calibrated focal length and principal point coordinates (i.e. those parameters that can be directly compared between the two procedures). Anyway, the differences between the corresponding standard deviations derived from the two calibration approaches are not metrically significant.
Finally, the automatic calibration procedure offered by the IGN tools was also used to process the Test-1 and Test-2 image datasets, by adopting the same above mentioned procedural choices.

5.3 Accuracy assessment

Starting from the results achieved in the tests, an accuracy assessment of the automatic self-calibration procedure provided by the IGN's suite of tools was performed. Initially, the parameters computed via the test-range approach were adopted as reference data and compared to the ones delivered by the self-calibrating bundle adjustment. This required some mathematical operations aimed at making the results comparable and homogeneous. First of all, the radial symmetrical distortion profiles (Δr) associated with the Tapas-derived calibration parameters were computed using the following odd-ordered polynomial series truncated at the seventh order term:

$$\Delta r = K_1 \times r^3 + K_2 \times r^5 + K_3 \times r^7 \qquad [5.2]$$

Then, resulting Gaussian distortion profiles were properly balanced through the introduction of a linear term, as mathematically described in (Fraser, 2011). An associated change, Δc, was added to the calibrated focal length in accordance with the above quoted formulation.
Finally, within the huge amount of available data and results, the following comparison strategies were adopted:

1. Comparisons between corresponding mean values[8] and associated standard deviations[8].
   The statistical parameters associated to calibrated focal length and principal point coordinates, derived from the tests performed with the two calibration procedures (simply termed MicroMap and MicMac) and the test-field datasets, are summarized in Table 5.5. The corresponding symmetrical radial distortion profiles are shown in Figure 5.6. Besides the above mentioned discussion on the repeatability of each measure, these direct comparisons show a good match between the automated simplified calibration results (Tapas procedure) and those computed with the test-range calibration of MicroMap.

[8] $\bar{n}$ is the mean value associated with the generic parameter n, whereas σ_n is its corresponding standard deviation

PARAMETER	MICROMAP	MICMAC
$\bar{c}$ (mm)	33.941035	34.061128
σ_c (mm)	0.029250	0.073464
$\overline{x_{PP}}$ (pixels)	2772.478	2741.743
$\sigma_{x_{PP}}$(pixels)	1.009	8.677
$\overline{y_{PP}}$ (pixels)	1865.569	1876.766
$\sigma_{y_{PP}}$(pixels)	1.049	6.588

Table 5.5 Comparison between mean values and corresponding standard deviations (test-field datasets)

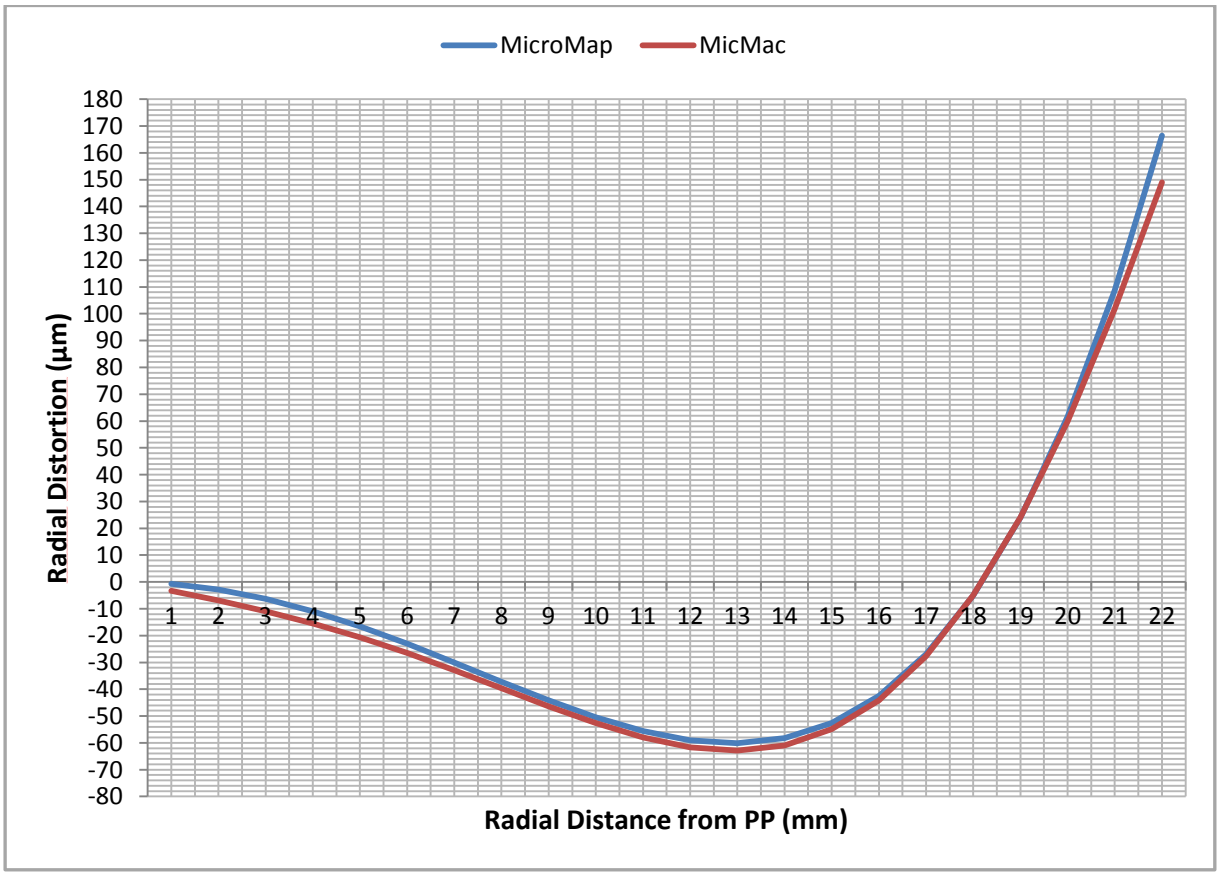


Figure 5.6 Radial distortion profiles computed from mean values (test-field datasets)

2. Comparisons between results achieved in two specific image combinations (starting from the test-field datasets) and those delivered by the Test-1 and Test-2 datasets.
   Among the 20 different test-field image combinations, combination no.11 was selected for the MicroMap procedure, whereas combination no.8 was chosen for the Tapas calibration approach (simply termed MicMac). These datasets were picked up among the others because their values of calibrated focal length, c, are closest to the mean value computed for the 20 combinations: in fact, at first analysis, c can be considered as an affordable and stable calibration parameter. Afterwards, results achieved with the self-calibrating bundle adjustment performed with Tapas and the Test-1 and Test-2 datasets were included within the compared entity. Table 5.6 and Figure 5.7 summarize the described comparisons.

	MICROMAP: COMBINATION NO. 11	MICMAP: COMBINATION NO. 8	MICMAC: TEST-1 DATASET	MICMAC: TEST-2 DATASET
c (mm)	33.940610	34.056203	34.122037	33.893903
x_{PP} (pixels)	2772.459	2735.911	2763.841	2762.607
y_{PP} (pixels)	1866.548	1876.906	1850.951	1852.287

Table 5.6 Comparison between results achieved with two image combinations of the test-field dataset and Test-1, Test-2 image datasets

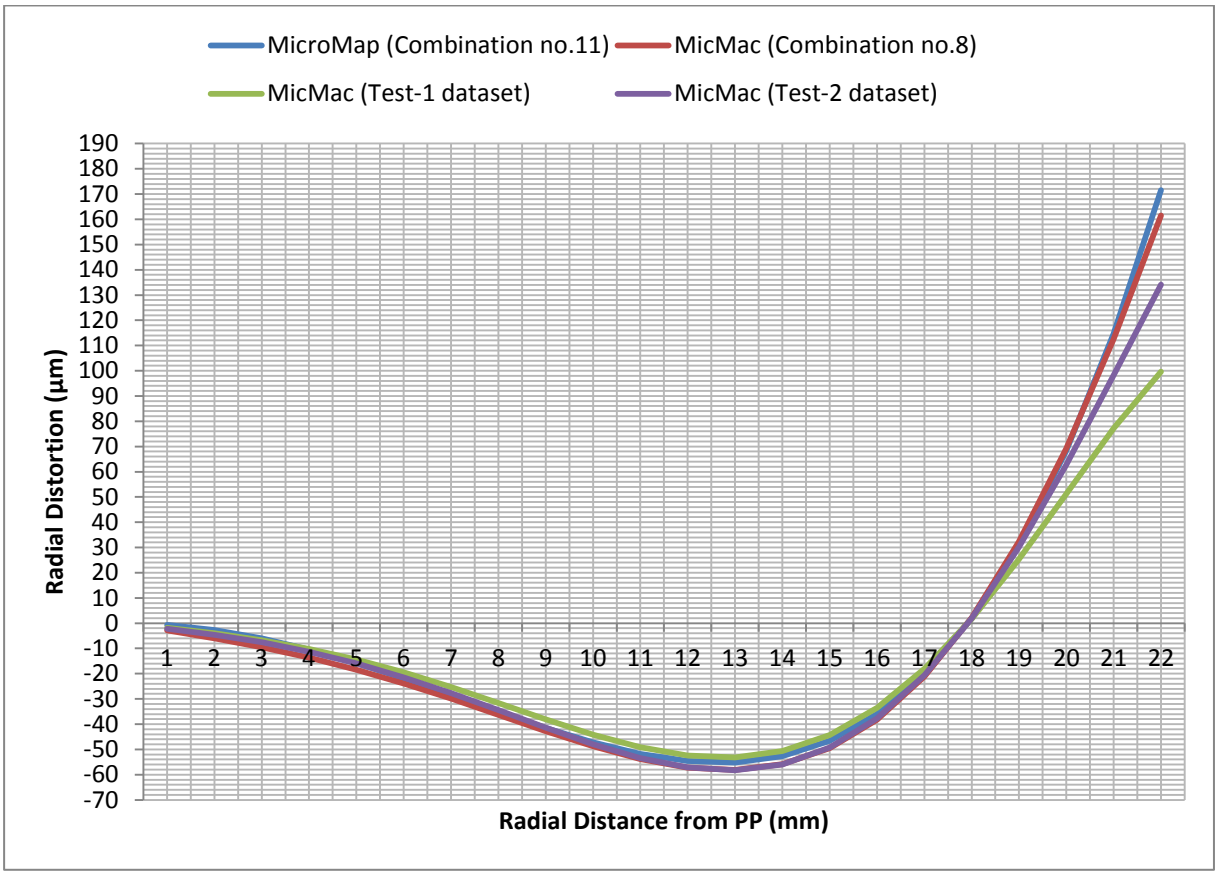


Figure 5.7 Radial distortion profiles computed from results achieved with two image combinations of the test-field dataset and Test-1, Test-2 image datasets

Looking at all the compared results, it is possible to point out again the same evidence discussed within the previous comparison approach: the values computed through the automated self-calibration procedure and with the three different image datasets are metrically comparable to the ones derived from the stand-alone calibration approach. Furthermore, by analysing the achievements of the MicMac procedure, their reproducibility can be discussed, i.e. the property that, generally speaking, describes the agreement within a set of measurements acquired under different conditions in terms of equipment, place, time and human factors (NPL, 2010). In this case, the three tests performed with the IGN's suite of tools involved the use of datasets characterized by different acquired 3D scene, number of images and boundary environmental conditions.

Results thereby achieved don't show significant differences, showing the good stability of the procedure even if noisy 3D objects are included in the scene (e.g. low-textured or reflective surfaces) and a lower number of images is employed.

In order to verify these preliminary findings using externally-measured observations as adequate reference data, an additional validation strategy was finally carried out. This approach was based on the assumption that, even if distortion profiles and calibration parameters can be metrically analysed and compared to some sort of reference values, from a photogrammetric standpoint the main quality indication of a calibration procedure is the object point accuracy and residuals computed from the derived calibration parameter set (Remondino and Fraser, 2006). The laboratory test-field was thus employed as adequate control field, whose measured targets assumed the role of independent check points. These *a-posteriori* validation tests were performed with the software MicroMap using two new images of the lab test-field acquired in stereoscopic-mode and adopting the same photographic parameter setup selected for the "calibration" image datasets. This stereo-pair was then oriented with seven well-distributed targets of the test-field, chosen as Ground Control Points (GCPs) and manually collimated on the two images.

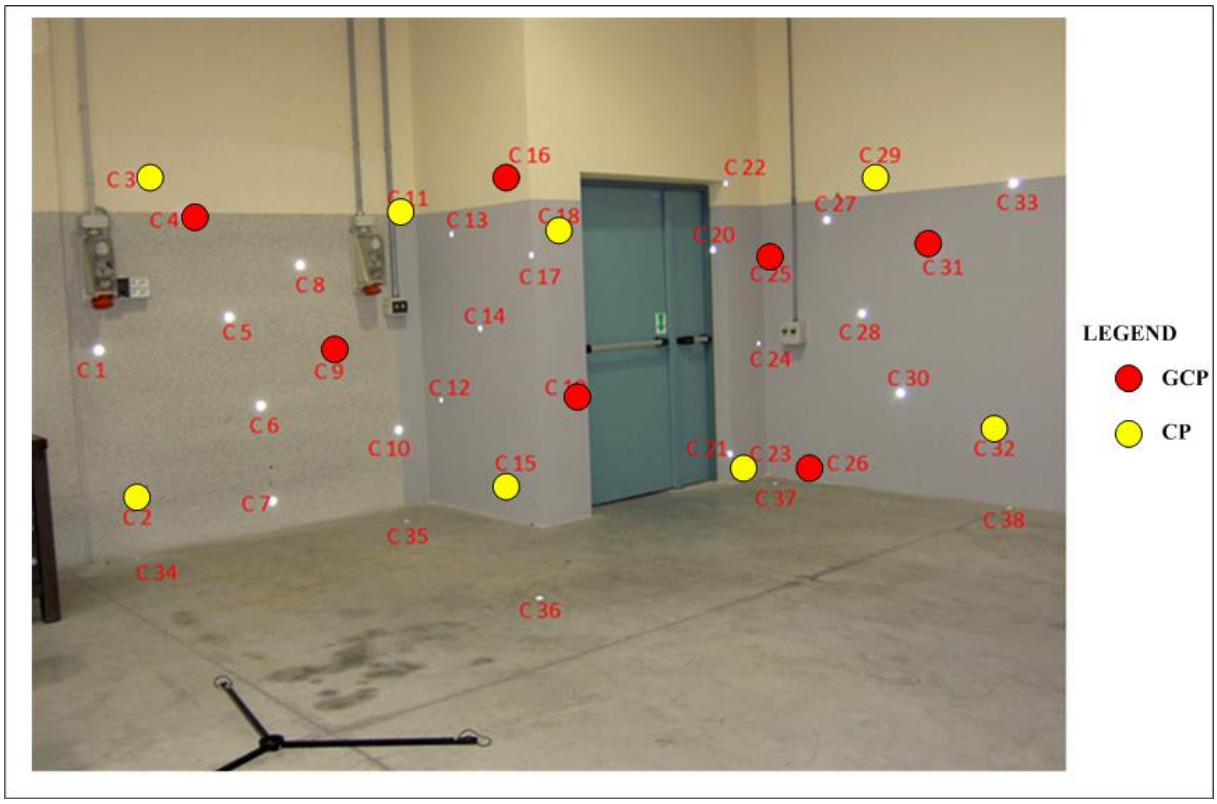


Figure 5.8 Configuration of the 7 GCPs and 8 CPs

The interior orientation was performed adopting as input the most significant sets of calibration parameters previously computed with the two procedures: both mean values and results delivered in combinations no.11 (MicroMap) and no.8 (MicMac) were selected for the test-field dataset; results achieved with Test-1 and Test-2 datasets were employed as well. A

further test was carried with non-calibrated images, i.e. by using as calibration input values corresponding to the ideal pinhole camera model (nominal focal length, no principal point offsets, no lens distortion). Within the bundle adjustment processes, all the above mentioned parameter values were always constrained.
Validation assessments were finally carried out by comparing residuals computed on eight well distributed targets of the test-field, chosen as Check Points (CPs), and their Standard Deviations (Std. Dev.). Residuals were calculated as the differences between the XYZ coordinates retrieved from the oriented stereoscopic model and those measured with total station. Figure 5.8 shows the configuration of selected GCPs and CPs, whereas Table 5.7 summarizes the main results.

***A-POSTERIORI* VALIDATION TESTS**			
Calibration input	Std. Dev. X (m)	Std. Dev. Y (m)	Std. Dev. Z (m)
Non-calibrated images	0.033	0.130	0.112
MicroMap: mean values	0.006	0.033	0.032
MicMac: mean values	0.031	0.051	0.065
MicroMap: combination no. 11	0.003	0.029	0.025
MicMac: combination no. 8	0.019	0.041	0.041
MicMac: Test-1 dataset	0.013	0.074	0.066
MicMac: Test-2 dataset	0.004	0.069	0.070

Table 5.7 Standard deviations of the residuals computed on 8 CPs

Results point out the important role played by calibration in metric information recovery: the use of computed calibration parameters, in fact, greatly reduces the errors if compared to the test where nominal values have been employed. Furthermore, both the MicroMap calibration and the MicMac procedure deliver standard deviations of few centimetres: although the use of GCPs coordinates within a test-range calibration approach (MicroMap) appears to achieve the best results, the automated self-calibration procedure (MicMac) is anyway able to reach a comparable level of accuracy. Moreover, even with reflective objects (Test-1 dataset), low textured surfaces (test-filed and Test-2 datasets) and different number of images, the automated calibration procedure offered by the IGN's tools shows a good stability and metric accuracy, as it was already pointed out from direct comparisons.

6. 3D MODELLING FROM TERRESTRIAL IMAGERY

6.1 Introduction

After a validation assessment of open-source procedures for digital camera calibration, validation tests of these photogrammetric and computer vision-based methods will be presented in this chapter, in order to show their metric potentiality in dealing with terrestrial image-based 3D modelling. The term "terrestrial" is here used to characterize the employed image data, that were all acquired from terrestrial platforms and directed to close-range applications. Among them, this chapter will especially focus on Cultural Heritage applications, performed using mainly the IGN's suite of tool. Furthermore, in order to produce comparative data and best practices relevant to different spatial scales and case studies, three different applications will be here discussed:

- Image-based 3D modelling of small and detailed sculptural elements (Section 6.2). As test-objects, three different types of sculptural elements were chosen among those constituting the well-known sculptural heritage of the Cathedral of Modena, Italy. These experimental tests were performed within a project supported by the United Nations Educational, Scientific and Cultural Organization (**UNESCO**), that, among its principal purposes, aims especially at building intercultural understanding, through protection of heritage and support for cultural diversity. In order to perform this task, the "intellectual" agency of the United Nations created the idea of World Heritage Sites, whose outstanding universal value should be protected. The deriving World Heritage List currently comprises 962 sites (745 cultural, 188 natural and 29 mixed) in 157 States Parties. Among the 49 Italian recognized sites, the "Cathedral, Torre Civica and Piazza Grande, Modena" was inscribed within the list in 1997, since it represents a masterpiece of human creative genius, that created a new dialectical relationship between architecture and sculpture in the Romanesque art.
- Image-based 3D modelling of multi-shape and complex architectural elements (Section 6.3). As test-object, the main entrance of the "Cathédrale de la Major", Marseille (France), was chosen for its impressive content of different geometric, material and texture peculiar features. This project was performed in collaboration with the UMR (Unité Mixte de Recherche) 3495 **CNRS** (Centre National de la Recherche Scientifique) / MCC (Ministère de la Culture et de la Communication), **MAP** (Modèles et simulations pour l'Architecture et le Patrimoine) **GAMSAU** (Groupe de Recherche pour l'Application des Méthodes Scientifiques à l'Architecture et à l'Urbanisme) Laboratory in Marseille (UMR 3495 CNRS/MCC MAP-Gamsau). This laboratory, directed by Livio de Luca, is collaborating with the IGN in many projects related to the suite of tools Apero/MicMac, such as: the development of specific interface for the entire 3D reconstruction process, the development of a web-viewer for image-based 3D navigation and point cloud visualization, the development

of an informatics cloud platform for 3D digitization, documentation, conservation and diffusion of cultural heritage (CULTURE 3D CLOUDS), the development of free solutions, such as software, methodologies, guidelines and best practices, within the Tools and Acquisition Protocols for Enhancing the Artifact Documentation project (TAPEnADe project). The application that will be presented in this chapter was carried out within a period spent at the MAP Laboratory during the three year of PhD studies.

- Image-based 3D modelling of a laboratory test-object, specifically designed in order to exhibit external surfaces characterized by different orientations, shapes, materials, textures and roughness (Section 6.4). This project was performed in collaboration with the National Research Council Canada, Ottawa, unit of Measurement Science and Standards (**NRC – MSS**). In particular, the tests were carried out within an environmentally controlled metrological laboratory for research and developments activities in non-contact 3D imaging metrology. This activity was performed within a two-week period at the NRC-MSS Laboratory during the three year of PhD studies.

6.2 Sculptural elements (Cathedral of Modena, Italy)

Figure 6.1 The UNESCO World Heritage Site "Modena. Cathedral, Civic Tower and The Piazza Grande" (Modena, Italy) – © Ghigo Roli www.ghigoroli.com

The Cathedral of Modena is part of the UNESCO World Heritage Site "Modena. Cathedral, Civic Tower and The Piazza Grande" inscribed in 1997. This monumental and architectural complex (Figure 6.1) was selected on the basis of its cultural value: it was in fact recognized as possessing those unique, authentic and emblematic characteristics that make it part of our universal cultural heritage. The whole complex, symbol of the City of Modena, stands for an exceptional example of urban settlement linked with the values of communal civilization, with its peculiar and complex interweaving of religious and civic functions. The Cathedral is an impressive and detailed masterpiece of the Romanesque style, built by the Lanfranco architect on two pre-existing places of worship. Its history is significantly characterized by the succession of very different construction phases, whose architectural limits are still well-recognizable. The overall architectural system dates from the beginning of the eleventh century and was realized through the combined efforts made by Lanfranco, from the architectural point of view, and Wiligelmo, from the sculptural one. Many recovery materials were used during this first building phase, especially by exploiting some ancient Roman ruins. Starting from the second half of the twelfth century, the Campionesi masters succeeded in the completing of the Cathedral. Their activities, lasted for three generations, led to realization of the inner decorations, together with that of some structural changes (e.g. the "Porta Regia" on the South façade, the two lateral doors on the Wets façade and its gothic rose window). Afterwards, few less significant actions took place up to the last century. Finally, starting from 2007, a restoration activity was carried out in order to reinforce some architectural elements and remove the degradation of the external stone facing.

The sculptures of the Cathedral represent an integral part of the overall monumental complex, being an exceptional example of the Romanesque Italian sculpture. Its sculptural cycles include human figures of different types (saints, prayers, soldiers, workers, etc…), together with animals, natural motifs and monstrous symbols. Within this rich sculptural heritage, three different types of elements were selected as test-objects and employed for a metric purpose. Figures 6.2-4 show these sculptures, that include a capital, a small figure-like corbel and a medieval relief; their main dimensions are reported too.

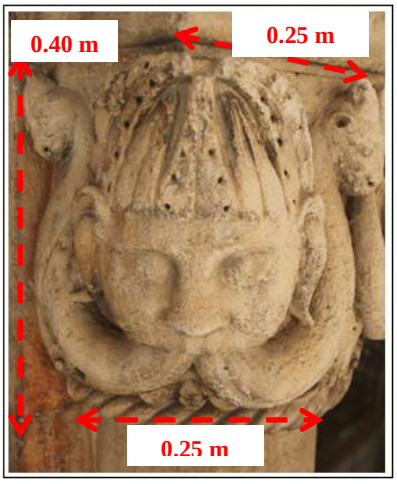


Figure 6.2 The capital

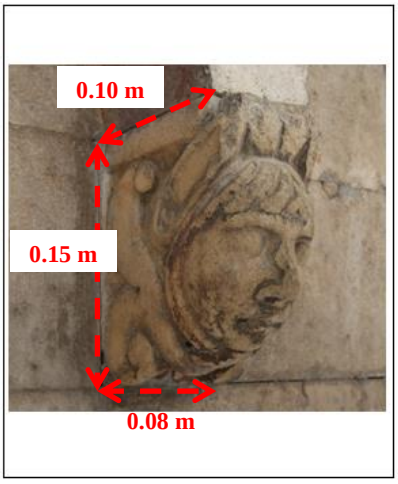


Figure 6.3 The figure-like corbel

Figure 6.4 The medieval relief

This application was carried out within an activity performed after the restoration and aimed at acquiring detailed three-dimensional models of the most precious external sculptures of the Cathedral. The final goal of the project was the provision of a metric digital archive for accurate detailed reconstruction, digital preservation, prototyping and virtual tourism purposes. Besides the classical range-based modelling methods, 3D passive imaging was tested as well, in order to assess the potentiality of the IBM (Image-Based Modelling) approach in dealing with small sculptural test-objects. In particular, the possibility of achieving accurate 3D models from the image-extracted dense point clouds was deepen analyzed. The tests were performed mainly with the IGN's suite of tools and the choice of three different types of sculptures aims at evaluating the potential of these algorithms in dealing with elements characterized by various extensions and depths. The sculptures are made of light natural stones, with fine decorative details. They adorn the external facades of the Cathedral, thus providing an example of an outdoor application, characterized by rapid illumination changes during the course of the day. Furthermore, in order to reach the height needed for the survey (around 9 m ASL), a metal scaffolding was employed and its resulting vibration problems had to be dealt with. All these external conditions represent some common problems, typical of outdoor surveys in the Cultural Heritage field: therefore they gave the possibility of testing the metric performances of the IBM methods within these unfavourable and uncontrollable boundary conditions.

6.2.1 Procedural workflow

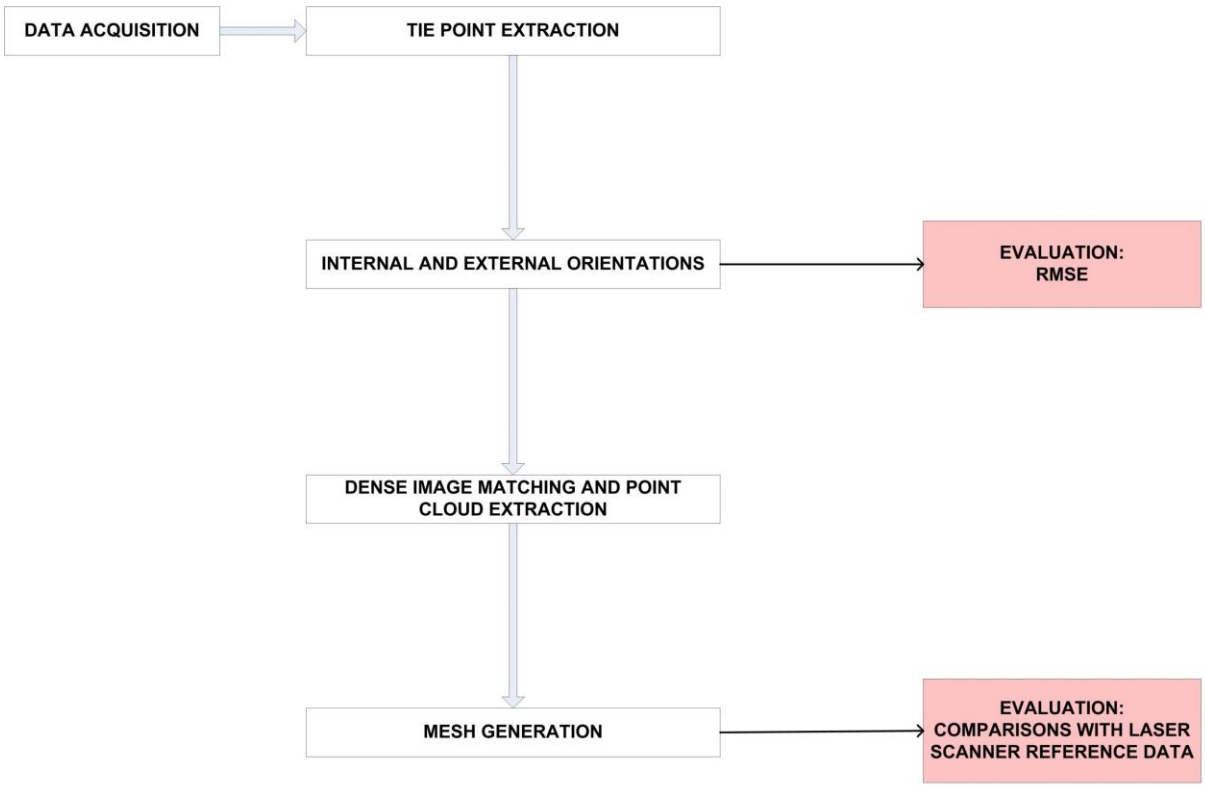


Figure 6.5 Procedural workflow (Sculptural elements – Cathedral of Modena)

Given the general purpose of the project that includes the present case study, the attention was here paid mainly on the metric accuracy assessment of the final products, i.e. the 3D models of the three sculptural elements. Thus, while each procedural step performed in the later discussed applications will be deepen evaluated, in this case the analysis is especially aimed at defining the accuracy achievable by the last modelling phase, by comparing its outputs to adequate reference data, acquired with high resolution triangulation laser scanners.
The entire photogrammetric and computer-vision based pipeline has been performed with the IGN's suite of tools, starting from the detection of homologous points between the images and ending up with the dense 3D point clouds, that reconstruct the scene of interest. Afterwards, a commercial software was used in order to perform the modelling process, using automatic algorithms of mesh construction. Comparisons were finally carried out after a post-processing phase performed on both LS- and IBM-derived 3D models and aimed at providing final products suitable to be read by numerical control machines for rapid prototyping.
Besides the IGN's suite of tools, the free web-service 123D Catch by Autodesk was tested as well, in order to extract the 3D model from the corbel dataset. The resulting output was finally compared with the same reference model, that was kept as reference data also within the IGN-tool assessment analysis. For the algorithmic and operative aspects related to the IGN's suite of tools the reader is referred to Chapter 3, where a brief description of the employed web-service is provided too. Finally, Toschi et al., 2003 provide a summary of the whole experiment, together with a discussion on the results thereby achieved

6.2.2 Image acquisition

Figure 6.6 Canon EOS 5D Mark II equipped with the zoom lens Canon EF 16-35mm f2.8L USM

The digital image acquisition phase was performed using a Canon EOS 5D Mark II digital camera (5616 x 3744 pixels), equipped with a wide-angle zoom lens, Canon EF 16-35mm f2.8L USM: they are both depicted in Figure 6.6. The related technical specifications are listed in Table 6.1.

Canon EOS 5D Mark II	
BODY TYPE	Mid-Size SRL
SENSOR RESOLUTION	21 Mpixel
SENSOR SIZE	Full Frame
SENSOR TYPE	CMOS
ISO	Auto, 100-6400 in 1/3 stops, plus 50, 12800, 25600 as option
MIN SHUTTER SPEED	30 sec
MAX SHUTTER SPEED	1/8000 sec
Canon EF 16-35mm f2.8L USM	
FOCAL LENGHT	16-35 mm
MAX APERTURE	f2.8
MIN APERTURE	f22.0
MIN FOCUS	0.28 m

Table 6.1 Technical specifications of Canon EOS 5D Mark II and lens employed

The acquisition phase was carried out just after the end of restoration activities, thus employing the metal scaffolding that had enclosed and partially hidden the eternal facades of the Cathedral for the entire restoration period, in order to reach the required height. Of course, this reduced the available space in front of each element and required the acquisition to be performed within a maximum focusing distance of around 1 m. Only the medieval relief was acquired directly from the ground, thus eliminating the over mentioned practical difficulties. A photographic tripod was always employed in order to reduce as much as possible the vibration effects, that were especially significant for the acquisitions performed from the scaffolding; furthermore, images were acquired as quickly as possible in order to reduce the illumination changes over the scene. Each element was acquired following the general protocol recommended for terrestrial modelling (Pierrot-Deseilligny and Clery, 2011). For each point of view and, consequently, for each desired point cloud, a central "master" image was taken, together with other 3-4 closed associated images. A sufficient number of convergent shots were taken in order to assure the connection between each master image, with an overlap of around 80% between each pair of images. A reasonable base-to-depth ratio was adopted to guarantee a good tie point detection without an excessive reduction of the final reconstruction accuracy. The number of acquired images depended consequently on the dimensions of the object and on external conditions, like the presence of obstacles. For

example, 15 images constitute the relief dataset, for which Figure 6.7 provides a sparse 3D reconstruction with the corresponding camera relative poses computed with the tool AperiCloud. The three crosswise acquisition configurations are here well-recognizable and correspond to the three points of view needed to achieve a complete 3D reconstruction of the relief. 12 images were instead sufficient to cover almost the entire surface of the corbel; nevertheless, the presence of the scaffold (above) and of the wall (sideways) prevented the shooting of the top and of some small side portions of the object. These practical difficulties rose up for the capital acquisition too, whose bigger dimensions and depth variations required the recovery of 15 images.

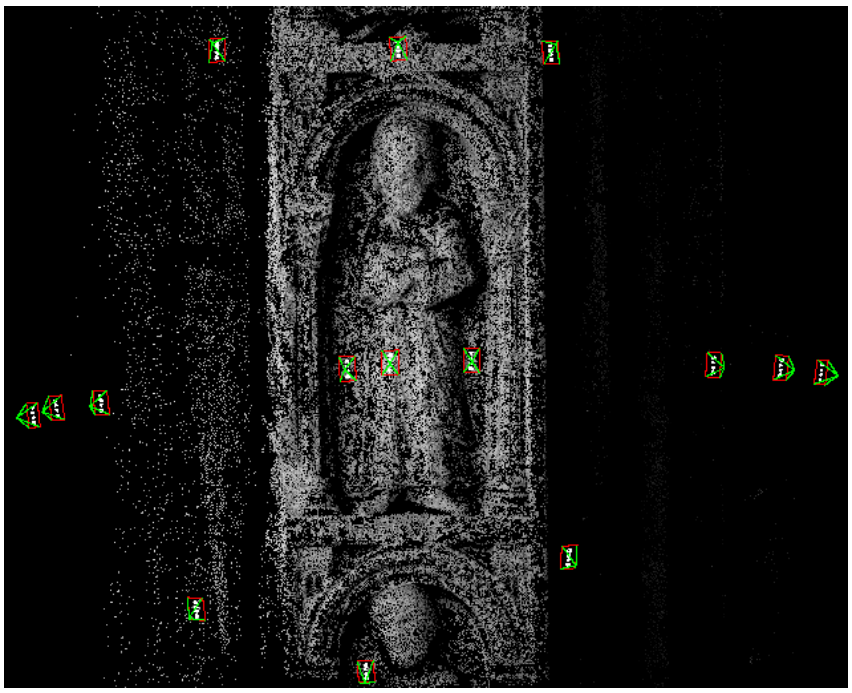


Figure 6.7 Image acquisition layout (relief dataset)

	CAPITAL	CORBEL	RELIEF
No. of images	15	12	15
Focal length	35 mm	35 mm	35 mm
f-number	f-5.0	f-22.6	f-5.7
ISO	100	100	100

Table 6.2 Acquisition setups and number of acquired images for the three selected test-objects

Table 6.2 summarizes the acquisition setups, kept fixed during the whole phase, and the number of necessary images. The focusing distances were selected and fixed according to the specific requirements of each case study, especially in terms of allowable camera-object distances, and according to the main characteristics of each sculptural element (e.g. dimensions and depth extension).

6.2.3 Laser scanner survey

The reference data were acquired with three different types of triangulation laser scanners, whose main technical specifications are listed in Tables 6.3-5. Figures 6.8-10 depict the employed instruments.

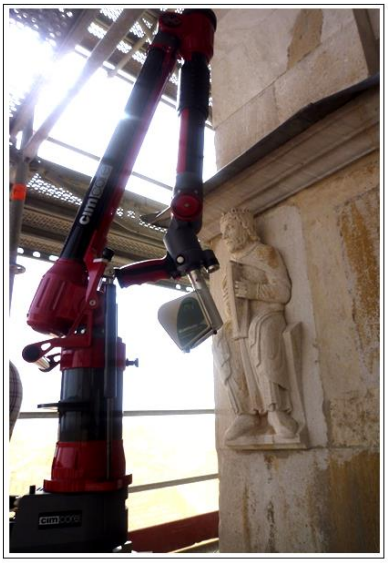


Figure 6.8 Romer CMM Infinite 2.0, with ScanWorks System by Perceptron

Figure 6.9 Faro CAM2 Platinum ScanArm

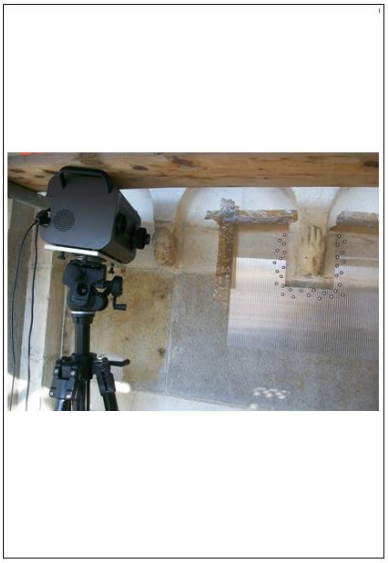


Figure 6.10 Konica Minolta Range 7

Romer Infinite 2.0 Portable Arm CMM (ScanWorks System by Perceptron)	
SCAN PRINCIPLE	Triangulation – Tactile Probe + Laser Head
MEASURING FIELD	2.5 m (spherical diameter)
NOMINAL ACCURACY	0.07 mm

Table 6.3 Technical specifications of Romer Infinite 2.0 Portable Arm CMM with ScanWorks System by Perceptron

Faro CAM2 Platinum Scan Arm	
SCAN PRINCIPLE	Triangulation – Tactile Probe + Laser Head
MEASURING FIELD	2.0 m (spherical diameter)
NOMINAL ACCURACY	0.07 mm

Table 6.4 Technical specifications of Faro CAM2 Platinum Scan Arm

Konica Minolta Range 7	
SCAN PRINCIPLE	Triangulation by light sectioning method
MEASUREMENT DISTANCE	450 to 800 mm
NOMINAL ACCURACY	0.04 mm

Table 6.5 Technical specifications of Konica Minolta Range 7

Besides the three selected test-objects, the above mentioned laser scanners were employed in order to digitize almost 200 sculptural elements, adorning the four external facades of the Cathedral and its main doors. Most of these surveys, such as the ones carried out for the selected capital and corbel, were performed by exploiting the same metallic scaffolding, that was also used within the digital image acquisition. These external conditions gave rise to many difficulties, that had to be coped with and can be mainly summarized as follows:

- Vibrations, produced by the synergistic action of several factors, i.e., in particular, the movements of the operators working together on the scaffold and the presence of the wind. The former issue was mainly due to the contemporary execution of different activities, all performed on the same scaffolding, such as: the final cleaning of the sculptural elements, the laser scanner relief of those that had already been restored and the dismantling of the scaffold itself. Of course, this situation required an effective collaboration between all the involved professionals and a careful planning of the timeline, in order to meet the needs of each specific job. As regards the wind-derived vibrations, this effect was particularly significant on the upper floors of the scaffold. In order to reduce as much as possible these oscillations, the two articulated portable scan arm were anchored to a metal plate and fixed to the uprights of the scaffold through the use of metallic or wooden walkways. A solid tripod was, instead, employed for the acquisitions with the Konica Minolta laser scanner.
- Material-related and lighting/temperature problems, especially caused by the outdoor environmental conditions. The former effect was mainly connected to the presence of a thin layer of powder material on the top of some sculptural elements, derived from the previously performed restoration activities: the particular reflectance properties of these surfaces turned out to have a negative interaction with the laser light scan. Furthermore, the acquisition was often performed under difficult external conditions, such as very hot temperatures and strong direct illumination in the summer, replaced by cold temperatures and sometimes foggy weather in winter time. Many measures were taken in order to limit the negative influence of these factors by employing ,for example, adequate sun barriers and heating/cooling systems.

All these particular working conditions didn’t allow the 3D models to reach the maximum nominal accuracy of the instrumentations; anyway a final 3D modelling resolution of around 0.3 mm was achieved and considered to be adequate for the purposes of the general project.

For the simplest and smallest elements, such as the relief and the corbel, the range maps were acquired from a single scanning position, i.e. without moving the base of the instrument during the acquisition of each 3D model. On the other hand, the larger and more detailed sculptures, e.g. the capital, required two or more measure stations in order to reconstruct their complete 3D geometry. Thus, the acquisition time depended on the detected sculptural element and was, on average, equal to 2.5 hours per sculpture.

6.2.4 Image processing: IGN's suite of tools

The first phase, i.e. the tie point extraction, was performed by employing the tool Tapioca (Subsection 3.3.2), with its SIFT^{++} implementation of SIFT algorithm. The homologous point search mode was set to "All" in order to consider all possible pairs of images; furthermore, this research was carried out without any previous image shrinking, i.e. the original image resolution (5616 x 3744 pixel) was always used.

The retrieval of internal and external orientations was carried out with two approaches:

1. Simultaneously computation of both calibration parameters and relative camera poses with the tool Tapas (Subsection 3.3.3). The FraserBasic distortion formulation was employed with its 10 degrees of freedom; the RMSE (Root Mean Square Error) of all re-projection residuals was always around half a pixel, i.e. equal to 0.44 pixels for the capital, 0.50 pixels for the corbel and 0.56 pixels for the relief.
2. Refinement of pre-computed calibration parameters and computation of absolute camera poses with the tool Apero. This second procedure was only performed with the corbel dataset, since it represents a simple example of a typical sculptural element. In this test, absolute orientation was computed using 12 well-distributed GCPs (Ground Control Points), whose 3D coordinates were directly derived from the laser scanner model.

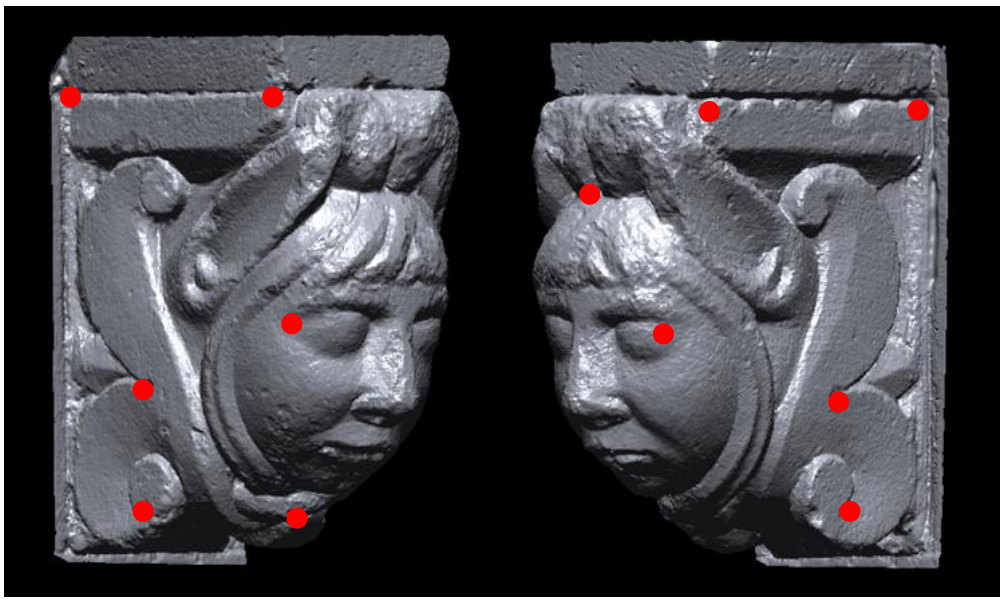


Figure 6.11 The selected GCPs shown on the LS-derived corbel 3D model

Figure 6.11 shows two views of the corbel LS-derived 3D model with the position of the selected GCPs. For the calibration XML-section, results computed via the previously performed Tapas calibration were declared as initial values; they were kept frozen at the beginning of the compensation and then re-evaluated within the final phases of the bundle adjustment procedure. The final RMSE of the bundle adjustment was 1.2 pixels.

Results achieved in the over-mentioned tests show that a self-calibration approach during the bundle adjustment step represents a good strategy, if the employed image dataset is favourable to the calibration recovery. Since the figure-like corbel exhibits significant depth variations and surface textures and was imaged according to the main calibration-favourable acquisition requirements (different heights, angles of view, etc…), the resulting dataset can be employed for a simultaneous computation of both calibration parameters and relative camera poses. Of course, orientations computed with the first approach are simply relative and will deliver a scaled-version of the real object (hereinafter termed "Tapas-derived" model). On the contrary the second approach, performed with the corbel dataset and the LS-derived GCPs, is able to achieve an Euclidean 3D reconstruction (hereinafter termed "Apero-derived" model), that will be already geo-referenced in the LS reference frame.

The third step of the procedure is, finally, the dense matching computation from oriented images. This phase was performed using the MicMac software with its multi-scale, multi-resolution and pyramidal matching approach (Subsection 3.3.4). A depth map was extracted for each point of view of each dataset, running the computation in image-ground geometry. The central "master" images collected during the acquisition phase were selected for the correlation procedure; the research area on each of them was then defined through a masking process. The depth of field interval to be explored was set as well, starting from the orientation results. Table 6.6 lists the main parameter setups selected for the three case studies.

	CAPITAL	**CORBEL**	**RELIEF**
Regularization factor	0.2	0.2	0.2
Z-Quantification factor	0.5	0.5	0.5
Depth of field interval	$[0.5 * D_0; 1.1 * D_0]$	$[0.5 * D_0; 1.1 * D_0]$	$[0.5 * D_0; 1.1 * D_0]$
Final Z-Resolution	1	1	1

Table 6.6 Some of the main parameters selected in the dense image matching procedure (D_0 is the average depth computed by Apero)

The computed depth maps were finally converted into 3D point clouds with the tool Nuage2Ply (Subsection 3.3.4): this algorithm projects each pixel of the master image in the object space, using image orientation parameters and depth values. RGB attribute from master images is also assigned to each 3D point. The capital reconstruction resulted in more the 16 million of extracted points, whereas both the corbel and the relief were described by more than 7 million of points.

6.2.5 Mesh generation

Since the overall project was aimed at achieving detailed 3D surface models suitable for various applications, such as computer aided restoration and rapid prototyping, an adequate final post-processing phase was performed therefor. Both image- and range-derived 3D data were imported into the commercial software Geomagic Design X by Geomagic (Geomagic Design X), formerly Rapidform XOR, where the entire 3D modelling pipeline was performed (Remondino and El-Hakim, 2006).

Point clouds extracted with the IBM approach into a pure relative reference frame were first scaled and geo-referenced by using the corresponding LS models as references. Each alignment process was performed in two subsequent steps: first, a roughly alignment was achieved through manually recognized pair points; secondly, an automatic refinement of the previous alignment was performed, applying the well-known ICP (Iterative Closest Point) algorithm (Besl and McKay, 1992). The latter step was also employed in order to improve the alignment between the corbel Apero-derived 3D point clouds and the corresponding LS 3D model. The refinement processes showed always a millimeter-level accuracy: in particular, the RMSE computed for the registration of the corbel 3D model was equal to 1.09 mm. Point clouds were then converted into polygonal models through an automatic algorithm of mesh construction (3D Systems), starting from the well-known principle of Delaunay triangulation (Delaunay, 1934): the surface of each element was thereby discretized in a suitable number of flat triangular surfaces (Triangulated Irregular Network, TIN). Afterwards small holes, corresponding to those parts that were not directly acquired by the digital camera due to the presence of obstacles, were detected and filled by running both automatic and manual processes. Finally, defective flat surfaces were automatically identified and corrected, taking into account four main defect types:

- non – manifold faces, i.e. those sharing three or more edges with the neighboring ones;
- redundant faces, i.e. those violating the rule that states that, with the exception of boundary vertices, each vertex must have an identical number of faces and edges incident to it;
- crossing faces, i.e. those intersecting each other's;
- reversed faces, i.e. those having a normal direction opposite to the one of neighboring faces.

The so-achieved 3D models were finally exported in STereoLithography (STL) file format, that is supported my most of the CAD, modelling and rendering software packages and is widely used for rapid prototyping and computer-aided manufacturing. The over mentioned procedure was also carried out starting from the LS-derived 3D data. Figures 6.12-14 show the final photo-textured 3D models delivered by the capital, corbel and relief image datasets.

Figure 6.12 Capital 3D Model (IBM approach – IGN's tools)

Figure 6.13 Corbel 3D Model (IBM approach – IGN's tools)

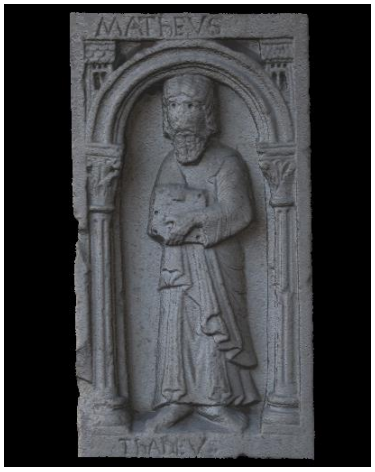


Figure 6.14 Relief 3D Model (IBM approach – IGN's tools)

6.2.6 Image processing: 123D Catch web-service

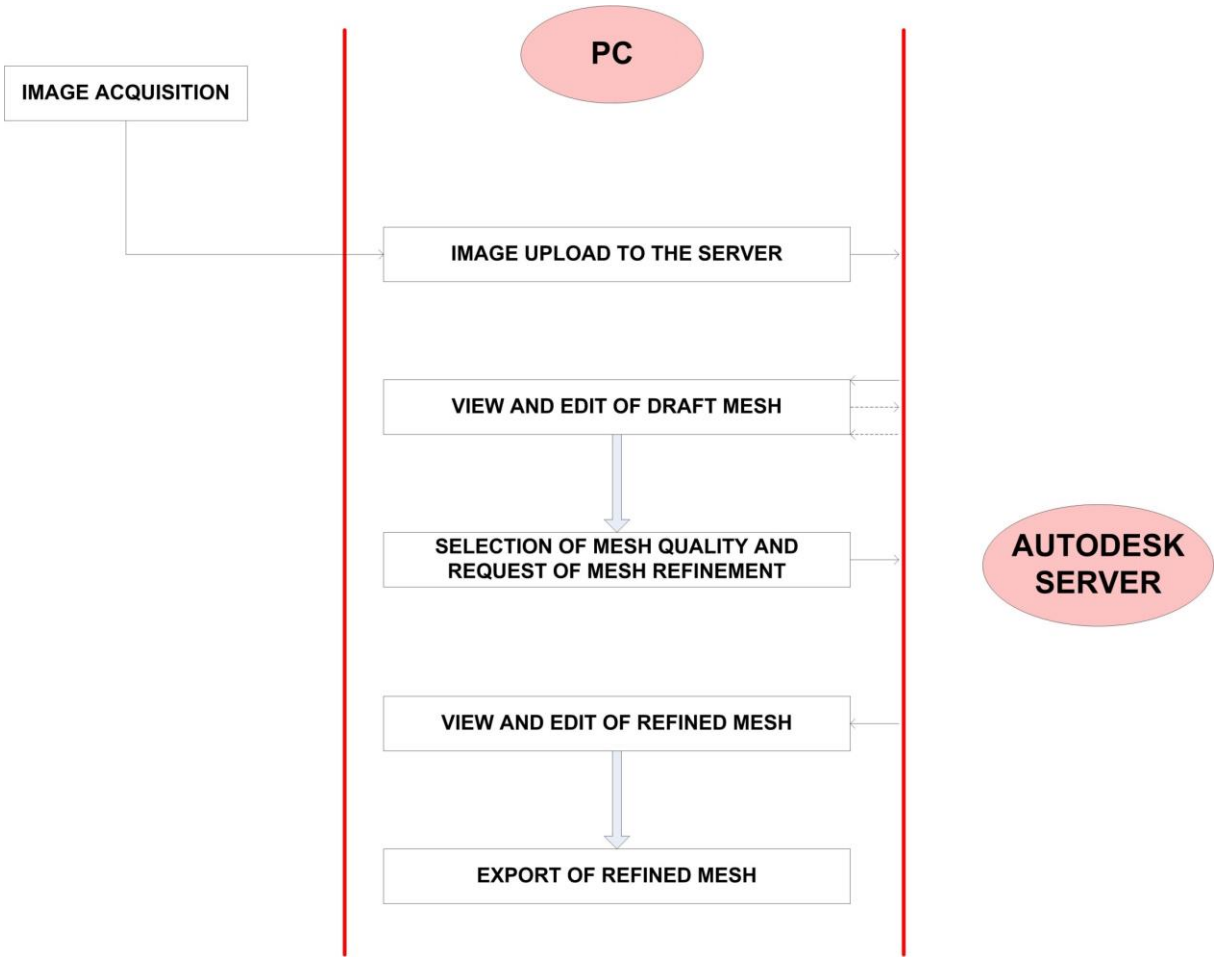


Figure 6.15 123D Catch procedural workflow

The corbel dataset was also processed through the free web-service 123D Catch, whose main functionalites were described in Section 3.2 and are again summarized in Figure 6.15. This cloud-based application creates and delivers photo-textured 3D models from images in a completely automatic mode. Once the 12 images were uploaded to the remote services in the cloud, the process was completed in about 25 minutes, opening the resulting 3D scene in a basic 3D editor interface. After having selected the best mesh quality option, the automatically refined 3D model was finally exported and three views of it are shown in Figure 6.16.

Figure 6.16 Corbel 3D Model (IBM approach – 123D Catch web service)

6.2.7 Comparisons with LS reference data

The metric accuracy of the final 3D models generated from point clouds extracted with the image-based approaches was assessed through geometric comparisons with the above mentioned reference data. These tests were all performed within the open-source software CloudCompare vs 2.4 (CloudCompare). Geometric distances between the vertices of the image-based models and the range-based ones were thereby computed with associated statistics. All comparisons delivered a standard deviation of the differences between the datasets of millimetre-level. In particular, some statistical parameters obtained for the corbel dataset are listed in Table 6.7: both 3D models derived from point clouds oriented with the relative simplified procedure (Tapas-derived 3D model) and with the absolute one (Apero-derived 3D model) were metrically evaluated. Furthermore, Tables 6.8-9 list the corresponding statistics delivered by the tests performed with the capital and relief datasets.
Figure 6.17 shows three views of the deviation map obtained from the comparison between the Tapas-derived 3D model and the reference data, for the corbel test-object. The corresponding color-coded maps extracted for the capital and relief 3D models are then presented in Figures 6.18-19. The colour scale intervals are always chosen in accordance with the standard deviations delivered by the comparisons. All points of distance values falling outside these ranges are colored in grey.

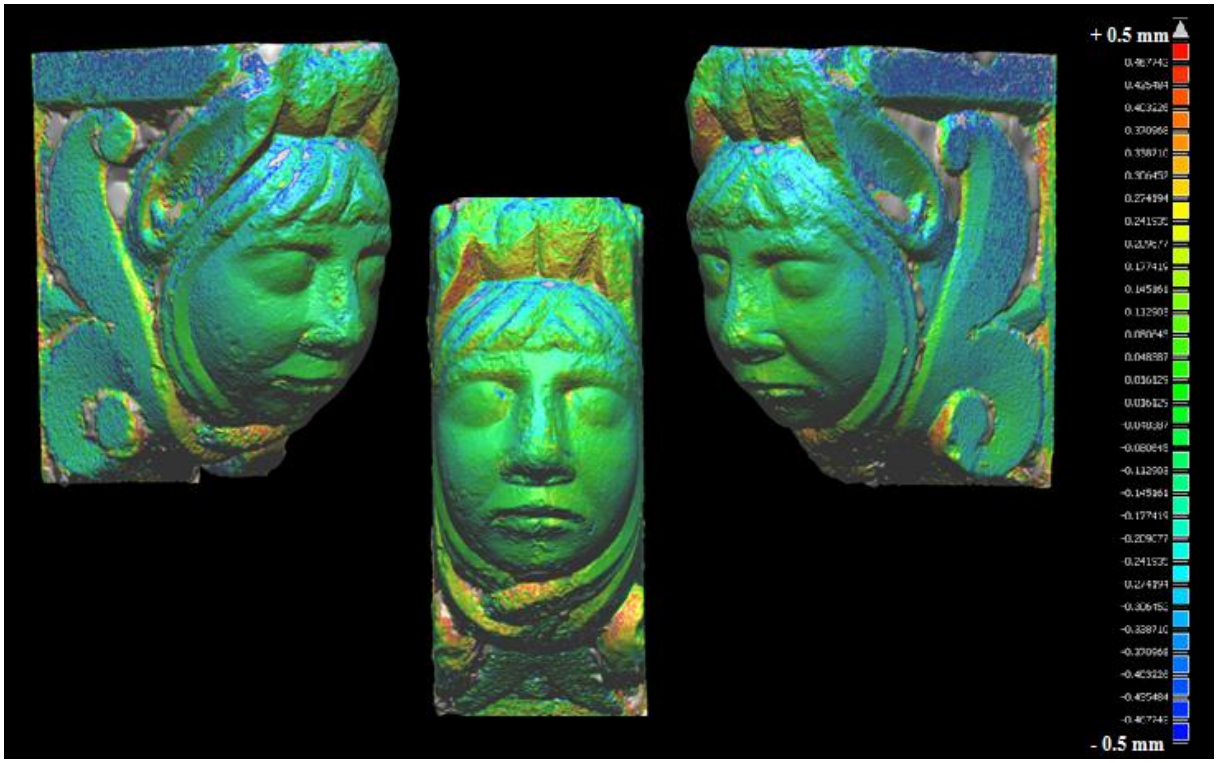


Figure 6.17 Comparison between the corbel IBM Tapas-derived model (IGN's tools) and the LS model. The colour scale ranges from -0.5 mm (blue) to +0.5 mm (red)

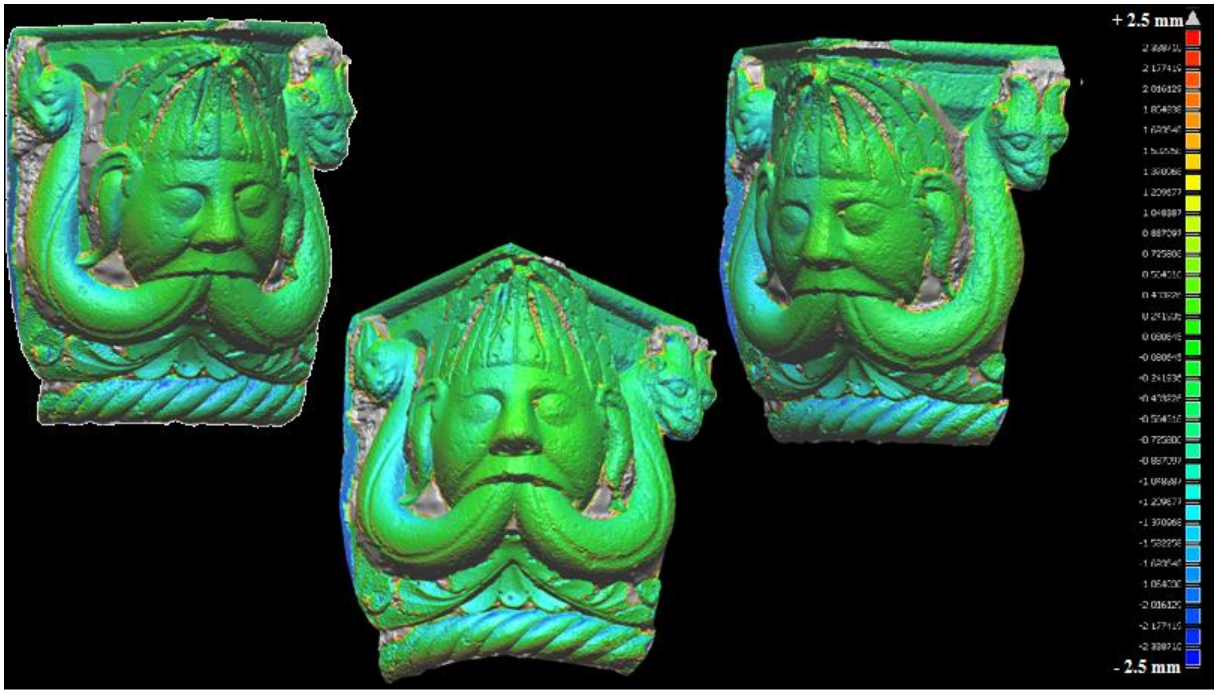


Figure 6.18 Comparison between the capital IBM model (IGN's tools) and the LS model. The colour scale ranges from -2.5 mm (blue) to +2.5 mm (red)

Figure 6.19 Comparison between the relief IBM model (IGN's tools) and the LS model. The colour scale ranges from -0.4 mm (blue) to +0.4 mm (red)

	CORBEL 3D MODEL (TAPAS-DERIVED)	CORBEL 3D MODEL (APERO-DERIVED)
Mean distance (mm)	-0.04	-0.06
Std. Dev. (mm)	0.51	0.46
Positive maximum (mm)	4.90	4.91
Negative maximum (mm)	-3.44	-3.48

Table 6.7 Comparison between the corbel IBM models (IGN's tools) and the LS model: statistical results

CAPITAL 3D MODEL	
Mean distance (mm)	0.72
Std. Dev. (mm)	2.53
Positive maximum (mm)	21.43
Negative maximum (mm)	-6.38

Table 6.8 Comparison between the capital IBM model (IGN's tools) and the LS model: statistical results

RELIEF 3D MODEL	
Mean distance (mm)	0.00
Std. Dev. (mm)	0.38
Positive maximum (mm)	6.12
Negative maximum (mm)	-7.15

Table 6.9 Comparison between the relief IBM model (IGN's tools) and the LS model: statistical results

Table 6.7 shows that both 3D models extracted with the IGN's tools from the corbel dataset show the same level of metric accuracy, if compared with the same reference model: this evidence points out that for simple and small objects even the simplified procedure of calibration and relative orientation (Tapas) works successfully and is sufficient to achieve a suitable accuracy level. For both models, i.e. the Tapas- and Apero-derived ones, the portions showing greatest deviations from the reference data correspond to those parts that were not directly acquired by the digital camera due to the presence of obstacles (top and some side portions). Those small areas were instead acquired with the triangulation laser scanner, making the reference model complete everywhere.

The same evidences can be deduced from the comparisons performed with the capital and the relief 3D models. The former delivers an higher value of standard deviation due to the presence of outlier distances (positive maximum equal to 21.43 mm), that are especially localized on the upper parts of the object: those portions, in fact, were almost totally *a-posteriori* reconstructed, since they were not imaged due to the presence of the scaffold walkways. These problems didn't occur for the relief acquisition, that was performed directly on the ground, without the use of the scaffolding: this strongly reduced both the vibration effects and the presence of occluded area. Results retrieved from the relief dataset are, consequently, more accurate, as the corresponding statistics point out. Finally, by analysing all the extracted color-coded maps, one can deduce that, besides occluded area, most of the problems are especially localized along boundaries and at sharp surface gradients.

The corbel 3D model delivered by the 123D Catch was compared with the LS reference data too; resulting statistics (Table 6.10) and color-coded map (Figure 6.20) show the same accuracy level achieved with the IGN's tools, even if the model resolution is much lower (142,000 faces against almost 4 million faces for the Tapas-derived 3D model).

CORBEL 3D MODEL	
Mean distance (mm)	0.03
Std. Dev. (mm)	0.46
Positive maximum (mm)	4.41
Negative maximum (mm)	-3.49

Table 6.10 Comparison between the corbel IBM model (123D Catch) and the LS model: statistical results

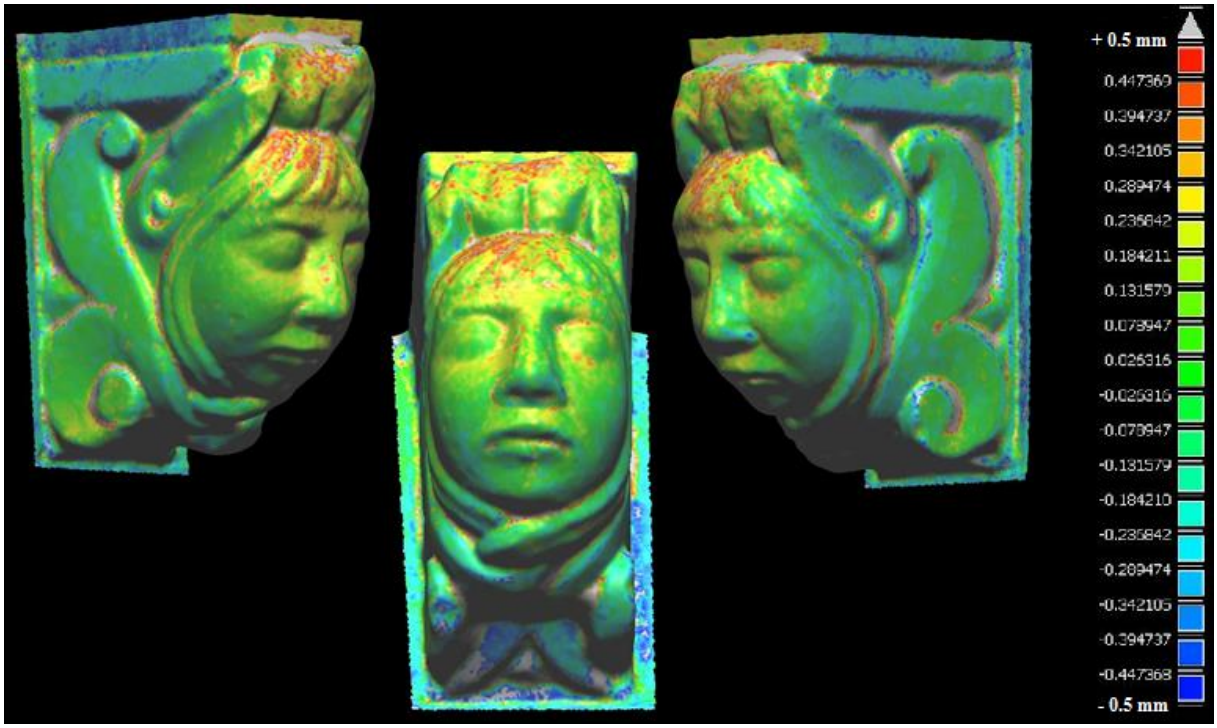


Figure 6.20 Comparison between the corbel IBM model (123D Catch) and the LS model. The colour scale ranges from -0.5 mm (blue) to +0.5 mm (red)

6.3 Cathédrale de la Major (Marseille, France)

The Cathedral of Marseille (Cathédrale Sainte-Marie-Majeure de Marseille or Cathédrale de la Major) is a Roman Catholic church and seat nowadays of the Archdiocese of the city. The structure appears now as it is depicted in Figure 6.21 and is the result of a long-lasting building process. The first part of the cathedral was built starting from the 12th century in a simple Romanesque style. The general plan of the structure was characterized by a simple rectangular base, composed by a central nave and two side aisles, without any transept. Unfortunately, no iconographic document of this first monument has been handed down to the present day; anyway, historians agree that the old cathedral was still in service during the erection of the nearby baptistery, that was built between the 14th and the 15th centuries. Nowadays, only a small part of this earlier and much smaller cathedral (termed the Vieille Major) still remains, alongside the new cathedral: in particular, only the choir and one bay of the nave are still visible. The new cathedral (or Nouvelle Major as it is called) was built in Byzantine-Roman style from 1852 to 1896 by the architects Léon Vaudoyer and Henri-Jacques Espérendieu. It dominates the north part of Marseille, where it is overlooking the port of Juliette; the verb "to dominate" refers to the enormous dimensions characterizing this structure, in terms of length (146 m), maximum width (54 m) and main dome height and diameter (68 m and 18m). The style recalls the maritime and Mediterranean vocation of the city, with its originality and dynamism. Thus, beside Romanesque elements (semi-circular

arches, domes…), one can find an explicit reference to the oriental Byzantine-style monuments, especially through the alternation of dark stones (the "Golfalina" of Tuscany) and white stones (the "Calissane" taken from the banks of the "étang de Berre"). Hence the term "Byzantine-Roman" is employed in order to describe this original and composite style.

Figure 6.21 Cathédrale de la Major, Marseille (France)
(en.wikipedia.org)

Since its huge dimensions and significant architectural complexity, only a part of the entire structure has been chosen for the application described below and this choice was taken according to the specific purposes of the application itself. This case study, in particular, aims at investigating the influence of different procedural parameters in each step of the image-based 3D modelling pipeline carried out with the IGN's suite of tools. Furthermore, different image acquisition protocols were tested and their impact on the subsequent photogrammetric and computer vision-based phases was analysed. In order to meet all these expectations, the **main entrance** of the cathedral was chosen as test-object: an image of the interested area with its main dimensions is shown in Figure 6.22. The selected volume of interest fully satisfies the following key-features:

- Significant depth variations, with the presence of many consecutive depth levels;
- Presence of surfaces characterized by different textures and colours, with the recurring alternation of dark and white stones in the lateral pillars and the succession of grey detailed motifs and planar pink surfaces in the main semi-circular arch; also the impressive red colour of the door is very interesting from this point of view, together with the pink veins of the marble columns. Finally, the two quasi-triangular

decorations framing the upper arch are characterized of precious golden, blue and green motifs, completing the ample colorimetric scale offered by the object.

- Presence of surfaces characterized by different materials and, consequently, roughness properties, such as stones of different origins, marble (columns) and wood (door).
- Presence of primitive geometric shapes, such as planes (main beam and pillars) and cylinders (columns);
- Presence of portions characterized by a significant amount of fine details, such as the central semi-circular relief, the decorations of the overcoming arches and the intermediate horizontal cornices.

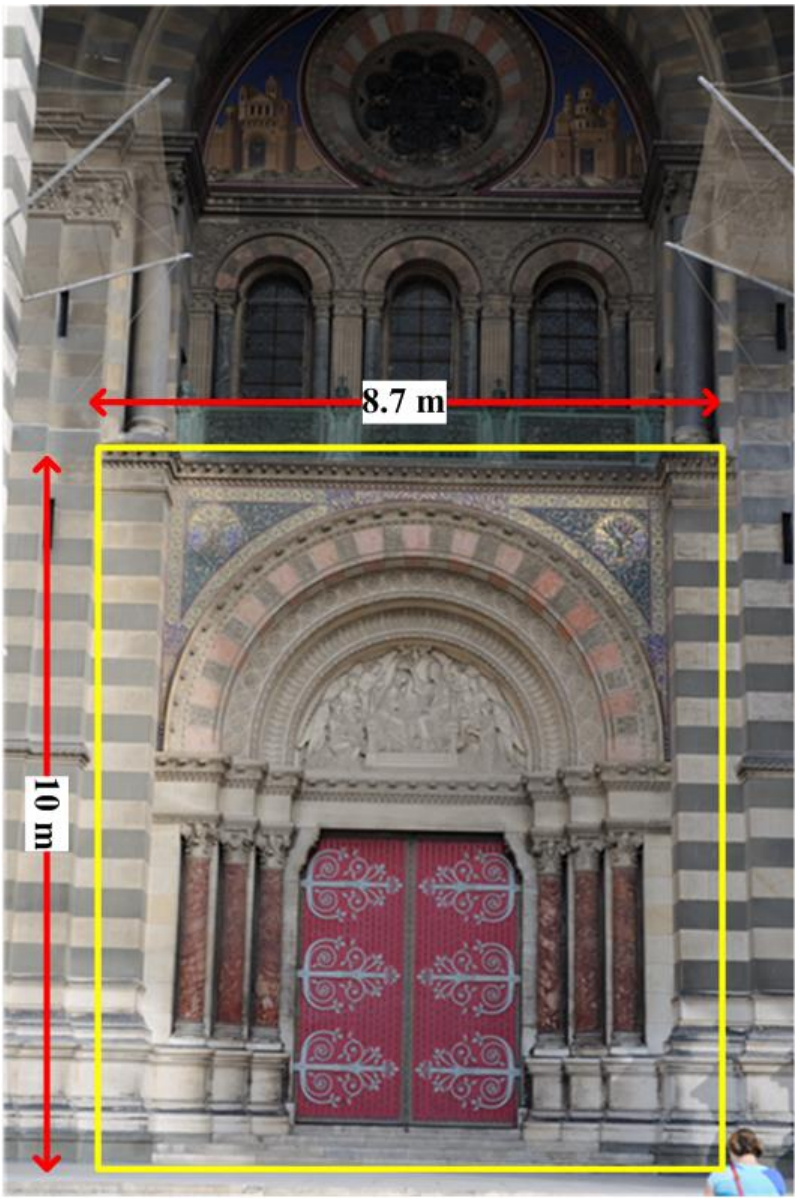


Figure 6.22 The acquired 3D scene (yellow rectangular) and its dimensions

- High availability of open space in front of the scene, allowing the testing of different image acquisition protocols. Furthermore, during the acquisition phase, the corridor in

front of the entrance was delimited with a red and white strip, keeping it out of the passage and stationing of visitors. The two lateral entrances were used to convey the normal flow of tourists and schoolchildren. This represented a necessary conditions, especially in order to fulfil eye-safety requirements, that have a paramount importance when working with coherent light sources (laser scanner survey).

- Presence of weather conditions typical of an outdoor application, characterized by rapid illumination changes during the course of the day. Furthermore, the acquisition phase was performed under a very windy wheatear, that required the employment of adequate solutions and strategies.

6.3.1 Procedural workflow

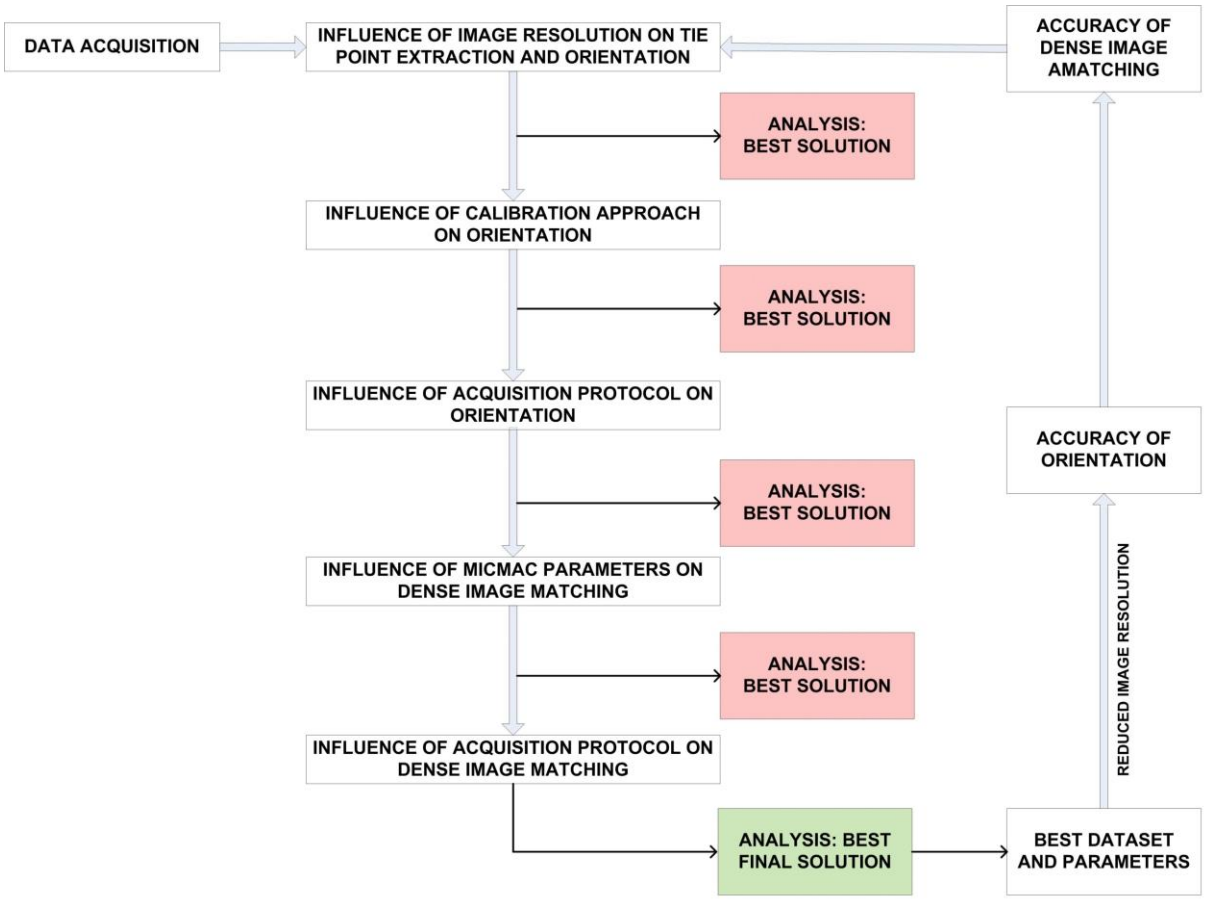


Figure 6.23 Procedural workflow (Cathédrale de la Major)

In Figure 6.23 the synthetic procedural workflow is shown. As previously mentioned (Chapter 3), the IGN's suite provides very parametrical tools, that offer to the user the possibility of finely controlling each processing phase through a huge amount of attributes and parameters. Of course, this generally represents a significant advantage, allowing the adaption of the pipeline to the specific requirements of any particular application. On the other hand, this flexibility has to deal with the lack of clear rules and good practices, that often leads the user to make a "blind" choice. Furthermore, also the acquisition protocol doesn't usually follow

specific rules, especially in terms of convergence angles and employed focal setting: in other words, it still represents a personal and, often accidental, choice. Starting from these observations, the detailed studies that will be described further in this section have been carried out in order to analyse the influence of different parametrical choices on each step of the image-based reconstruction procedure. The effects of the employed image acquisition protocol have been studied too, examining their influence on the orientation and dense matching phases. All these tests have been performed with a metrological approach, that implements the following strategy:

- For each phase of the procedural pipeline, a set of most significant parameters has been examined, starting from a specific set of acquisition protocols; these latters have always been selected according to the particular objective of the performed analysis. Both choices will be illustrated at the beginning of each subsection.
- The analysis carried out within each evaluation step has always employed adequate reference data in order to perform metrological comparisons; as a rule of thumb, the measurements acquired by the reference instruments were always an order of magnitude more accurate than the analysed ones.
- According to the results achieved at the step N, the best parameter set was then employed in the subsequent step, (N+1).
- The above listed approaches have been carried out for each step of the pipeline, up to the dense image matching process. At the end of the assessment, a final best solution, in terms of both procedural parameters and acquisition protocol, has been identified.
- This best solution was finally re-processed, starting from a reduced image resolution selected during the tie point extraction phase.

Of course, the results achieved within this experimental application have not a general validity, since they are influenced by the specific operative conditions affecting this case study, in terms, for example, of datasets, hardware/software means, ambient and operators. Nevertheless, these studies offer a possible reference procedural workflow, that can be further applied in different case studies and operative conditions in order to set their specific best practices. Each phase of the workflow described in Figure 6.23 will be deepen discussed in the following subsections; starting from the topic of data acquisition. For the algorithmic and operative aspects related to the IGN's suite of tools, the reader is referred to Chapter 3.

6.3.2 Image acquisition

The digital image acquisition phase was performed using a Nikon D3X digital camera (6080 x 4044 pixels) and two different lenses: a fixed focal length lens (Nikon AF Micro-Nikkor 60mm f/2.8D) and a zoom-lens (Nikon AF Nikkor 24-85 mm f/2.8 4D IF). Both lenses are not equipped with optical image stabilizers that would represent a critical factor reducing the camera rigidity. The related technical specifications are listed in Table 6.11, whereas Figure 6.24 provides a view of the digital camera.

Nikon D3X	
BODY TYPE	Large SRL
FORMAT	Nikon FX
SENSOR RESOLUTION	24.5 Mpixel
SENSOR SIZE	Full Frame
SENSOR TYPE	CMOS
ISO	100-200-400-600-800-1600
MIN SHUTTER SPEED	30 sec
MAX SHUTTER SPEED	1/8000 sec
Nikon AF Micro-Nikkor 60mm f/2.8D	
FOCAL LENS	60 mm
MAX APERTURE	f2.8
MIN APERTURE	f32.0
MIN FOCUS	0.22 m
Nikon AF Nikkor 24-85 mm f/2.8 4D IF	
FOCAL LENGHT	24-85 mm
MAX APERTURE	f2.8 – f4.0
MIN APERTURE	f22.0
MIN FOCUS	0.50 m

Table 6.11 Technical specifications of Nikon D3X and lenses employed

Figure 6.24 Nikon D3X

Two focal lengths were tested and employed during the acquisition: besides the fixed focal length lens (60 mm), the zoom lens was used at its lower zoom scale (24 mm). Furthermore, for each lens, images were acquired with three different values of convergence angles (3°, 5°, 10°), following the suggested crosswise convergent configuration. Of course, in order to compare the results achieved with these different acquisition protocols, their corresponding range accuracy and lateral resolution (ΔZ and ΔX) were initially evaluated using the

following formulas, derived from mathematical equations presented in Chapter 2 and from geometrical observations:

$$\Delta Z = \frac{Z^2}{f_{eq} B} \times \Delta P = \frac{Z^2}{f_{eq}\left(2Z \tan\left(\frac{\alpha}{2}\right)\right)} \times \frac{0.5 \text{ pix}}{\sqrt{N-1}} \quad [6.1]$$

$$\Delta X = \text{pix} \times \frac{Z}{f_{eq}} \quad [6.2]$$

Where:

- $f_{eq} = f\frac{Z}{z-f}$, with f is the nominal focal length;
- Z is the camera-object distance (focusing distance);
- α is the angle of convergent images;
- B is the baseline, expressed in function Z and α;
- ΔP is the detector pixel size, expressed in function of the pixel size (pix) and number of images employed during the image matching process ($N = 5$ considering the crosswise configuration).

Thus, considering the two selected focal lengths and the three different convergence angles, Equations [6.1] and [6.2] were employed to identify the camera-object distances that minimize the difference between the lens performances, in terms of range and lateral accuracies. According to this requirement, 14 m and 26 m were, thus, chosen as mean acquisition distances for, respectively, the 24mm-lens and the 60mm-lens, providing the uncertainty values listed in Table 6.12.

Range Accuracy (mm)		Focal Length	
		24 mm	60 mm
Angle of Convergent Images	3°	13.6	10.1
	5°	8.2	6.0
	10°	4.1	3.0
Lateral Resolution (mm)		Focal Length	
		24 mm	60 mm
Angle of Convergent Images	3°	2.1	2.1
	5°		
	10°		

Table 6.12 Expected range accuracy and lateral resolution

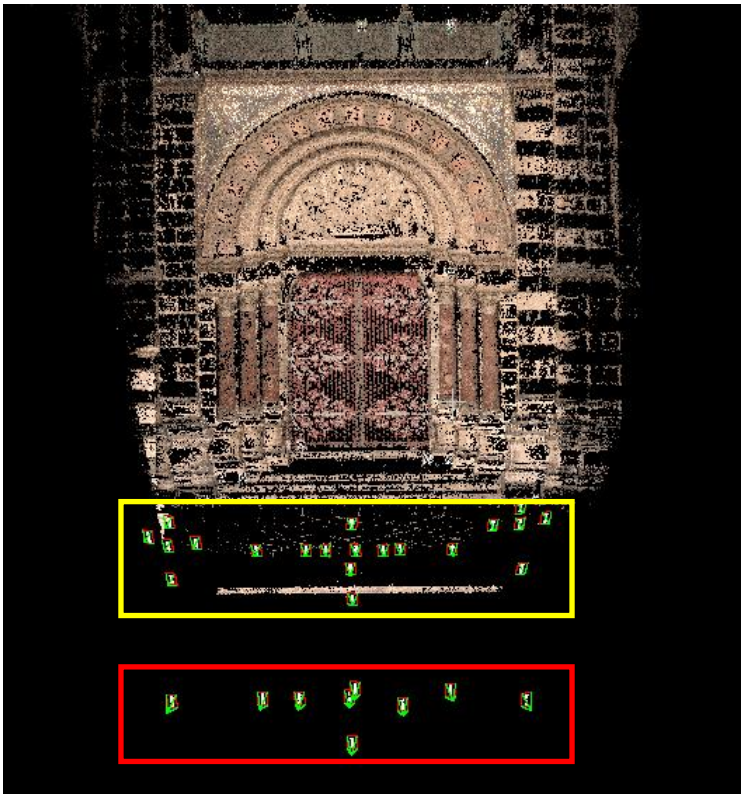


Figure 6.25 The spatial configuration of the 24mm-layout (yellow rectangle) and 60mm-layout (red rectangle)

Focal Length: 24 mm Camera-Object Distance: 14 m			**Focal Length: 60 mm Camera-Object Distance: 26 m**		
Point of View	Angle of Convergent Images (α)	Number of Images	Point of View	Angle of Convergent Images (α)	Number of Images
Central	3 °	5	Central	3 °	5
	5 °	4		5 °	3
	10 °	3		10 °	3
Left	3 °	5			
Right	3 °	5			

Table 6.13 Summary of the different acquisition protocols

Table 6.13 summarizes the different acquisition protocols, that are also graphically represented in Figure 6.25: this latter represents a sparse 3D reconstruction of the scene with the camera relative poses, computed with the tool AperiCloud.
As pointed out in Table 6.13, the complete crosswise configuration was adopted only for the 3°-datasets, since the other convergence angles required unrealizable vertical baselines.
Since both focusing distances were beyond the corresponding lens hyperfocal distances (Kraus, 1994), the focus was always set at infinity. The f-number and ISO sensibility were both kept fixed to, correspondingly, f8 and 200. Furthermore, the digital image acquisition was performed using a photographic tripod in order to reduce the vibration effects; the baselines and distances were measured through a Leica DISTO Plus laser rangefinder and a yardstick. Finally, in order to reduce as much as possible the illumination changes over the scene, the acquisition was carried out as quickly as possible, after having clearly identified the station positions on the ground.

6.3.3 Laser scanner tests and survey

FARO Focus^{3D}120	
SCAN PRINCIPLE	Time of Flight – Phase Measurement
SCAN RANGE	0.6 – 120 m
MEASUREMENT SPEED	Up to 976,000 points/second
RANGING ERROR	+/- 2 mm at 10 m and 25 m
LASER CLASS	Laser Class 1
WEIGHT	5.0 kg
DUAL-AXES INCLINATION SENSOR	Accuracy 0.015°; Range +/- 5°

Table 6.14 Technical specifications of FARO Focus^{3D}120

Figure 6.26 Faro Focus^{3D}120

The reference data required to evaluate the metric accuracy of the dense image matching algorithm, were acquired with a Time of Flight (Amplitude Modulation) Laser Scanner, the FARO Focus^{3D}120 (Figure 6.26). The related main technical specifications are listed in Table 6.14.

In order to perform a low-level characterization of the instrument (Chapter 4) and verify the specifications stated by the vendor, two experimental tests were performed in laboratory, before the on-the-field acquisition.

In the first test, four spheres of known diameters and a planar surface were put into the scene at different heights and depth levels; this set-up was then acquired with the laser scanner, after having selected its maximum level of resolution and quality (noise removal). Afterwards, the resulting point cloud (Figure 6.27) was imported and analysed into the software PolyWorks® (PolyWorks), module IMSurvey™, vs 12.1.18.

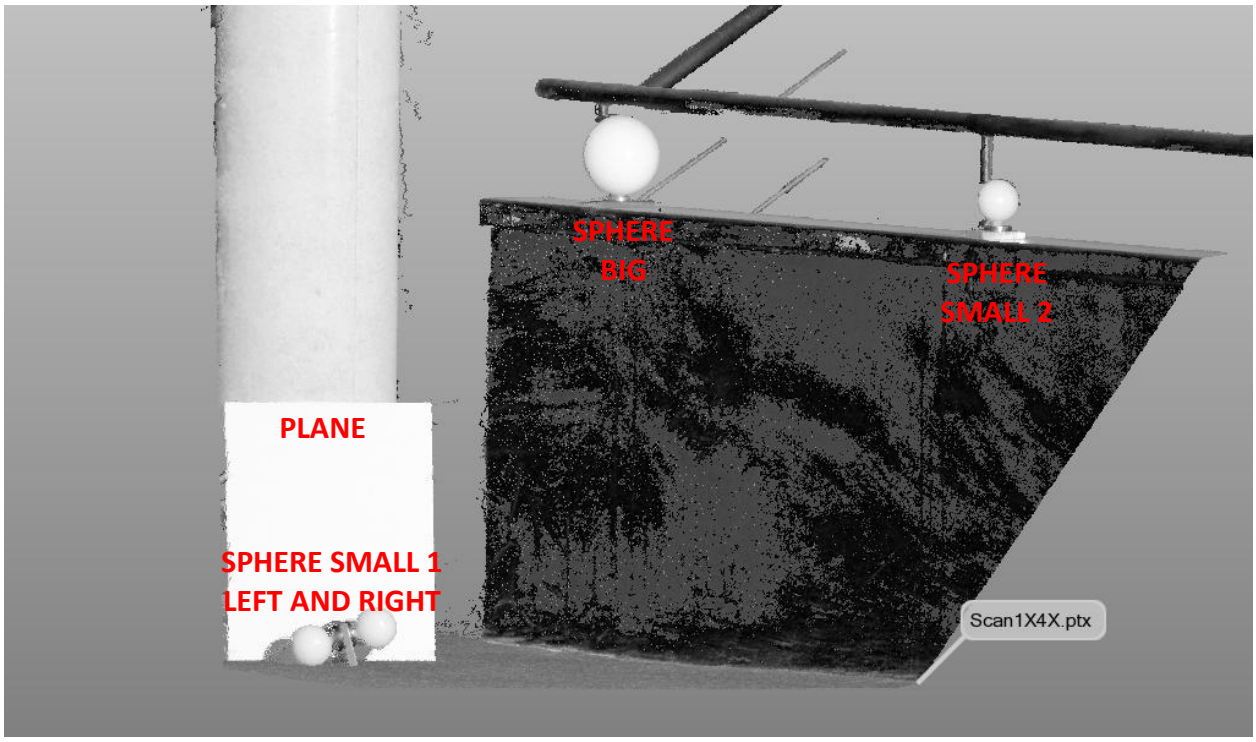


Figure 6.27 The primitive best-fitting experimental test

Object ID	Primitive best-fitting: Standard Deviation (mm)
Sphere small 1 left	0.5
Sphere small 1 right	0.6
Sphere small 2	0.5
Sphere big	0.8
Plane	0.8

Table 6.15 Primitive best-fitting experimental test

The corresponding best-fitting plane and spheres were then created: the displacement of each 3D point from these best-fitting primitives allowed finally to evaluate the standard deviations listed in Table 6.15. The results are all below one millimeter and show that the measurements are not affected by significant systematic errors.

The second experimental test allowed to confirm the results achieved in the previous one and, additionally, to evaluate the optical resolution of the instrument (propriety associated to the laser propagation, as described in Chapter 4). In addition to a plane, an *ad-hoc* created resolution chart was put into the scene. This arrangement was then acquired with the laser scanner, after having selected its maximum level of resolution and quality (noise removal): unlike the previous case, however, the acquisitions were here performed at four different instrument-object mean distances, i.e. 5, 10, 15, 20 m. Afterwards, the resulting point clouds (Figure 6.28) were imported and analysed into the software PolyWorks IMSurvey™.

Figure 6.28 The best-fitting and optical resolution estimation experimental test

The estimation of the optical resolution was performed with a qualitative approach, by analysing the reconstruction of the resolution chart in the different range maps. By knowing the exact distance between the lines depicted in the chart, it was possible to estimate the

capability of the scanner to discriminate two adjacent structures on the acquired surface (i.e., its lateral resolution). The results shown in Figure 6.29 give a clear, although only qualitative, idea of how this capability is influenced by the distance and, hence, by the angular resolution; they also provides a means of interpretation of the technical specifications declared in the datasheet, in terms of quality and spatial resolution.

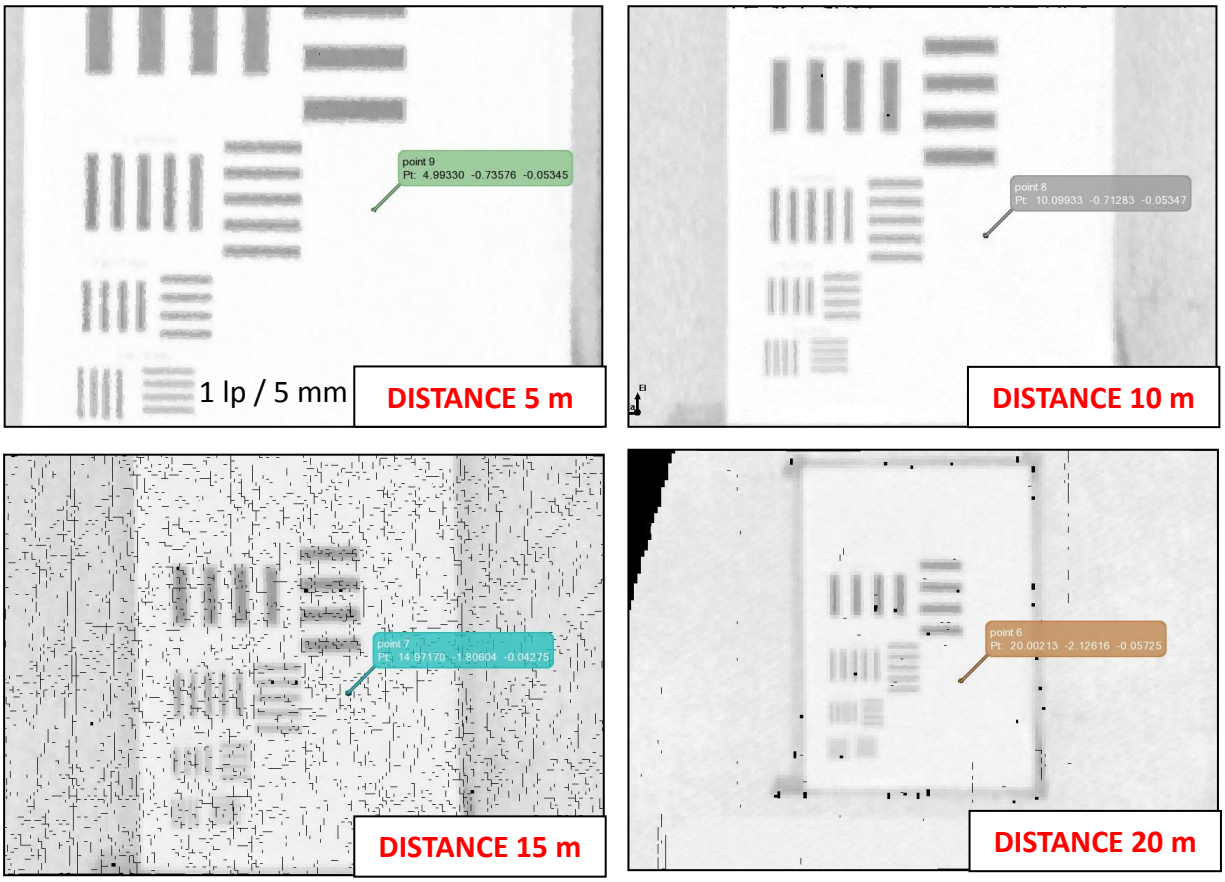


Figure 6.29 Estimation of optical resolution (Laser propagation)

Distance (m)	Plane best-fitting: Standard Deviation (mm)
5	0.8
10	0.7
15	0.9
20	0.8

Table 6.16 Plane best-fitting experimental test

The best-fitting plane was then extracted from each point cloud, and the resulting standard deviations (Table 6.16) were analysed and compared to each other's. The recovered flatness measurement errors are characterized by the same magnitude level of the previously computed data. In this case, however, it was possible to evaluate their trend as a function of the acquisition distance: a sweet spot of lowest range noise was thereby identified at an instrument-object distance of 10 m. Of course, all the results achieved in these experimental tests are only strictly valid in the laboratory conditions (hardware/software means, method, material, ambient and people) in which they were performed.

Starting from the results of the previously described low-level characterization, the distances and deriving lateral resolutions to be adopted during the on-the-field acquisition were thus decided. The main entrance of the Cathedral was acquired by setting both the quality and resolution parameters at their higher possible values. Furthermore, three different instrument-object distances were adopted and measured with a Leica DISTO Plus laser rangefinder, i.e. about 5–10–15 m. One of the three point clouds is shown in Figure 6.30, whereas Table 6.17 reports the values of lateral mean resolution corresponding to each of the three range maps.

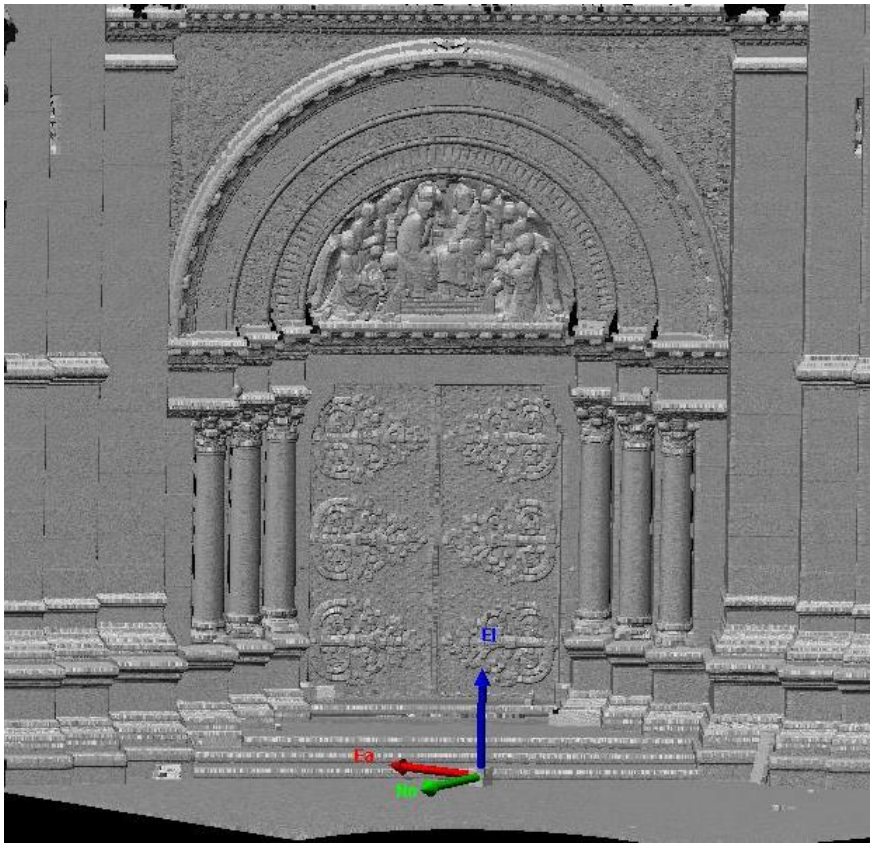


Figure 6.30 The laser scanner point cloud

A suitable number of spheres, provided with the instrument and identical to the ones used during the laboratory test, were put into the scene and acquired: their role will be explained in Subsection 6.3.8 and refers to the point cloud registration phase. The acquisition phase was performed using a tripod and delimiting the necessary eye-safety distance with a red and

white strip, keeping the dangerous area out of the passage and stationing of visitors. The eye-safety distances were computed through the datasheet specifications. Furthermore, to reduce as much as possible the vibrations effects caused by the strong wind, adequate barriers of protections were raised around the area of interest.

Instrument-object distance (m)	Lateral mean resolution (mm)
4.81	1
9.74	6
14.95	9

Table 6.17 Lateral mean resolution at the three different acquisition distances

Finally, a plane best-fitting analysis was performed on the main beam above the door using the software PolyWorks IMSurvey™ 12.1.18. The resulting Root Mean Square Errors (RMSE) are listed in Table 6.18, as a function of the three acquisition distances. The values reflect the results achieved within the experimental tests, confirming the presence of a sweet spot of lowest range noise at an instrument-object mean distance of about 10 m. Starting from these analyses, a local measurement uncertainty[9] $u_R(LS)$=1 mm was assumed as realistic representation of the noise level present in the 3D point clouds.

Distance (m)	Plane best-fitting on the main beam RMSE (mm)
4.81	1.0
9.74	0.8
14.95	1.0

Table 6.18 Flatness measurement errors at the three acquisition distances (main beam analysis)

6.3.4 Total station survey

The reference data required to evaluate the metric accuracy of the orientation algorithm, were acquired with a Total Station, TS LEICA Plus Ultra 3'' by Leica Geosystems (Figure 6.31).

[9] According to (VIM3), uncertainty should be used to express the accuracy of a measurement. Hereinafter the symbol *u* will be used to identify the standard uncertainty (1σ) , whereas the symbol U will be used in the case of an expanded uncertainty, i.e. the product of a standard uncertainty and a coverage factor k larger then 1 (here, k=2). The subscript R refers to the radial uncertainty, whereas the subscript T identify the transverse one.

The related main technical specifications are listed in Table 6.19. Three different types of target were previously prepared in the laboratory and glued on planar rigid supports, made of different materials, such as plywood, metal and granite. Figure 6.32 show the three typologies of adopted target.

TS LEICA Plus Ultra 3''	
ANGLE MEASUREMENT PRECISION	
WITH REFLECTOR	1" (Hz) – 2" (V)
WITHOUT REFLECTOR	3" (Hz) – 5" (V)
DISTANCE MEASUREMENT PRECISION	
WITH REFLECTOR	1.5 mm + 2 ppm
WITHOUT REFLECTOR	2 mm + 2 ppm

Table 6.19 Technical specifications of TS LEICA Plus Ultra 3''

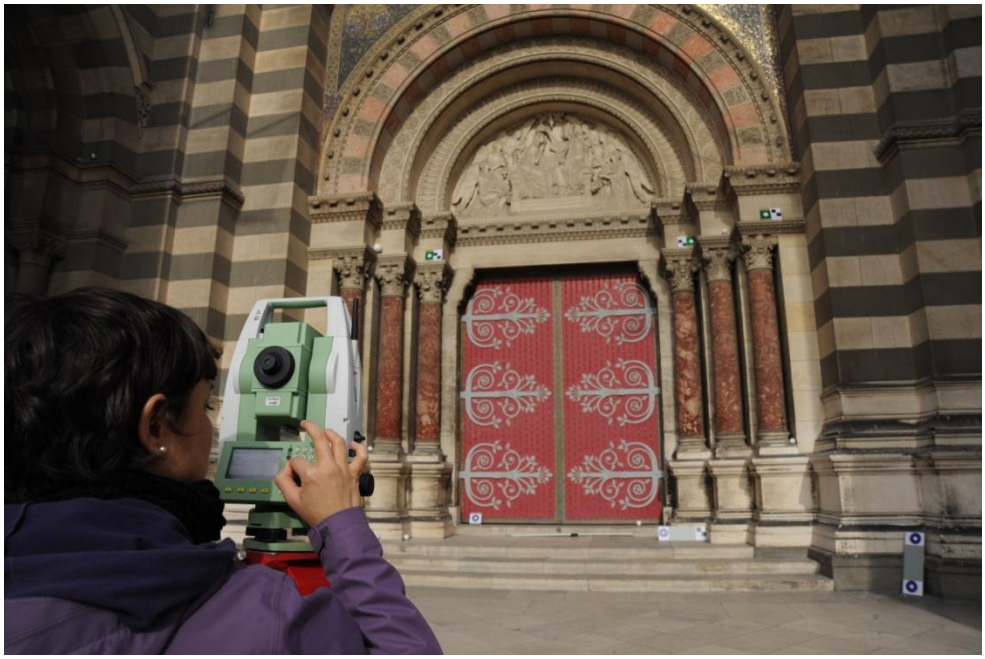


Figure 6.31 The Total Station survey (TS LEICA Plus Ultra 3'')

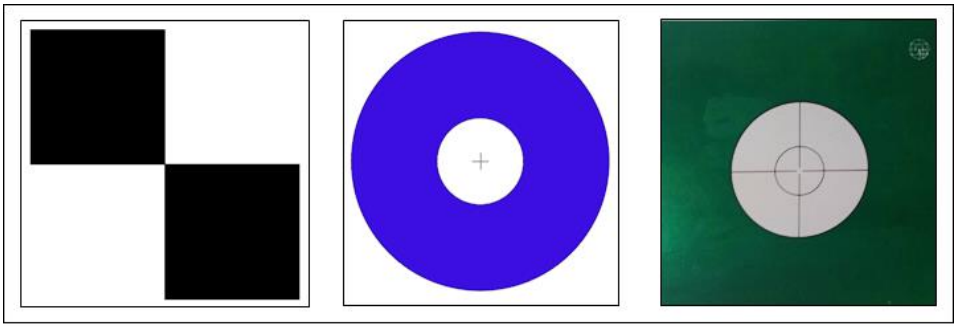


Figure 6.32 The three different typologies of target employed

A total of 19 targets were positioned in the scene of interest, so that their resulting spatial configuration was as well-distributed as possible throughout the entire acquired 3D volume: in particular, this configuration was designed in order to have an adequate number of known points well distributed along the edges of the scene (since the image borders are always characterized by the highest values of radial geometrical distortion) and in its central part. Both ladders and ropes were used to place all the targets in their designed location. Besides the targets, 4 "natural" points, i.e. not pre-signalized, were also measured and chosen within the areas where no targets were placed due to practical reasons. Of course, well recognizable details were selected, favouring the "natural" intersections of linear elements and corner points: the central relief and the decorations above the principal arch represented two optimal regions where to look for such a points. Finally, a graduated bar was placed horizontally on the floor at the foot of the door: two small targets were also glued on this bar and measured with the total station. Although all targets were placed on rigid supports and in protected positions, some of them have been moved by the power of the wind: of course, these points were not used in the following orientation phase. A total of 16 points, comprising both targets and natural details, were considered stable and used as Ground Control Points (GCPs) and Check Points (CPs).

The maximum acquisition distance was about 16.6 m. By means of this information and the ISO-17123-4, the uncertainty calculation shows that the main uncertainty comes from the range distance estimation. The elevation and azimuth angular uncertainties at 16.6 m are, in fact, about one quarter of the range uncertainty. The combined uncertainty for the range is about $u_R(TS)$=2.18 mm (1σ) and expanded uncertainty $U_R(TS)$=4.36 mm (k=2). One can thus see that the TS becomes a limiting factor in the present accuracy assessments.

6.3.5 Influence of image resolution on tie point extraction and orientation

<table>
<tr><th colspan="3">Focal Length: 24 mm</th><th colspan="3">Focal Length: 60 mm</th></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.20 The datasets used in the procedural step "Influence of image resolution on tie point extraction and orientation"

Before discussing the first step of the evaluation procedure, the structure of the Table 6.20 will be here explained, as it will be reported at the beginning of each following section. The table will always list all the possible available image datasets, corresponding to the different acquisition protocols: the rows highlighted in green will identify the datasets employed in each specific assessment step. Three observations should be detailed:

1. Even if a total of 4 images have been acquired with the 24mm-lens at 5° of convergence angle, only 3 of them (the master central image and the horizontal stereo-pair) have been used when this specific dataset is employed alone or together with the other two datasets acquired from a central point of view. In this way, the comparisons with the corresponding 60mm-lens datasets will be performed starting from the same number of images.
2. For each focal lens and from each point of view, the central master image is always the same for all the three datasets corresponding to the different convergence angles.
3. When considering all the images acquired with the 24mm-lens, the image removed at point 1 is again re-added and re-used, since the reason explained at point 1 no longer applies; the same observation is also valid when considering all the images acquired with the two lenses together.

This first evaluation test aims at assessing the impact of different "working" resolutions on the tie point extraction and relative orientation phases. The two datasets acquired with a convergence angle equal to 3° have been here employed, since they constitute a simple, but "complete" example of the crosswise configuration protocol. Both datasets have been first processed with the tool Tapioca (Subsection 3.3.2) in order to extract the homologous points between the input images. The search mode has always been set to the value "All", i.e. all possible image pairs have always been compared to each other's. The parameter "Size" has been iteratively adjusted in order to evaluate different "desired" widths for shrinking the images during this phase: in particular, its value has been increased from 1000 to 6080 (corresponding to the original image width), with an intermediate step of 500. After each process, the extracted tie points have then been used as input for the subsequent orientation phase, carried out with the tool Tapas (Subsection 3.3.3). The FraserBasic calibration model was selected for all these first tests. In order to evaluate and compare the effects deriving from the different image shrinking resolutions, three variables have been finally considered:

- The mean number of extracted tie points as a function of the selected image width; for each test, this value has been computed by averaging the number of tie points extracted from each of the five images (Figure 6.33);
- The Root Mean Square Error (RMSE) of the orientation procedure as a function of the selected image width; for each test, this value was computed by the tool and appeared at the end of the computation as "Résidu Liaison Moyen" (Figure 6.34);
- The computational time required to run the entire procedure, from tie point extraction up to the end of the relative orientation step; also this value is reported as a function of the selected image width (Figure 6.35). The hardware information on the employed processing environment is detailed in Table 6.21.

The results are shown below; in order to make them more generally applicable, the normalized image width is always reported along the x-axis and computed as a percentage of the original image full width.

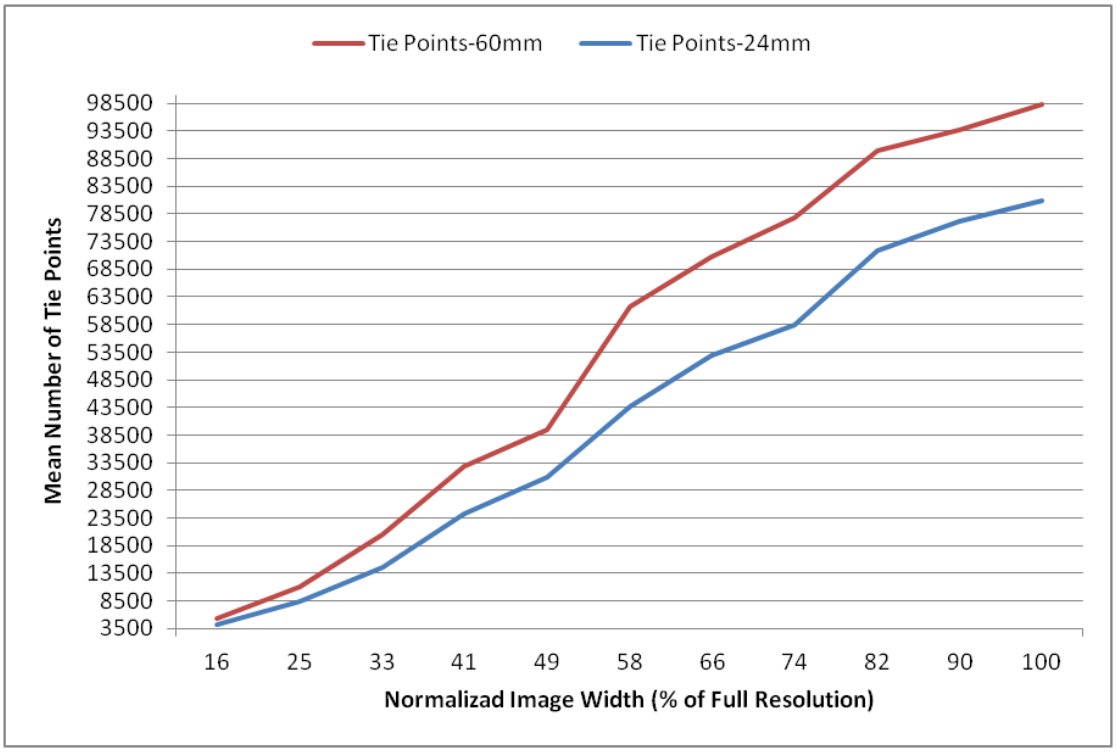


Figure 6.33 The mean number of extracted tie points as a function of the selected image width

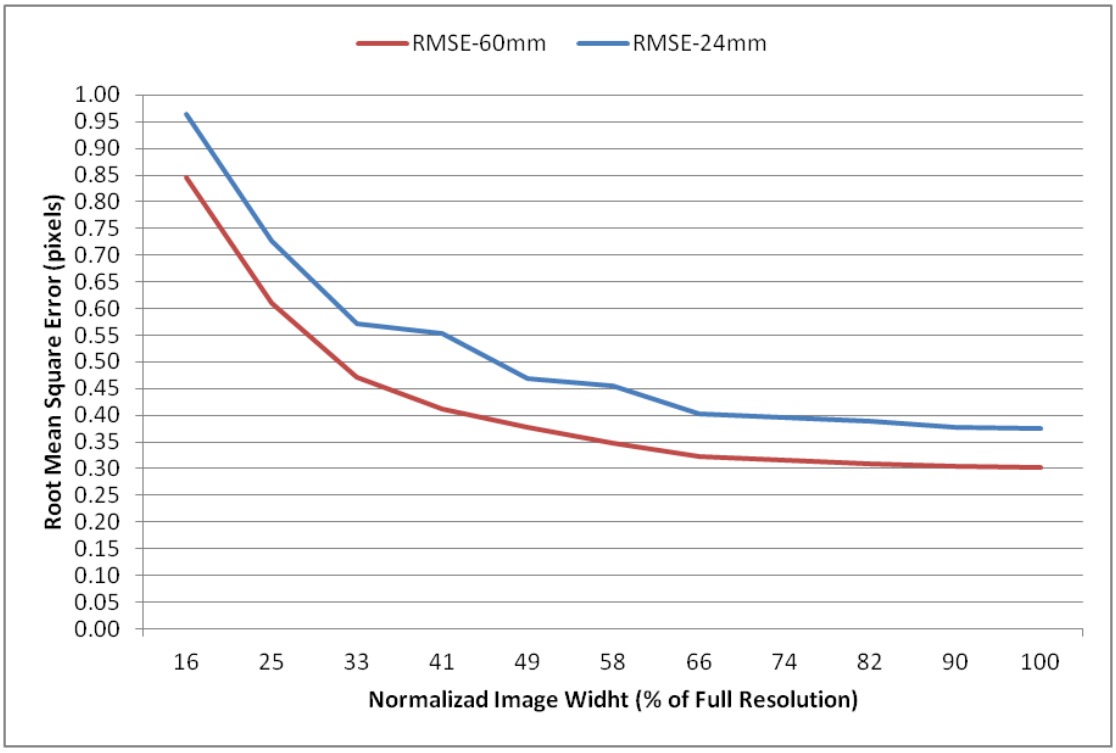


Figure 6.34 The RMSE of the orientation procedure as a function of the selected image width

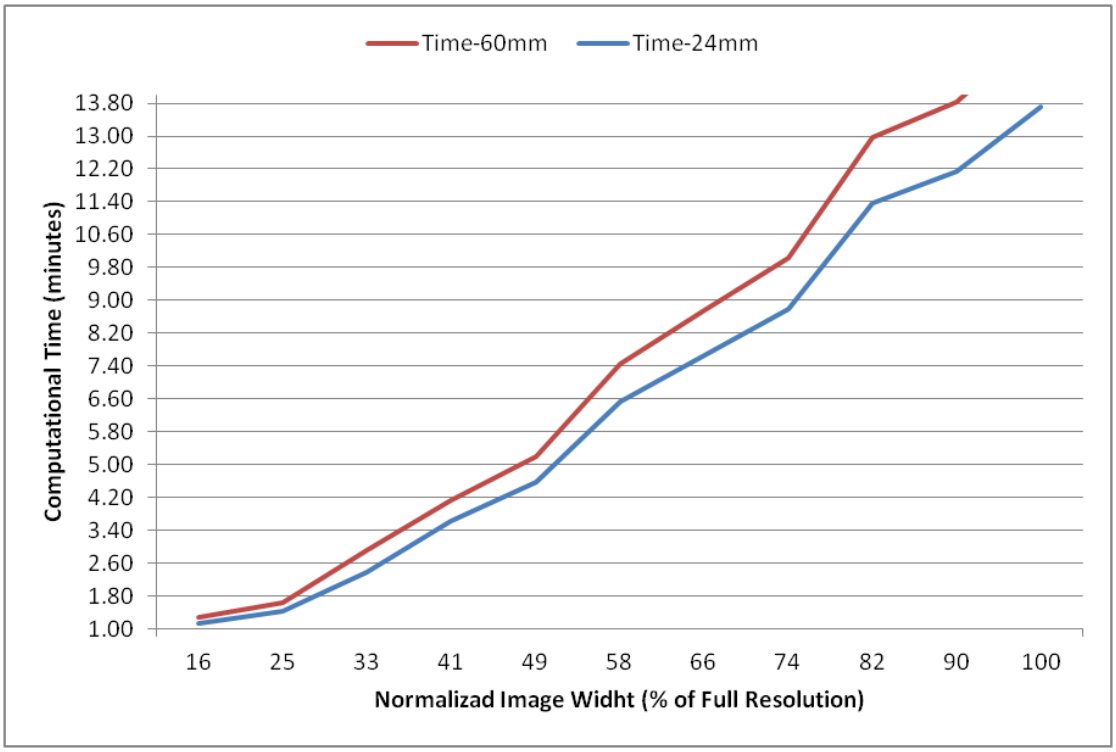


Figure 6.35 The computational time as a function of the selected image width

HARDWARE INFORMATION	
NUMBER OF PROCESSORS	12
PROCESSOR ARCHITECTURE	AMD64
PROCESSOR DESCRIPTION	Intel(R) Xeon(R) CPU E5-1650 0 @ 3.20GHz
FILE SYSTEM	NTFS
MAIN MEMORY	
TOTAL PHYSICAL	17.9277 GB
AVAILABLE PHISICAL	9.26563 GB
FIXED DISK DRIVES	
TOTAL SPACE	882.894 GB
AVAILABLE SPACE	525.249 GB

Table 6.21 Hardware information on the employed processing environment

Both lenses exhibit the same general trends, that can be summarized as follows:

- The mean number of extracted tie points increases with the increase of the working image width, displaying an almost linear trend; in other words, as expected, the image resolution plays a significant role in the tie point extraction phase, in terms of the amount of extracted homologous points;
- Of course, by increasing the number of extracted tie points, the RMSE of the corresponding relative orientation decreases: homologous points represent, in fact, the predominant information (the only one in this case), according to which the initial

solution is computed and then refined by the orientation algorithm. It is, however, interesting to observe that this decrease doesn't follow a linear trend, but an exponential one and is, thus, no longer significant beyond a certain level of image resolution. In other words, by increasing the desired image width over a threshold equal approximately to the 60% of its original value, one cannot achieve a significant improvement of the orientation accuracy.

- The computational time, as expected, follows the same trend shown by the number of extracted homologous points, i.e. it increases almost linearly with the increase of the working image width.

Furthermore, by comparing the behaviours of the two lenses, one can easily infer that the 60mm-dataset achieves always the best results, both in terms of extracted tie points and final RMSE. The 60mm-lens is, in fact, characterized by a lower level of geometrical deformations and distortions, since it is more "close" to the ideal case of a 50mm-lens on a full-frame sensor. The lens aberrations and distortions, are, in fact, lower and more easily correctable if the focal length of the employed lens is close to its "normal" value, i.e. equal to the diagonal of the camera sensor. Such a lens, in fact, is able to achieve a perspective view more similar to the one produced by the human eyes, thus reducing the distortion effects. Furthermore, range and lateral accuracies are improved by the use of a longer focal length, as shown by triangulation equations [6.1] and [6.2].

Since it is now feasible to parallelize several computations across multiple cores, the computational time is no longer an urgent issue; moreover, the original image width seems to reach the best results, especially in terms of the amount of extracted homologous points among the images Thus, next steps of the evaluation procedure will be performed using the full image resolution during the tie point detection phase; nevertheless, at the end of the process, the entire 3D reconstruction pipeline will be performed again using a reduced image size, in order to evaluate the advantages of such a choice, especially in terms of the required overall computational time.

6.3.6 Influence of calibration approach on orientation

This second evaluation step aims at assessing the impact of different calibration strategies on the accuracy of the relative orientation phase. A self-calibration approach (Chapter 2) was initially adopted: thus, calibration was carried out directly during the bundle adjustment procedure, using the same images that would have been further employed in the image matching phase. Secondly, a different set of images was specifically acquired for each lens, in order to be "favorable" to the calibration procedure; in particular, the general rules deduced in Subsection 2.1.3 were considered, trying to fulfill the following requirements:

- All images converging to the same part of the scene, to facilitate the computation of external orientation;
- A scene characterized by significant depth variations, in order to facilitate the focal length estimation;

- A scene characterized by well-textured surfaces, in order to facilitate the homologous point extraction;
- Multiple photo stations with different roll angles (horizontal, vertical, oblique), great image point density and covering the entire image format with scene points.

A building corner was chosen as favorable 3D scene, since it displays significant texture and different depth levels. It was thus acquired with each of the two lenses, keeping an infinity focusing and fixing the f-number and ISO parameters to the same values employed during the acquisition of the cathedral datasets. A total of 10 images were taken for each focal length (Figure 6.36) and then processed with the IGN tools. After having computed a set of calibration parameters for each of the two lenses, they were then used as initial values for the calibration and orientation procedure of the cathedral datasets. In particular, two strategies were carried on, by:

- re-evaluating the calibration parameters using the additional information given by the processed cathedral datasets, in terms of extracted tie points; the option AutoCal was here selected;
- keeping the calibration parameters frozen during the bundle adjustment procedure of the cathedral datasets; the option Figee was here selected.

Both the self-calibration approach and the pre-calibration one, performed with the corner datasets, were carried out using the tool Tapas and selecting the FraserBasic calibration model.

Figure 6.36 The corner dataset acquired with the 24mm-lens

As usually, the employed cathedral datasets are shown in the following Table 6.22, together with the corresponding number of images.

<table>
<tr><td colspan="3">Focal Length: 24 mm</td><td colspan="3">Focal Length: 60 mm</td></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.22 The datasets used in the procedural step "Influence of calibration approach on orientation"

In order to analyse the metric accuracy of the orientation procedure performed with the three different calibration strategies (i.e. self-calibration, pre-calibration & re-evaluation, pre-calibration & freezing), the RMSEs of the re-projection residuals computed by Tapas were finally evaluated and compared. Results are listed in Table 6.23.

RMSE (pixel)	**Dataset**		
Calibration strategy	24 mm, 3° (5 Images)	60 mm, 3° (5 Images)	24 mm (All 20 Images)
Self-Calibration	0.38	0.31	0.50
Pre-Calibration & Autocal	0.38	0.31	0.50
Pre-Calibration & Figee	0.418	0.33	0.52

Table 6.23 RMSE of the orientation procedure as a function of the image dataset and calibration strategy

Three observations can thereby be inferred:

- The self-calibration approach achieves an accuracy level equal to the one gathered by the strategy based on a pre-calibration procedure, followed by a refinement phase. The cathedral datasets represent, in fact, a calibration-favorable image dataset, since they exhibit significant depth variations and surface textures.

- The Figee option reduces the metric performance of the orientation procedure, showing that a parameter re-evaluation process during the bundle adjustment is able to refine the first calibration estimate and to improve the metric accuracy.
- As already shown in the previous subsection, the 60mm-dataset achieves the best results in terms of relative orientation accuracy; also in this case, the same motivations explained in Subsection 6.3.5 can be easily advanced.

In order to employ always the same calibration input, the following tests will be performed starting from the calibration parameters computed by processing the two corner datasets: these parameters will be then re-evaluated and refined during the bundle adjustment phase performed with each cathedral dataset.

6.3.7 Influence of acquisition protocol on orientation

This third evaluation step performs an analysis of the influence played by the acquisition layout on the orientation procedure. Of course, all the different acquisition protocols have been here tested, as highlighted in Table 6.24.

<table>
<tr><th colspan="3">Focal Length: 24 mm</th><th colspan="3">Focal Length: 60 mm</th></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.24 The datasets used in the procedural step “Influence of acquisition protocol on orientation”

Initially, all datasets were oriented with the pipeline selected in the previous two evaluation steps. Note that the two datasets acquired from the left and right points of view, although oriented, will not be here discussed separately, but only together with the images acquired from a central point of view, defining thereby the dataset termed “central-left-right”.
An ad-hoc procedure should be defined for the last dataset, including all the 29 images acquired with the two lenses. Four different approaches were testes, i.e.:

- Orientation of the 60mm-dataset based on the canvas of the already oriented 24mm-dataset; the 24mm-poses are kept frozen;

- Orientation of the 60mm-dataset based on the canvas of the already oriented 24mm-dataset; the 24mm-poses are re-evaluated during the bundle adjustment;
- Orientation of the 24mm-dataset based on the canvas of the already oriented 60mm-dataset; the 60mm-poses are kept frozen;
- Orientation of the 24mm-dataset based on the canvas of the already oriented 60mm-dataset; the 60mm-poses are re-evaluated during the bundle adjustment.

The RMSEs of the relative orientation procedures performed with the datasets acquired with the two distinct lenses are listed in Table 6.25; furthermore, Table 6.26 shows the accuracy of the four different strategies carried on in order to orientate the biggest dataset, including both 24mm and 60mm-images.

	24 mm		**60 mm**	
Dataset	No. Images	RMSE (pixel)	No. Images	RMSE (pixel)
3°	5	0.38	5	0.30
5°	3	0.31	3	0.23
10°	3	0.30	3	0.22
3+5+10°	9	0.43	9	0.34
Central-left-right	20	0.50	-	-

Table 6.25 Accuracy of the relative orientation (distinctive lenses)

Strategy (24 mm + 60 mm)	**Description**	**RMSE (pixel)**	**Time (h)**
0	• Orientation of 24mm images; • Orientation of 60mm on the 24mm "canvas" (24mm-poses are frozen)	0.46	1.10
1	• Orientation of 24mm images; • Orientation of 60mm on the 24mm "canvas" (24mm-poses are re-evaluated)	0.47	1.08
2	• Orientation of 60mm images; • Orientation of 24mm on the 24mm "canvas" (60mm-poses are frozen)	0.47	1.10
3	• Orientation of 60mm images; • Orientation of 24mm on the 60mm "canvas" (60mm-poses are re-evaluated)	0.47	1.10

Table 6.26 Accuracy of the relative orientation (both lenses together)

Starting from the results listed in Table 6.25, it is possible to deduce that the 60mm-dataset behaves always better in terms of relative orientation accuracy, if compared with the corresponding 24mm-datasets. In particular, the 10° angle of convergent images seems to represent the optimal choice. The four different strategies carried out with the group of all 29 images are, in this case, equivalent in terms of both RMSE and computational time. The first one, termed strategy 0, will be further employed in the following evaluation steps.

In order to perform an "*a-posteriori* evaluation" of the orientation accuracy using some external known reference data, relative orientations were then converted into absolute ones employing the measurements acquired with the Total Station. Among the 16 points of known 3D coordinates, 7 of them were chosen as Ground Control Points (GCPs) during the bundle adjustment procedure (Figure 6.37). They were always manually collimated on the same three images for each focal length (tool SaisieAppuisInit); then the global transformation from a purely relative orientation to the one "included" in the reference frame of GCPs was carried out with the tool GCPBascule (Subsection 3.3.3). Finally, the tool Campari (Subsection 3.3.3) was run in order to perform a compensation (bundle adjustment) of all the provided heterogeneous observations, i.e. tie points and GCPs.

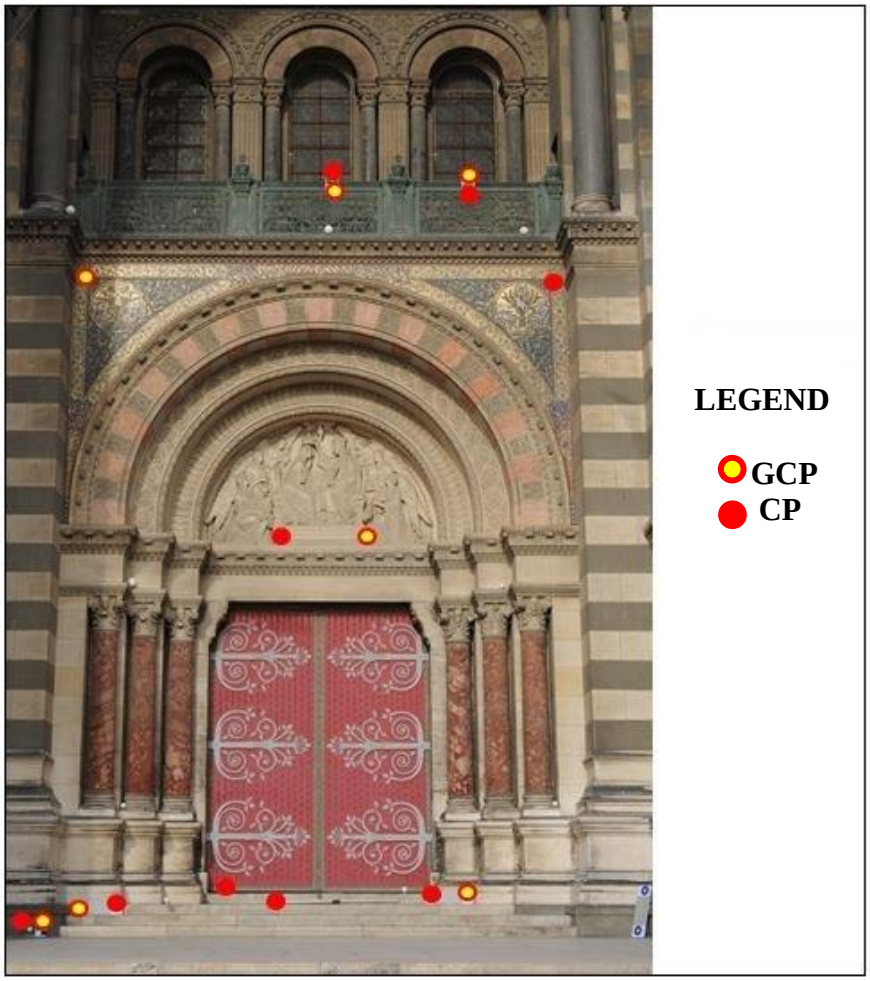


Figure 6.37 Configuration of the 7 GCPs and 9 CPs

Once the absolute external parameters have been computed, it is possible to re-project each pixel (visible in at least two images) into the 3D absolute space, and compute its 3D position by simply intersecting the homologous rays. This can be done with the tool SaisieAppuisInit, by "clicking" a point in two or more oriented images and looking for its computed 3D coordinates. This approach was performed in order to assess the metric accuracy of the absolute orientation procedure. In particular, the 9 points (Figure 6.37) that had not been used during the orientation procedure as GCPs were here employed as Check Points (CPs): they were all manually collimated on the same three images that were also chosen for the previous step. The computed 3D coordinates were later compared to the ones measured with the Total Station and corresponding residuals were finally calculated for each of the three absolute coordinates (X,Y,Z). The resulting Standard Deviations (Std. Dev.) are listed in Table 6.27 as a function of the initial acquisition protocol.

	24 mm			**60 mm**		
Dataset	Std. Dev. X (m)	Std. Dev. Y (m)	Std. Dev. Z (m)	Std. Dev. X (m)	Std. Dev. Y (m)	Std. Dev. Z (m)
3°	0.012	0.015	0.013	0.009	0.014	0.006
5°	0.021	0.029	0.014	0.019	0.017	0.021
10°	0.019	0.028	0.024	0.015	0.016	0.014
3+5+10°	0.009	0.013	0.011	0.005	0.002	0.003
Central-left-right	0.002	0.002	0.003	-	-	-
	24 mm + 60 mm					
	Std. Dev. X (m)		Std. Dev. Y (m)		Std. Dev. Z (m)	
Strategy 0	0.003		0.002		0.003	

Table 6.27 *A-posteriori* validation of the absolute orientation

As expected, also the *a-posteriori* validation confirms that the most accurate results can be achieved with the 60mm-lens (highlighted in red), if they are compared with the corresponding ones gathered by the 24mm-lens. Furthermore, the role played by the bundle adjustment phase is here more evident: while in the relative orientation procedure (where only tie points are used) results are not so influenced by the number of processed images, in the absolute one the compensation of more numerous and heterogeneous observations (tie points

and GCPs) achieves the best results with the bigger datasets (highlighted in yellow), i.e. when more images are processed. For these datasets a millimetre-level accuracy is achieved for all the three coordinates, showing an impressive metric potentiality of the orientation algorithm.

6.3.8 Influence of MicMac parameters on dense image matching

As shown in Chapter 3, MicMac is a very parametrical tool, allowing the user to perform an effective control over almost all the geometrical and mathematical aspects that characterize the implemented algorithm. This fourth evaluation step aims at analysing the influence of such a parameterization on the dense image matching procedure. Within the huge amount of available and adjustable parameters, three main variables have been here tested: Regularization Factor (Regul), Z-Quantification Factor (ZPas) and Final Z-Resolution (ZoomF). The role played by these parameters within the image matching procedure (Subsection 3.3.4) is, in fact, particularly significant and should be studied from a metric point of view. The influence of the other available parameters, since less substantial, is not studied in this present research work: thus, their default value will always be adopted during the following tests.

In order to test and compare many different parameter values and their resulting metrological impact, a small image dataset, allowing reasonable computational efforts, should be chosen. Starting from the results achieved in the previous step (Table 6.27) and according to the above mentioned requirement, the optimal choice should lie within the group of 60mm-datasets highlighted in red. A preliminary test was thus carried out in order to choose one of the three options. The tool Malt (Subsection 3.3.4) was employed to extract a depth map from each dataset, using the image-ground geometry and its default parametrical values. The resulting depth maps were then converted into point clouds (tool Nuage2Ply) and analysed with the software PolyWorks, module IMInspect™, vs 12.1.18. In particular, the flatness error associated with the main beam over the door was computed, by extracting a best-fitting plane from the three point clouds. The resulting Standard Deviations (Std. Dev.) and Root Mean Square Errors (RMSE) are listed in Table 6.28.

	Best-fitting plane – Main beam		
	60mm – 3°	60mm – 5°	60mm – 10°
Std. Dev. (m)	0.008	0.006	0.004
RMSE (m)	0.008	0.006	0.004

Table 6.28 Best-fitting planes on the three 60mm-datasets

The third dataset, corresponding to a 10° convergence angle and highlighted in yellow, was thus selected and employed for the following tests. Table 6.29 summarizes, as usual, the dataset selected for the evaluation of the Image-Based Modelling (IBM) approach.

<table>
<tr><th colspan="3">Focal Length: 24 mm</th><th colspan="3">Focal Length: 60 mm</th></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.29 The dataset used in the procedural step "Influence of MicMac parameters on dense image matching"

In order to perform a metric evaluation of the dense image matching accuracy, two different strategies were designed and adopted, i.e.:

1. Direct comparisons with known reference data, considering the point clouds acquired with the Laser Scanner (LS) to be adequate reference data of known uncertainty. Of course, this approach requires all compared entity to be registered in the same external reference frame (the Total Station one, in this case). Thus, a pre-processing phase was performed on the LS point clouds: at first, a relative external orientation of the three range maps was carried out using the FARO Laser Scanner software SCENE (SCENE) and the provided spheres. Then, relative orientations were converted into absolute ones employing the measurements acquired with the Total Station and the software PolyWorks IMSurvey™. The same GCPs used for the image datasets were also exploited for the LS ones. The registered point clouds were later manually filtered, in order to remove gross errors and noisy clusters. For these tests, the open-source software CloudCompare vs 2.4 (CloudCompare) was used, in order to calculate the distances between the reference data (LS clouds) and the compared entities (IBM clouds). The mean values of the computed distances (Mean. Dist.) and the corresponding standard deviations (Std. Dev.) were then analysed and will be further reported. Before each comparison, the registration between the compared entities was always refined with the ICP (Iterative Closest Point) algorithm (Besl and McKay, 1992) implemented in CloudCompare, in order to reduce as much as possible any misalignment effect. Three significant portions of the acquired 3D scene (Figure 6.38) were selected for the comparison tests; in particular, by exploiting the vertical symmetry of the object, the following elements were analysed:
    - The right half of the portal;
    - The right half of the relief on top of the door;
    - The right half of the door.

The last two entities were selected in order to perform a specific analysis on finely detailed surfaces. Finally, the LS range map acquired at about 10 m was here employed as reference data. In fact, in these kinds of evaluation tests range accuracy plays a more significant role than lateral resolution.

2. In order to perform a metric assessment that may be independent of LS data and their accuracy, a second type of tests was also carried out. The IBM point clouds were analysed with the software PolyWorks IMInspect™ and best-fitting geometrical primitives were extracted from significant portions of the acquired 3D scene. In particular, four regions were considered (Figure 6.38):
    - Part of the main beam over the door (best-fitting plane);
    - The column to the right of the door (best-fitting cylinder);
    - A dark pattern of the pillar to the right of the door (best-fitting plane);
    - A light pattern of the pillar to the right of the door (best-fitting plane).

    Deriving Standard Deviations (Std. Dev.) and Root Mean Square Errors (RMSE) have finally be compared and will be further listed.

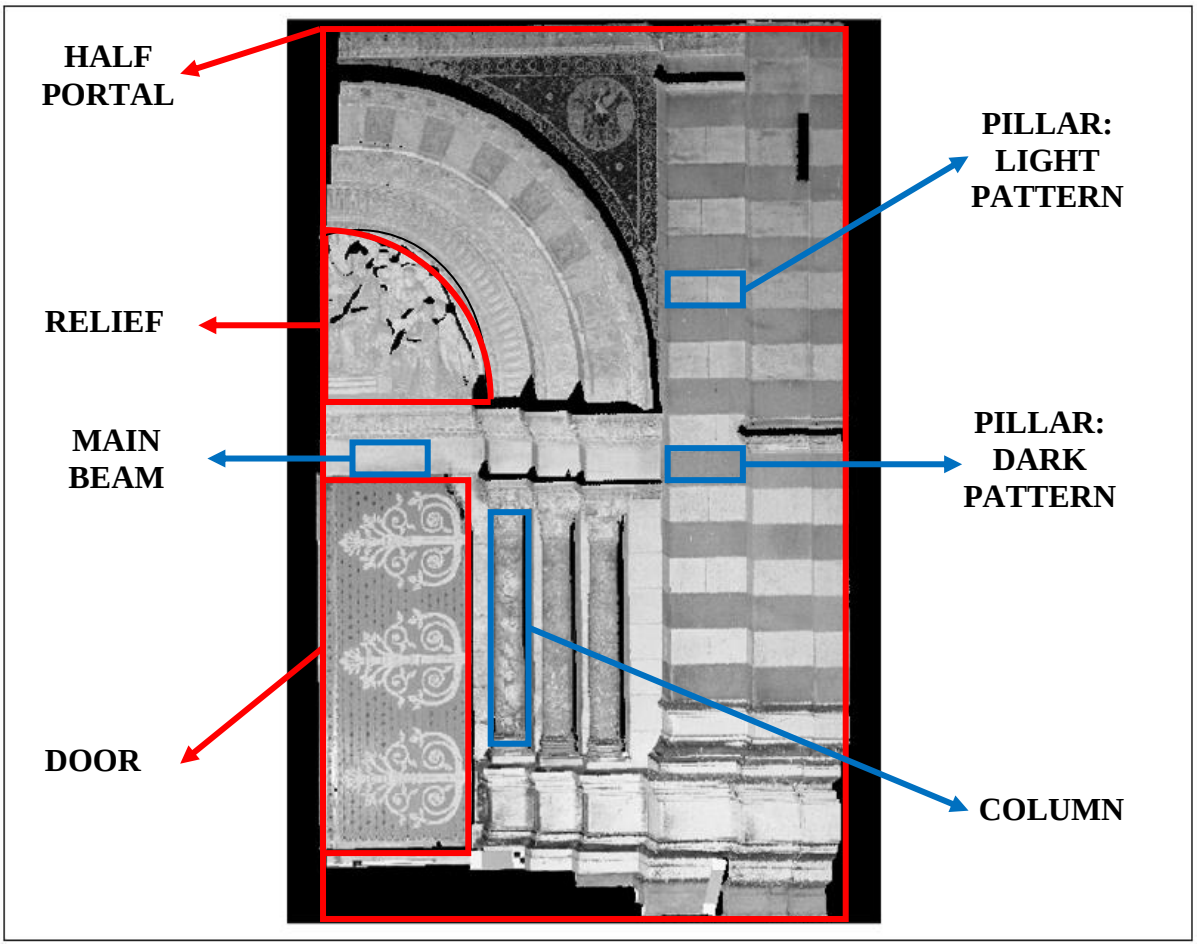


Figure 6.38 The significant regions considered by the two types of evaluation tests: LS-IBM comparison (red) and primitive best-fitting (blue)

With regards to the IBM procedure, the tool Malt, with its image-ground geometry, was always used to extract the depth maps, that were finally converted into point clouds with the tool Nuage2Ply. In each computation set, only one of the three selected parameters (Regul, ZPas and ZoomF) was varied and forced to assume five different values, that always included the default one (Def.) too. The remaining two parameters were kept frozen and set to their default values. Results are presented in Table 6.30-32, where they are grouped according to the evaluated parameter. Considering the LS accuracy, the geometric resolution of the digital camera and its pixel size on the object, four significant digits are always reported. Finally, Table 6.33 gives an idea of the computational effort required by the different processes, that were all run in the hardware environment previously detailed in Table 6.21.

PARAMETER: REGUL					
Half Portal (direct comparison)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Mean Dist. (m)	0.0114	0.0113	0.0045	0.0108	0.0106
Std. Dev. (m)	0.0106	0.0105	0.0009	0.0101	0.0103
Relief (direct comparison)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Mean Dist. (m)	0.0115	0.0114	0.0112	0.0112	0.0106
Std. Dev. (m)	0.0085	0.0083	0.0082	0.0082	0.0075
Door (direct comparison)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Mean Dist. (m)	0.0130	0.0129	0.0123	0.0119	0.0094
Std. Dev. (m)	0.0111	0.0110	0.0106	0.0101	0.0070
Column (primitive best-fitting)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Std. Dev. (m)	0.0044	0.0047	0.0037	0.0036	0.0039
RMSE (m)	0.0044	0.0047	0.0037	0.0036	0.0039

Pillar – Dark pattern (primitive best-fitting)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Std. Dev. (m)	0.0086	0.0084	0.0077	0.0071	0.0033
RMSE (m)	0.0086	0.0084	0.0077	0.0071	0.0033
Pillar – Light pattern (primitive best-fitting)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Std. Dev. (m)	0.0029	0.0026	0.0021	0.0017	0.0009
RMSE (m)	0.0029	0.0026	0.0021	0.0017	0.0009
Main beam (primitive best-fitting)					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Std. Dev. (m)	0.0047	0.0041	0.0034	0.0028	0.0014
RMSE (m)	0.0047	0.0041	0.0034	0.0028	0.0014

Table 6.30 Influence of Regularization Factor on dense image matching

PARAMETER: ZPAS					
Half Portal (direct comparison)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Mean Dist. (m)	0.0111	0.0112	0.0113	0.0112	0.0111
Std. Dev. (m)	0.0103	0.0104	0.0105	0.0105	0.0104
Relief (direct comparison)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Mean Dist. (m)	0.0112	0.0111	0.0114	0.0110	0.0111
Std. Dev. (m)	0.0079	0.0080	0.0083	0.0080	0.0081
Door (direct comparison)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Mean Dist. (m)	0.0120	0.0123	0.0129	0.0127	0.0123
Std. Dev. (m)	0.0095	0.0103	0.0110	0.0109	0.0105

Column (primitive best-fitting)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Std. Dev. (m)	0.0041	0.0040	0.0047	0.0040	0.0039
RMSE (m)	0.0041	0.0040	0.0047	0.0040	0.0039
Pillar – Dark pattern (primitive best-fitting)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Std. Dev. (m)	0.0076	0.0080	0.0084	0.0079	0.0074
RMSE (m)	0.0076	0.0080	0.0084	0.0079	0.0074
Pillar – Light pattern (primitive best-fitting)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Std. Dev. (m)	0.0027	0.0026	0.0026	0.0025	0.0025
RMSE (m)	0.0027	0.0026	0.0026	0.0025	0.0025
Main beam (primitive best-fitting)					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Std. Dev. (m)	0.0041	0.0041	0.0041	0.0041	0.0054
RMSE (m)	0.0041	0.0041	0.0041	0.0041	0.0054

Table 6.31 Influence of Z-Quantification Factor on dense image matching

PARAMETER: ZOOMF			
Half Portal (direct comparison)			
ZoomF	1 (Def.)	2	4
Number of Points	7,368,419	1,884,199	465,525
Mean Dist. (m)	0.0113	0.0110	0.0117
Std. Dev. (m)	0.0105	0.0107	0.0114

Relief (direct comparison)			
ZoomF	1 (Def.)	2	4
Mean Dist. (m)	0.0114	0.0112	0.0115
Std. Dev. (m)	0.0083	0.0082	0.0087
Door (direct comparison)			
ZoomF	1 (Def.)	2	4
Mean Dist. (m)	0.0129	0.0114	0.0102
Std. Dev. (m)	0.0110	0.0100	0.0085
Column (primitive best-fitting)			
ZoomF	1 (Def.)	2	4
Std. Dev. (m)	0.0047	0.0040	0.0053
RMSE (m)	0.0047	0.0040	0.0053
Pillar – Dark pattern (primitive best-fitting)			
ZoomF	1 (Def.)	2	4
Std. Dev. (m)	0.0084	0.0053	0.0047
RMSE (m)	0.0084	0.0053	0.0047
Pillar – Light pattern (primitive best-fitting)			
ZoomF	1 (Def.)	2	4
Std. Dev. (m)	0.0026	0.0014	0.0019
RMSE (m)	0.0026	0.0014	0.0019
Main beam (primitive best-fitting)			
ZoomF	1 (Def.)	2	4
Std. Dev. (m)	0.0041	0.0025	0.0032
RMSE (m)	0.0041	0.0025	0.0032

Table 6.32 Influence of Final Z-Resolution Factor on dense image matching

COMPUTATIONAL TIME					
Regul Parameter – Half Portal					
Regul	0.005	0.02 (Def.)	0.05	0.1	0.5
Time (minutes)	5.30	5.15	5.03	4.90	4.93
ZPas Parameter – Half Portal					
ZPas	0.1	0.2	0.4 (Def.)	0.5	0.8
Time (minutes)	19.20	8.17	5.15	4.62	3.90
ZoomF Parameter – Half Portal					
ZoomF	1 (Def.)		2	4	
Time (minutes)	5.15		1.73	0.78	

Table 6.33 Required computational time

Through a detailed analysis of the above listed results, the optimal value for each considered parameter was finally selected. These choices (highlighted in yellow in Table 6.30-33) can be argued with the following observations:

- The Regularization Factor has no significant impact on the computational time. From a metric point of view, giving more weight to the regularization term in the energetic formulation seems to slightly improve the metric performance of the algorithm, especially if the best-fitting tests are analysed. Thus the Regul parameter will be set to 0.5, that seems to represent the optimal choice for this kind of dataset.
- The Z-Quantification Factor is proved to have a very significant influence on the required computational time. As could be expected, if the desired altitude resolution increases (i.e. the ZPas parameter decreases), the necessary computational efforts grow up significantly. Since both types of tests don't show a stable and clearly identifiable trend, in terms of metric uncertainty connected with the parameter variation, the ZPas factor will be set to 0.5, that seems to represent the best compromise choice for this kind of dataset.
- Also the Final Z-Resolution factor seems to play a significant role, if one considers the required computational time and its variation within the tests. As could be expected, if the desired final resolution of the pyramidal image matching approach increases (i.e. the ZoomF parameter increases), the necessary computational efforts grow up significantly and the extracted point cloud is denser (i.e. a greater number of points is matched). From a metric point of view, both types of tests show that the best accuracy may be achieved "stopping" the image matching algorithm at the second to last level, even if the spatial resolution of the final point cloud will be thereby penalized. This

can be explained by referring to the de-quantification process, that is performed at the higher resolution step: this post-processing phase eliminates the quantification artefacts and acts more significantly if it is carried out on a less-dense depth map, delivering a less noisy result. For these reasons, the ZoomF parameter will be set to 2, that seems to represent the optimal choice for this kind of dataset.

6.3.9 Influence of acquisition protocol on dense image matching

This last evaluation step aims at analysing the role played by the acquisition protocol on the dense image matching process, in terms of metric accuracy of its final results. Thus, all available datasets have been here processed, as shown in Table 6.34. As already pointed out in Subsection 6.3.7, the two datasets acquired from the left and right points of view, although oriented and matched, will not be here discussed separately, but only together with the images acquired from a central point of view, defining thereby the dataset termed "central-left-right".

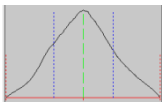
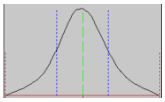
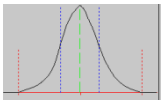
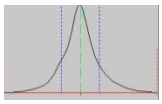
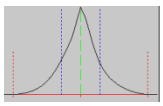
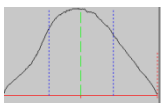
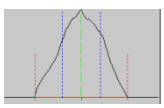
<table>
<tr><th colspan="3">Focal Length: 24 mm</th><th colspan="3">Focal Length: 60 mm</th></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.34 The datasets used in the procedural step "Influence of acquisition protocol on dense image matching"

Each dataset, after the orientation procedure explained in Subsection 6.3.7, was finally processed with the tool Malt, using its Image-Ground geometry and the set of parameter values selected in the previous evaluation step. The central image of each dataset was always declared as the master one. Extracted depth maps were later converted into point clouds using the tool Nuage2Ply. In order to perform a metric assessment of these results and study their behavior as a function of the initial acquisition protocol, the same two evaluation strategies described and performed in Subsection 6.3.6 were here carried out. Moreover, also the scene portions that were selected to be detailed analyzed correspond to the ones highlighted in Figure 6.38. Thus, the observations that were above explained for both approaches, i.e. LS-IBM direct comparison and geometric primitive best-fitting, remain still valid here; only two different procedural choices were adopted:

- The LS point cloud selected as appropriate reference data for the comparison approach was represented by the range map acquired at about 5 m of instrument-object distance. In these tests, in fact, the spatial resolution of the reference model plays a significant role, thus leading to the choice of the most dense point cloud as the reference one.
- The comparisons between the LS and IBM outputs were carried out, in this case, with the software PolyWorks, module IMAlign[10], vs 12.1.18. Furthermore, besides the Standard Deviations (Std. Dev.) of the distances between the two datasets, the corresponding histograms were computed and analyzed too: the geometric shape of the resulting error distribution, in fact, gives an idea of how much it is fitting to the optimal possible configuration, represented by the Gaussian curve. The latter is in fact the ultimate shape of any error distribution, since it corresponds to the ideal case of no gross or systematic errors. Thus a good histogram should be as close as possible to the Gaussian bell curve.

The results are summarized in the following tables (Table 6.35-41); Table 6.42, finally, gives an idea of the computational effort required by the different processes, that were all run in the hardware environment previously detailed in Table 6.21.

HALF PORTAL		
Acquisition Protocol	Std. Dev. (m)	Histograms
24 mm – 3 °	0.0077	
24 mm – 5 °	0.0067	
24 mm – 10 °	0.0062	
24 mm – 3+5+10 °	0.0048	
24 mm – “central-left-right”	0.0057	
60 mm – 3 °	0.0084	
60 mm – 5 °	0.0083	

[10] The alignment is based on an iterative algorithm that computes an optimal alignment by minimizing the 3D distances between surface overlaps in a set of 3D images acquired from unknown viewpoints. If the alignment process is set to perform no iteration, a comparison between 3D point clouds is achieved. This feature is useful when the two 3D point clouds being compared are already registered in the same coordinate system.

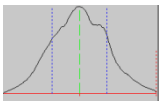
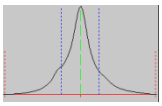
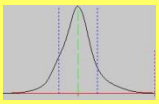
60 mm – 10 °	0.0071	
60 mm – 3+5+10 °	0.0051	
24 mm + 60 mm	0.0037	

Table 6.35 LS-IBM comparisons – Half Portal

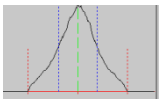
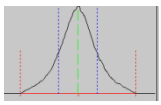
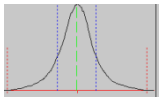
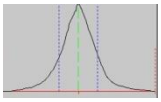
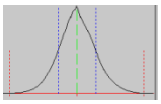
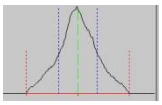
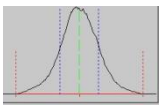
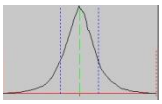
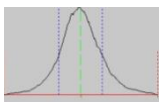
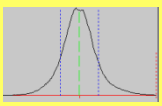
RELIEF		
Acquisition Protocol	Std. Dev. (m)	Histograms
24 mm – 3 °	0.0077	
24 mm – 5 °	0.0067	
24 mm – 10 °	0.0055	
24 mm – 3+5+10 °	0.0049	
24 mm – “central-left-right”	0.0056	
60 mm – 3 °	0.0075	
60 mm – 5 °	0.0060	
60 mm – 10 °	0.0046	
60 mm – 3+5+10 °	0.0056	
24 mm + 60 mm	0.0042	

Table 6.36 LS-IBM comparisons – Relief

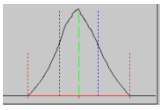
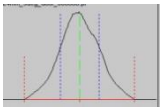
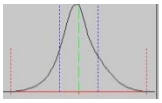
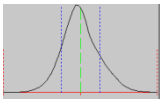
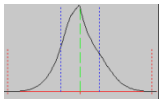
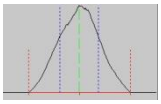
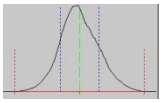
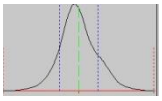
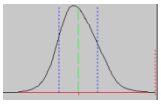
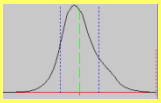
DOOR		
Acquisition Protocol	Std. Dev. (m)	Histograms
24 mm – 3 °	0.0076	
24 mm – 5 °	0.0070	
24 mm – 10 °	0.0056	
24 mm – 3+5+10 °	0.0047	
24 mm – “central-left-right”	0.0053	
60 mm – 3 °	0.0076	
60 mm – 5 °	0.0059	
60 mm – 10 °	0.0051	
60 mm – 3+5+10 °	0.0046	
24 mm + 60 mm	0.0043	

Table 6.37 LS-IBM comparisons – Door

COLUMN		
Acquisition Protocol	Std. Dev. (m)	RMSE (m)
24 mm – 3 °	0.0039	0.0039
24 mm – 5 °	0.0034	0.0034

24 mm – 10 °	0.0040	0.0040
24 mm – 3+5+10 °	0.0074	0.0074
24 mm – “central-left-right”	0.0078	0.0078
60 mm – 3 °	0.0084	0.0084
60 mm – 5 °	0.0092	0.0092
60 mm – 10 °	0.0041	0.0041
60 mm – 3+5+10 °	0.0063	0.0063
24 mm + 60 mm	0.0029	0.0029

Table 6.38 Geometrical primitive best-fitting – Column

MAIN BEAM		
Acquisition Protocol	Std. Dev. (m)	RMSE (m)
24 mm – 3 °	0.0058	0.0058
24 mm – 5 °	0.0036	0.0036
24 mm – 10 °	0.0027	0.0027
24 mm – 3+5+10 °	0.0041	0.0041
24 mm – “central-left-right”	0.0281	0.0281
60 mm – 3 °	0.0033	0.0033

60 mm – 5 °	0.0027	0.0027
60 mm – 10 °	0.0017	0.0017
60 mm – 3+5+10 °	0.0022	0.0022
24 mm + 60 mm	0.0011	0.0011

Table 6.39 Geometrical primitive best-fitting – Main Beam

PILLAR – DARK PATTERN		
Acquisition Protocol	Std. Dev. (m)	RMSE (m)
24 mm – 3 °	0.0068	0.0068
24 mm – 5 °	0.0050	0.0050
24 mm – 10 °	0.0037	0.0037
24 mm – 3+5+10 °	0.0030	0.0030
24 mm – “central-left-right”	0.0099	0.0099
60 mm – 3 °	0.0061	0.0061
60 mm – 5 °	0.0041	0.0041
60 mm – 10 °	0.0032	0.0032
60 mm – 3+5+10 °	0.0016	0.0016
24 mm + 60 mm	0.0012	0.0012

Table 6.40 Geometrical primitive best-fitting – Pillar (Dark Pattern)

PILLAR – LIGHT PATTERN		
Acquisition Protocol	Std. Dev. (m)	RMSE (m)
24 mm – 3 °	0.0025	0.0025
24 mm – 5 °	0.0016	0.0016
24 mm – 10 °	0.0013	0.0013
24 mm – 3+5+10 °	0.0017	0.0017
24 mm – “central-left-right”	0.0045	0.0045
60 mm – 3 °	0.0019	0.0019
60 mm – 5 °	0.0012	0.001235
60 mm – 10 °	0.0010	0.0010
60 mm – 3+5+10 °	0.0008	0.0008
24 mm + 60 mm	0.0008	0.0008

Table 6.41 Geometrical primitive best-fitting – Pillar (Light Pattern)

Acquisition Protocol	**Time** (min.)	Acquisition Protocol	**Time** (min.)
24 mm – 3 °	1.68	60 mm – 3 °	1.73
24 mm – 5 °	1.30	60 mm – 5 °	1.42
24 mm – 10 °	1.43	60 mm – 10 °	1.55
24 mm – 3+5+10°	2.72	60 mm – 3+5+10 °	2.90
24 mm – “central-left-right”	10.29	24 mm + 60 mm	13.29

Table 6.42 Required computational time

The above listed data clearly show that the best results, in terms of metric accuracy, can be achieved starting from the most complete and big dataset, i.e. the one including both 24mm and 60mm-images (results highlighted in yellow). Although this requires, of course, the most onerous computational effort, the image matching procedure is thereby able to deliver a more accurate (millimetre-level) 3D reconstruction of the scene, showing an impressive metric potentiality of the implemented multi-view stereo reconstruction algorithm.

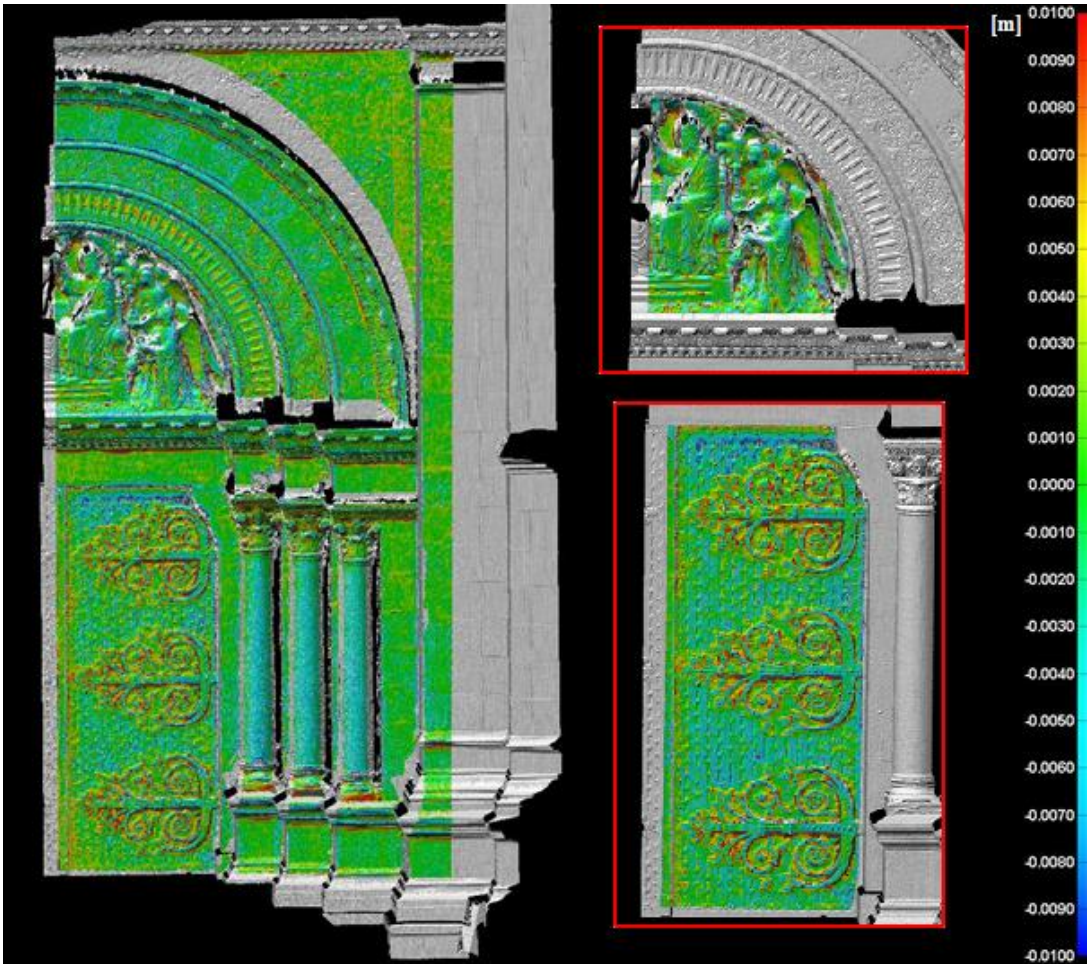


Figure 6.39 Comparison between the IBM point cloud (selected optimal dataset) and the LS point cloud. The colour scale ranges from -0.01 m (blue) to 0.01 m (red)

Figure 6.39 shows the colour-coded results of the LS-IBM comparison, performed with the selected optimal acquisition protocol. Besides the result delivered by the comparison carried out with the half portal datasets, the figure includes also the colour-coded maps concerning the relief and the door. The main concerns encountered with the IBM approach and associated with its greatest deviations from the reference data are listed below and showed in Figure 6.40.

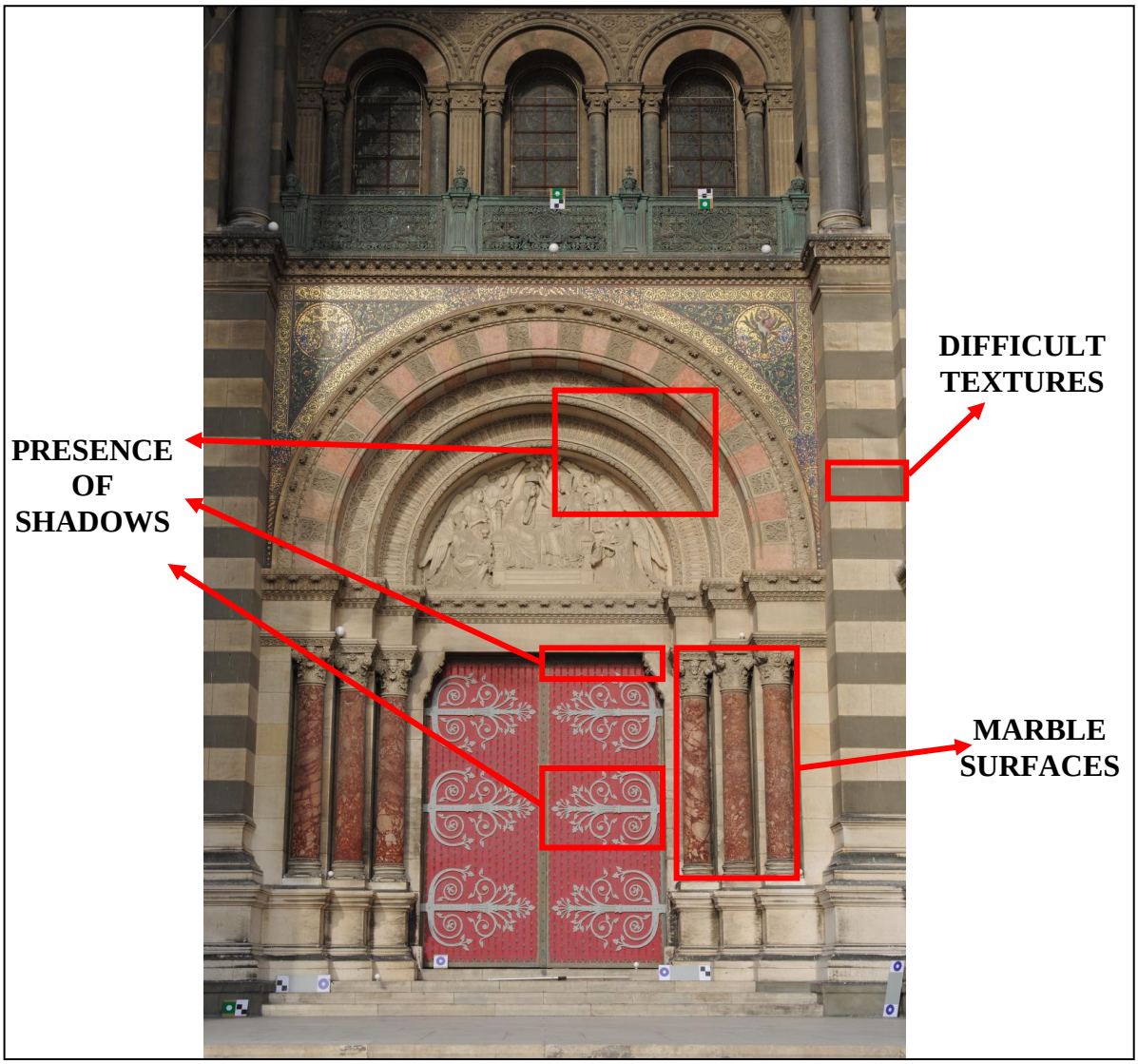


Figure 6.40 Localization of the main deviations between the compared point clouds (LS-IBM)

- Presence of shadows in the digital images, as, for example, within the portions located under the arches and the main beam. These regions represents "difficult" patterns to be matched, showing that it is always advisable to acquire images without any shadows, thus with overcast sky if the scene is outside.
- Presence of difficult textures, as, for example, the dark pattern of the pillar. If one compare the results achieved by both evaluation approaches with the two different pillar patterns, i.e. the dark and the light ones, it is possible to observe a significant difference in terms of metric accuracy. The errors associated with the dark pattern are, in fact, always about twice the ones delivered by the light pattern. This suggests that a preliminary study of the textures characterizing the scene surfaces is necessary in order to achieve accurate results with the IBM approach.
- Presence of marble surface, as, for example, the columns. In these regions the delivered LS-IBM differences are negative, showing that the LS surfaces are below

the IBM ones. This problem is connected with LS performances, when they have to deal with surfaces that depart from the following underlying hypothesis of active optical measurements (Beraldin, 2004): the acquired surface should always be opaque and diffusely reflecting. Marble, instead, exhibits two different optical properties: translucency and non-homogeneity at the scale of the measurement process. This structure causes a bias in the distance measurement and an increase in noise level, that represent two key concerns affecting the geometric measurement.

6.3.10 Influence of image resolution on the entire pipeline

The analysis described in Subsection 6.3.5 shows that the use of a reduced image resolution for the tie point extraction phase saves valuable computational time during this step and the subsequent orientation one, without significantly compromising their accuracy. In particular, some kind of an image width threshold, equal to about the 60% of the original image resolution, was proved to be effective for this kind of datasets, defining a limit over which a significant improvement of the orientation accuracy cannot be achieved.

<table>
<tr><td colspan="3">Focal Length: 24 mm</td><td colspan="3">Focal Length: 60 mm</td></tr>
<tr><td>Point of View</td><td>α</td><td>Number of Images</td><td>Point of View</td><td>α</td><td>Number of Images</td></tr>
<tr><td rowspan="3">Central</td><td>3 °</td><td>5</td><td rowspan="3">Central</td><td>3 °</td><td>5</td></tr>
<tr><td>5 °</td><td>3</td><td>5 °</td><td>3</td></tr>
<tr><td>10 °</td><td>3</td><td>10 °</td><td>3</td></tr>
<tr><td colspan="3">All 9 central images</td><td colspan="3">All 9 central images</td></tr>
<tr><td>Left</td><td>3 °</td><td>5</td><td colspan="3">Focal Length: 24 + 60 mm</td></tr>
<tr><td>Right</td><td>3 °</td><td>5</td><td colspan="3" rowspan="2">All 29 images acquired with the two lenses</td></tr>
<tr><td colspan="3">All 20 images (central-left-right)</td></tr>
</table>

Table 6.43 The dataset used in the procedural step “Influence of image resolution on the entire pipeline”

Starting from this analysis and from the results delivered by all the previous evaluation steps, a final study has been carried out in order to assess which kind of influence the selected image resolution proves to have on the entire IBM pipeline, in terms of both metric accuracy and required computational time. Note that the term “image resolution” refers here to the image width selected during the tie point extraction phase: all subsequent steps are, then, performed using the initial images. Furthermore, since it was found to be the optimal choice, the dataset including both 24mm and 60mm-images was employing for this final analysis (Table 6.43). The entire photogrammetric and computer vision-based pipeline was thus performed again, adopting the approaches and parameterization selected as optimal choices during all the

previous assessment steps. Table 6.44 summarizes these implemented strategies, underlying in particular the desired initial image resolution.

TIE POINT EXTRACTION	ORIENTATION	DENSE IMAGE MATCHING
Tapioca	Tapas, GCPBascule, Campari	Malt
• All possible pairs of images; • **Image width: 3500 pixel (58% of the original image width).**	• Calibration Input : parameters calculated with the "corner" datasets; • Relative Orientation of 24mm- images; • Orientation of 60mm-images on the 24mm-"canvas" (24 mm-poses are frozen); • Absolute orientation with GCPs (Pose Calculation and Bundle Compensation).	• Image-Ground Geometry; • Regul = 0.5; • ZPas = 0.5; • ZoomF = 2.

Table 6.44 The adopted strategies in the IBM pipeline

Each procedural step was analysed using the same strategies explained in Subsections 6.3.5, 6.3.7 and 6.3.9. Results are listed in the following tables (Table 6.45-49), where the performance achieved by the "reduced image resolution" approach is compared with the one delivered by the previously described "original image resolution" strategy. Computational times required by both processes are reported and compared too. Table 6.50 completes this analysis giving an overview of how long the entire IBM pipeline lasted in the two different cases. The results achieved by the "reduced image resolution" strategy are always highlighted in yellow.

TIE POINT EXTRACTION		
Evaluated Parameters	Reduced Image Resolution	Original Image Resolution
Mean number of Tie Points	56462	98712

Table 6.45 Performances achieved in the phase of "Tie point extraction" (Tool: Tapioca)

RELATIVE ORIENTATION		
Evaluated Parameters	Reduced Image Resolution	Original Image Resolution
RMSE (pixel)	0.53	0.46

Table 6.46 Performances achieved in the phase of "Relative orientation" (Tool: Tapas)

ABSOLUTE ORIENTATION		
Evaluated Parameters	Reduced Image Resolution	Original Image Resolution
Std. Dev. X (m)	0.004	0.003
Std. Dev. Y (m)	0.003	0.002
Std. Dev. Z (m)	0.002	0.003

Table 6.47 Performances achieved in the phase of "Absolute orientation" (Tools: GCPBascule and Campari)

DENSE IMAGE MATCHING – PRIMITIVE BEST-FITTING		
Selected Regions	Standard Deviation (m)	
	Reduced Image Resolution	Original Image Resolution
Column	0.0029	0.0029
Main Beam	0.0011	0.0011
Pillar – Dark Pattern	0.0013	0.0012
Pillar – Light Pattern	0.0009	0.0008

Table 6.48 Performances achieved in the phase of "Dense image matching" (Tool: Malt) - primitive best-fitting approach

DENSE IMAGE MATCHING – LS-IBM COMPARISON		
Selected Regions	Standard Deviation (m)	
	Reduced Image Resolution	Original Image Resolution
Half Portal	0.0037	0.0037

Relief	0.0042	0.0042
Door	0.0042	0.0043

Table 6.49 Performances achieved in the phase of "Dense image matching" (Tool: Malt) – LS-IBM comparison approach

COMPUTATIONAL TIME (minutes)		
Process	Reduced Image Resolution	Original Image Resolution
Tie Point Extraction	110.10	228.65
Relative Orientation	49.89	80.20
Absolute Orientation	7.52	11.67
Dense Image Matching	11.00	13.29
Complete Procedure	178.51 (≈3h)	333.81 (≈5.30h)

Table 6.50 Computational time required by the processes

The use of a reduced image resolution during the first procedural step, i.e. the tie point extraction phase, reduces significantly the number of detected homologous points. This reduction is linear, as previously shown in Subsection 6.3.5: a 58% of the original image width leads to a number of tie points equal to the 57% of the same parameter computed through the "original image resolution" approach. The relative orientation RMSE shows a slight increase, that, however, is not metrically significant. Furthermore, both absolute orientation and dense image matching phases are proved to achieve the same metric accuracy with the two different approaches, showing once again the impressive metric potentiality of the implemented algorithms. Moreover, the choice of a reduced image resolution during the first procedural step leads to a considerable advantage in terms of time saving: the computational time is almost half the one required by operating at full image resolution. For all these reasons, thus, the selection of a partially reduced image width (≈ 60% of its original value) during the Tapioca process, seems to represent an optimal compromise choice between the metric accuracy of the IBM procedure and its time requirements.

6.4 ISO1 Laboratory (Ottawa, Canada)

The NRC MSS Metrological Laboratory is an environmentally (ISO 1:2002) controlled facility for research and development activities in non-contact 3D imaging metrology. In particular, it is dedicated to research in the areas of calibration, certification and evaluation of

3D imaging systems. The laboratory was built at NRC Canada in 2006 and has a usable surface of about 80 square meters. The adjective "controlled" refers to the continuous monitoring of both air temperature and relative humidity, that are kept almost constant throughout the year (except when power failure occurs). In particular, the temperature of a laminar air flow is always maintained at 20° (+/- 0.1° accuracy), whereas its relative humidity is kept fixed at 45% (+/- 5% accuracy). Moreover, temperature and humidity are monitored by independent instrumentation, while multiple backup systems ensure a continuous tracking of laboratory conditions. Furthermore, the total volume of air in the laboratory is changed twice a minute and adequately filtered, so that the air cleanliness is exceptionally good, belonging to Class 1000 (ISO1). Figure 6.41 gives a general overview of the metrological laboratory and some of the its facilities.

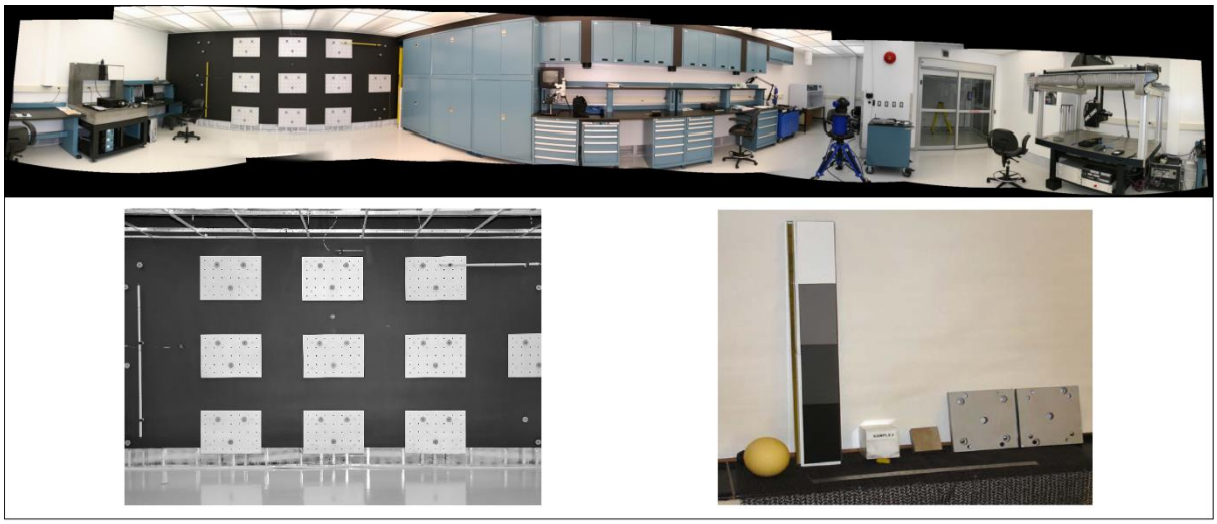


Figure 6.41 The controlled (ISO1) metrological laboratory: a panoramic view of the facility (up), a view of the back wall (bottom, left) and some available 3D artefacts (bottom, right)
(Beraldin et al., 2007; Beraldin, 2009)

The laboratory's equipment consists mostly of 3D active optical imaging systems, commercial or NRC-developed; these instruments are based on various measurement principles (laser triangulation, time-of-flight, structure light projection, holographic and confocal microscopy), thus covering a wide range of measurement spatial scales, from few micrometres up to 100 m. Moreover, ancillary measuring systems complete the capabilities of the laboratory, through the following instruments: a laser tracker, a touch-probing arm, an optical interferometer, a multi-spectral video camera, spectrum-photometers, wavelength meters and sonometers. Finally, reference artefacts and targets (prismatic objects, rogue objects, 2D targets, etc.) are usually employed for calibration activities, in order to perform length and reflectivity measurements. The properties and performances of the laboratory are detailed described in (Beraldin et al., 2007), where also a partial list of the available equipment is provided.

The tests performed within this metrological laboratory aim at assessing the metric potentiality and associated measurement uncertainty of the orientation and dense image

matching algorithms implemented into the IGN suite of tools; in particular, this case study differs from the ones previously described, since it was specifically designed in order to provide the following opportunities:

1. The possibility of carrying out all measurements in a controlled laboratory environment. All data acquisition processes were, in fact, performed under the same environmental conditions, allowing the comparisons between evaluated and reference measurements to be realized within the same metrological environment background. Furthermore, both image acquisition and reference data survey were performed according to their specific optimal environmental requirements, in terms of illumination, temperature, humidity and absence of vibrations (such as the ones caused, for example, by the wind). These choices will be better detailed in the following subsections.
2. The possibility of carrying out all measurements on verifiable, *ad-hoc* designed 3D artefact, that is not changing over time and is not subject to weather ravages. In particular, the 3D object chosen as case study (Figure 6.42) was specifically designed in order to fully satisfy the following key features:

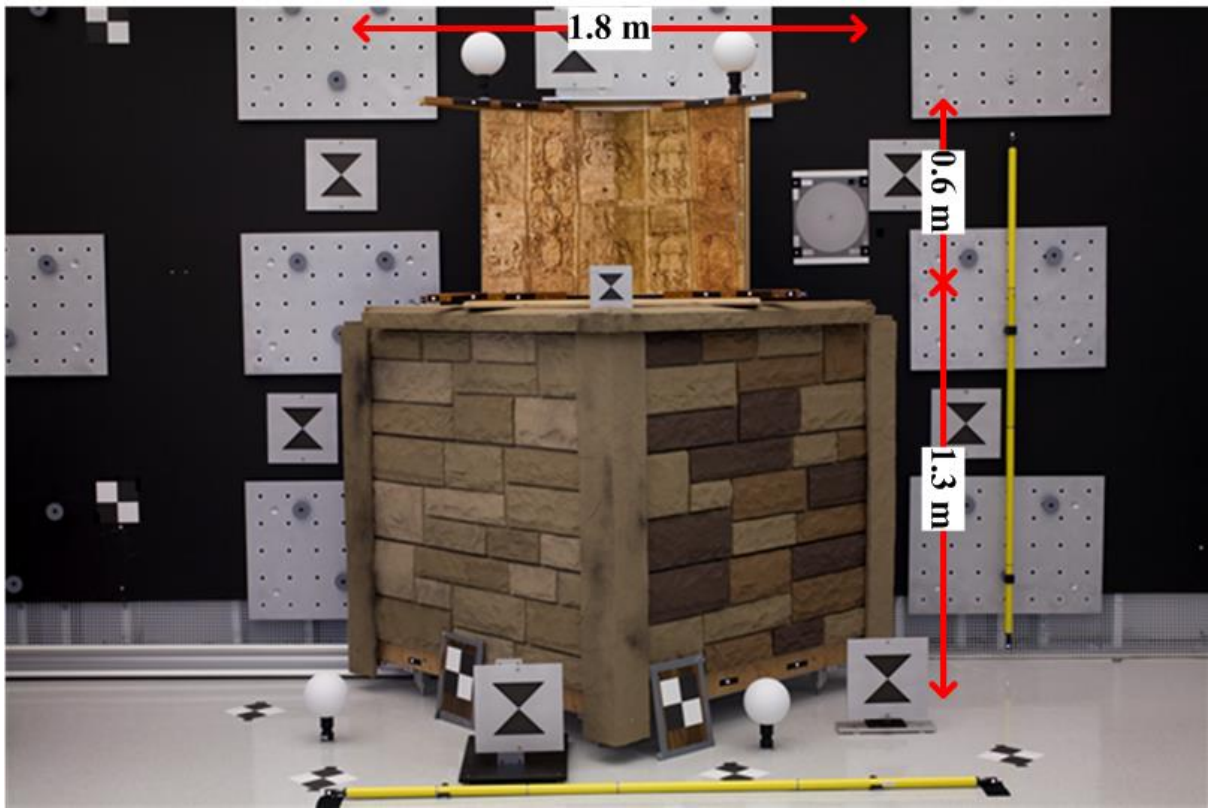


Figure 6.42 The 3D test-object and its main dimensions

- Significant depth variations, with the presence of different depth levels and convergence angles of intersecting surfaces;
- Presence of surfaces characterized by different textures and colours, with the interesting alternation of darker and lighter patterns in the vertical walls of the lower corner and the succession of different decorative motifs in the upper one.

- Presence of surfaces characterized by different materials and, consequently, roughness properties, such as paint-coated plywood, wallpaper and metallic surfaces;
- Presence of primitive geometric shapes, such as planes and spheres;
- Presence of tricky surfaces to be matched, like, for example, the one corresponding to the small upper corner. This element, in fact, was built by gluing a sheet of paper on a rigid support made of plywood. Few images of reliefs are depicted on this paper, offering the possibility of evaluating the algorithm performance in dealing with a "false" reliefs and optical tricks. Furthermore, some small creases are present on the glued surface, making it not perfectly planar: the capability of the evaluated algorithm to image and reconstruct small structural details and defects can be thereby assessed.
- High availability of open space in front of the scene, allowing the testing of different image acquisition protocols. Furthermore, during the acquisition phase, no people were in the laboratory, with the exception of three trained operators. This situation, necessary to maintain the environmental conditions stable during the entire acquisition phase, represented an advantage also in terms of respecting the eye-safety requirements, that have a paramount importance when working with coherent light sources (laser scanner survey). No difficult operations of interest area-delimitation were thereby necessary.

3. The possibility of having many different kinds of references within the acquired 3D scene, such as targets of different sizes, shapes (rectangular, circular, etc…) and typologies (2D target, 3D target, etc…), together with graduated bars and spheres. All these elements were firmly "fixed" in the scene in order to avoid any perturbation of the measured configuration. Furthermore, given the well-accessibility of the acquired 3D scene, an optimal spatial configuration of the reference targets could be achieved.
4. The possibility of acquiring the reference measurements with a huge amount of different optical 3D imaging instruments, that cover a wide range of measurement principles and spatial scales. Many different surveying methods were thus carried out and only some of them will be further described: the others will be processed in the future, in order to perform further and more detailed analyses, starting from the so far obtained results.

6.4.1 Procedural workflow

The entire photogrammetric and computer-vision based pipeline has been performed with the IGN's suite of tools, starting from the detection of homologous points between the images and ending up with the dense 3D reconstruction of the scene of interest.
Figure 6.43 shows a synthetic overview of the procedural workflow followed within these experimental tests.

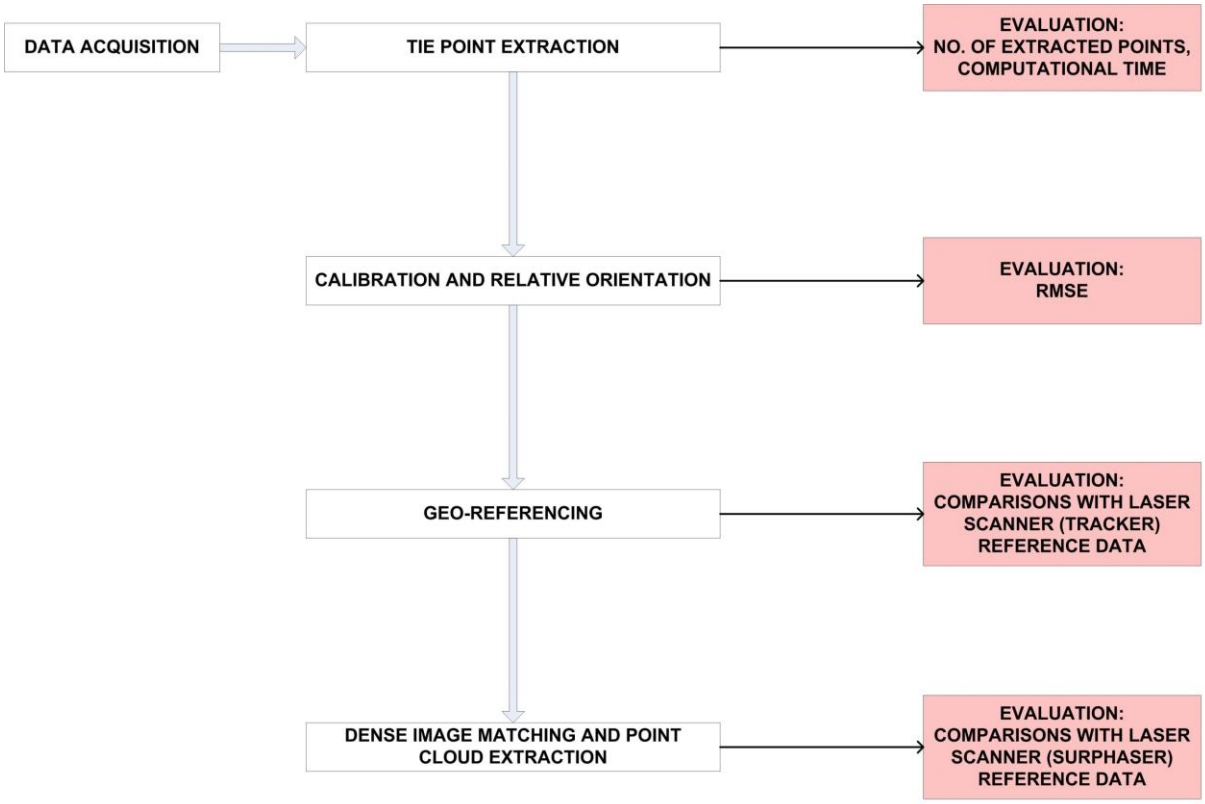


Figure 6.43 Procedural workflow (Laboratory ISO1)

Each procedural step was metrically evaluated, using two different strategies:

- Analyses of internally-computed parameters, such as the mean number of detected homologous points ("Tie point extraction" phase) and the orientation Root Mean Square Error ("Calibration and relative orientation" phase). These tests will be described in Subsections 6.4.4-5.
- Comparisons between results achieved with the Image-Based Modelling (IBM) approach and adequate reference data. The latters were represented by the 3D coordinates of known Check Points (CPs), that were measured with the Laser Tracker ("Geo-referencing" phase), and by dense point clouds, that were acquired with the Surphaser Laser Scanner ("Dense image matching and point cloud extraction" phase). These tests will be described in Subsection 6.4.5-6.

Both evaluation approaches were completed by reporting the computational time required to run each process. In order to compare these measures of computational effort, all steps were carried out in the same processing environment, whose main hardware information is listed in Table 6.51.

As in the previous experimental case study, the effects of different image acquisition protocols have been studied too, examining their influence on the orientation and dense matching phases. In this application, however, only the angles of convergent images were varied among the different protocols, whereas the focal setting was maintain the same and set to its optimal configuration: the only one lens employed in the tests, in fact, was a fix 50mm-

focal length lens, that represents, on a full frame 36x24 mm sensor, the better choice in terms of lens distortions and aberrations (Subsection 6.3.5). The acquisition of both datasets, i.e. digital images and LS measurements, will be deepen discussed in Subsections 6.4.2-3. For the algorithmic and operative aspects related to the IGN's suite of tools, the reader is referred to Chapter 3.

HARDWARE INFORMATION	
NUMBER OF PROCESSORS	16
PROCESSOR ARCHITECTURE	IA64
PROCESSOR DESCRIPTION	Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz
FILE SYSTEM	EXT4
MAIN MEMORY	
TOTAL PHYSICAL	128 GB
FIXED DISK DRIVES	
TOTAL SPACE	2 TB

Table 6.51 Hardware information on the employed processing environment

6.4.2 Image acquisition

Figure 6.44 Canon EOS 5D and the glued lens

Figure 6.45 Flash Speedlite 430EX

The digital image acquisition phase was performed using a Canon EOS 5D digital camera (4368 x 2912 pixels), equipped with a fixed focal length lens (Canon EF 50mm f2.5 Compact Macro Lens). No automatic optical image stabilization is present. The related technical specifications are listed in Table 6.52.

Canon EOS 5D	
BODY TYPE	Mid-Size SRL
SENSOR RESOLUTION	12.8 Mpixel
SENSOR SIZE	Full Frame
SENSOR TYPE	CMOS
ISO	100-1600 in 1/3 stops
MIN SHUTTER SPEED	30 sec
MAX SHUTTER SPEED	1/8000 sec
Canon EF 50mm f2.5 Compact Macro Lens	
FOCAL LENGHT	50 mm
MAX APERTURE	f2.5
MIN APERTURE	f32.0
MIN FOCUS	0.23 m

Table 6.52 Technical specifications of Canon EOS 5D and lens employed

The scene was acquired at a focusing distance of 4.75 m, gluing the lens (Figure 6.44) so that its focus setting would be maintained during the movements between each shoot position and the subsequent one. Moreover, this setup was necessary in order to be able in the future to employ again this lens at the same focus setting, for possible further tests or calibration purposes. Also the f-number and ISO sensibility were both kept fixed and set to, correspondingly, f8 and 100. Furthermore, as in the previous case study, different angles of convergent images were tested: in particular, images were acquired at 5° and 10° of convergence, following the suggested crosswise convergent configuration. For each point of view, the same central image will be used as the "master" one for both 5° and 10°-datasets. Table 6.53 summarizes all the different acquisition protocols, whereas Figure 6.46 shows the one that was performed with a convergence angle of 10°: this image represents a sparse 3D reconstruction of the scene with the corresponding camera relative poses, computed with the tool AperiCloud. Finally, Table 6.54 provides the resulting values of range accuracy and lateral accuracy, that were computed using Equations [6.1] and [6.2].

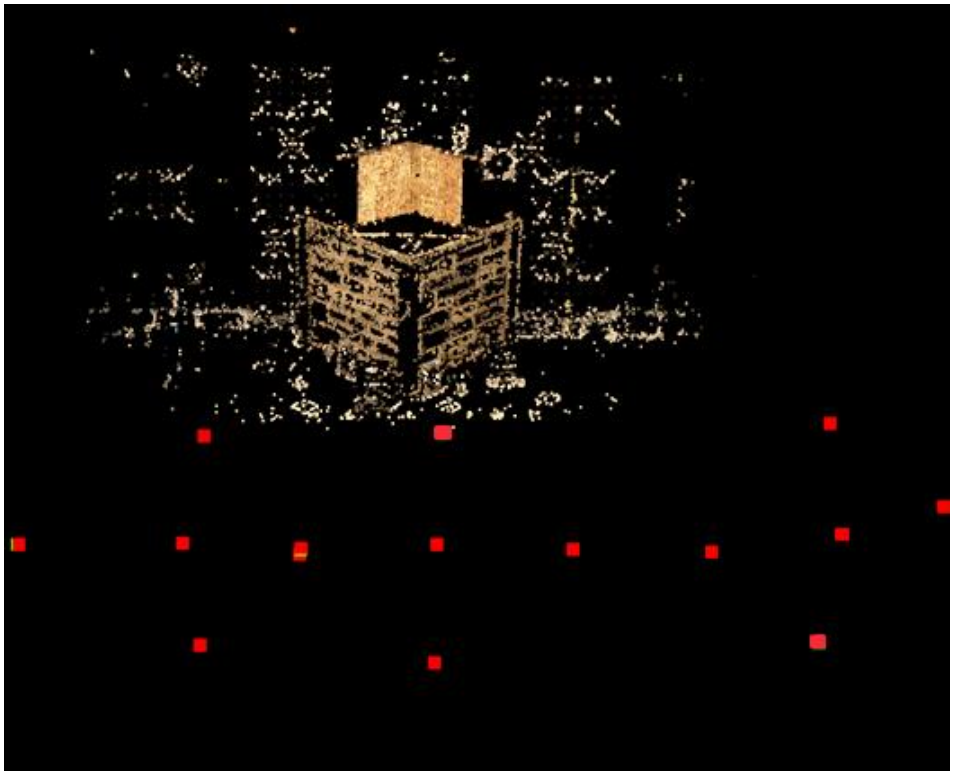


Figure 6.46 Image acquisition layout (α = 10°)

Focal Length: 50 mm **Camera-Object Distance: 4.75 m**					
Point of View	Angle of Convergent Images (α)	Number of Images	Point of View	Angle of Convergent Images (α)	Number of Images
Central	5°	5	Central	10°	5
Left	5°	5	Left	10°	5
Right	5°	5	Right	10°	5

Table 6.53 Summary of the different acquisition protocols

Range Accuracy (mm)		Focal Length
		50 mm
Angle of Convergent Images	5°	2.2
	10°	1.1
Lateral Accuracy (mm)		Focal Length
		50 mm
Angle of Convergent Images	5°	0.8
	10°	

Table 6.54 Range and lateral accuracies

Figure 6.47 A view of the digital image acquisition phase

Finally, images were acquired (Figure 6.47) using a photographic tripod, in order to reduce as much as possible any vibration effect; both ladders and rigid supports were employed to reach the necessary heights. A diffused and controlled ambient light (fluorescent) was provided to illuminate the scene, so that no shadows were present within it. Of course, illumination didn't change during the entire image acquisition phase. A flash Speedlite 430EX (Figure 6.45) was used in order to illuminate the retro-reflective targets: since it was set at 1/16 of its normal power, it didn't cause any significant illumination change.

6.4.3 Laser scanner survey

The reference data were acquired with two types of Laser Scanner. The Faro Laser Tracker Model X (Figure 6.48) from FARO Technologies Inc. was used to measure the 3D coordinates of 8 targets, that will act as GCPs and CPs in the following tests.

Figure 6.48 Faro Laser Tracker Model X
(Beraldin et al., 2009)

Faro Laser Tracker Model X	
SCAN PRINCIPLE	Time-of-Flight (AM) + Interferometer (IM)
WORKING RANGE	0 – 70 m
DISTANCE MEASUREMENT ACCURACY	10 μm + 0.4 μm/m
ANGLE MEASUREMENT ACCURACY	18 μm + 3 μm/m
LASER CLASS	Laser Class 1
HEAD WEIGHT	20 kg

Table 6.55 Technical specifications of Faro Laser Tracker (LT) Model X

This instrument acquires an absolute distance and two direction angles, using a spherically mounted retro-reflector, that must touch the surface of interest. The equipped distance measurement device, termed Absolute Distance Measurement (ADM), is based on a proprietary time-of-flight technology. The resulting 3D coordinate measurement system is able to measure up to 35 m (diameter 70 m) and was metrically certified by the manufacturer

using the procedure reported in the ASME B89.4.19 standard (ASME B89.4.19). For the employed working distance (4.75 m), the radial expanded (k=2) uncertainty is approximately U_R(LT)=23.8 μm and the transverse expanded (k=2) uncertainty is approximately U_T(LT)=64.5 μm. One can thus see that the LT will not be a limiting factor in the present accuracy assessments.The main technical specifications are listed in Table 6.55. Figure 6.49 shows a view of the acquisition phase, that was performed by an NRC metrology expert.

Figure 6.49 A view of the Laser Tracker survey

The second scanner used in the tests was another commercially-available system: Surphaser Model HS25X from Basis Software (Figure 6.50).

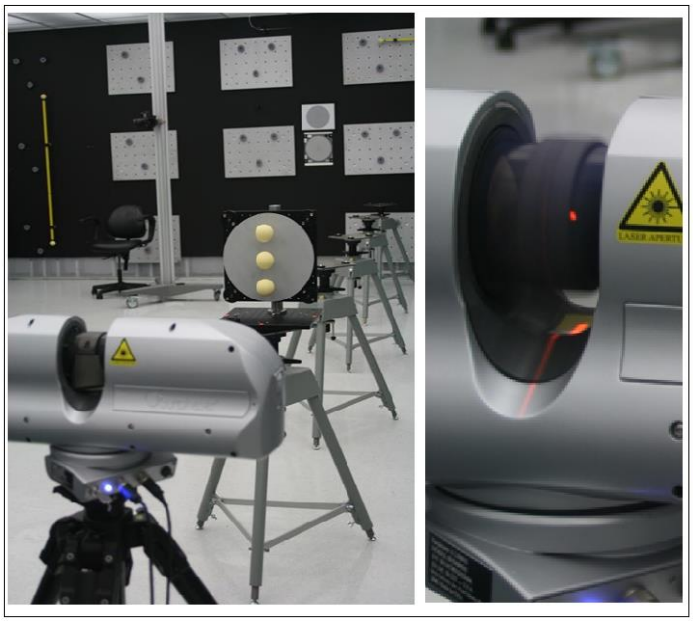


Figure 6.50 Surphaser Model HS25X

Surphaser Model HS25X	
SCAN PRINCIPLE	Time-of-Flight – Phase Measurement
SCAN RANGE	Up to 70 m
MEASUREMENT SPEED	Up to 1.2 million points/second
RANGE NOISE (1σ)	0.1 mm at 3 m
LASER CLASS	Laser Class 1

Table 6.56 Technical specifications of Laser Scanner Surphaser Model HS25X

This instrument was employed to acquire three dense point clouds of the scene of interest, including some 2D targets distributed within it. Thus, its measurements will be set as reference ones in order to evaluate both absolute orientation (together with CPs measured by Laser Tracker) and dense image matching. The scanner can be described as a hemispherical time-of-flight phase-shift laser scanner: as discussed in Chapter 4, this means that distance measurements are related to the difference in phase between the laser light reflected from the surface and the reference signal. Furthermore, multiple frequencies are used to achieve high accuracy and to reduce interval ambiguities. The instrument provides a field-of-view of 360°x270°: while the horizontal angular direction can be limited according to the specific acquisition requirements, the vertical angle cannot be modified, since the rotating mirror operates only in continuous mode. Many metrological tests have been carried out at NRC, in order to determine the instrument performances when dealing with different operative conditions. Some results can be found in (Beraldin et al., 2009; Beraldin, 2009; Beraldin et al., 2011); in the present application, a local measurement uncertainty of u_R(LS)=0.3 mm was assumed: this value is a realistic representation of the noise level present in the 3D point clouds. The main technical specifications of the laser scanner are summarized in Table 6.56.

6.4.4 Tie point extraction

The SIFT^{++} implementation of SIFT algorithm employed by the tool Tapioca (Subsection 3.3.2) was used in order to detect and match homologous points between the images. In this step, as in all the subsequent ones, both 5° and 10° datasets will be processed, orienting the three different points of view of each dataset within the same reference system. Moreover, the dataset including both 5° and 10° images was processed too. Table 6.57 summarizes and details these choices, that will be maintained during the entire evaluation procedure.

The tie point search mode was set to “All” in order to consider all possible pairs of images; furthermore, this research was carried out without any previous image shrinking, i.e. the original image resolution (4368 x 2912 pixel) was always employed. In order to evaluate the results, the mean number of homologous points matched in each image was analysed,

together with the computational time required by each process. The outcome is reported in Table 6.58.

Focal Length: 50 mm Camera-Object Distance: 4.75 m					
Point of View	α	Number of Images	Point of View	α	Number of Images
Central	5°	5	Central	10°	5
Left	5°	5	Left	10°	5
Right	5°	5	Right	10°	5
Central-Left-Right: 15 images			Central-Left-Right: 15 images		
All 27 images (5° dataset + 10° dataset)					

Table 6.57 The image datasets used in this step and in all the subsequent ones

TIE POINT EXTRACTION			
Dataset	5°	10°	5° + 10°
No. of Images	15	15	27
No. of Tie Points	13542	11558	14565
Time (minutes)	2.22	1.92	4.77

Table 6.58 Performances achieved in the phase of "Tie point extraction" (Tool: Tapioca)

No significant difference can be detected between the 5° dataset and the one acquired with 10° of convergence angle. The combined use of both datasets leads to a slight increase in homologous point number, since more images (i.e. more different points of view) are processed together.

6.4.5 Calibration and relative orientation

Starting from the detected tie points, the calibration parameters and relative poses of the camera were then computed with the tool Tapas (Subsection 3.3.3). Results achieved in the previous case study (Subsection 6.3.6) showed that a self-calibration approach during the bundle adjustment step represents a good strategy, if the employed image dataset is

favourable to the calibration recovery. Since the test-object chosen for the present application exhibits significant depth variations and surface textures and it was imaged according to the main acquisition requirements (different heights, angles of view, etc…), no pre-calibration inputs were provided to the process. Thus, calibration was performed starting from the usual initialization values (EXIF-derived focal length, no principal point offsets, no distortions) and then refined using the homologous point information in the bundle adjustment phase. Camera relative poses were computed and compensated within the same procedure; the Root Mean Square Error (RMSE) of all re-projection residuals was finally computed for each process. The outcome is shown in Table 6.59, together with the computational time required by the processes.

CALIBRATION AND RELATIVE ORIENTATION			
Dataset	5°	10°	5° + 10°
No. of Images	15	15	27
RMSE (pixel)	0.32	0.31	0.33
Time (minutes)	3.75	3.18	10.88

Table 6.59 Performances achieved in the phase of "Calibration and relative orientation" (Tool: Tapas)

As in the previous procedural step, no significant difference in terms of process performances can be pointed out: all three datasets achieve an optimal accuracy level, if one consider the mean "Tie Residual" (mean value of all re-projection errors, highlighted in yellow) as a good indicator of the proper system convergence.

6.4.6 Geo-referencing

In order to achieve a geo-reference of the photogrammetric results and perform a "*a-posteriori* evaluation" of the orientation accuracy using some external known reference data, relative orientations were then converted into absolute ones employing the measurements acquired with the two Laser Scanners. First of all, a common reference frame should be adopted and the Laser Tracker data were employed to define it. Thus, 4 well distributed targets measured with the Laser Tracker were chosen as Ground Control Points and set as reference for the registration of the three point clouds acquired with the second Laser Scanner. This process was performed within the software PolyWorks IMAlign™, where the 3D coordinates of the Surphaser-measured 2D targets, once registered in the Laser Tracker reference frame, were picked up too. At the end of this procedure, hence, the complete GCP and CP configuration shown in Figure 6.51 was achieved and used within the IBM pipeline. Besides the previously mentioned 4 GCPs, a total of 10 points were employed as CPs, 4 of which measured by the Laser Tracker and the remaining ones acquired with the Surphaser

Laser Scanner. The resulting layout of the measured targets provides an optimal network of known 3D points, since they are well distributed in all directions within the entire volume of interest.

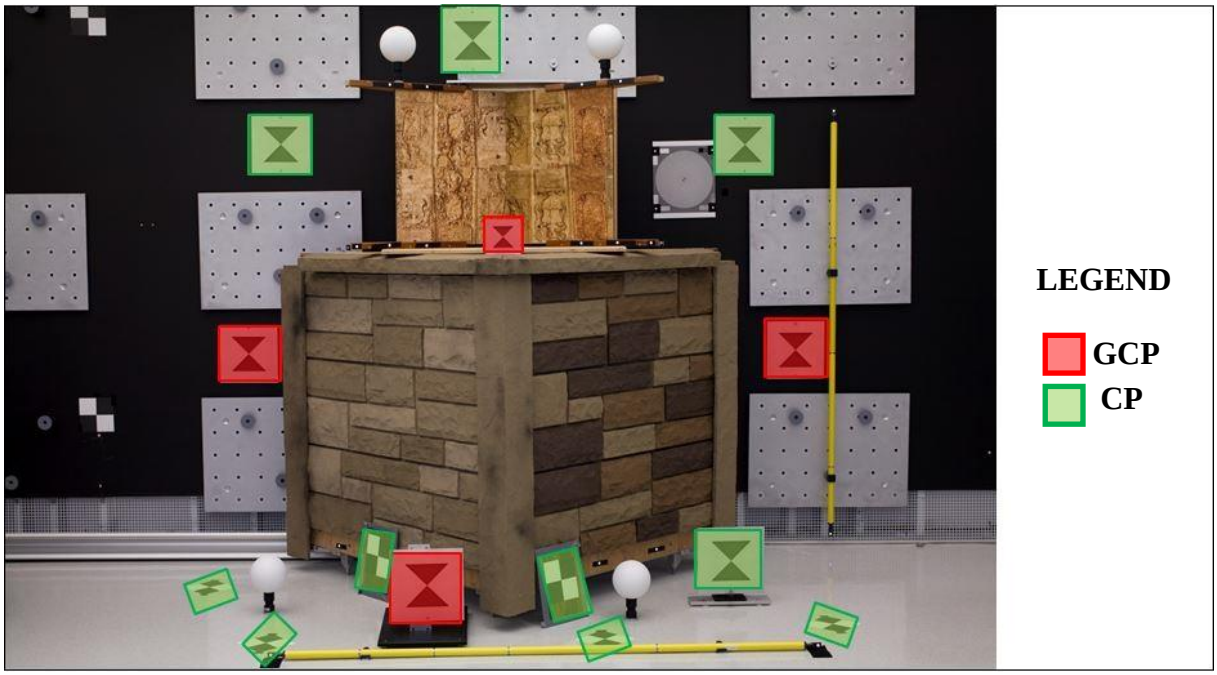


Figure 6.51 Configuration of the 4 GCPs and 10 CPs

Three images were chosen within each dataset and the 4 measured GCPs were manually collimated on each of them, using the tool SaisieAppuisInit; starting from the collimated 2D coordinates (image reference frame) and the corresponding 3D coordinates (absolute reference frame), the global transformation from a purely relative orientation to the one "registered" in the Laser Tracker reference frame was then carried out with the tool GCPBascule (Subsection 3.3.3). Finally, the tool Campari (Subsection 3.3.3) was run in order to perform a compensation (bundle adjustment) of all the provided heterogeneous observations, i.e. tie points and GCPs.

Once the datum ambiguity was solved, i.e. the geo-referencing of IBM results was achieved, the remaining 10 targets were assumed as independent check points and matched in the same three images, that were previously employed. Their 3D coordinates were then computed in the photogrammetric pipeline as intersections of homologous rays, with the tool SaisieAppuisInit. These measurements were later compared to the ones acquired with the laser scanners and corresponding residuals were finally calculated for each of the three absolute coordinates (X,Y,Z). The resulting Standard Deviations (Std. Dev.) are listed in Table 6.60, together with the computational time required by the processes carried out with the tools GCPBascule and Campari.

The results show that the orientation algorithm is able to achieve a considerable accuracy level, when the acquisition is performed under the over mentioned controlled environmental conditions: all standard deviations (highlighted in yellow) are, in fact, below 1 mm, with the only exception of one value, related to the third dataset, that shows, anyway, a millimetre

order of magnitude. Furthermore, by analysing the individual residuals (i.e. the differences between the photogrammetric-computed coordinates and the corresponding reference ones), it is possible to note that the higher deviations correspond to the 2D targets on the floor, whose position requires an unfavourable acquisition direction for all the employed instruments.

GEO-REFERENCING			
Dataset	5°	10°	5° + 10°
No. of Images	15	15	27
Std. Dev. X (mm)	0.42	0.49	0.47
Std. Dev. Y (mm)	0.60	0.80	1.34
Std. Dev. Z (mm)	0.36	0.32	0.41
Time (minutes)	0.74	0.60	2.01

Table 6.60 Performances achieved in the phase of "Geo-referencing" (Tools: GCPBascule and Campari)

6.4.7 Dense image matching and point cloud extraction

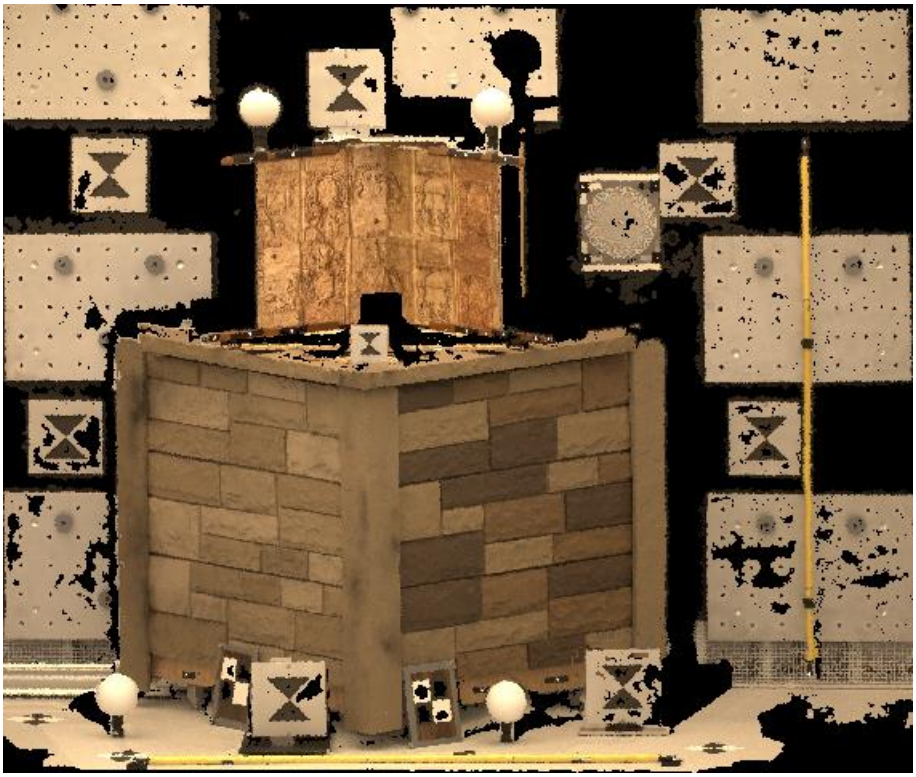


Figure 6.52 The point cloud extracted from the 5°-dataset

The tool Malt (Subsection 3.3.4) was then employed to extract a depth map for each point of view of each dataset. The computations were all run in image-ground geometry, after having selected each central image as the master one. Starting from the results achieved in the previous case study, the same parameter set, selected as the optimal one, was adopted also in this case. Depth maps were finally converted into point clouds with the tool Nuage2Ply (Subsection 3.3.4): the algorithm is also able to deliver photo-textured results, by assigning to each triangulated 3D point the corresponding RGB attribute from the selected master image. Figure 6.52 shows the point cloud achieved by starting from the 5° dataset, whereas Figure 6.53 zooms in on a smaller area of the same point cloud. This second view is also reported in shading-mode (Figure 6.54), as it was delivered by the IGN's tool Grshade. As expected, the dark back wall was not matched by the algorithm, since it represents a textureless surface.

Figure 6.53 Zoom view (point cloud)

Figure 6.54 Zoom view (shading)

In order to assess the metric accuracy of the IBM results, the raw and not-edited image-based point clouds were compared with adequate reference data. The point clouds acquired with the Surphaser Laser Scanner and registered within the Laser Tracker reference frame were chosen as reference model. Comparisons were carried out with the software PolyWorks IMAlign™, by setting the alignment process to perform no iteration (a comparison between the 3D point clouds is thereby achieved). For each dataset, the following results are provided:

- Standard Deviation (Std. Dev.) of the distances between the two compared datasets (LS and IBM), with the corresponding histogram (Table 6.61);
- Color-coded map of the comparison (Figure 6.55-57). The colour scale always ranges from -5 mm (violet) to +5 mm (red).

Finally, Table 6.62 summarizes the computational time required by the dense image matching processes and the overall computational time required by the complete image-based pipelines.

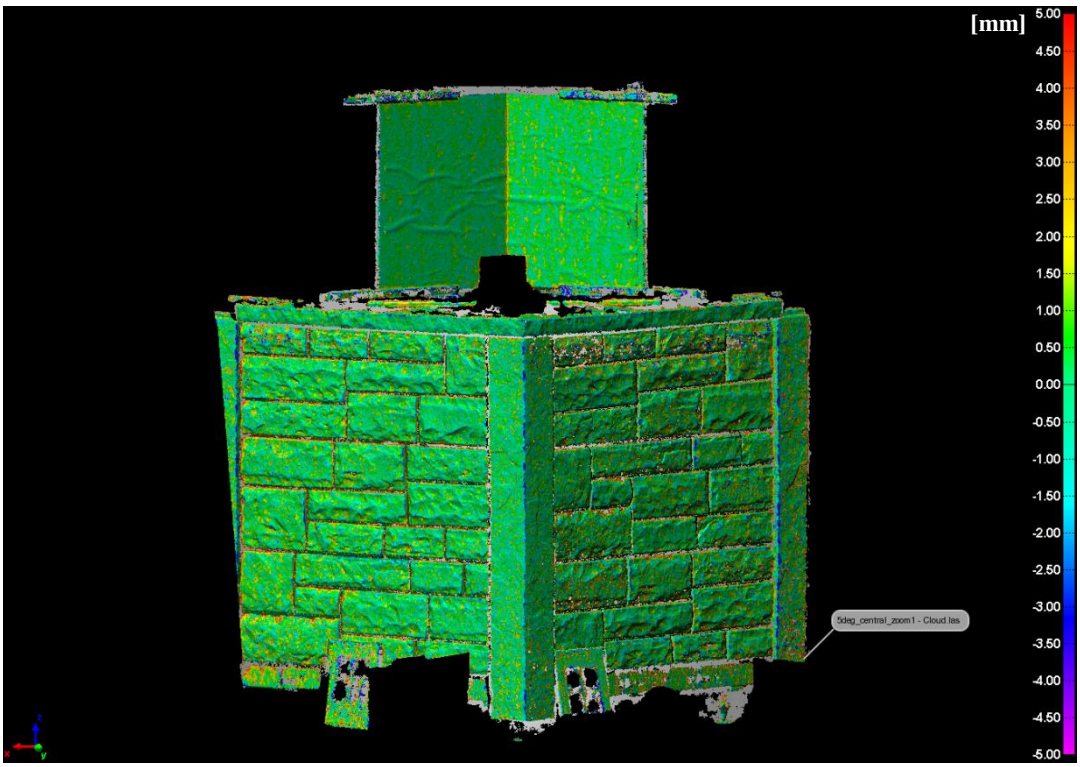


Figure 6.55 Comparison between the IBM point cloud (5°-dataset) and the LS point cloud. The colour scale ranges from -5 mm (violet) to +5 mm (red)

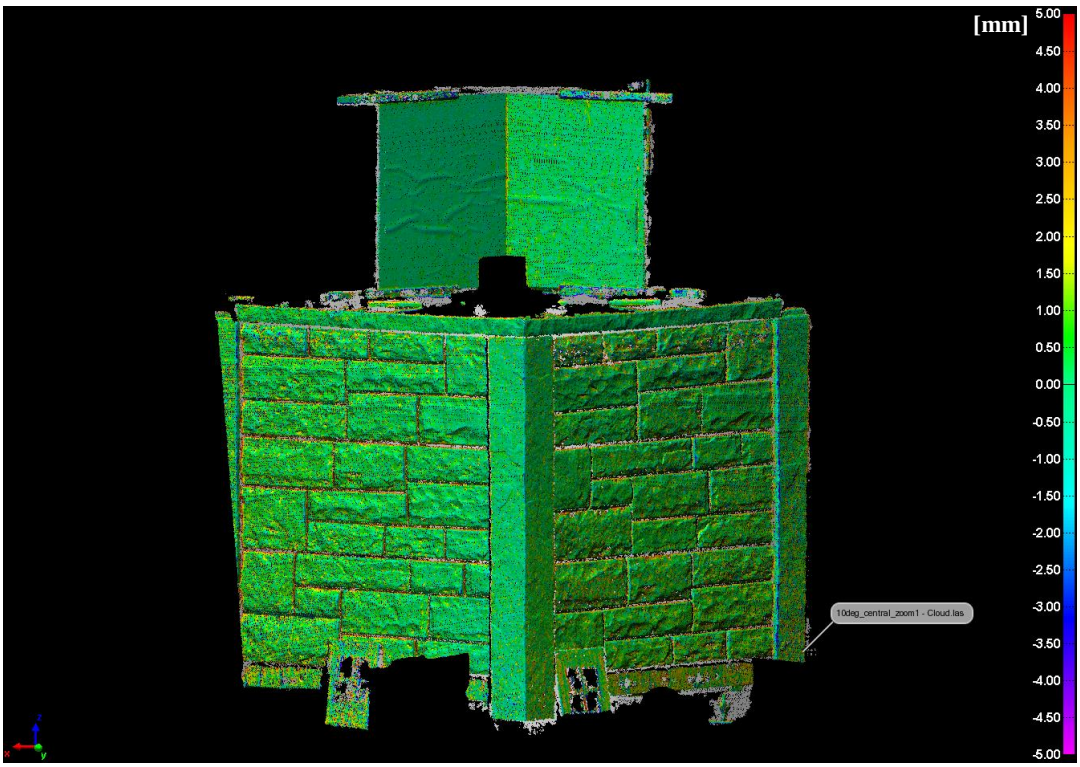


Figure 6.56 Comparison between the IBM point cloud (10°-dataset) and the LS point cloud. The colour scale ranges from -5 mm (violet) to +5 mm (red)

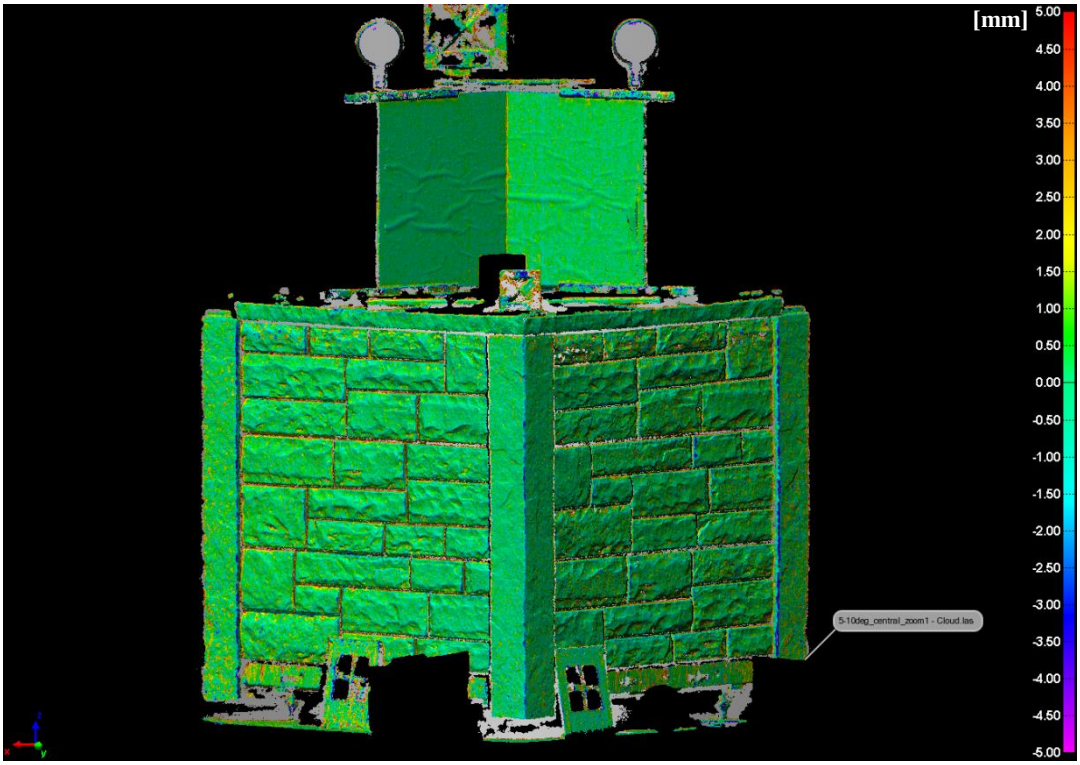


Figure 6.57 Comparison between the IBM point cloud (5°+10°-dataset) and the LS point cloud. The colour scale ranges from -5 mm (violet) to +5 mm (red)

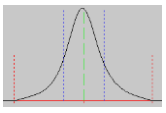
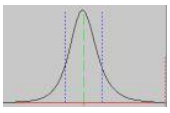
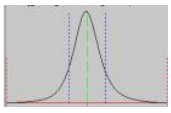
DENSE IMAGE MATCHING AND POINT CLOUD EXTRACTION			
Dataset	5°	10°	5° + 10°
Std. Dev. (mm)	0.88	0.67	0.67
Histogram			

Table 6.61 Performances achieved in the phase of "Dense image matching and point cloud extraction" (Tools: Malt and Nuage2Ply)

COMPUTATIONAL TIME (minutes)			
Dataset	5°	10°	5° + 10°
Dense Matching	9.15	9.96	12.69
Complete Procedure	15.86	15.66	30.35

Table 6.62 Computational time required by the dense matching processes and by the entire IBM procedures

The dense image matching algorithm is able to reach the same accuracy level previously pointed out within the orientation metrological assessment: all tests, in fact, deliver a standard deviation of the differences between the reference entity (Surphaser LS point cloud) and the compared one (image-based point cloud) that is always below 1 mm. The results provided by the three datasets are comparable from a metric point of view, since the differences between the accuracy levels achieved within the three tests are not metrically significant. This is also proved by the regular bell shape of the corresponding three Gaussian error distributions. Nevertheless, the 5°-dataset seems to achieve a slightly lower accuracy result.
Furthermore, by looking at the delivered color-coded maps, the following observations can be finally added:

- The three error distributions show that the largest errors, in terms of deviations from the reference data, are mainly located at sharp surface gradients, such as the ones corresponding to the edges between the vertical walls of the corners and to the small grooves among the bricks. These sharp edges are problematic for active laser scanners when the spot diameter is large compared to the structural (lateral) resolution being analysed. In the present situation, there may be a mismatch between the structural resolution of the LS and the image-based 3D point clouds.
- The 3D image-based reconstruction is so detailed that also the small creases on the glued surfaces (upper corner-object) are correctly identified. This evidence proves the capability of the dense image matching algorithm to reconstruct small structural details and defects on the imaged surfaces.

7. 3D RECONSTRUCTION FROM UAV-BASED IMAGERY

7.1 Introduction

In order to create 3D models, both active and passive optical sensors can be efficiently applied. Nevertheless, the reconstruction of a 3D scene should deal, always more frequently, with the complexity of the object that must be modelled: thus, the use of integrated techniques, based on the logic of the Multi-Sensor Data Fusion (Chapter 1), is often the only way to get today a both complete and cost-efficient 3D digitization. In the field of Cultural Heritage this issue is especially related to objects or portions of structure that cannot be directly reached without the use of expensive or cumbersome equipment: this is for example the case of valuable objects placed on the top of columns, pillars and domes, or, in architectural applications, the roofing of historical buildings together with the upper part of towers. In all these circumstances, the multi-sensor integration may concern not only the fusion of data acquired with different instruments, but also the combined use of different acquisition platforms. In particular, **UAV (Unmanned Aerial Vehicle) systems**, purpose-fitting and metrically calibrated, can be efficiently used in order to integrate surveys performed from classical terrestrial platforms.

UAVs are to be understood as uninhabited and reusable motorized aerial vehicles (Van Blyenburgh, 1999) or, according to the UVS (Unmanned Vehicle Systems) International definition, as generic aircrafts designed to operate with no human pilot on-board. Many different terms can be found in the literature to define these systems, such as: RPV (Remotely Piloted Vehicle), ROA (Remotely Operated Aircraft), RC (Remote Controlled) Helicopter, UVS (Unmanned Vehicle Systems) and Model Helicopter. Independently from the employed terminology, the UVS International provides a categorization of these systems, based on their size, weight, endurance, range and flying altitude:

- Tactical UAVs, that include micro, mini, close-, short-, medium-range, medium-range endurance, low altitude deep penetration, low altitude long endurance, medium altitude long endurance systems. Their weight ranges from few kilograms up to 1000 kg, their range from few kilometres up to 500 km, their flight altitude from few kilometres up to 500 km and their endurance from some minutes up to 2-3 days;
- Strategical UAVs, that include high altitude long endurance, stratospheric and exo-stratospheric systems which can fly higher than 20,000 m altitude and have an endurance of 2-4 days;
- Special tasks UAVs, that include unmanned combat autonomous vehicles, lethal and decoys systems.

All these vehicles are remotely controlled, semi-autonomous, autonomous, or have a combination of these capabilities (Eisenbeiss, 2009). Thus, the fact that no pilot is physically present in the aircraft does not necessary imply that the UAV flies autonomously: the pilot

responsible for the system is controlling the flight from a remote station and the UAV crew (operator, backup-pilot, etc…) is, in many cases, larger than the one of a conventional aircraft (Everaerts, 2008).

From the application point of view, the development of UAV systems was initially motivated by military goals: unmanned inspections, surveillance, reconnaissance and mapping of inimical areas were the primary military applications. In the last years, however, UAV platforms have begun to be increasingly common even in the geomatics field, after the first experiences carried out by (Przybilla and Wester-Ebbinghaus, 1979). The so called **UAV photogrammetry** (Eisenbeiss, 2008) represents now a new photogrammetric measurement tool, introducing low-cost alternatives to the classical manned aerial photogrammetry and opening various new applications in the close range domain. In particular, the typical geomatics application domains that employ UAVs images and photogrammetry- or Computer Vision-derived 3D data can be summarized as follows (Remondino et al., 2011):

- Archaeology and Cutural Heritage, for the 3D documentation of sites and structures with low-altitude image-based surveys (Lambers et al., 2007; Sauerbier and Eisenbeiss, 2010);
- 3D reconstruction of man-made structures with image-based approach (Wang and Li, 2007; Irschara et al., 2010);
- Environmental surveying, for land and water monitoring (Thamm and Judex, 2006), Digital Surface Model extraction of coastal environment (Mancini et al., 2013) and post-disaster mapping (Baiocchi et al., 2013);
- Agriculture, in order to take adequate decisions to save money and time (e.g. precision farming) and to record possible damages or problems in the filed with a quick and accurate approach (Newcombe, 2007);
- Forestry, for woodlot assessments, fire surveillance, species identification and silviculture (Grenzdörffer et al., 2008);
- Traffic monitoring, for the retrieval of different data, such as travel time estimation, lane occupancies, incidence response and surveillance (Puri et al., 2007).

The development of UAV-based applications in so many fields can be mainly explained by the availability of low-cost platforms, combined with cost-effective measurement systems, such as amateur or SRL (Single-Lens Reflex) digital cameras and GNSS/INS systems (Global Navigation Satellite and Inertial Navigation Systems). Starting from the photogrammetric measurement equipment, off-the-shelf cameras can be also replaced by video cameras, thermal and infrared camera systems, or the platform can be equipped with a combination thereof (Eisenbeiss, 2009). The GNSS/INS systems, necessary to stabilize or support the flight and pilot the UAV with high precision to the predefined acquisition points (waypoints), are limited by the small size and reduced pay-load of common UAV platforms: thus, GNSS is mainly used in code-based positioning mode and is not sufficient for an accurate direct sensor orientation (Remondino et al., 2011). For this reasons, external orientation is generally computed with a Bundle Adjustment procedure, starting from automatically extracted tie pints

and using measured GCPs for the scene geo-reference. For a deeper study and overview of UAV systems, the reader is referred to (Eisenbeiss, 2009; Remondino et al., 2011).

The case study presented in this chapter is an example of a multi-sensor and multi-platform application in the Cultural Heritage architectural field. The use of a UAV system, with its image-derived products, is here integrated with data collected by an active optical sensor (Terrestrial Laser Scanner, TLS) from terrestrial acquisition stations: the combined use of active/passive sensors and UAV/terrestrial platforms was necessary in order to reconstruct the complete 3D geometry of an architectural element with a significant vertical extension. Next subsections will describe the test-object, the TLS survey and the general procedural workflow, whereas the following sections will provide a detailed discussion on the image-based UAV-derived 3D modelling pipeline and on its integration with the range-based data.

7.1.1 Basilica Santo Stefano (Bologna, Italy)

Figure 7.1 An overhead view of the Basilica Santo Stefano, Bologna (Italy)

Basilica Santo Stefano (or "Seven Churches Complex", as it is generally termed) is one of the main symbols of Bologna and is located at the intersection of its principal lines in the historical centre of the city. Figure 7.1 shows the significant development, both in plane and in altitude, of the religious estate, that includes seven main churches, differently connected to each other's. The origin of the buildings are very old: the legend states that the first development should be attributed to San Petronio, who, during its episcopacy between 431 and 450 DC, led the construction of the oldest part above a pre-existing temple of Isis. Afterwards, the expansion went on until the twentieth century, when the final restoration activities were completed. Two inner courtyards, the museum and the friar residences

integrate the complex, that is articulated into three main levels, enclosed by a basement and an attic. The bell tower stands above all: built in the thirteenth century, it was lifted up in the nineteenth century.

In 2012 the interdepartmental centre e-GEA (e-GEA) was charged to reconstruct the 3D model of the entire complex, both internal and external parts, with BIM (Building Information Modelling) approach. An active 3D imaging sensor, the Time-of-Flight pulsed ScanStation C10 laser scanner by Leica Geosystems, was adopted for the geometrical survey. Given the complexity of the object, the project should take on many different challenges, among which the roofing survey probably represented the most difficult one. The design and adoption of a favourable network of acquisition stations, that were partially performed from terraces and attics of the buildings surrounding the complex, have enabled the 3D complete reconstruction of the Basilica, including the upper portions. Unfortunately, this was not the case of the **bell tower**, which is so high that it stands out above all the other buildings. Its vertical facades, up to the highest mullioned window level, were acquired with the laser scanner, by exploiting the favourable height provided by the surrounding building roofing (Figure 7.2): this expedient allowed the scans to be performed without excessively sloping acquisition views.

Figure 7.2 Views of the LS acquisition phase

For the upper part of the tower, especially for its roofing, a different acquisition approach should be employed, in order to effectively deal with the following problems:

- Height of the architectural element, whose higher levels were not directly reachable without the use of very expensive and cumbersome equipment;
- Location of the architectural element, that is placed in the heart of the historical city centre, thus requiring special measures to secure the area;

- Impossibility of designing and realizing a pre-signalized target network in the upper part of the architectural element, since it is hard to be reached and has a significant artistic value. Moreover, no natural points, such as brick corners, were measured with Total Station survey: an accurate and adequate station network was, in fact, practically prevented by the over mentioned problems.

In order to adequately ride out these difficulties, images acquired from a UAV platform were employed in this application to extract the point cloud of the upper part of the tower, that was then integrated with LS data, delivering a complete 3D model of the structure.

7.1.2 Procedural workflow

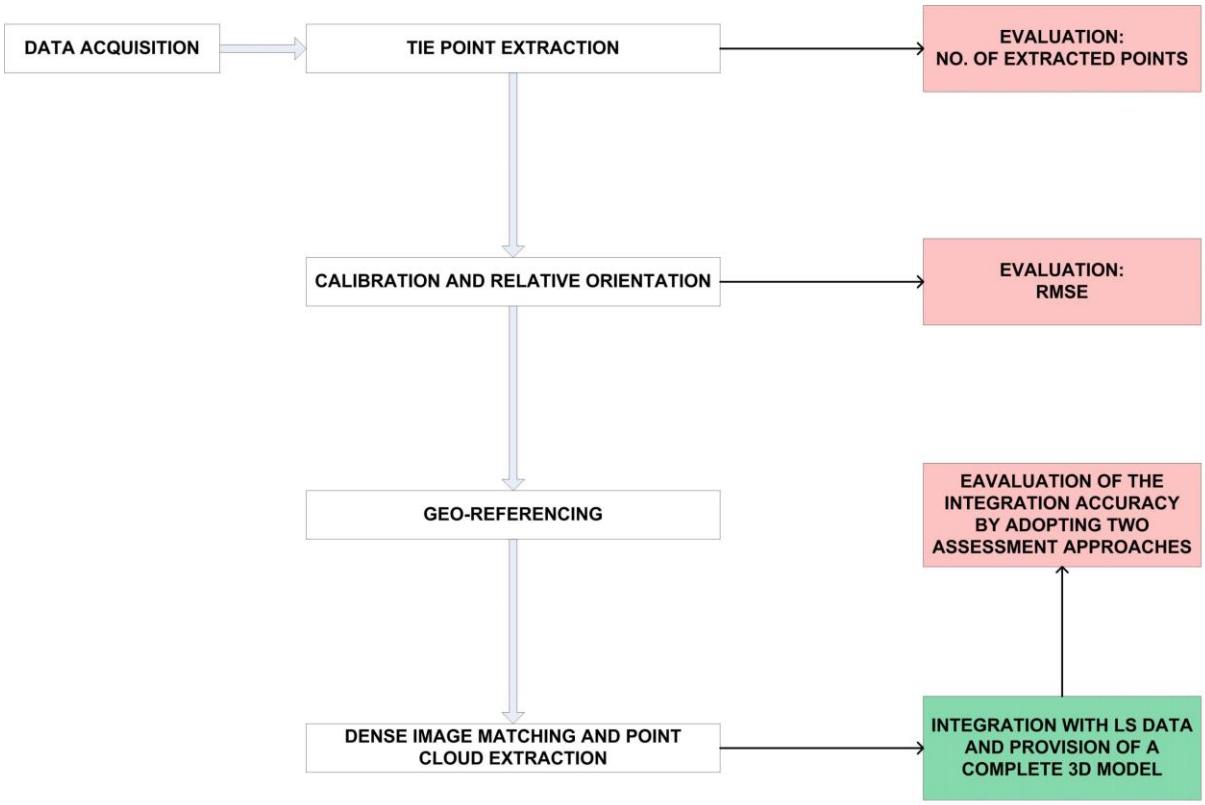


Figure 7.3 Procedural workflow (UAV-based application)

Figure 7.3 shows a synthetic overview of the procedural workflow followed within this experimental test. The entire photogrammetric and computer-vision based pipeline has been performed with the IGN's suite of tools, starting from the detection of homologous points between the images and ending up with the dense 3D reconstruction of the scene of interest. While during the previously described applications each procedural step was metrically evaluated, in this case study the attention was mainly paid to the metric assessment of the integration between the data acquired with the multi-sensor and multi-platform approach. Two different evaluation analyses will be presented at the end of this Chapter, after a detailed discussion of each image-based modelling phase, starting from data acquisition. For the

algorithmic and operative aspects related to the IGN's suite of tools, the reader is referred to Chapter 3.

7.2 Image acquisition

The UAV system employed for the digital image acquisition was a VTOL (Vertical Take Off and Landing) multi-rotor hexacopter, designed and manufactured by Sea Air Land Engineering (SAL Engineering) and equipped with a Canon EOS model 550D digital camera. A Canon Zoom Lens EF-S 18-55 mm was adopted. Table 7.1 lists the main technical specifications of the UAV system whereas Table 7.2 provides those of the on-board photogrammetric equipment. Figure 7.4 finally depicts the entire equipment.

UAV System	
TYPE	Micro-drone Multi-rotor Hexacopter
ENGINE POWER	6 electric brushless
DIMENSION	100 cm (diameter), 30 cm (height)
WEIGHT	3.3 kg (including batteries)
MAX. PAYLOAD	2.5 kg
FLIGHT MODE	Automatic based on waypoints / manual based on wireless control
ENDURANCE	Standard 20 minutes (+5 minutes safety)
FLEXIBLE CAMERAS CONFIGURATIONS	Digital gimbal Bi-axial roll and pitch control
GROUND CONTROL SYSTEMS	8-channels, UHF modem, telemetry for real time flight control, and path tracking on video within 5 km

Table 7.1 Some key specifications of the Unmanned Aerial Vehicle (UAV) system

Canon EOS 550D	
BODY TYPE	Compact SRL
SENSOR RESOLUTION	18 Mpixel
SENSOR SIZE	APS-C (22.3 – 14.9 mm)
SENSOR TYPE	CMOS
ISO	Auto,100-200-400-800-1600-3200-6400-

	12800 (with boost)
MIN SHUTTER SPEED	30 sec
MAX SHUTTER SPEED	1/4000 sec
Canon Zoom Lens EF-S 18-55 mm	
FOCAL LENS	18-55 mm
MAX APERTURE	f3.5-f5.6
MIN APERTURE	f22.0-f38.0
MIN FOCUS	0.28 m

Table 7.2 Some key specifications of the on-board photogrammetric equipment

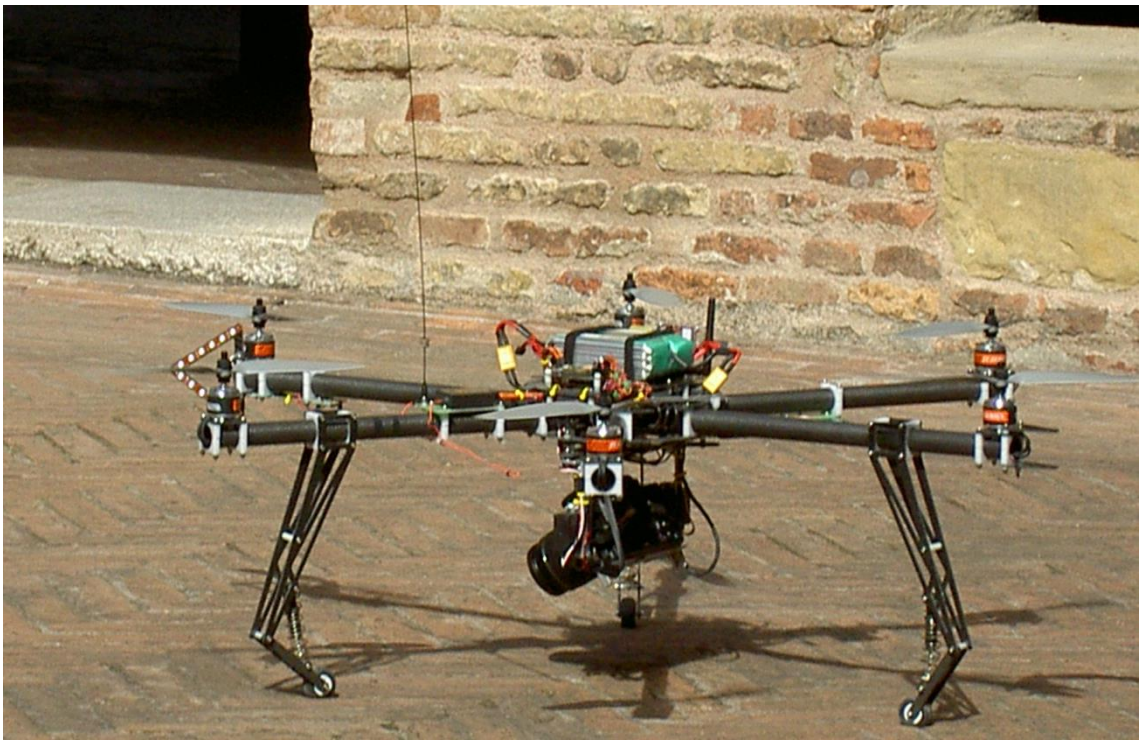


Figure 7.4 The UAV Hexacopter with the embedded equipment

The acquisition was performed through two consecutive flights, that were both carried out in manual mode, i.e. by remotely controlling and piloting the vehicle with an ad-hoc wireless management system. During the first flight images were acquired using a remotely controlled shooting system, that is independent by the one driving the drone: two operators were hence involved with the UAV management. The second flight, vice versa, was conducted with an automatic sequential shooting mode. Along the whole acquisition phase, the following setup was adequately kept fixed: focal lens 18 mm, infinity focusing, aperture f-4.5 and sensibility ISO 100. In order to adopt an acquisition configuration that may be, as much as possible,

favourable to the photogrammetric and computer vision-based image processing, the following measures were taken:

- Different angles of convergent images and different camera-object distances. The flights were carried out with a minimum distance to the object of 6 m, whereas the maximum distance was about 24 m. Thereby the GSD (Ground Sample Distance) on the surface of interest was, on average, equal to 3 mm, that was considered a good compromise value, considering the spatial resolution of the LS point cloud.
- Short duration of the image acquisition phase. A huge amount of images was acquired as quickly as possible in order to avoid illumination changes over the 3D scene. Nevertheless, the two flights were performed under sunny weather conditions, thus some "moving" shadows were anyway imaged.

Figure 7.5 Some images acquired during the flights

A total amount of 156 images was shot and recorded in Canon raw format (.CR2 files); the flights were piloted from one of the inner courtyards, that offered a favourable point of view and control. The vehicle was always controlled at sight: thus, only the portions of the tower that were directly visible from the courtyard were imaged for the 3D reconstruction purpose. This resulted in a lack of data describing the North-East and North-West parts of the object, whereas its roofing, South-East and South-West portions were adequately imaged. Moreover, the South-West side was partially hidden from the piloting point, thus it was acquired with a less favourable configuration, i.e. a reduced number of different shootings and points of view was taken. Finally, for security reasons, the entire religious complex and the place in front of

it were both kept closed during the whole acquisition phase. Figure 7.5 shows four images of the total acquired amount, in order to provide a more clear idea of the portions effectively recorded and, thus, later modelled.

7.3 Image processing

Within a post-analysing phase, 25 images were selected for the 3D reconstruction purpose among the total amount of acquired data: the selection criterion was based on a compromise choice between a good coverage of the object and the computational time required by the post-processing steps. Figure 7.6 depicts the acquisition layout of the selected 25 images.

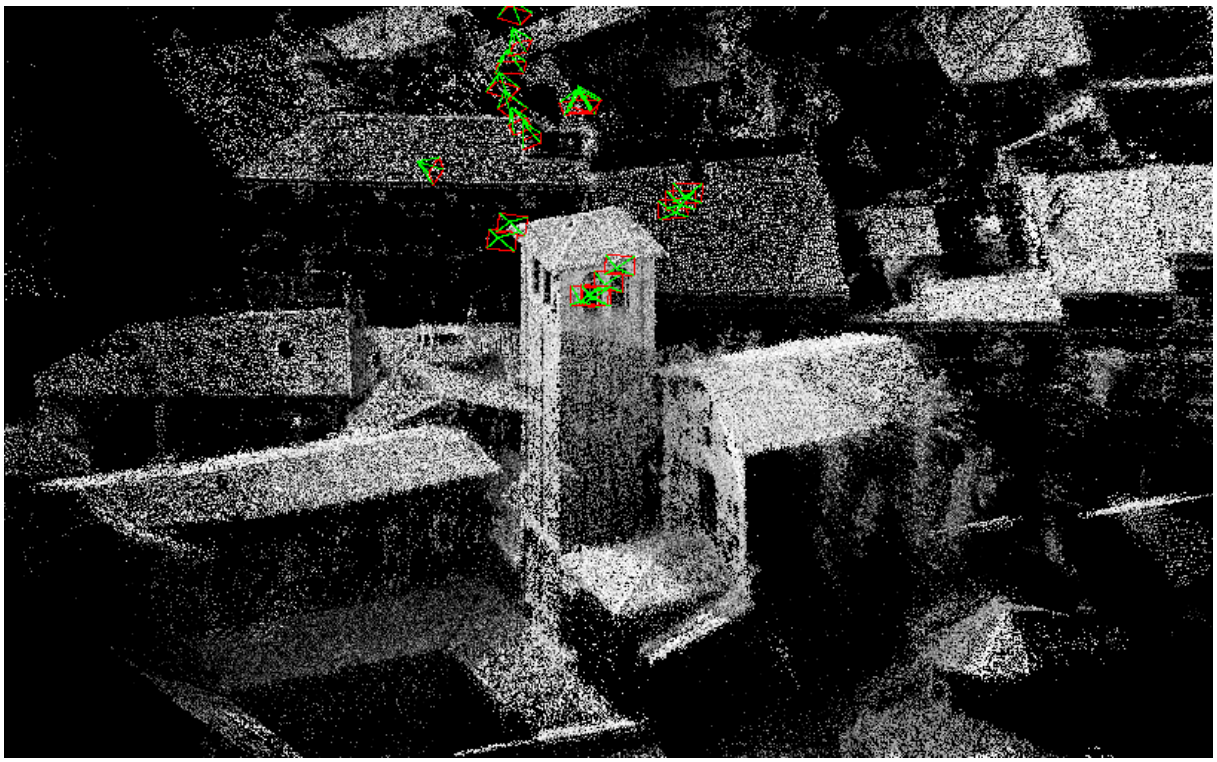


Figure 7.6 Acquisition layout of the selected images

In addition to these 25 images, 6 additional pictures were chosen for the pre-calibration procedure. The entire image-based modelling pipeline is described in the following subsections.

7.3.1 Tie point extraction

The SIFT^{++} implementation of SIFT algorithm employed by the tool Tapioca (Subsection 3.3.2) was used in order to detect and match homologous points between the 25 selected images. The tie point search mode was set to "All" in order to consider all possible pairs of images; furthermore, this research was carried out without any previous image shrinking, i.e. the original image resolution (5184 x 3456 pixel) was always used. In order to evaluate the

results, the mean number of homologous points matched in each image was finally computed: it amounts to 41,142 correspondences extracted, on average, per image.

7.3.2 Calibration and relative orientation

Since the 3D scene imaged in the selected dataset was not so favourable for the calibration parameter recovery, six additional images were chosen for a pre-calibration phase. Figure 7.7 shows three of these pictures.

Figure 7.7 Some images selected for the pre-calibration phase

This smaller dataset exhibits significant depth variations and surface textures and it was imaged according to the main calibration-favourable acquisition requirements (different heights, angles of view, etc…). Tie points were extracted with the tool Tapioca, following the same strategy adopted in the previous step (i.e. for the 25-image dataset). Calibration was then performed, starting from the usual initialization values (EXIF-derived focal length, no principal point offsets, no distortions), that were then refined using the homologous point information in the bundle adjustment phase. The tool Tapas (Subsection 3.3.3) was employed for this processing and the FraserBasic calibration model was adopted, with its 10 degrees of freedom.

Afterwards, the computed calibration parameters were given as input for the relative orientation of the 25-image dataset. This choice provided a "good" initialization of all the intrinsic calibration parameters, thus favouring the convergence of the global orientation computation. Starting from the results achieved in the previous case studies (Subsection

6.3.6), the calibration parameters were kept free and re-evaluated during the bundle adjustment procedure, using the additional information given by the tie points previously extracted from the 25-image dataset. Camera relative poses were computed and compensated within the same procedure; the Root Mean Square Error (RMSE) of all re-projection residuals was equal to 0.64 pixel: this value was considered a positive clue of the orientation accuracy.

7.3.3 Geo-referencing

In order to deliver a complete 3D model of the Tower, the two reconstruction outputs, i.e. the one delivered by the UAV-based image data processing and the one acquired with laser scanner technology, should be adequately registered in the same reference frame. Since neither pre-signalized targets nor natural points could be measured due to practical impossibilities, the metrically-consistent LS point cloud provided the reference data, necessary for the image-based model geo-referencing. Four points well recognizable in both LS data and images were selected (Figure 7.8) and used as GCPs.

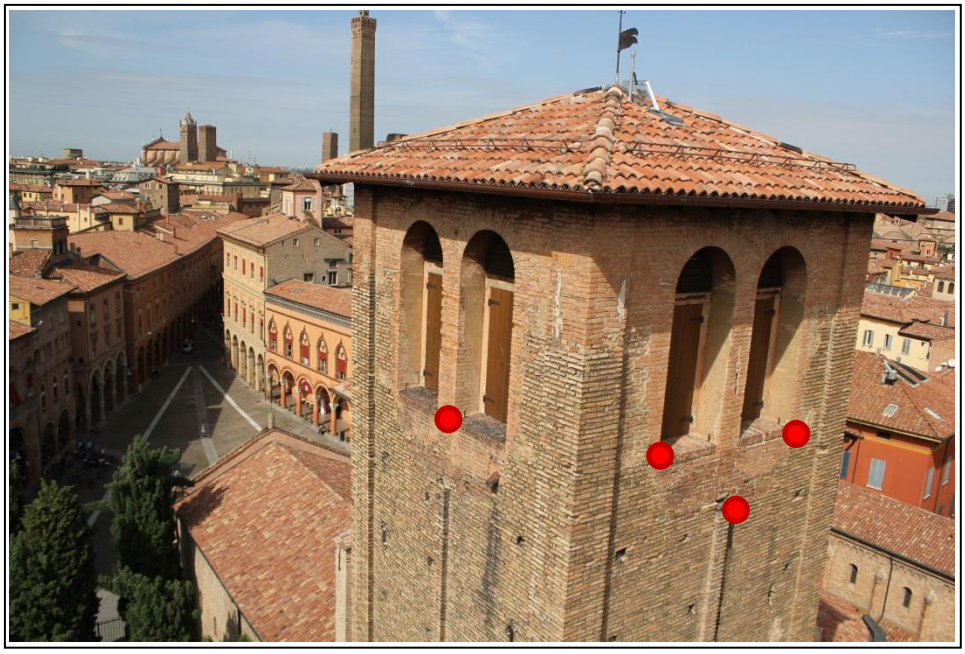


Figure 7.8 The GCPs used in the geo-reference procedure

The 3D coordinates of the selected GCPs were "included" in the bundle adjustment procedure in order to compensate them together with the other input observations. Of course, they have a limited metric accuracy, since they were extracted directly from the LS point cloud acquired with a sloping ray of view in the very upper part of the tower. For this reason, an adequate confidence level was associated with these observations in the compensation procedure.

Five images were chosen within the dataset and the four selected GCPs were manually collimated on each of them, using the tool SaisieAppuisInit. Starting from the collimated 2D coordinates (image reference frame) and the corresponding 3D coordinates (absolute reference frame), the global transformation from a purely relative orientation to the one

"registered" in the Laser Scanner reference frame was then carried out with the tool GCPBascule (Subsection 3.3.3). Finally, the tool Campari (Subsection 3.3.3) was run in order to perform a compensation (bundle adjustment) of all the provided heterogeneous observations, i.e. tie points and GCPs. In this application, no "*a-posteriori* evaluation" of the orientation accuracy using some external known reference data was feasible, due to the lack of reliable and externally measured Check Points.

7.3.4 Dense image matching and point cloud extraction

The object and, consequently, its 3D reconstruction were then divided into three main different points of view, corresponding to the following portions:

- Roofing;
- South-West upper façade;
- South-East upper façade.

The corresponding three master images were so selected (Figure 7.9) among the acquired dataset.

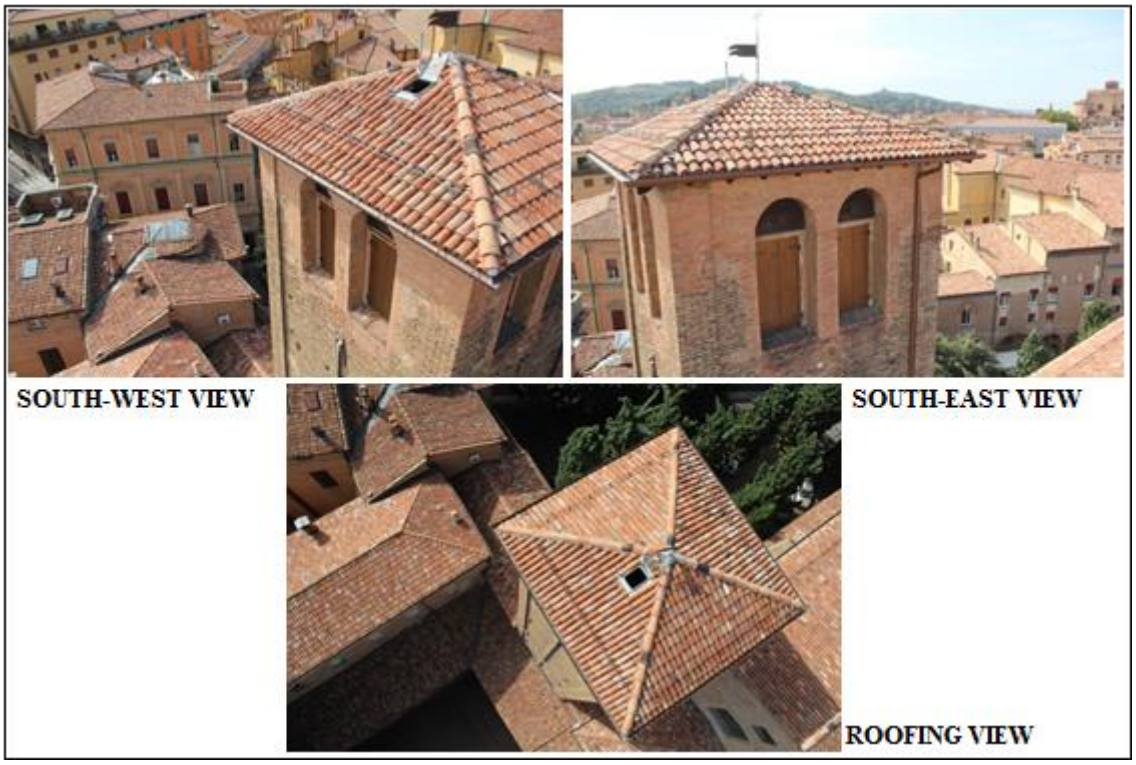


Figure 7.9 The master images selected for the three different points of view

Afterwards, the tool Malt (Subsection 3.3.4) was employed to extract a depth map for each point of view, using the image-ground geometry approach. Starting from the results achieved in the previous case studies and analysing the specific characteristics of the present dataset, the following parameter setup was finally adopted:

- Regularization factor equal to 0.2;

- Z-Quantification factor equal to 0.5;
- Depth of field interval to be explored equal to $[0.3 * D_0; 0.4 * D_0]$;
- Final Z-Resolution equal to 1 (higher pyramidal level).

Starting from the orientation results, the three depth maps were finally converted into point clouds with the tool Nuage2Ply (Subsection 3.3.4): the algorithm is able to deliver photo-textured results, by assigning to each triangulated 3D point the corresponding RGB attribute from the selected master image. Figure 7.10 shows the three point clouds achieved by the procedure.

Figure 7.10 The three image-based extracted point clouds

The roofing and the South-East point clouds resulted in, correspondingly, 4.6 and 4 million points. The South-West façade reconstruction was, on the contrary, less complete, as the number of dense extracted points (around 1,4 million) proves: this was of course due to the "poor" acquisition configuration adopted for that portion and forced by safety practical reasons.

7.4 Data integration and assessment

The integration between the 3D data acquired from the two platforms (terrestrial/UAV) with the two optical sensors (LS/digital camera) was performed within a commercial software, JRC 3D Reconstructor by Gexcel Software Solutions (JRC 3D Reconstructor). A refinement

of the registration among the point clouds was performed through the ICP (Iterative Closest Point) algorithm implemented in the software (Besl and McKay, 1992). Figure 7.11 shows the original point clouds and the integration between them.

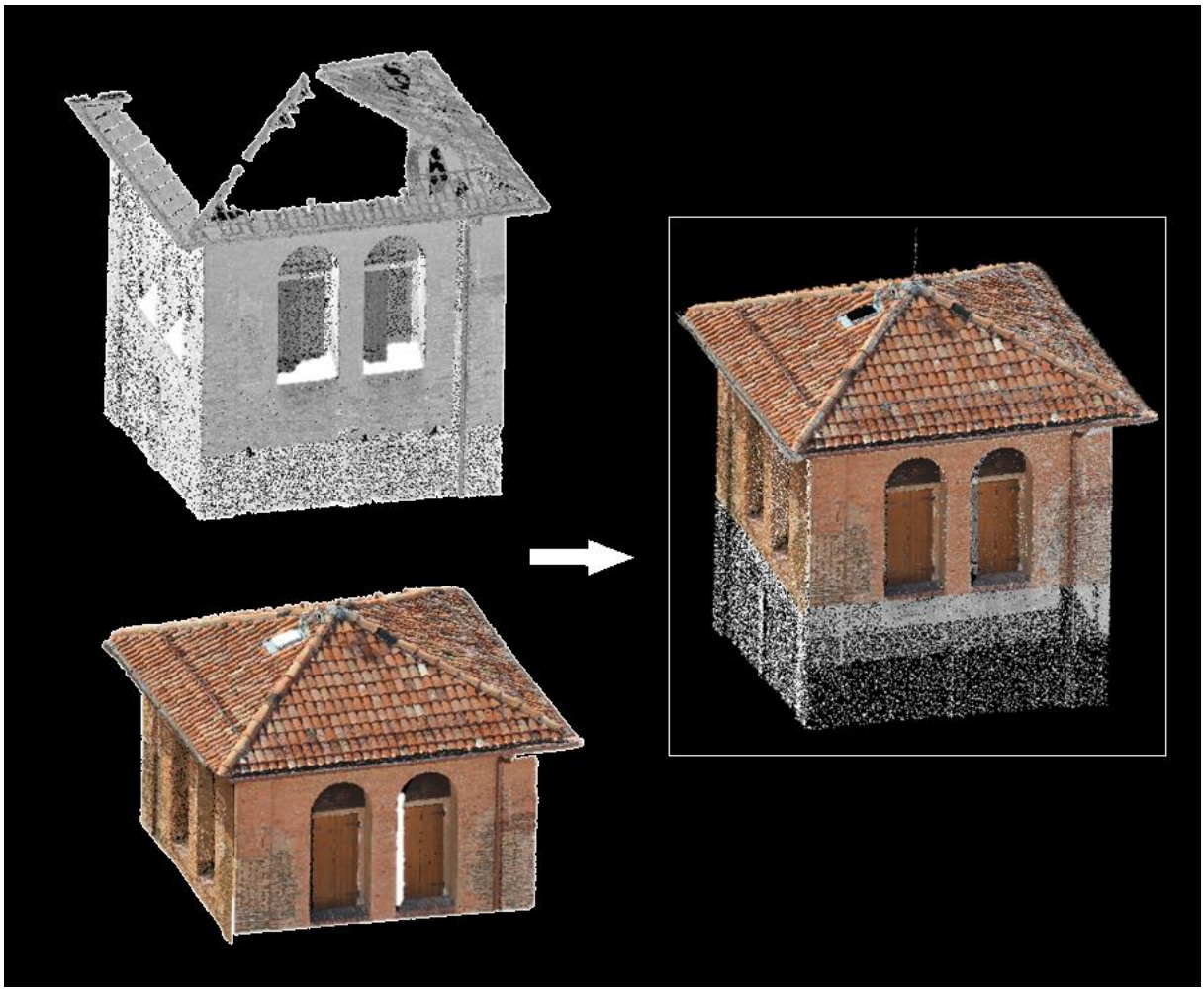


Figure 7.11 The original point clouds (on the left) and the integration between them (on the right)

In order to assess the accuracy of the integration between the two different 3D reconstructions, their common parts, corresponding to the upper portions of the South-West and South-East facades, were deepen analysed. Since no external reference information (e.g. Check Points) was available, the metric evaluation was based only on adequate comparisons between the dataset, after having developed appropriate methodologies, necessary to deal with data characterized by different levels of completeness. Two different assessment approaches were carried out and a detailed description of their results is below provided.

At first, a "localized" analysis was performed, by extracting many different horizontal and vertical cross sections from the image-based and range-based point clouds. Three of the planes employed to cut the 3D datasets are shown in Figure 7.12 and correspond to an horizontal section plane and two vertical section planes; the latters were chosen to be parallel to the Sooth-East and South-West facades of the Tower and their names are therewith accordant.

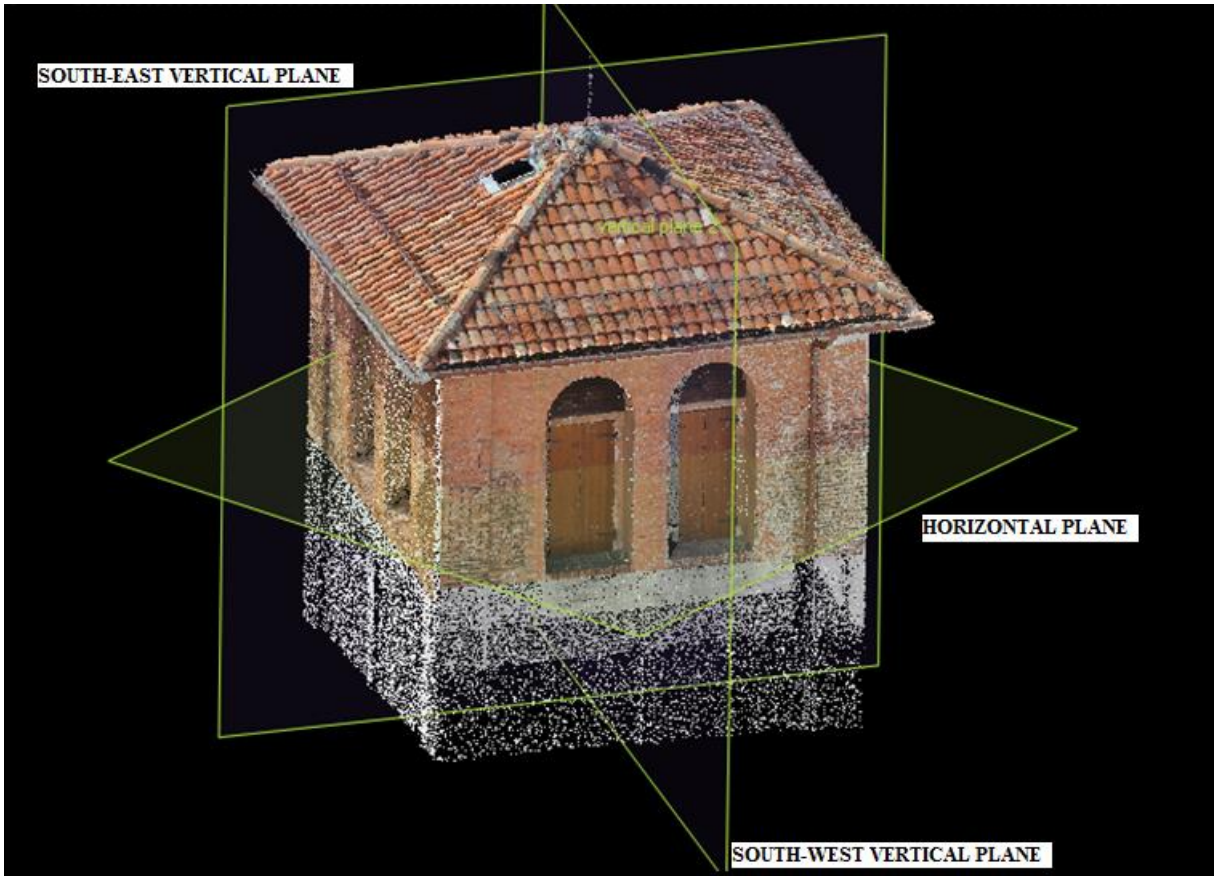


Figure 7.12 The horizontal and vertical cross section planes

The extracted sections were vectorized and analysed in CAD (Computer-Aided Drafting) environment, in order to estimate the local displacements between the point clouds. The major differences, in terms of section displacements, are reported below: Figure 7.13-15 show the analysed sections, whereas Table 7.3 lists the corresponding maximum displacement values.

MAX. DISPLACEMENTS (mm)		
Horizontal plane	South-West vertical plane	South-East vertical plane
18	13	21

Table 7.3 Cross section analysis: maximum displacements

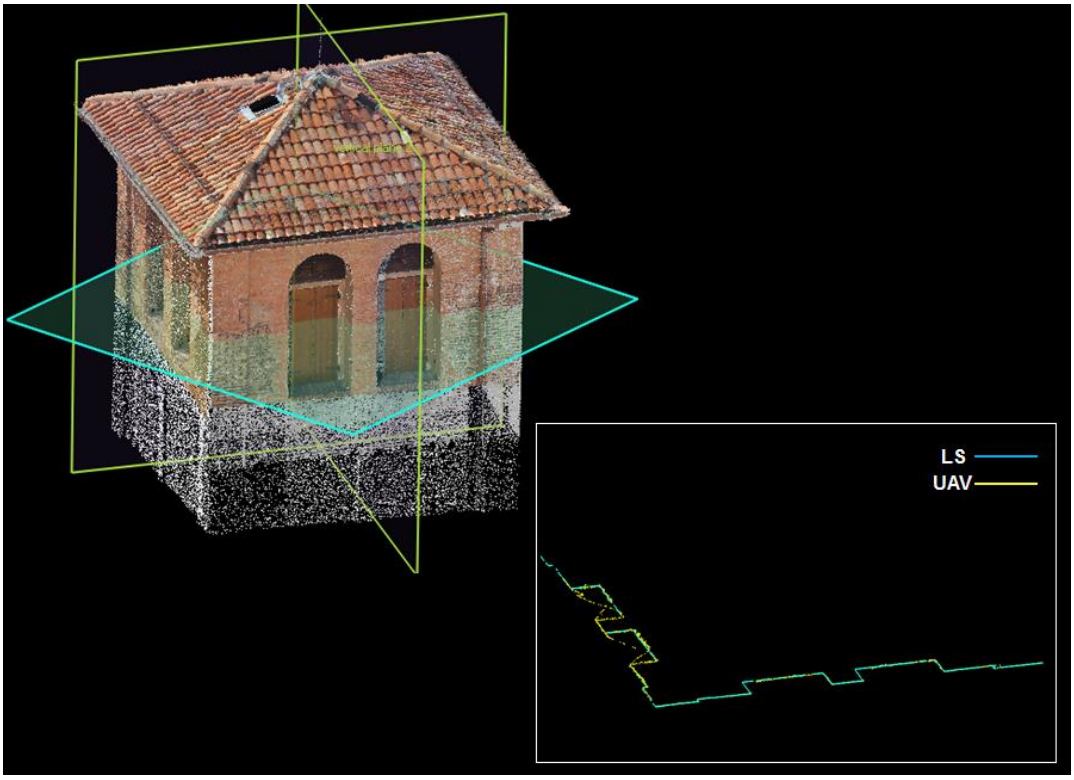


Figure 7.13 Cross section analysis: horizontal section

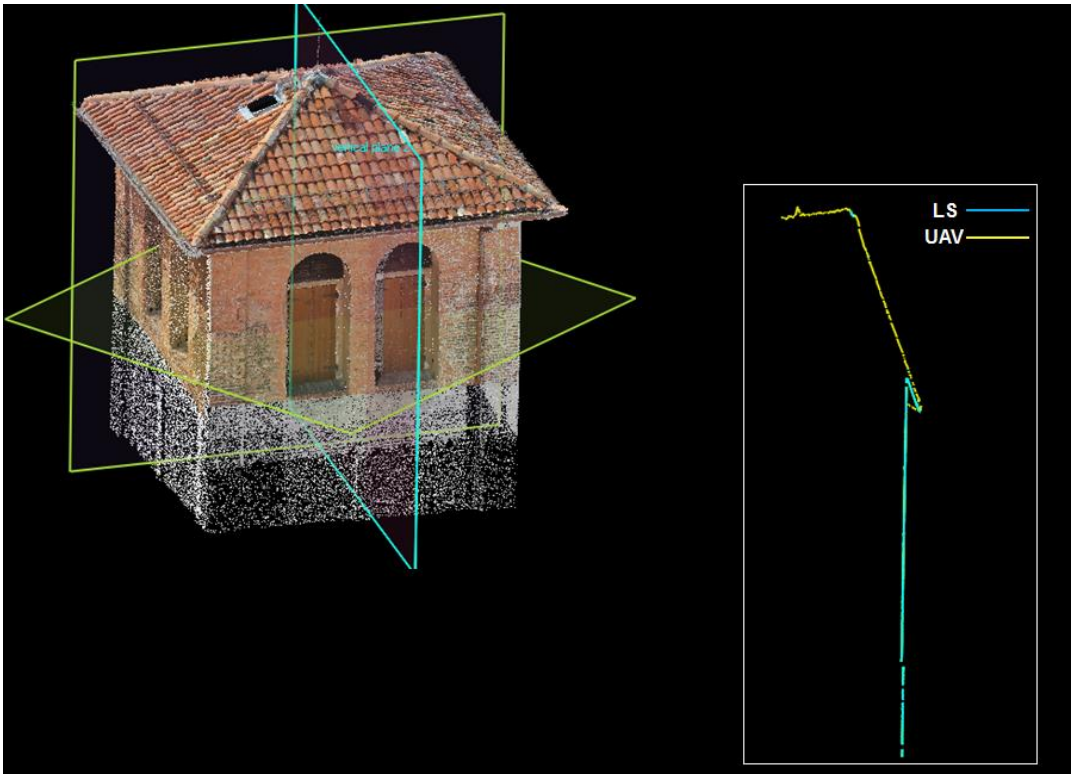


Figure 7.14 Cross section analysis: South-West vertical section

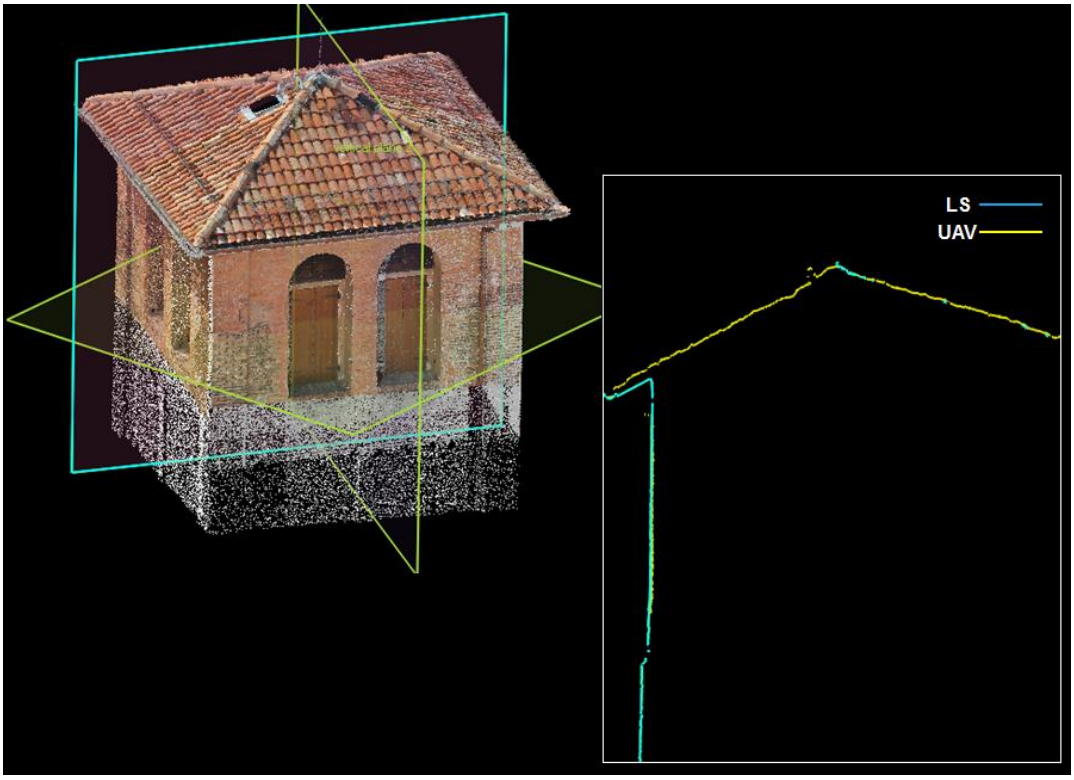


Figure 7.15 Cross section analysis: South-East vertical section

The second assessment approach was performed with the open-source software CloudCompare vs 2.4 (CloudCompare). The 3D model achieved with the software JRC 3D Reconstructor starting from the LS point cloud was set as the "reference model", whereas the point cloud extracted within the image-based pipeline was considered as "compared entity". Thereby the Euclidian distance between each point of the UAV-derived reconstruction to the LS model was computed, adding a new scalar value to each image-based coordinate-triplet. These measures were then filtered, limiting the maximum and minimum acceptable values of displacement within a pre-defined interval: many tests were performed, selecting different threshold values for the filtering process. Each filtered point cloud was then re-compared with the same LS model, until a good compromise choice was defined: this occurred when the filtering result was such that only the common parts between the two datasets could be effectively compared. This procedure was necessary in order to automatically eliminate outlier values from the computed distances, caused by comparisons between datasets characterized by different level of completeness. In other words, some portions of the object were visible only in one of the two delivered 3D reconstructions and only their common portions should be correctly compared. The implemented procedure offered a solution to this problem, without requiring a manual filtering of the two datasets.
Results are listed in Table 7.4, where the mean values of the computed distances (Mean. Dist.), the corresponding standard deviations (Std. Dev.) and the percentage of filtered points are reported. The resulting color-coded maps with the histograms of error distributions are shown in Figure 7.16-17.

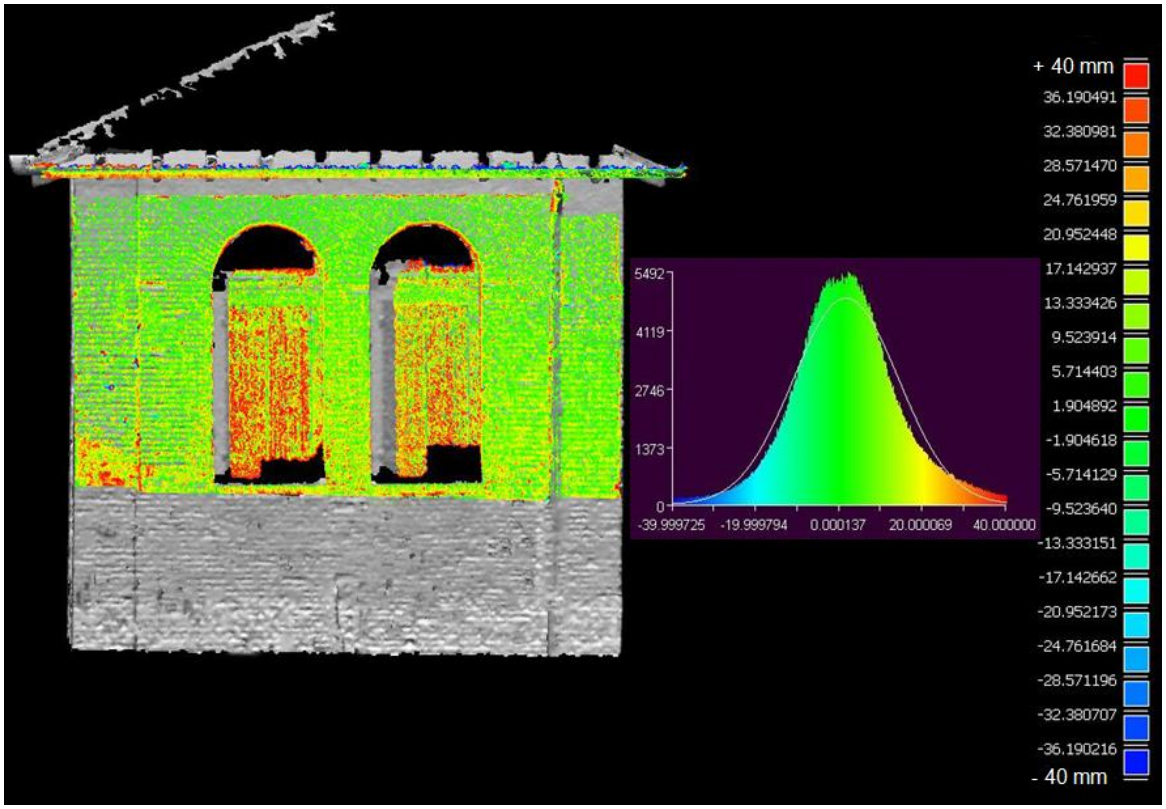


Figure 7.16 Comparison between the IBM point cloud (South-East façade) and the LS model. The colour scale ranges from -40 mm (blue) to 40 mm (red)

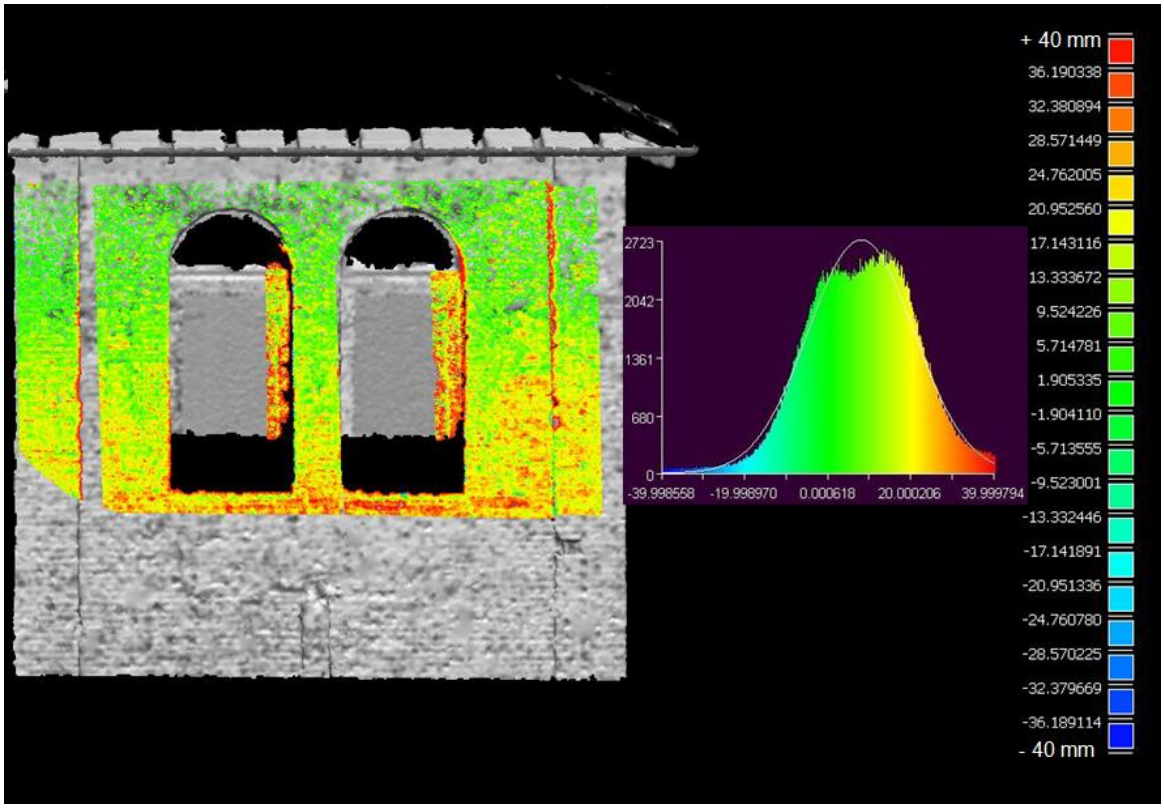


Figure 7.17 Comparison between the IBM point cloud (South-West façade) and the LS model. The colour scale ranges from -40 mm (blue) to 40 mm (red)

RESULTS OF COMPARISONS			
Compared portion	Mean Dist. (mm)	Std. Dev. (mm)	% of filtered points
South-East façade	1.5	12.7	9.53
South-West façade	7.7	12.7	17.43

Table 7.4 Comparison analysis: statistical results and percentage of filtered points

Both analyses show a significant metric accordance between the two integrated outputs, proving that the adopted multi-sensor and multi-platform approach has been successfully realized. By analysing the extracted color-coded maps, one can deduce that the most problematic areas, delivering the highest deviations between the dataset, are mainly the following ones:

- Boundary areas of the modelled 3D scene;
- Sharp surface gradients, such as those connected to the window sills and to the two lateral prominent bands;
- Shaded area, like the ones produced on the windows by the surrounding prominent masonry;
- Areas characterized by difficult textures, as the dark wooden window shutters.

All these performance limits were already pointed out within the previously described case studies, and are typical drawbacks of the image-based modelling approach (Remondino et al., 2008; Haala, 2013). Of course, errors within the LS-derived 3D model should be considered as well, especially since the range data acquisition was here performed under difficult external conditions due to the height and location of the element. A proper error budget calculation was not here performed; the use of external reference data would have supported this process.
Finally, as expected, the 3D reconstruction of the South-West façade is less complete and accurate from a metric point of view. This results in higher localized maximum displacements and higher mean value of distances, computed correspondingly within the cross section and the comparison analyses. The irregular shape of the histogram of error distribution, together with the more significant percentage of filtered points are two additional evidences thereof, proving that the influence of the acquisition protocol on the accuracy of the image-based approach cannot be ignored.

8. 3D MODELLING FROM SPACEBORNE IMAGERY

8.1 Introduction

A Digital Elevation Model (DEM) is a general term referring to a digital representation of a terrain's surface, created from terrain elevation data (Wei and Bartels, 2012). If the extracted surface model represents all sensor-detected heights, thus including visible objects on the top of the surface, it is termed DSM (Digital Surface Model); a DTM (Digital Terrain Model) is, on the contrary, a terrain representation showing only the bare ground surface topography. Such 3D models play a very important role in many applications requiring the recovery of the surface topography, in order to monitor changes in the Earth's surface as a function of time: they are, for example, employed within monitoring systems for post-disaster action planning (e.g. volcanoes, earthquakes and tsunamis). Furthermore, DEMs are also required by those scientific research disciplines involving studies of the Earth's land surface, e.g. cartography, climate modelling, geology, biogeography and soil science (Hutchinson, 1993). Especially thanks to the improvements in extracted data accuracy, DEMs are today employed within an increasingly wide range of applications, including for example suitability and sustainability studies for urban developments, telecommunication base stations, intelligent transportation and travel systems, together with floodplain mapping, land erosion analysis and applications in agriculture and forestry (Lohr, 1998). In this scenario, many relevant actors, such as public administrations and governances (Baiocchi et al., 2007), require DEMs to be adequate tools for description of the land surface and of its changes, both man-made and natural. In the first case, the three-dimensional information needed for urban area planning should provide a correct identification of the buildings: some literature studies showed the possibility of extracting these data from high resolution DSM (Weidner e Förstner, 1995; Lafarge et al., 2008; Tournaire et al., 2010) and spaceborne stereoscopic imagery (Fraser et al., 2002). If the requirement, on the contrary, refers to the modelling of the land morphology, including the natural phenomena that take place on it, DEMs can represent an adequate answer therefor only if they are able to accurately represent the bare ground surface, removing the visible objects on the top of it.

DEMs can be generated with different methods, that can be categorized as follows (Wei and Bartels, 2012):

- Passive remote sensors, relying on natural energy sources like the sun. This is also referred to as classical photogrammetry and is based on airborne or spaceborne multispectral/panchromatic images acquired in stereo-pairs in order to extract 3D information;
- Active remote sensors, detecting artificial energy sources transmitted to a target. They include RADAR (RAdio Detection And Ranging) stereo-pairs, InSAR (Interferometric Synthetic Aperture Radar) and LIDAR (Light Detection and Ranging), that involves laser scanning.

- Geodetic measurements through geodetic instruments, that collect planar and altitude coordinates (and resulting contour lines) starting from measures of lengths, angles and levels of land surface. This traditional method turns an hardcopy map to a digital data information by digitizing the surveyed contour lines and gridding them, if required.

The present chapter will be mainly devoted to spaceborne remote sensing through optical stereo imaging, by describing an application performed with a satellite stereo-pair. This case study especially aims at defining adequate procedures for the extraction of purpose-fitting three-dimensional information, i.e. the accurate description of all sensor-detector heights (DSM) or of only the bare ground surface (DTM). The application was carried out through two main experimental analyses. The first one delivers an accuracy assessment of both orientation and DSM extraction procedures, whereas the second one offers a comparison between two automatic approaches aimed at achieving a DTM through a building extraction phase. Both studies were performed with a stereo-pair captured by the WorldView-1 satellite over the Colli Albani (Rome), in collaboration with Planetek Italia s.r.l (Planetek Italia s.r.l). The employed dataset and the procedural workflow are described in Subsections 8.1.2 and 8.1.3, whereas subsection 8.1.1 provides a brief discussion on DEM generation from stereoscopic imagery.

8.1.1 DEM extraction from stereoscopic imagery

Started from photographic film cameras, passive remote sensing employs today spaceborne digital cameras with selective sensing bands, such as multispectral, thermal, hyperspectral and radar. It represents a good example of the passive, multi-view imaging technology discussed in Chapters 2 and 3, thus following the main rules of stereo 3D reconstruction systems. First of all, in fact, two or more overlapping images are required, and they can be acquired by along-track or across-track arrangements of sensors. In the first case, the acquisition is defined by the forward motion of the satellite along its orbital path, thus reducing the time interval between the imaged data: in this case, the radiometric changes between the images are limited and the correlation procedure within the image matching phase is thereby improved (Toutin, 2000). Across-track, on the contrary, refers to an image acquisition performed from different orbits, with a consequent more relevant influence of variable weather conditions.

The main steps of the DEM extraction procedure can be summarized as follows (Gabet et al., 1997; Hashimoto, 2000):

- Image pre-processing, in order to mitigate the effects of noise introduced by the image sensors;
- Image matching, in order to find correspondences between the images by either area-based or feature-based matching methods (or a combination thereof);
- Triangulation process, in order to convert the image coordinates of matched points into their corresponding ground coordinates. This procedure employs the camera

interior and exterior parameters and requires a geometric modelling of the satellite camera system;
- Accuracy evaluation of the extracted DEM, usually by means of measured check points.

Since most of the over-mentioned processes have already been discussed in Chapters 2 and 3, the attention is here focused especially on the orientation procedure, that requires a suitable image correction process in order to model the unavoidable geometric distortions. Of course, these latters depend on various different factors, but their main sources can be grouped into two categories (Toutin, 2004b): the acquisition system (platform, optical sensor and other measuring instruments) and the observed object (atmosphere and Earth). In order to perform the geometric correction of an image, models and mathematical functions are thus required. Two main correction models can be adopted (Toutin et al., 2002):

- Rigorous 3D parametric model, based on collinearity conditions and on a comprehensive understanding of the imaging geometry. In this case, if available, GCPs (Ground Control Points) can be employed in order to refine the ephemeris information provided by the image metadata;
- 2D/3D non-parametric models, such as Rational Polynomial Functions (RPF) including the Rational Polynomial Coefficients (RPC) provided within the image metadata. The RPC model is usually based on a mathematical ratio between third-order polynomials, thus requiring 80 coefficients per image: this formulation is employed to model the relationship among corresponding image and ground coordinates. In order to improve the computation, by incorporating the bias caused by residual image shift and drift effects, a bias-compensated formulation of the RPC model is proposed in (Fraser and Hanley, 2005) and solved through a multi-image and multi-point bundle adjustment approach (Fraser e Hanley, 2003; Grodecki e Dial, 2003). In this case, if available, GCPs can be used in order to refine the orientations by computing the coefficients of an affine transformation with a least-mean square iterative procedure.

After the first space mission providing stereoscopic imagery of the Earth's surface, the American CORONA spy satellite program (Galiatsatos et al., 2008), an increasingly number of different platforms have been launched to the space over the past decades, equipped with high resolution imaging systems. The lists includes, for example, Landsat (1972), IKONOS (1999), QUICKBIRD (2001), SPOT-5 (2002), ENVISAT (2002), ALOS (2006), WorldView-1 (2007), Geo-Eye-1 (2008) and WorldView-2 (2009). This rapid growth required a simultaneous effort in terms of software solutions, i.e. the development of new methodologies aimed at improving both the DEM accuracy and the level of automation. More robust computer vision algorithms were so developed for stereo image matching (Lowe, 2004) and, at the same time, many commercial software appeared on the market with special modules for automated DEM generation from stereoscopic imagery, such as: PCI Geomatica by PCI Geomatics Inc. (Geomatica 2013), ERDAS IMAGINE Photogrammetry by Intergraph (IMAGINE Photogrammetry) and ENVI by Exelis Visual Information Solutions (ENVI).

8.1.2 Procedural workflow

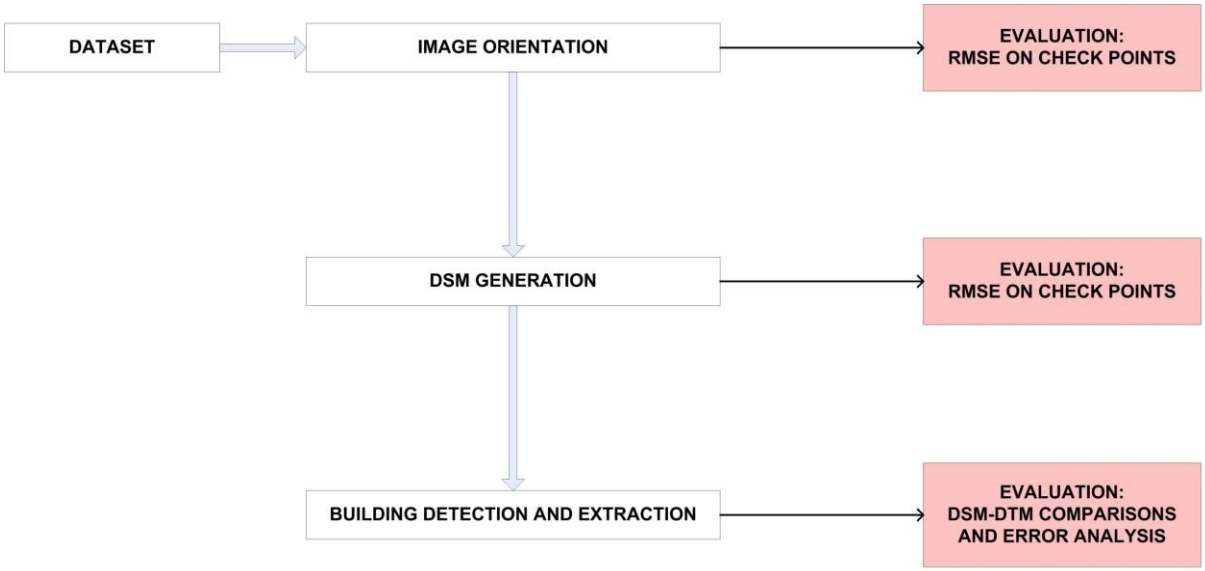


Figure 8.1 Procedural workflow (WorldView-1 stereoscopic imagery)

The application aims at analyzing and validating all the main steps that constitute the procedure for the extraction of digital elevation models from satellite images. The first part of the analysis is designed to compare the accuracy achievable by using both the physically based and the Rational Polynomial models. For this purpose, the orientation, the image matching and the final DSM extraction have been performed with the commercial software ERDAS Imagine Suite 2011, using the two different correction methodologies. An accuracy assessment was carried out at the end of each procedural step by employing a suitable number of Check Points (CPs), in order to evaluate the metric performance of the algorithms.
Secondly, the analysis is focused on the possibility of achieving a three-dimensional model of the area without the presence of buildings: two alternative automatic procedures were applied on adequate subset of the acquired 3D scene. Results were finally compared through two evaluation studies, i.e. DSM-DTM altitude differentiation and visual error analysis.

8.1.3 Dataset

The investigation has been carried out on an along-track panchromatic stereopair acquired over the Colli Albani area (Rome) by the DigitalGlobe's WorldView-1 satellite. The choice of such a sensor, in orbit since September 2007, was mainly due to its significant metric potentialities, especially in terms of spatial resolution (Ground sample Distance, GSD, 0.50 m) and stereoscopic imagery acquisition mode (along-track). The two images constituting the stereo-pair, hereinafter termed "North-Image" (Figure 8.3) and "South-Image", were delivered in GeoTIFF (Tagged Image File Format) format file and product level "Stereo-1B". Tables 8.1 lists the main characteristics of the imagery dataset.

CHARACTERISTICS OF THE STEREO-PAIR		
Acquisition Date	July 25, 2009	
Product Level	Stereo 1B	
Band	Panchromatic	
Radiometric Level	Corrected	
Bits Per Pixel	16	
Acquisition Mode	Full Swath	
Scan Direction	Forward	
File Format	GeoTIFF	
	North-Image	South-Image
GSD	0.523 m	0.674 m
Sun Elevation	64.8°	64.9°
Sun Azimuth	146.2°	146.6°
Satellite Elevation	76.4°	54.3°
Satellite Azimuth	6.4°	195.3°
Cloud Cover	0.003%	0.000%

Table 8.1 Main characteristics of the imagery dataset

Figure 8.2 The test-area

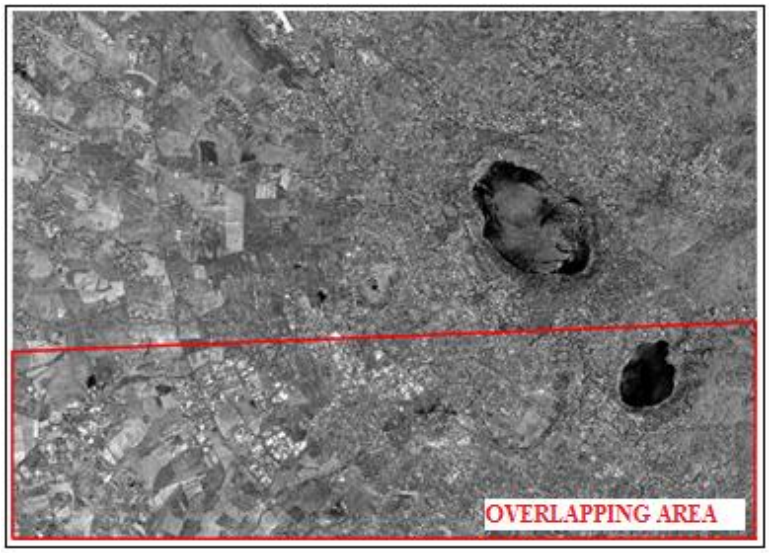


Figure 8.3 The overlapping area shown on the North-Image

The stereo-pair covers an area of about 100 km^2 over Colli Albani, south-east of Rome (Figure 8.2). The test-area is characterized by a complex and varied soil cover, including rural-woody lands, industrial parks, small towns and stand-alone buildings. The surface morphology is varied too, with altitude ranges from 0 to 700 m ASL.

Both ground control and check points, employed for the geometric correction of the images (GCPs) and for the *a-posteriori* accuracy assessment (CPs), were measured with GPS system (Global Positioning System) in RTK mode (Real Time Kinematic): a Leica Viva GNSS GS08 Receiver with CS10 Controller constituted the equipment. By averaging five measures one-second spaced, the UTM (Universal Transverse Mercator) ETRF2000 (European Terrestrial Reference Frame, epoch 2008.0) coordinates were delivered with accuracies on the order of few centimetres. Absolute altitudes were then computed with the software Verto2mila delivered by the IGM (Military Geographic Institute) with the ITALGEO99 geoid model, whose accuracy (σ=0.16 m) was considered suitable, if compared with the image spatial resolution. A total of 48 points were acquired, after having selected the best compromise choice in terms of spatial configuration: in particular, an homogeneous point distribution over the images and the possibility of easily pointing them out in both image and ground spaces were searched for (Cilloccu et al., 2009).

8.2 Image orientation

The orientation phase was carried out with the module LPS (Leica Photogrammetry Suite) of the ERDAS Imagine Suite 2011. Both rigorous 3D parametric model and the non-parametric RPC-based model were adopted in order to correct the image geometric distortions. The rigorous error formulation was based on the Orbital Pushbroom parametrical model implemented in LPS (Wang et al., 2008): a Lagrangian interpolation of the ephemeris parameters included in the image metadata was added by employing the available GCPs. The bias-compensated RPC-based non-parametrical model (simply termed RPC model) was applied by adding an orientation refinement through an affine transformation computed via GCPs.
The image orientation was performed starting from four different GCP datasets, i.e. including 5, 10, 15 and finally 20 points. Each dataset was chosen in order to fulfil the over mentioned requirements in terms of point distribution and selection. A quantitative evaluation of the orientation accuracy was conducted starting from the reports automatically delivered by the software (intrinsic assessment). In particular, the RMSE (Root Mean Square Error) of the orientation residuals computed on 14 CPs was considered a good accuracy measure. Four orientation tests were thus performed for each correction model (Orbital Pushbroom and RPC), using the four different GCP datasets. The same CP dataset was always employed, including 14 points distinct from those used as GCPs. All these setups were defined in order to analyse the metric performance of the two correction models when a different amount and distribution of GCP is available. Table 8.2 lists the RMSE values achieved by the tests, whereas Figure 8.4 shows the GCP and CP configurations employed in each test.

Both correction models show the same accuracy trend, if the three RMSE components (along the East, North and Up directions) are separately considered: residuals computed along the East direction show an increasingly smaller RMSE if the number of GCPs grows up; residuals computed along the North direction are more stable, i.e. less influenced by the number of

employed GCPs; finally, residuals on altitude coordinates show a slight increase if an higher amount of GCPs is employed. Furthermore, even if only 5 GCPs are employed, both correction models can achieve an accuracy level comparable with the image GSD; nevertheless, only the last test, performed with 20 GCPs, shows an optimal stabilization of the RMSE (around 0.40 m). Finally, the Orbital Pushbroom model always delivers the best results, even if the differences between the metric performances of the two models decrease with the increment of available GCPs.

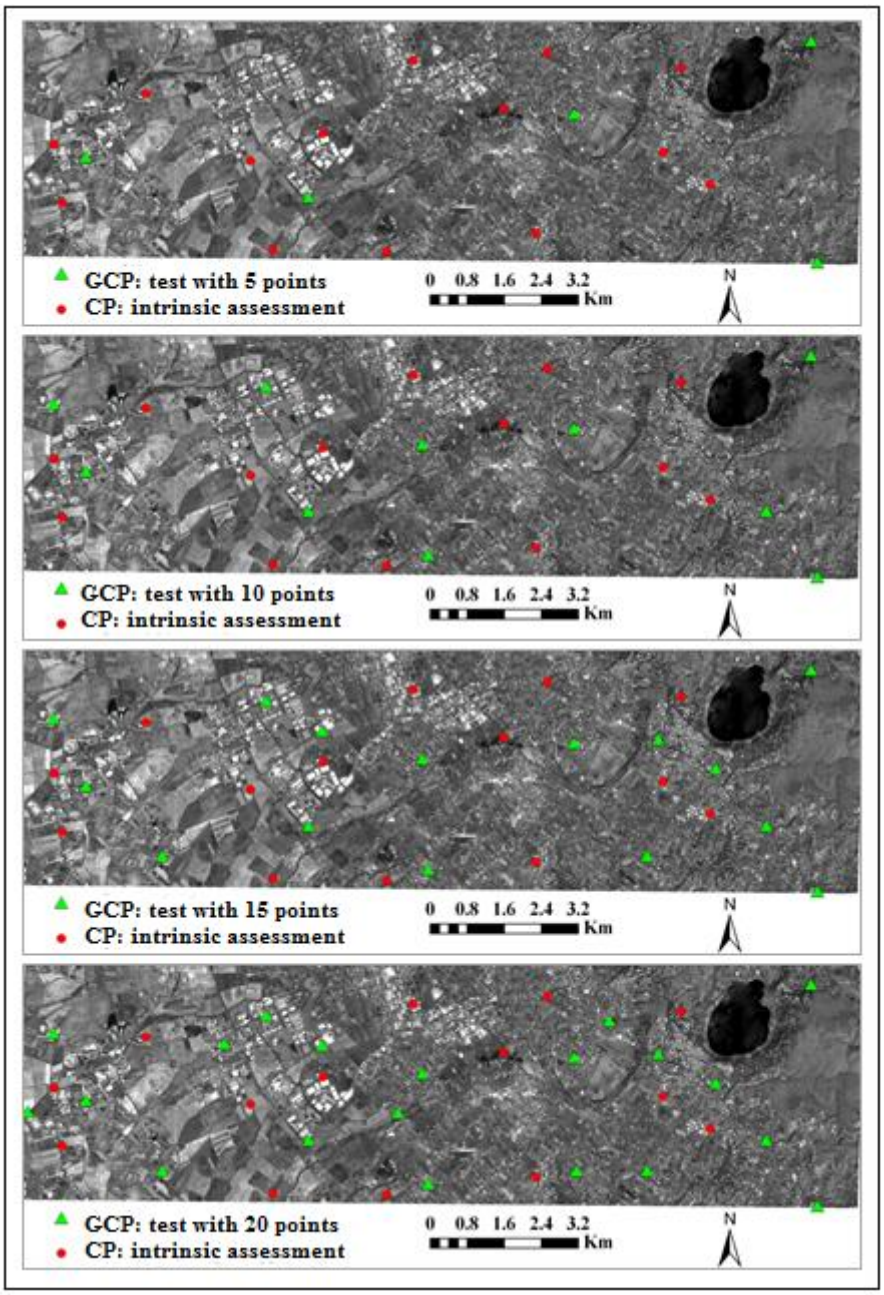


Figure 8.4 The GCP and CP configuration in the four tests: test with 5 GCPs, test with 10 GCPs, test with 15 GCPs and test with 20 GCPs (from top to bottom)

	TEST WITH 5 GCPS			TEST WITH 10 GCPS		
	RMSE $_{East}$ (m)	RMSE $_{North}$ (m)	RMSE $_{Up}$ (m)	RMSE $_{East}$ (m)	RMSE $_{North}$ (m)	RMSE $_{Up}$ (m)
Orbital Pushbroom	0.56	0.54	0.35	0.49	0.46	0.39
RPC	0.62	0.56	0.41	0.52	0.55	0.39
	TEST WITH 15 GCPS			**TEST WITH 20 GCPS**		
	RMSE $_{East}$ (m)	RMSE $_{North}$ (m)	RMSE $_{Up}$ (m)	RMSE $_{East}$ (m)	RMSE $_{North}$ (m)	RMSE $_{Up}$ (m)
Orbital Pushbroom	0.47	0.46	0.43	0.40	0.47	0.41
RPC	0.49	0.47	0.44	0.41	0.48	0.45

Table 8.2 RMSE of the residuals computed along the East, North and Up directions: intrinsic assessment

In order to perform an accuracy assessment that is independent from the control performed by the software itself, an additional analysis (termed *a-posteriori* assessment) was carried out on the two orientation results achieved with 20 GCPs. A new dataset of 14 GPS-measured points was selected among the available data that were not previously employed neither as GCPs nor as CPs. These points were then collimated in stereoscopic mode on the previously orientated stereo-models and their resulting 3D coordinates were finally compared with the ones measured with GPS. Figure 8.5 shows the spatial distribution of these new check points, whereas Table 8.3 lists the results achieved within this *a-posteriori* assessment.

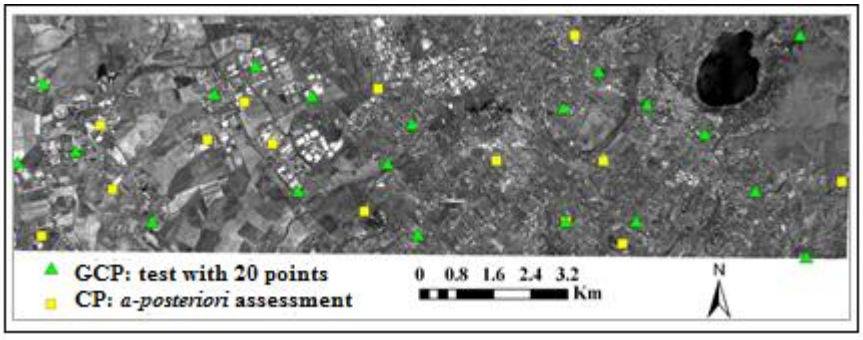


Figure 8.5 The GCP and CP configurations in the *a-posteriori* accuracy assessment

By comparing the results delivered by the two analyses (i.e. intrinsic and *a-posteriori* assessments), the same evidence can be pointed out for both correction models: the accuracy estimate computed along the East and North directions by the *a-posteriori* assessment is higher than the one provided in the internally-created reports; on the contrary, the external analysis delivers RMSE values along the altitude direction that are slightly worse than the ones declared by the software. Finally both analyses show that the highest level of metric accuracy is achievable by applying the rigorous 3D parametric model, whose average RMSE

is around 0.6 pixels, if 20 GCPs are used. With the same 20-GCP dataset, the RPC-based model is in fact able to reach a mean accuracy of around 0.7 pixels.

	***A-POSTERIORI* ANALYSYS (20-GCP DATASET)**		
	$RMSE_{East}$ (m)	$RMSE_{North}$ (m)	$RMSE_{Up}$ (m)
Orbital Pushbroom	0.29	0.32	0.50
RPC	0.35	0.42	0.57

Table 8.3 RMSE of the residuals computed along the East, North and Up directions: *a-posteriori* assessment

8.3 DSM generation

The automatic DSM extraction was carried out with the module eATE (enhanced Automatic Terrain Extraction) of the suite ERDAS Imagine 2011. The software provides "on-the-fly" quasi-epipolar image generation, thus reducing the computational time required by the procedure (Rozycki and Wolniewicz, 2007). DSM generation is then automatically performed through the following consecutive steps (Toutin, 2004a):

- Elevation parallaxes are first extracted by applying a multi-scale mean-normalized cross-correlation method, that computes the maximum of the correlation coefficient. This approach was proven to have a good metric performance with satellite images (Gülch, 1991).
- The XYZ coordinates are finally computed, by performing a 3D least square stereo-intersection with the previously computed orientation parameters and the elevation parallaxes.

Two DSMs were so generated starting from the stereo-models oriented with the 20-GCP dataset and the two correction approaches (Orbital Pushbroom and RPC-based). The elevation models were extracted in GRID raster format, i.e. the XYZ coordinates were computed in a regular grid spacing: its spatial resolution, equal to 1 m, was selected in accordance with the image mean GSD value (Cilloccu et al., 2009). Starting from the test-area characteristics and considering the resulting computational effort, the following parameter setup was selected for the DSM extraction procedure:

- Size of the correlation window equal to 5x5 pixels;
- Spike interpolation for non-correlated points;
- PCA analysis (Principal Component Analysis) for redundant data filtering (Daultrey, 1976);
- Moderate smoothing for peak removal.

Once the two DSMs were extracted, a quantitative assessment of their final accuracy was conducted. This is generally performed by comparing the computed elevation values with adequate reference data, such as:

- Control points collected by ground survey;
- Corresponding DEMs generated by higher accuracy technologies (e.g. LIDAR).

In both cases, the statistical parameter RMSE is usually employed and formulated as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \Delta h_i^2}{n}} \qquad [8.1]$$

Where:

- n is the number of evaluated points;
- Δh is the difference between the DEM elevations and the corresponding reference values.

Within the present application, since no reference DSM of the test-area is available, all 28 CPs measured with GPS in RTK mode were employed as reference data (Cilloccu et al., 2009). The comparisons were performed with the software Surfer 9.0 by Golden Software Inc. (Surfer), by computing the differences between the altitude values extracted from the DSM with bi-linear interpolation and the corresponding altitude values measured with GPS. Statistics (Table 8.4) show an altitude accuracy equal to 2.3 pixels for the rigorous correction model and equal to 3 pixels for the RPC-based one. A study of the computed residuals was then performed, in order to identify possible outliers. In particular, two points were more deepen analysed, since they delivered significantly higher (> 4 m) residual values. This evidence was linked to the position of the two points: they are in fact localized along the image boundaries, that usually represent problematic areas in terms of extracted DSM accuracy (Crespi et al., 2009). After having removed the image boundaries from the area subjected to DSM extraction, two new DSMs were generated and metrically evaluated through the remaining 26 CPs. Both correction models show, in this case, the same accuracy level, with a RMSE along the Up direction equal to 1.3 pixels (Table 8.4).

	ORBITAL PUSHBROOM (20 GCP)		**RPC-BASED MODEL (20 GCP)**	
No. of CPs	28 CPs	26 CPs	28 CPs	26 CPs
Mean Error $_{Up}$ (m)	-0.33	-0.42	-0.16	-0.24
RMSE $_{Up}$ (m)	1.36	0.79	1.75	0.73

Table 8.4 DSM accuracy assessment. The outputs extracted with the 20-GCP dataset and the two correction models (Orbital Pushbroom and RPC-based) are evaluated

A final test was performed with the DSM extracted from the stereo-model orientated with the RPC-based formulation and the 5-GCP dataset. The same parameter setup employed within the previous processes was adopted also in this case; as previously, image boundary areas were first included in the computation and then removed. Table 8.5 lists the accuracy results achieved in the subsequent accuracy assessment phase, that was performed with the software Surfer 9.0 and the two CP datasets already selected for the previous DSM evaluation tests.

	RPC-BASED MODEL (5 GCP)	
No. of CPs	28 CPs	26 CPs
Mean Error $_{Up}$ (m)	-0.23	-0.43
RMSE $_{Up}$ (m)	1.76	0.78

Table 8.5 DSM accuracy assessment. The output extracted with the 5-GCP dataset and the RPC-based correction model is evaluated

By comparing the statistical results listed in Tables 8.4-5 and considering the performances achieved by the RPC-based models, one can first point out that the accuracy of the extracted altitudes is the same, even if only 5 GCPs are used in the orientation refinement phase. As shown in (Cheng e Chaapel, 2008), DSMs generated from WorldView-1 stereo images with RPC-correction model can achieve a significant accuracy level, even if a minimum amount of measured GCPs is available. Furthermore, the RPC data provided by the latest generation sensors, such as WorldView-1, allow now a metric performance that is comparable with the one achieved by the rigorous correction model.

8.4 Building detection and extraction

The possibility of detecting the building information and removing it from the final 3D digital model was then analysed. This study was performed on two smaller test-areas (Figure 8.6), selected within the entire image overlapping region. The "West-Area" and "East-Area", as they will be hereinafter termed, have the following characteristics:

- The West-Area has an extent of about 2 km^2 and shows a main industrial vocation. Its buildings are mostly warehouses of regular and compact shapes (e.g. rectangular bases), with a significant development both in plane and height.
- The East-Area has an extent of about 4 km^2 and is an example of a typical residential area, whose small and irregular buildings are variously arranged over it.

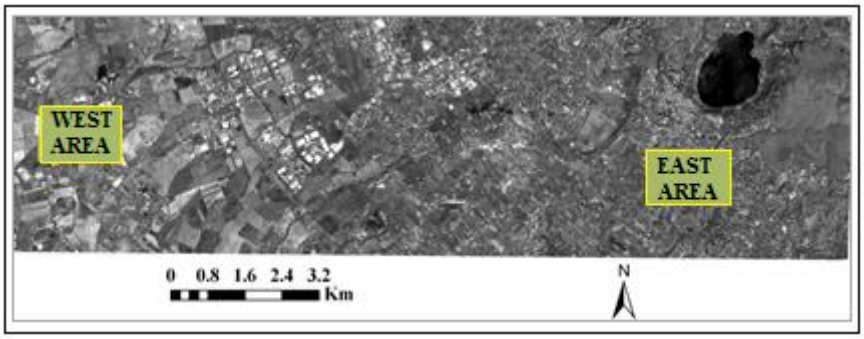


Figure 8.6 The two test-areas selected for the analysis (Building detection and extraction)

Figure 8.7 shows the two selected test-areas as they are reconstructed by the DSM generated with the 20-GCP dataset and the rigorous correction model.

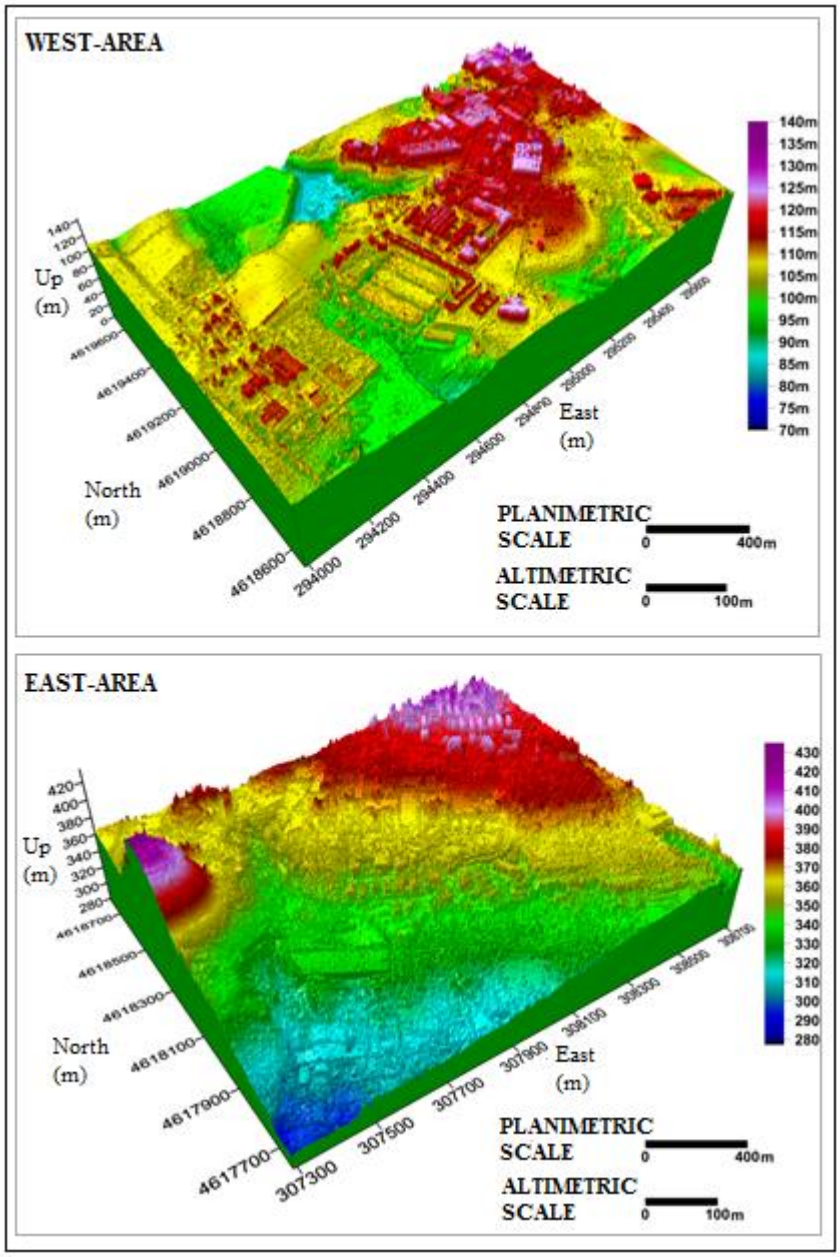


Figure 8.7 The DSMs of the two selected areas
(extracted with the 20-GCP dataset and the rigorous correction model)

Two different procedures were tested; their workflows can be briefly described as follows:

1. The first procedure (hereinafter termed "Procedure-1") was performed by applying the automatic classification algorithm implemented in ERDAS-eATE; the stereo-model oriented with the 20-GCP dataset and the rigorous correction approach was employed. During the DSM extraction, the algorithm allows the automatic detection of points belonging to buildings and vegetation: these points can then be removed from the final

output. The user can adjust the process to his/her specific needs (especially in accordance with the spatial and radiometric available information), by selecting a set of adequate parameters. In case of building extraction procedure, the more significant options are the following ones:

- Slope threshold, i.e. the minimum angle between the object boundary and the adjacent terrain.
- Minimum and maximum object area, i.e. the size of the smallest/largest area to be classified as a building;
- Minimum object height, i.e. the lowest height of an object to be classified as a building.

Since the two selected test-areas show significant differences in terms of type, dimension and distribution of their buildings, a specific parameter setup was adopted for each of them, as shown in Table 8.6. In particular, different values of maximum object area were chosen, in order to better identify the larger industrial buildings of the West-Area and the smaller residential ones of the East-Area. The minimum object height was selected in order to detect possible fences too.

	PARAMETER SETUP: BUILDING CLASSIFICATION			
	Slope threshold (°)	Minimum Area (m^2)	Maximum Area (m^2)	Minimum height (m)
West-Area	40	1	10,000	1
East-Area	40	1	2,000	1

Table 8.6 Parameter setup selected for the two test-areas in the building classification procedure

2. The second procedure (hereinafter termed "Procedure-2") was carried out in two subsequent phases. At first, an automatic classification of the two original images was performed with the software eCognition Developer vs. 8.64.1 by Trimble (eCognition Developer). This is the first object-oriented image analysis software on the market, i.e. its classification algorithms are based on the novel approach described in (Benz et al., 2004). The basic processing units of this so-called "object-oriented image analysis" are not the single image pixels, but homogenous groups of them, termed image objects. The latters are created by image segmentation, that represents the subdivision of an image into separate regions. At first, a segmentation based on primary features (gray tone and shape) is performed starting from the single image pixels. For the present application, this phase was carried out by favoring those pixel groups with the highest homogeneity in terms of shape, and especially regularity. Then a more advanced classification-based segmentation was carried out, by employing both spectral and geometrical information as higher order object features, such as area, shape and compactness of the objects. This segmentation approach (Baatz and Shäpe, 2000) is based on a hierarchical network of image objects: besides to its neighbors, in fact, each object is topologically connected to its sub-objects and super-objects too, forming a strict hierarchical structure. This object-oriented image

analysis offers many advantages (Benz et al., 2004), if compared with the traditional pixel-oriented approach: in particular, the close relation between real-world objects and image objects strongly improves the value of the final classification.

A specific segmentation strategy was selected for each test-area and performed on both images. Many tests have been carried out, in order to identify the parameter setup that correctly fitted with the specific characteristics of each case; in particular "Area", "Compactness", "Border Index" and "Density" features (Trimble, 2011) were mostly employed. At the end of the procedure, the "building" class was masked and the two images were processed with the ERDAS Imagine Suite 2011. In particular, the orientation was performed with ERDAS-LPS by using the 20-GCP dataset and the rigorous correction model. The DSM of the two selected test-areas was finally extracted with ERDAS-eATE.

In order to perform a metrical assessment of the results achieved with the described procedures, two different analyses were carried out. At first, the altitudes reconstructed at the end of the two classification procedures were subtracted from the ones computed by the traditional procedure of DSM generation described in Section 8.3. The DSM generated with the 20-GCP dataset and the rigorous correction approach was employed; furthermore, all compared models were generated with the same spatial resolution, equal to 1 m. Each differentiation, performed in GIS (Geographic Information System) environment, produced a new raster GRID, where each altitude value is the difference between the corresponding values of surface and ground altitudes: in other words, the elevation models delivered by this analysis should point out the buildings that were, correctly or not, identified and removed by the two classification procedures. In particular, the differentiation-derived altitudes (h) were grouped into four ranges:

- $h \leq 0$ m, that represents a qualitative indication of the bare ground;
- $h \in [0; 2]$ m, that represents a qualitative indication of short vegetation and fences;
- $h \in [2; 20]$ m, that represents a qualitative indication of high vegetation and buildings;
- $h > 20$ m, that represents a qualitative indication of towers and tall buildings.

Results are presented in Figures 8.8-9 as follows: for each test-area, the two raster GRID outputs of the differentiation procedures are provided, together with the main mistakes made by the two classification approaches and the mask produced on the North-Image by the object-oriented classification. These two latter results are visualized over the orthophoto of the North-Image.

The qualitative analysis carried out on the West-Area (Figure 8.8) shows a good potentiality of the Procedure-1, that was able to correctly identified most of the buildings. Furthermore, the classification algorithm implemented in eATE provided also a partial elimination of the vegetation. On the contrary, it failed in recognizing those buildings that are not well-textured or have an irregular shape. As pointed out by the applied mask, also the eCognition classification could correctly recognize the buildings, but this information was not adequate

exploited by the subsequent DSM extraction procedure: in the final output, in fact, only some portions of the masked buildings were actually removed.

The qualitative analysis performed on the East-Area (Figure 8.9) shows the same over-mentioned evidences. In this case, however, both procedures encountered some problems in dealing with the small and irregularly-distributed buildings, that are typical of a residential area. A detailed discussion on the results achieved by this qualitative analysis can be found in (Bertacchini et al., 2012).

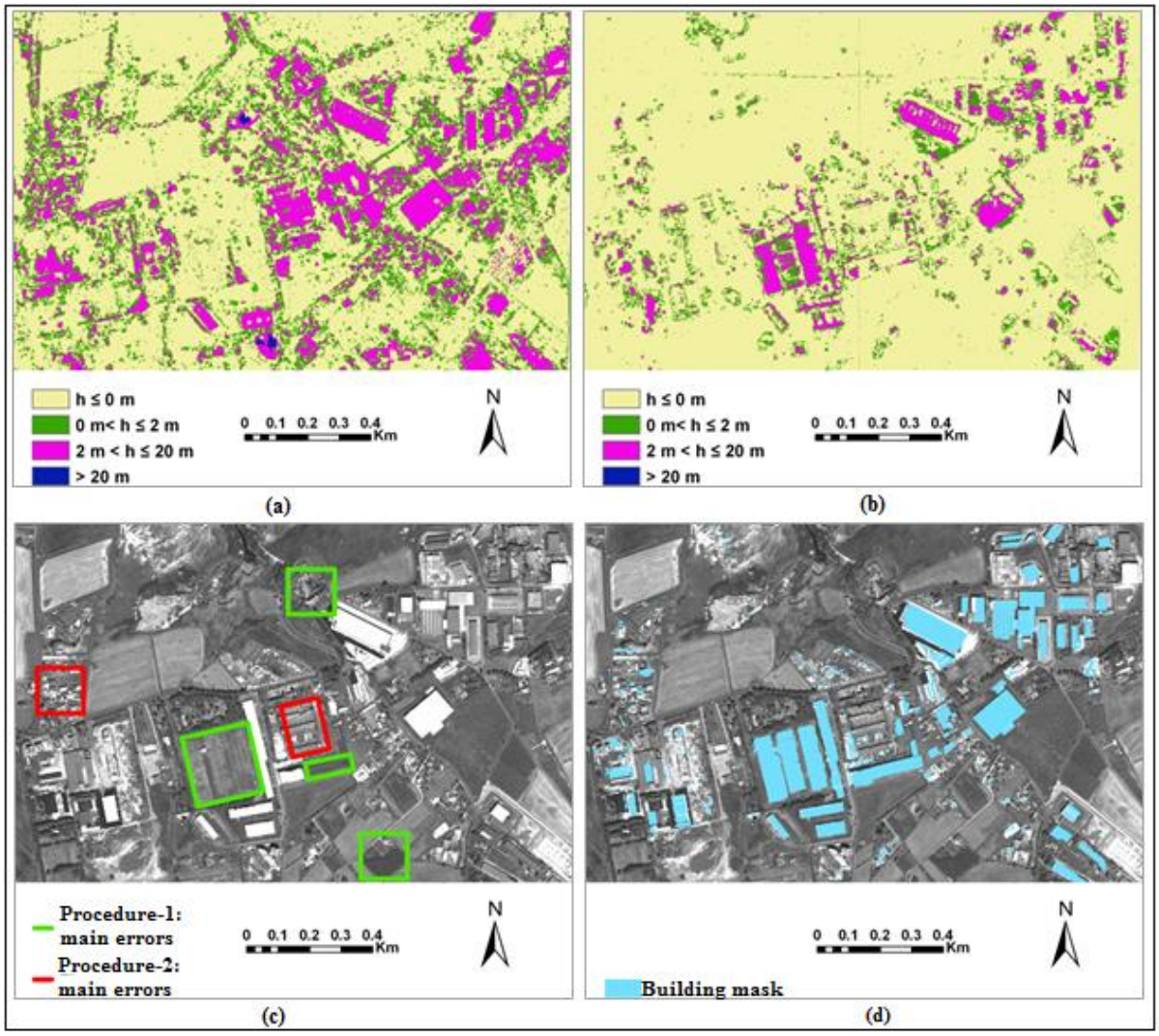


Figure 8.8 Results of the qualitative analysis performed on the West-Area:
(a) Procedure-1, DSM-DTM altitude differentiation;
(b) Procedure 2, DSM-DTM altitude differentiation;
(c) Examples of problems encountered by the two procedures;
(d) The mask generated by the object-oriented procedure on the North-Image

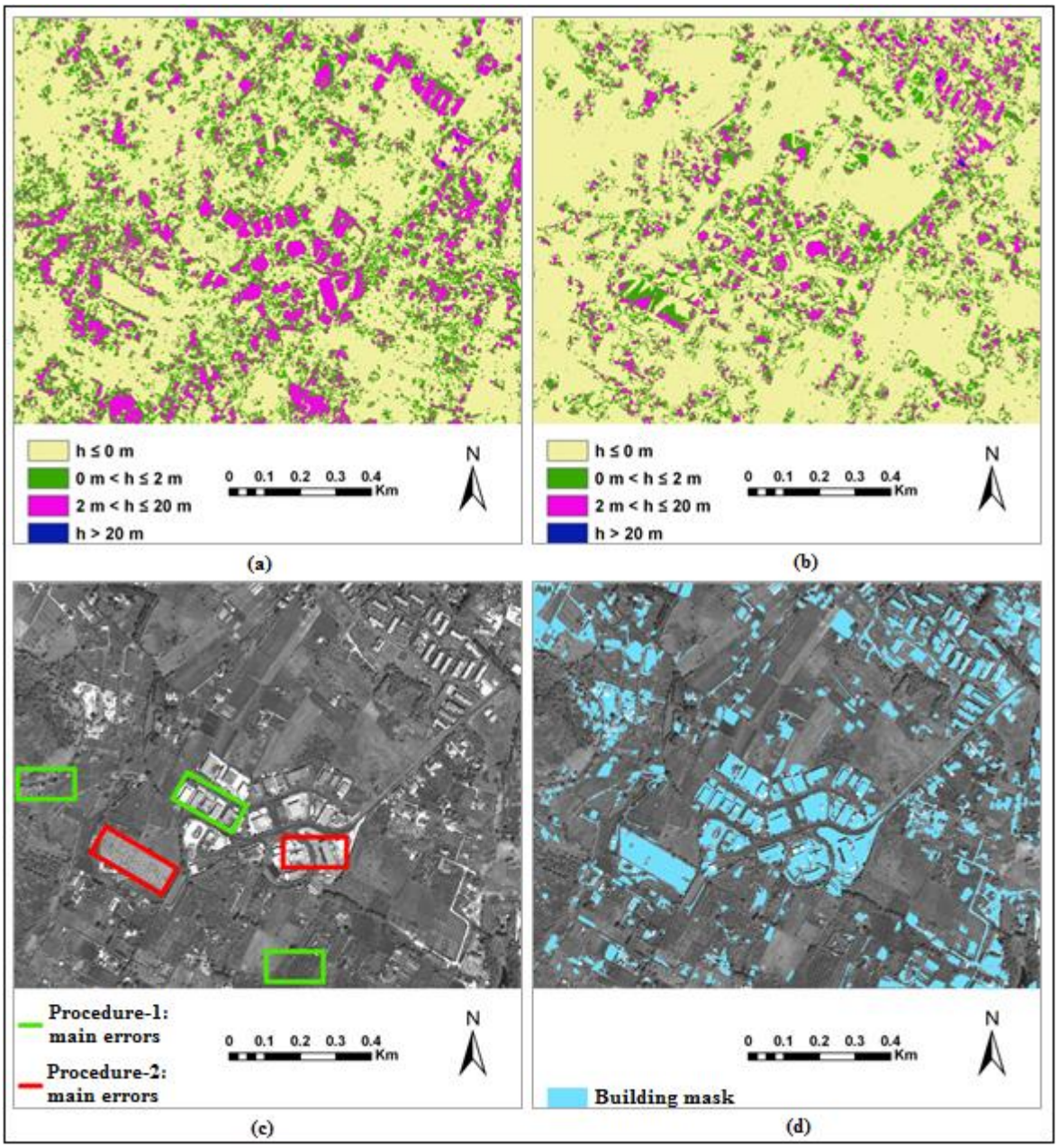


Figure 8.9 Results of the qualitative analysis performed on the East-Area:
(a) Procedure-1, DSM-DTM altitude differentiation;
(b) Procedure 2, DSM-DTM altitude differentiation;
(c) Examples of problems encountered by the two procedures;
(d) The mask generated by the object-oriented procedure on the North-Image

A second type of study was carried out by super-imposing the raster GRID outputs achieved with the two classification procedures on the orthophoto of the North-Image. This analysis was performed with the software ArcGIS by esri (ArcGIS); in particular, its instruments of linear/areal measure together with a manual count were employed in order to identify the following parameters:

- The total amount of buildings presented in the subset of the ortophoto corresponding to the two selected test-areas;
- The number of buildings that were correctly removed by the two procedures (a correctly classified building should have been recognized for at least the 50% of its areal extension);
- The number of "false" buildings removed by the two procedures (an incorrectly classified building should have been recognized for at least the 50% of its areal extension). These errors are mainly localized on field and woody areas.

Table 8.7 lists the results achieved within this second analysis.

	WEST-AREA		**EAST-AREA**	
	Correct Buildings (No.)	False Buildings (No.)	Correct Buildings (No.)	False Buildings (No.)
Ortophoto	45	-	76	-
Procedure-1	35	11	42	15
Procedure-2	23	4	38	18

Table 8.7 The number of correctly and incorrectly recognized buildings

This study confirms the evidences pointed out within the previous qualitative assessment. The automatic classification algorithm implemented in ERDAS-eATE is able to correctly identify a greater number of buildings: about the 78% and 55% of the total amount of buildings were in fact detected in, correspondingly, the West-Area and the East-Area. As afore-mentioned, Procedure-1 provides also the elimination of some vegetation, such as the large woody areas: these classifications, here counted within the "false" building group, can also represent an advantage if the complete bare ground should be extracted from panchromatic image data (i.e., if no radiometric information can be used therefor). On the contrary, Procedure-2 is less accurate in defining the correct geometry of the buildings and is thus able to identify a smaller number of them: about the 51% and 50% of the total amount of buildings were in fact correctly detected in, correspondingly, the West-Area and the East-Area. These problems seem to be introduced by the final phase of the procedure, i.e. the DSM extraction from the previously masked images, since the eCognition outputs are largely correct. A detailed discussion on the results achieved by this quantitative analysis can be found in (Bertacchini et al., 2012).

CONCLUSION

The work described in this research thesis has resulted in a metrological approach, that attempts to fill the void created by a lack of internationally recognized standards in the field of 3D imaging. In particular, the assessment tests performed in the three years of PhD studies have produced a significant amount of comparative data and validation evidences that may support the young and growing sector of automated procedures for image-based 3D modelling. In order to sustain this market growth, in fact, the user confidence should be increased, especially by fighting the risk of offering methods that produce too much data, but only little information: since "a measurement result is meaningful only if its uncertainty is known" (Sir William Thomson, Lord Kelvin, 1824-1907), the importance of understanding how a 3D system recovers physical information is a critical factor that should be coped with. The work here presented tries to contribute to this process, by analysing the metrological aspects of the problem and by proposing a possible solution therefor, especially in terms of inter-comparison or fit-to-purpose integration between dissimilar technologies. Results thereby achieved have not a general validity, since they are, of course, influenced by the specific operative conditions affecting each case study, in terms, for example, of datasets, hardware/software means, ambient and operators. Nevertheless, these studies offer a possible reference procedural workflow in order to evaluate the accuracy, resolution, repeatability and measurement uncertainty of novel 3D imaging techniques for the automatic generation of textured dense 3D point clouds from a set of un-oriented images.

In particular, Chapter 5 discussed the importance of performing an accurate digital camera calibration, especially in high-accuracy close-range measurement applications. An automated self-calibrating approach turns out to be comparable to the classical stand-alone photogrammetric calibration procedure, only if the recovery of camera parameters is supported by an adequate geometry of the image network.

Chapter 6, then, addressed the problem of 3D modelling form terrestrial imagery, by discussing results achieved in three different applications, performed both in outdoor environments and in environmentally controlled facilities. Each step of the photogrammetric and computer vision-based procedural pipeline was metrically investigated, by focusing the attention on a recently developed open-source suite of tools (Apero/MicMac); in particular, the influence of the employed image acquisition protocol was deepen studied, by examining its effects on the orientation and dense matching phases. Results show, at several spatial scales, the impressive and significant metric potentialities of the tested automated procedures in dealing with different 3D scene and boundary conditions; the current limitations of these methods are pointed out too, especially in terms of problematic surface textures, sharp surface gradients and cast shadows. The image acquisition layout is identified as a critical issue, affecting the overall measurement accuracy: although it actually depends upon several factors, it can be favoured by adopting image convergence angles close to 10°, corresponding to a reasonable base-to-depth ratio. The use of multiple-lens configurations can be successfully managed too, especially in case of complex and large 3D scene. Furthermore, the possibility

of finely controlling each processing phase through a huge amount of attributes and parameters can be successfully exploited if a specific analysis on the role played by each procedural choice within the several algorithmic solutions is performed.

In Chapter 7 the use of UAVs as photogrammetric acquisition platforms was discussed, especially as part of the multi-sensor data fusion research topic. The case study described in the chapter shows that UAV systems, purpose-fitting and metrically calibrated, can be efficiently used in order to integrate surveys performed from classical terrestrial platforms in those applications where the use of a single technique cannot provide a complete and detailed 3D model (e.g. for complex and large architectural objects). The information augmentation that can be thereby achieved, in fact, extends the spatial coverage of each system and avoids thereby the use of expensive and cumbersome equipment. Of course, the issue of uncertainty management should be adequately dealt with, especially if part of the 3D scene is acquired by both different sensors and platforms. In this case, as the described experiment proves, the lowest global uncertainty, that justifies the use of a multi-sensor solution, is reached only if one can manage all the uncertainties associated to each method. Comparative tests show that the use of external reference data would support this process; nevertheless, it can be anyway successfully performed even where the external conditions prevent the availability of a proper object space control field.

Finally, the main steps that constitute the procedure of DTM and DSM extraction from spaceborne imagery were metrically evaluated in Chapter 8. This application was carried out with a commercial software package (ERDAS Imagine Suite 2011) and compared the accuracies achievable by employing two different image correction formulations, i.e. a rigorous 3D parametric model and a RPC-based non parametric one. Results show that DSMs generated from WorldView-1 stereo images with RPC-correction model can achieve a purpose-fitting accuracy level, even if a minimum amount of measured GCPs is available: the RPC data provided by the latest generation sensors, in fact, allow now a metric performance that is comparable with the one delivered by the rigorous correction model. Finally, the possibility of detecting the building information and removing it from the digital 3D output was analysed and successfully realized through two different procedures. Both approaches encounter some problems in dealing with small and irregularly-distributed buildings, that are typical of residential areas; nevertheless, they show good potentialities and a significant level of automation.

Waiting for the establishment of internationally recognized standards and guidelines, all these studies provide a metrological context for the accuracy assessment of image-based 3D modelling techniques. The proposed approach can be further applied to different case studies in order to define specifically-adoptable best practices and help the scientific community to set up *ad-hoc* methodologies for a traceable evaluation of 3D imaging systems. Such methodologies, as shown by this PhD research, should adequately addressed the following three metrological issues (VIM3):

- The definition of the quantity intended to be measured (the measurand issue in metrology). In particular, one should identify what the measurand is in a given comparison;
- The definition of an operation designed to evaluate if you can trust the measurement itself (the calibration issue in metrology). In particular, one should establish how calibration is performed and how often;
- The definition of a traceable measurement, i.e. a measurement result that can be connected to a reference through documented unbroken chain of comparisons (the traceability issue in metrology). In particular, one should evaluate what the traceability route is and how the measurement uncertainty of a system can be assessed.

These three aspects of metrological nature should be further developed in the future, offering great opportunities for research and development for academia, national measurement institutes and industry. In particular, starting from the results presented here, it would be desirable for future projects to develop an adequate understanding of the uncertainty components related to each measurement, in order to evaluate 3D imaging systems by relying on metrological inter-comparisons of results from dissimilar instruments. The error budget calculation will be, thus, of paramount interest for the success of the projects and will become a crucial step in a metrological approach.

Finally, the creation of new benchmark projects, that measure the performance of state-of-the-art algorithms through shared datasets and platforms, should be favoured in the future too. These actions, in fact, will further support the definition of standards and guidelines that are critical for market growth and contribute to the development of efficient working groups in collaboration with both photogrammetric and computer vision communities.

REFERENCES

3D Systems. Rapidform XOR - User Guide.

Abate, D., Furinia, G., Migliori, S., Pierattini, S., 2010. Project Photofly: New 3D Modeling Online Web Service (Case Studies and Assessments). *ENEA Research Centre, UTICT, Bologna, Italy*.

Abdel-Aziz, Y. I., Karara, H. M., 1971. Direct linear transformation into object space coordinates in close-range photogrammetry. In: *Proc. Symposium on Close-Range Photogrammetry, Urbana, Illinois*, pp. 1-18.

Ahmadabadian, A. H., Robson, S., Boehm, J., Shortis, M., Wenzel, K., Fritsch, D., 2013. A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. *ISPRS Journal of Photogrammetry and Remote Sensing*, *78*, pp. 157-167.

Akbarzadeh, A., Frahm, J. M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., Nistér, D., Pollefeys, M., 2006. Towards urban 3D reconstruction from video. In: *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pp. 1-8. IEEE.

Alby, E., Smigiel, E., Assali, P., Grussenmeyer, P., Kauffmann-Smigiel, I., 2009. Low cost solutions for dense point clouds of small objects: photomodeler scanner vs. david laserscanner. In: *22nd CIPA Symposium*, Kyoto, Japan.

Alshawabkeh, Y., Haala, N., 2004. Integration of digital photogrammetry and laser scanning for heritage documentation. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXV, Part. B5, pp. 537-546.

Amann, M. C., Bosch, T., Myllyla, R., Rioux, M., Lescure, M., 2001. Laser ranging: a critical review of usual techniques for distance measurement. *Optical Engineering*, *40*(1), pp. 10-19.

ANSI/ASME B89.4.19. 2006. Performance Evaluation of Laser-Based Spherical Coordinate Measurement Systems.

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., Wu, A. Y., 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, *45*(6), pp. 891-923.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation - an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*. Wichmann Verlag, Karlsruhe, pp. 12–23.

Baiocchi, V., Dominici, D., Mormile, M., 2013. UAV application in post-seismic environment, In: *Proceedings of UAV-g 2013*.

Baiocchi. V., Catalano, R,. Piano, M., 2007. Utilizzo di immagini satellitari ad alta risoluzione quale supporto al governo del territorio. *Atti della 11[a] Conferenza nazionale ASITA*, Torino, Italia.

Bandiera, A., Beraldin, J.-A., Gaiani, M., 2011. Birth and use of 3D imaging, modeling and visualization digital techniques for architecture and cultural heritage applications: a short history, *Annale di analisi grafica e storia della rappresentazione*, Ed. Lombardi, pp. 81-170.

Baltsavias, E. P. ,1999. A comparison between photogrammetry and laser scanning. *ISPRS Journal of photogrammetry and Remote Sensing*, *54*(2), pp. 83-94.

Barazzetti, L., Scaioni, M., Remondino, F., 2010. Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. *The Photogrammetric Record*, *25*(132), pp. 356-381.

Barfoot, T., Se, S., Jasiobedzki, P., 2006. Vision-based localization and terrain modelling for planetary rovers. In: Howard, A., Tunstel, E. (eds.) *Intelligence for Space Robotics*, pp. 71–92. TSI Press, Albuquerque.

Baribeau, R., Rioux, M., 1991. Influence of speckle on laser range finders. *Applied Optics*, 30(20), pp. 2873-2878.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: *Computer Vision–ECCV 2006*, pp. 404-417. Springer Berlin Heidelberg.

Benedetti B., Gaiani M., Remondino F. (eds.), 2010. Mesurés, dessinés et décrits avec la plus grande exactitude. Una metodologia per l'acquisizione e la restituzione di siti archeologici complessi ai fini della costruzione di sistemi informativi basati su modelli digitali 3D. Il caso dell'area archeologica di Pompei, Scuola Normale Superiore di Pisa, Pisa, Italy.

Benz, U.C., Hofmann, P., Willhauck., G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 58(3-4), pp. 239-258

Beraldin, J. –A, Blais, F., Cournoyer, L., Rioux, M., El-Hakim, S.H., Rodella, R., Bernier, F., Harrison, N., 1999. Digital 3D imaging system for rapid response on remote sites. In: *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, pp. 34-43. IEEE.

Beraldin, J. A., Guidi, G., Ciofi, S., Atzeni, C., 2002. Improvement of metric accuracy of digital 3D models through digital photogrammetry. a case study: Donatello's Maddalena. *International Symposium on 3D Data Processing Visualization and Transmission*, Padova, Italy, pp. 758-761.

Beraldin, J.-A., Picard, M., El-Hakim, S. F., Godin, G., Latouche, C., Valzano, V. and Bandiera, A.,2002. Exploring a Byzantine crypt through a high-resolution texture mapped 3D model: combining range data and photogrammetry. *Proceedings of the CIPA WG6 International Workshop on Scanning for Cultural Heritage Recording*, Corfu, Greece, pp. 65–72.

Beraldin, J. A., Picard, M., El-Hakim, S., Godin, G., Latouche, C., Valzano, V., Bandiera, A. , 2002. Exploring a Byzantine crypt through a high-resolution texture mapped 3D model: combining range data and photogrammetry. In: *Proceedings of International Workshop on Scanning for Cultural Heritage Recording - Complementing or Replacing Photogrammetry - Commission V - Symposium ISPRS 2002* – Corfu (Grecia), pp. 65-70.

Beraldin, J-A, 2004. Integration of Laser Scanning and Close-Range Photogrammetry – The Last Decade and Beyond. In: *Proceedings of the XXth ISPRS Congress*, Istanbul, Turkey.

Beraldin, J. A., Picard, M., El-Hakim, S. F., Godin, G., Valzano, V., Bandiera, A., 2005. Combining 3D technologies for cultural heritage interpretation and entertainment. In: *Proc. SPIE*, Vol. 5665, pp. 108-118.

Beraldin, J. A., Rioux, M., Cournoyer, L., Blais, F., Picard, M., Pekelsky, J., 2007. Traceable 3D imaging metrology. In: *Electronic Imaging 2007*, pp. 64910B-64910B. International Society for Optics and Photonics.

Beraldin, J. A., 2009. Basic theory on surface measurement uncertainty of 3D imaging systems. In *IS&T/SPIE Electronic Imaging*, pp. 723902-723902. International Society for Optics and Photonics.

Beraldin, J. A., Cournoyer, L., Picard, M., Blais, F., 2009. Proposed procedure for a distance protocol in support of ASTM-E57 standards activities on 3D imaging. In: *IS&T/SPIE Electronic Imaging*, pp. 72390S-72390S. International Society for Optics and Photonics.

Beraldin, J.A., Picard, M., Valzano, V., Bandiera, A., Negro, F., 2011. Best practices for the 3D documentation of the Grotta dei Cervi of Porto Badisco, Italy. In: *IS&T/SPIE Electronic Imaging*, pp. 78640J-78640J. International Society for Optics and Photonics.

Bergen, J. R., Anandan, P., Hanna, K. J., Hingorani, R., 1992. Hierarchical model-based motion estimation. In: *Computer Vision—ECCV'92*, pp. 237-252. Springer Berlin Heidelberg.

Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., Taubin, G., 2002. Building a digital model of Michelangelo's Florentine Pieta. *Computer Graphics and Applications, IEEE*, *22*(1), pp. 59-67.

Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., Taubin, G., 2002. Building a digital model of Michelangelo's Florentine Pieta. *Computer Graphics and Applications, IEEE*, *22*(1), pp. 59-67.

Bertacchini, E., Toschi, I., Rivola, R., Castagnetti, C., Capra, A., 2012. Estrazione di DTM da una stereocoppia WorldView-1: procedure di orientamento, image-matching e rimozione degli edifici. *Bollettino SIFET 2-2012*, pp. 49-68, ISSN: 1721-971X.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-D shapes. In: *Robotics-DL tentative*, pp. 586-606. International Society for Optics and Photonics.

Blais, F., 2004. Review of 20 years of range sensors development. *Journal of Electronic Imaging*, 13(1), pp. 231-243.

Blais, F., Beraldin, J. A., 2006. Recent developments in 3D multi-modal laser imaging applied to cultural heritage. *Machine Vision and Applications*, *17*(6), pp. 395-409.

Böhler, W., Marbs, A., 2004. 3D scanning and photogrammetry for heritage recording: a comparison. In: *Proceedings of the XIIth International Conference on Geoinformatics*, Gävle, Sweden, pp. 291-198.

Böheler, W., 2005. Comparison of 3D laser scanning and other 3D measurement techniques. *Recording, Modeling and Visualization of Cultural Heritage, London, Taylor and Francis*, pp. 89-100.

Borg, C. E., Cannataci, J. A., 2002. Thealasermetry: a hybrid approach to documentation of sites and artefacts. In: *Proceedings of the CIPA WG6 International Workshop on Scanning for Cultural Heritage Recording*, pp. 93-104.

Brown, D. C., 1966. Decentering distortion of lenses. *Photometric Engineering*, *32*(3), pp. 444-462.

Brown, D.C., 1971, Close-range camera calibration. *PE&RS*, Vol. 37(8), pp.855-866.

Brown, M. Z., Burschka, D., Hager, G. D., 2003. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(8), pp. 993-1008.

Bruschweiler, W., Braun, M., Dirnhofer R., 2003. Analysis of patterned injuries and injury-causing instruments with forensic 3D/CAD supported photogrammetry (FPHG): an instruction manual for the documentation process. *Forensic Sci. Int.*, *132*, pp. 130-138.

Cheng, P., Chaapel, C., 2008. Automatic DEM Generation – Using WorldView-1 Stereo Data with or without Ground Control points. *GeoInformatics*, 11(7), pp. 34-39.

Cheok, G. S., Lytle, A. M., Saidi, K. S., 2008. ASTM E57 3D imaging systems committee: an update. In: *SPIE Defense and Security Symposium*, pp. 69500J-69500J. International Society for Optics and Photonics.

Cignoni, P., Scopigno, R., 2008. Sampled 3D models for CH applications: A viable and enabling new medium or just a technological exercise?. *Journal on Computing and Cultural Heritage (JOCCH)*, *1*(1), 2.

Cilloccu, F., Dequal., S., Brovelli, M.A., Crespi, M., Lingua, A., 2009. *Ortoimmagini 1:10.000 e Modelli Altimetrici - Linee Guida*. CISIS, Roma, Italia.

CIPA Working Group 6 & ISPRS Commission V. 2002. Corfu, Greece, Sept. 1-2.

Coxeter, H. S. M., 2003. *Projective geometry*. Springer.

Crespi, M., Colosimo, G., Fratarcangeli, F., Jacobsen, K., Pieralice, F., 2009. Valutazione dell'accuratezza del DSM estratto da una stereo coppia WorldView-1. *Atti della 13ª Conferenza nazionale ASITA*, Bari, Italia.

Cronk, S., Fraser, C., Hanley, H., 2006. Automated metric calibration of colour digital cameras. *The photogrammetric record*, *21*(116), pp. 355-372.

Daultrey, S., 1976. Principal components analysis. *Concepts and Techniques in Modern Geography*, Geo Abstracts Ltd., University of East Anglia, Norvich, UK, Vol. 8, pp. 1-51.

De Luca, L., Busayarat, C., Stefani, C., Véron, P., Florenzano, M., 2011. A semantic-based platform for the digital analysis of architectural heritage. *Computers & Graphics, 35*(2), pp. 227-241.

Delaunay, B., 1934. *Sur la sphère vide* . Bull. Acad. Science USSR VII: Class. Sci. Mat. Nat. 7:793-800.

Deseilligny, M. P., Clery, I., 2011. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images. In: *Proceedings of the ISPRS Symposium, 3DARCH11*, pp. 269-277.

Dorsch, R. G., Häusler, G., Herrmann, J. M., 1994. Laser triangulation: fundamental uncertainty in distance measurement. *Applied Optics*, 33(7), pp. 1306-1314.

Drouin, M. A., Beraldin, J. A., 2012. Active 3D Imaging Systems. In *3D Imaging, Analysis and Applications*, pp. 95-138. Springer London.

Ducke, B., Score, D., Reeves, J., 2011. Multiview 3D reconstruction of the archaeological site at Weymouth from image series. *Computers & Graphics, 35*(2), pp. 375-382.

Dyer, C. R., 2001. Volumetric scene reconstruction from multiple views. In: *Foundations of image understanding*, pp. 469-489. Springer US.

Eisenbeiss, H., 2008. UAV photogrammetry in plant sciences and geology, In: *6th ARIDA Workshop on "Innovations in 3D Measurement, Modeling and Visualization*, Povo (Trento), Italy.

Eisenbeiss, H., 2009. *UAV photogrammetry*. PhD Thesis, Institute of geodesy and Photogrammetry, ETH, Zurich, Switzerland.

El-Hakim, S. F., Beraldin, J. A., 1994. Integration of range and intensity data to improve vision-based three-dimensional measurements. In: *Photonics for Industrial Applications*, pp. 306-321. International Society for Optics and Photonics.

El-Hakim, S. F., Beraldin, J. A., 1995. Configuration design for sensor integration. *Proceedings of SPIE Videometrics IV*, Philadelphia, Pennsylvania, Vol. 2598, pp. 274–285.

El-Hakim, S., Beraldin, J. A., Blais, F., 2003. Critical factors and configurations for practical 3D image-based modeling. In: *Proceedings of VI Conference on Optical 3D Measurement Techniques*, Zurich, Switzerland, Vol. 2, pp. 159–167.

El-Hakim, S. F., Beraldin, J. A., Picard, M., Godin, G., 2004. Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. *Computer Graphics and Applications, IEEE, 24*(3), pp. 21-29.

El-Hakim, S., Beraldin, J. A., Remondino, F., Picard, M., Cournoyer, L., Baltsavias, E., 2008. Using terrestrial laser scanning and digital images for the 3D modelling of the Erechteion, Acropolis of Athens. In: *DMACH Conference Proceedings, Amman, Jordan*, pp. 3-16.

Everaerts, J., 2008. The Use of Unmanned Aerial Vehicles (UAVS) for Remote Sensing and Mapping, In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII. Part B1, pp.1187-1192, ISPRS Congress, Beijing, China.

Faugeras, O., 1993. *Three-dimensional computer vision: a geometric viewpoint*. MIT press.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), pp. 381-395.

Flack, P. A., Willmott, J., Browne, S. P., Arnold, D. B., Day, A. M., 2001. Scene assembly for large scale urban reconstructions. In: *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage*, pp. 227-234. ACM.

Flack, D., Hannaford, J., 2005. Measurement Good Practice Guide No. 80: Fundamental Good Practice in Dimensional Metrology. *Measurement Good Practice Guide*, National Physical Laboratory, NPL, United Kingdom.

Förstner, W., Bonn, U., 2009. Computer Vision and Remote Sensing-Lessons Learned. In: Fritsch, Dieter (Hg.)*: Photogrammetric Week*, pp. 241-249.

Fraser, C. S., Shortis, M. R., 1995. Metric exploitation of still video imagery. *The Photogrammetric Record*, *15*(85), pp. 107-122.

Fraser, C. S., 2001. Photogrammetric camera component calibration: A review of analytical techniques. In: Grün, A., Huang, T.S. (Eds.), *Calibration and Orientation of Cameras in Computer Vision*, pp. 95-121. Springer Berlin Heidelberg.

Fraser, C.S., Baltsavias, E., Gruen, A., 2002. Processing of *Ikonos* imagery for submetre 3D positioning and building extraction. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 56(3), pp. 177-194.

Fraser, C.S., Hanley, H.B., 2003. Bias compensation in rational functions for IKONOS satellite imagery. *Photogrammetric Engineering and Remote Sensing*, 69(1), pp 53-57.

Fraser, C. S., Hanley, H. B., 2004. Developments in close-range photogrammetry for 3D modelling: the iWitness example. In: *Proc. of Processing and Visualization using High-Resolution Imagery,* Pitsanulok, Thailand.

Fraser, C.S., Hanley, H.B., 2005. Bias-compensated RPCs for Sensor Orientation of High-resolution Satellite Imagery. *Photogrammetric Engineering and Remote Sensing*, 71(8), pp. 909-915.

Fraser, C. S., Al-Ajlouni, S., 2006. Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric engineering and remote sensing*, *72*(9), pp. 1017-1026.

Fraser, C. S., Cronk, S., Stamatopoulos, C., 2012. Implementation of zoom-dependent camera calibration in close-range photogrammetry. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX, Part. B5, pp. 15-19.

Fryer, J. G., Brown, D. C., 1986. Lens distortion for close-range photogrammetry. *Photogrammetric engineering and remote sensing*, *52*(1), pp. 51-58.

Fryer, J., 1996. Camera Calibration. In: *Close-range Photogrammetry and Machine Vision*, Atkinson (Ed.), Whittles Publishing, UK, pp.156-179.

Furukawa, Y.,Ponce, J., 2006. High-fidelity image-based modeling. Technical Report UIUC.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *32*(8), pp. 1362-1376.

Gabet, L., Giraudon, G., & Renouard, L., 1997. Automatic generation of high resolution urban zone digital elevation models. *ISPRS journal of photogrammetry and remote sensing*, *52*(1), pp. 33-47.

Galiatsatos, N., Donoghue, D. N., Philip, G., 2008. High resolution elevation data derived from stereoscopic CORONA imagery with minimal ground control: an approach using Ikonos and SRTM data. *Photogrammetric engineering and remote sensing.*, *74*(9), pp. 1093-1106.

Gårding, J., 1992. Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, *2*(4), pp. 327-350.

Georgantas, A., Brédif, M., Pierrot-Deseilligny, M., 2012. An accuracy assessment of automated photogrammetric techniques for 3D modelling of complex interiors. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. XXXIX, Part. B3, pp. 23-28.

Godin, G., Beraldin, J.-A., Rioux, M., Levoy, M., Cournoyer, L., Blais, F., 2001. An Assessment of Laser Range Measurement of Marble Surfaces, In: *Proc. Fifth Conference on optical 3-D measurement techniques*, Vienna University of Technology, Vienna, Austria.

Godin, G., Beraldin, J.-A., Taylor, J., Cournoyer, L., Rioux, M., El-Hakim, S., Baribeau, R., Blais, F., Boulanger, P., Domey, J., Picard, M., 2002. Active Optical 3D Imaging for Heritage Applications. *Computer Graphics and Applications*, 22(5), pp. 24-35.

Goesele, M., Curless, B., Seitz, S. M., 2006. Multi-view stereo revisited. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, pp. 2402-2409. IEEE.

Golub, G. H., Van Loan, C. F., 1996. Matrix computations. *Johns Hopkins University, Press, Baltimore, MD, USA*, pp. 374-426.

Grenzdörffer, G. J., Engel, A., Teichert, B., 2008. The photogrammetric potential of low-cost UAVs in forestry and agriculture. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXI, Part. B3, pp.1207-1214.

Grodecki, J., Dial, G., 2003. Block Adjustment of high resolution satellite images described by rational functions. *Photogrammetric Engineering and Remote Sensing*, 69(1), pp 59-68.

Grün, A., Beyer, H. A., 2001. System calibration through self-calibration. In: *Calibration and Orientation of Cameras in Computer Vision*, pp. 163-193. Springer Berlin Heidelberg.

Grün, A., Remondino, F., Zhang, L., 2004. Photogrammetric reconstruction of the great Buddha of Bamiyan, Afghanistan. *The Photogrammetric Record*, *19*(107), pp. 177-199.

Grussenmeyer, P., Landes, T., Voegtle, T., Ringle, K., 2008. Comparison methods of terrestrial laser scanning, photogrammetry and tacheometry data for recording of cultural heritage buildings. In: *ISPRS Congress Proceedings, Beijing*, pp. 213-18.

Guidi, G., Remondino, F., Russo, M., Menna, F., Rizzi, A., Ercoli, S., 2009. A multi-resolution methodology for the 3D modeling of large and complex archeological areas. *International Journal of Architectural Computing*, *7*(1), pp. 39-55.

Guidi, G., Remondino, F., Russo, M., Menna, F., Rizzi, A., Ercoli, S., 2009. A multi-resolution methodology for the 3D modeling of large and complex archeological areas. *International Journal of Architectural Computing*, *7*(1), pp.39-55.

Guidi, G., Russo, M., Beraldin, J. A., 2010. *Acquisizione 3D e modellazione poligonale*. McGraw-Hill.

Gülch, E., 1991. Results of test on image matching of ISPRS WGIII/4. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 46(1), pp. 1-8.

Haala, N., 2013. The Landscape of Dense Image Matching Algorithms. In: *Proc. Photogrammetric Week 2013*. Dieter Fritsch (Ed.), Stuttgart, pp. 271-284.

Hall, D. L., Llinas, J., 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*, *85*(1), pp. 6-23.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: *Alvey vision conference*, pp. 147-151.

Hartley, R. I., Mundy, J. L., 1993. Relationship between photogrammmetry and computer vision. In: *Optical Engineering and Photonics in Aerospace Sensing*, pp. 92-105. International Society for Optics and Photonics.

Hartley, R. I., 1997. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *19*(6), pp. 580-593.

Hartley, R. I., 1999. Theory and practice of projective rectification. *International Journal of Computer Vision*, *35*(2), pp. 115-127.

Hartley, R., Zisserman, A., 2004.. *Multiple view geometry in computer vision*. Cambridge university press.

Hashimoto, T., 2000. DEM generation from stereo AVNIR images. *Advances in Space Research*, *25*(5), pp. 931-936.

Healey, G., Binford, T. O., 1988. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, *42*(1), pp. 62-86.

Heikkila, J., Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 1106-1112. IEEE.

Hernández Esteban, C., Schmitt, F., 2004. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, *96*(3), pp. 367-392.

Heuchel, T., Köstli, A., Lemaire, C., Wild, D., 2011. Towards a next level of quality DSM/DTM extraction with Match-T. In: *Proc. Photogrammetric Week '11*, Vol. 11, pp. 197-202.

Hiebert, K. L., 1981. An evaluation of mathematical software that solves nonlinear least squares problems. *ACM Transactions on Mathematical Software (TOMS)*, *7*(1), pp. 1-16.

Hiep, V. H., Keriven, R., Labatut, P., Pons, J. P., 2009. Towards high-resolution large-scale multi-view stereo. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1430-1437. IEEE.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, pp. 807-814. IEEE.

Hong, L., 1999. Sense your world better: multisensor/information fusion. *IEEE Journal of Circuits and Systems*, *10*(3), pp. 7-8.

Horn, B. K. P., Brooks, M. J., 1989. Shape from shading. Cambridge, MA: MIT Press.

Hutchinson, M. F., 1993. Development of a continent-wide DEM with applications to terrain and climate analysis. *Environmental modeling with GIS*, pp. 392-399.

Ikeuchi, K., 2001. Modeling from reality. In: *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pp. 117-124. IEEE.

Ikeuchi, K., Miyazaki, D., 2007. Digitally Archiving Cultural Heritage. New York: Springer.

Irschara, A., Kaufmann, V., Klopschitz, M., Bischof, H., Leberl, F., 2010. Towards fully automatic photogrammetric reconstruction using digital images taken from UAVs. In: *Proceedings of the ISPRS TC VII Symposium—100 Years ISPRS*.

Jähne, B., Hauβecker, H., Geiβler, P., 1999. *Handbook of Computer Vision and Application*. Volume 2. Academic Publishers.

Jantos, R., Luhmann, T., Peipe, P., Schneider, C. T., 2002. Photogrammetric performance evaluation of the Kodak DCS Pro Back. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part. 5, pp. 42-47.

Kadobayashi, R., Kochi, N., Otani, H., Furukawa, R., 2004. Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 35, Part. 5, pp. 401-406.

Kasser, M., Egels, Y., 2002. *Digital photogrammetry*. Taylor & Francis.

Kender, J.R., 1981. Shape from Texture. Technical Report CMU-CS-81-102, Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA.

Koch, R., Pollefeys, M., Van Gool, L., 1998. Multi Viewpoint Stereo from Uncalibrated Video Sequences. *Proc. European Conference on Computer Vision*, pp. 55-71, Freiburg, Germany.

Kolmogorov, V., Zabih, R., 2002. Multi-camera scene reconstruction via graph cuts. In: *Computer Vision - ECCV 2002*, pp. 82-96. Springer Berlin Heidelberg.

Kraus, K., 1994. Photogrammetrie. *Dümmler, Bonn*.

Kraus, K., 1997, Photogrammetry, Dümmler Verlag, Vol. 2, Bonn, Germany.

Kutulakos, K. N., Seitz, S. M., 2000. A theory of shape by space carving. *International Journal of Computer Vision*, *38*(3), pp. 199-218.

Läbe, T., Förstner, W., 2004. Geometric stability of low-cost digital consumer cameras. In: *Proceedings of the 20th ISPRS Congress, Istanbul, Turkey*, pp. 528-535.

Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2008. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 63(3), pp. 365-381.

Lambers, K., Eisenbeiss, H., Sauerbier, M., Kupferschmidt, D., Gaisecker, Th., Sotoodeh, S., Hanusch, Th., 2007. Combining photogrammetry and laser scanning for the recording and modelling of the late intermediate period site of Pinchango Alto, Palpa, Peru. *Journal of Archaeological Science*, 34(10), pp. 1702-1712.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D., 2000. The digital Michelangelo project: 3D scanning of large statues. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 131-144. ACM Press/Addison-Wesley Publishing Co..

Little, C., Small, D., Carlson, J., 1999. 3D Imaging and Modeling for Crime Scene Documentation. In: *Proc. ICIP99*, Kobe, Japan, pp. 27-34.

Lohr, U., 1998. Laserscan DEM for various applications. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, pp. 353-356.

Longuet-Higgins, H. C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature, 293*, pp. 133-135.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2, pp. 1150-1157. IEEE.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), pp. 91-110.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), pp. 91-110.

Ma, Y., Soatto, S., Kosecka, J., Sastry, S., 2003. An Invitation to 3D Vision: From Images to geometric Models, Springer Verlag.

Maimone, M., Biesiadecki, J., Tunstel, E., Cheng, Y., Leger, C., 2006. Surface navigation and mobility intelligence on the Mars Exploration Rovers. In: Howard, A., Tunstel, E. (eds.) *Intelligence for Space Robotics*, pp. 45–69. TSI Press, Albuquerque.

Maimone, M., Cheng, Y., Matthies, L., 2007. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, *24*(3), pp. 169-186.

Mallon, J., & Whelan, P. F., 2005. Projective rectification from the fundamental matrix. *Image and Vision Computing*, *23*(7), pp. 643-650.

Mancini, F., Dubbini, M., Gattelli, M., Stecchi, F., Fabbri, S., & Gabbianelli, G., 2013. Using Unmanned Aerial Vehicles (UAV) for High-Resolution Reconstruction of Topography: The Structure from Motion Approach on Coastal Environments. *Remote Sensing*, *5*(12).

Martin-Beaumont, N., Nony, N., Deshayes, B., Pierrot-Deseilligny, M., De Luca, L., 2013. Photographer-friendly work-flows for image-based modelling of heritage artefacts. In: *The International Archives of the Photogrammetry, remote Sensing and Spatial Information Sciences*, Vol. XL, Part. 5/W2, pp. 421-424.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *27*(10), pp. 1615-1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. *International journal of Computer Vision*, *65*(1-2), pp. 43-72.

Moons, T., Vergauwen, M., Van Gool, L., 2008. *3D reconstruction from multiple images*.

Mordohai, P., Frahm, J. M., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., Nistér, D., 2007. Real-time video-based reconstruction of urban environments. *ISPRS Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures.*

Nayar, S. K., Nakagawa, Y., 1994. Shape from focus. *Pattern analysis and machine intelligence, IEEE Transactions on*, *16*(8), pp. 824-831.

Newcombe, L., 2007. Green fingered UAVs. *Unmanned Vehicle*, November 2007.

Nitzan, D., 1988. Three-dimensional vision structure for robot applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *10*(3), pp. 291-309.

NPL (National Physical Laboratory), 2010. *Good Practice Guide NO. 118 – A Beginner's Guide to Measurement*. © Queen's Printer and Controller of HMSO, United Kingdom.

Pavlidis, G., Tsiafakis, D., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C., 2006. Recording cultural heritage. In: *Proceedings of Third International Conference of Museology,* Mytilene, Greece.

Peipe, J., Stephani, M., 2003. Performance evaluation of a 5 megapixel digital metric camera for use in architectural photogrammetry, In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part. 5/W12, pp. 259-261.

Pentland, A. P., 1987. A new sense for depth of field. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4), pp. 523-531.

Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 1/w41, pp. 73-77.

Pierrot-Deseilligny, M., De Luca, L., Remondino, F., 2011. Automated image-based procedures for accurate artifacts 3D modeling and orthoimage generation. In: *23nd CIPA Symposium*, Prague, Czech Republic.

Pollefeys, M., Koch, R., Van Gool, L., 1999. A simple and efficient rectification method for general motion. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Vol. 1, pp. 496-501. IEEE.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R., 2004. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, *59*(3), pp. 207-232.

Pons, J. P., Keriven, R., Faugeras, O., 2005. Modelling dynamic scenes by registering multi-view image sequences. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, pp. 822-827. IEEE.

Przybilla, H. J., Wester-Ebbinghaus, W., 1979. Bildflug mit ferngelenktem Kleinflugzeug. *Bildmessung und Luftbildwesen*, *47*(5), pp. 137-142.

Puri, A., Valavanis, K. P., Kontitsis, M., 2007. Statistical profile generation for traffic monitoring using real-time UAV based video data. In: *Control & Automation, 2007. MED'07. Mediterranean Conference on*, pp. 1-6. IEEE.

Remondino, F., Guarnieri, A., Vettore, A., 2005. 3D modeling of close-range objects: photogrammetry or laser scanning?. In: *Electronic Imaging 2005*, pp. 216-225. International Society for Optics and Photonics.

Remondino, F., El-Hakim, S., 2006. Image-based 3D Modelling: A Review. *The Photogrammetric Record*, *21*(115), pp. 269-291.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part. 5, pp. 266-272.

Remondino, F. and Menna, F., 2008. Image-based surface measurement for close-range heritage documentation. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII, Part. B5, pp. 199–206.

Remondino, F., El-Hakim, S., Gruen, A., Zhang, L., 2008. Turning images into 3-D models. *Signal Processing Magazine, IEEE, 25*(4), pp. 55-65.

Remondino, F., El-Hakim, S., Girardi, S., Rizzi, A., Benedetti, S., Gonzo, L., 2009. 3D Virtual Reconstruction and Visualization of Complex Architectures-The" 3D-ARCH" Project. In: *Proceedings of the ISPRS Working Group V/4 Workshop 3D-ARCH" Virtual Reconstruction and Visualization of Complex Architectures"*.

Remondino, F., 2011. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sensing, 3*(6), pp. 1104-1138.

Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., Sarazzi, D., 2011. UAV photogrammetry for mapping and 3D modelling – Current status and future perspectives. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part. 1/C22, ISPRS Zurich 2011 Workshop, Zurich, Switzerland.

Rennison, B., Jacobsen, M., Scafuri, M., 2009. The Alabama Yardstick: testing and assessing three-dimensional data capture techniques and best practices. In: *Proc. 37th Annual Computer Applications and Quantitative Methods in Archaeology Conference in Williamsburg*.

Roman, A., Garg, G., Levoy, M., 2004. Interactive design of multi-perspective images for visualizing urban landscapes. In: *Proceedings of the conference on Visualization'04*, pp. 537-544, IEEE Computer Society.

Rönnholm, P., Honkavaara, E., Litkey, P., Hyyppä, H., Hyyppä, J., 2007. Integration of laser scanning and photogrammetry. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX, pp. 355-362.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: *Proceedings LC3D Workshop*, Berlin.

Rousseeuw, P. J., Leroy, A. M., 1987. Robust regression and outlier detection. *Wiley series in probability and mathematical statistics (Applied probability and statistics).*

Roy, S., & Cox, I. J., 1998. A maximum-flow formulation of the n-camera stereo correspondence problem. In: *Computer Vision, 1998. Sixth International Conference on*, pp. 492-499. IEEE.

Rozycki, S., Wolniewicz., W., 2007. Assessment of DSM accuracy obtained by high resolution stereo images. *ISPRS Hannover Workshop 2007: High-Resolution Earth Imaging for Geospatial Information*, Hannover, Germania.

Salvi, J., Armangué, X., Batlle, J., 2002. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern recognition*, *35*(7), pp. 1617-1635.

Salvi, J., Pages, J., Batlle, J., 2004. Pattern codification strategies in structured light systems. *Pattern Recognition*, *37*(4), pp. 827-849.

Sansoni, G., Trebeschi, M., Docchio, F., 2009. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, *9*(1), pp. 568-601.

Santagati, C., Inzerillo, L., 2013. 123D Catch: efficiency, accuracy, constraints and limitations in architectural heritage field. *International Journal of Heritage in the Digital Era* 2.2 (2013): pp. 263-290.

Sauerbier, M., Eisenbeiss, H., 2010: UAVs for the documentation of archaeological excavations. *IAPRS&SIS*, Vol. 38(5), Newcastle upon Tyne, UK.

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1-3), pp. 7-42.

Schindler, G., Dellaert, F., Kang, S. B., 2007. Inferring temporal order of images from 3D structure. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1-7. IEEE.

Schmid, C., Mohr, R., Bauckhage, C., 2000. Evaluation of interest point detectors. *International Journal of Computer Vision*, *37*(2), pp. 151-172.

Schmid, C., Zisserman, A., 2000. The geometry and matching of lines and curves over multiple views. *International Journal of Computer Vision*, *40*(3), pp. 199-233.

Schneider, R., Thu, P., Stockmann, M., 2001. Distance measurement of moving objects by frequency modulated laser radar. *Optical Engineering*, *40*(1), pp. 33-3

Se, S., Firoozfam, P., Goldstein, N., Dutkiewicz, M., Pace, P., 2010. Automated UAV-based video exploitation for mapping and surveillance. In: *Proceedings of International Society for Photogrammetry and Remote Sensing (ISPRS) Commission I Symposium*.

Se, S., Pears, N., 2012. Passive 3D Imaging. In: Pears, N., Liu, Y., Bunting, P. (Eds.), *3D Imaging, Analysis and Applications*, pp. 35-94. Springer London.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, Vol. 1, pp. 519-528. IEEE.

Seitz, P., 2007. Photon-noise limited distance resolution of optical metrology methods. In: *Optical Metrology*, pp. 66160D-66160D. International Society for Optics and Photonics.

Sequeira, V., Wolfart, E., Bovisio, E., Biotti, E., Goncalves, J. G., 2001. Hybrid 3D reconstruction and image-based rendering techniques for reality modeling. In: *Photonics West 2001-Electronic Imaging*, pp. 126-136. International Society for Optics and Photonics.

Slabaugh, G., Culbertson, B., Malzbender, T., Schafer, R., 2001. A survey of methods for volumetric scene reconstruction from photographs. In: *Proceedings of the 2001 Eurographics conference on Volume Graphics*, pp. 81-101. Eurographics Association.

Smith, S. M., Brady, J. M., 1997. SUSAN - a new approach to low level image processing. *International journal of Computer Vision*, *23*(1), pp. 45-78.

Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM transactions on graphics (TOG)*, *25*(3), pp. 835-846.

Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM transactions on graphics (TOG)*, *25*(3), pp. 835-846.

Snavely, N., Seitz, S. M., Szeliski, R., 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision*, *80*(2), pp. 189-210.

Snavely, N., Seitz, S. M., Szeliski, R., 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision*, *80*(2), pp. 189-210.

Stamos, I., Liu, L., Chen, C., Wolberg, G., Yu, G., Zokai, S., 2008. Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *International Journal of Computer Vision*, *78*(2-3), pp. 237-260.

Stewart, C. V., 1999. Robust parameter estimation in computer vision. *SIAM review*, *41*(3), pp. 513-537.

Stumpfel, J., Tchou, C., Yun, N., Martinez, P., Hawkins, T., Jones, A., Emerson, B.; Debevec, P., 2003. Digital reunification of the Parthenon and its sculptures. In: *Proceedings of the 4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (*VAST*), Brighton, UK, pp. 41-50.

Sun, J., Zheng, N. N., Shum, H. Y., 2003. Stereo matching using belief propagation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(7), pp.787-800.

Thamm, H. P., Judex, M., 2006. The "low cost drone"–an interesting tool for process monitoring in a high spatial and temporal resolution. In: *ISPRS Mid-term Symposium*, pp. 8-11.

Torr, P. H. S., 2002. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, *50*(1), pp. 35-61.

Toschi, I., 2010. *Processi di calibrazione e restituzione fotogrammetrica per il rilievo architettonico e archeologico*. Master's Thesis, Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Italy.

Toschi, I., Rivola, R., Bertacchini, E., Castagnetti, C., Dubbini, M., Capra, A., 2013. Validation tests of open-source procedures for digital camera calibration and 3D image-based modelling. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL, Part. 5/W2, pp. 647-652, Strasbourg, France.

Tournaire, O., Brédif, M., Boldo, D., Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 65(4), pp. 317-327.

Toutin, T., 2000. Elevation modeling from satellite data. *Enc. of Analtical Chemistry: Applications, Theory and Instrumentations*, 10 (2000), pp. 8543-8572.

Toutin, T., Chénier, R., Carbonneau, Y., 2002. 3D models for high resolution images: examples with Quickbird, Ikonos and EROS. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part. 4, pp. 547-551, Ottawa, Canada.

Toutin,T., 2004a. Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-A, IKONOS-II and Quickbird. *IEEE-TGARS*, 42(10), pp. 2121-2129.

Toutin, T., 2004b. Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing*, 10(2004), pp 1893-1924.

Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 2000. Bundle adjustment - a modern synthesis. In: *Vision algorithms: theory and practice*, pp. 298-372. Springer Berlin Heidelberg.

Trimble, 2011. *eCognition® Developer 8.64.1 Reference Book*. Trimble Germany GmbH, München, Germany.

Tsai, R. Y., 1987. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *Robotics and Automation, IEEE Journal of*, *3*(4), pp. 323-344.

Ulupinar, F., Nevatia, R., 1995. Shape from contour: Straight homogeneous generalized cylinders and constant cross section generalized cylinders. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *17*(2), pp. 120-135.

Van Blyenburgh, P., 1999. UAVs: and Overview, *Air & Space Europe*, 1(5), pp. 43-47.

Van Meerbergen, G., Vergauwen, M., Pollefeys, M., Van Gool, L., 2002. A hierarchical symmetric stereo algorithm using dynamic programming. *International Journal of Computer Vision*, *47*(1-3), pp. 275-285.

Velios, A., Harrison, J.P., 2002. Laser scanning and digital close range photogrammetry for capturing 3D archaeological objects: a comparison of quality and practicality. In: *G. Burenhult (ed.), Archaeological Informatics: Pushing the Envelope, CAA 2001*, Oxford, pp. 205-211.

Vergauwen, M., Van Gool, L., 2006. Web-based 3D reconstruction service, *Machine vision and applications*, 17(6), pp. 411-426.

Verhoeven, G., 2011. Taking computer vision aloft–archaeological three-dimensional reconstructions from aerial photographs with photoscan. *Archaeological Prospection*, *18*(1), pp. 67-73.

Vogiatzis, G., Torr, P. H., Cipolla, R., 2005. Multi-view stereo via volumetric graph-cuts. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, pp. 391-398. IEEE.

Vogiatzis, G., Torr, P. H., Cipolla, R., 2005. Multi-view stereo via volumetric graph-cuts. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, pp. 391-398. IEEE.

Vosselman, G., Maas, H-G., 2010. *Airborne and terrestrial laser scanning*. CRC, Boca Raton, pp. 318.

Wang, J., Li, C., 2007. Acquisition of UAV images and the application in 3D city modeling. In: *International Symposium on Photoelectronic Detection and Imaging: Technology and Applications 2007*, pp. 66230Z-66230Z. International Society for Optics and Photonics.

Wang, Y., Yang, X., Xu, F., Leason, A., Megenta, S., 2008. An operational system for sensor modeling and DEM generation of satellite pushbroom sensor images. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII, Part. B1, Beijing, Cina.

Wei, H., Bartels, M., 2012. 3D Digital Elevation Model Generation. In: *3D Imaging, Analysis and Applications*, pp. 367-415. Springer London.

Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammmetry and Remote Sensing*, 50(4), pp. 38-49.

Winkelbach, S., Wahl, F. M., 2001. Shape from 2D edge gradients. In: *Pattern recognition*, pp. 377-384. Springer Berlin Heidelberg.

Wu, C., 2013. Towards linear-time incremental structure from motion. In: *3DTV-Conference, 2013 International Conference on*, pp. 127-134. IEEE.

Yin, X., Wonka, P., Razdan, A., 2009. Generating 3d building models from architectural drawings: A survey. *IEEE Computer Graphics and Applications*, *29*(1), pp. 20-30.

Zhang, R., Tsai, P. S., Cryer, J. E., Shah, M., 1999. Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *21*(8), pp. 690-706.

Zhang, Z., 2000. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(11), pp. 1330-1334.

Zhang, B., Miller, S., De Venecia, K., Walker, S., 2006. Automatic terrain extraction using multiple image pair and back matching. *Proc. ASPRS Annual Conference*, Reno, USA.

Websites (Last accessed on 2014, Mar. 11):

123D Catch, Autodesk, www.123dapp.com/catch

3DF Zephyr Pro, 3DFLOW, www.3dflow.net/3df-zephyr-pro-3d-models-from-photos

Apero/MicMac, www.micmac.ign.fr

ARC3D, www.arc3d.be

ArcGIS, esri, www.esri.com/software/arcgis

B89.4.19, 2006 - Performance Evaluation of Laser-Based Spherical Coordinate Measurement Systems, www.asme.org

B89.7.5, 2006 - Metrological Traceability of Dimensional Measurements to the SI Unit of Length, www.asme.org

Bundler, www.cs.cornell.edu/~snavely/bundler

CeCILL, 2005, www.cecill.info/licences/Licence_CeCILL-B_V1-fr.html

CLORAMA, 4DiXplorer, www.4dixplorer.com/software_clorama.html

CloudCompare, www.danielgm.net/cc

Committee E57 (E57.01, E57.02, E57.03, E57.04, and E57.90) on 3D Imaging Systems, ASTM International, 2008.www.astm.org/COMMIT/COMMITTEE/E57.htm

CULTURE 3D CLOUDS, www.map.archi.fr/?p=201

Documentation MicMac, logiciels.ign.fr/?Telechargement,20

eCognition Developer, Trimble, www.ecognition.com/products/ecognition-developer

E-curator research project, www.museums.ucl.ac.uk/research/ecurator/objectives.html

e-GEA, www.egea.unimore.it/site/home.html

ENVI, Exelis Visual Information Solutions, www.exelisvis.com/ProductsServices/ENVI/ENVI.aspx

exiftool, www.sno.phy.queensu.ca/~phil/exiftool

exiv2, www.exiv2.org

Forum MicMac, forum-micmac.forumprod.com

Geomagic Design X, Geomagic, www.rapidform.com/products/xor/overview

Geomatica 2013, PCI Geomatics, www.pcigeomatics.com/software/geomatica2013

Heritage3D, www.heritage3d.org

www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=28086

ImageMagick, www.imagemagick.org

ImageStation, Intergraph, www.geospatial.intergraph.com/products/ImageStation/ProductLiterature.aspx

IMAGINE Photogrammetry, Intergraph, www.geospatial.intergraph.com/products/imagine-photogrammetry

INPHO, Trimble, www.trimble.com/imaging/inpho.aspx

Insight3d, www.insight3d.sourceforge.net

ISO 1:2002, "Geometrical Product Specifications (GPS) – Standard reference temperature for geometrical product specification and verification", International Organization for Standardization (ISO), www.iso.org

ISO/IEC GUIDE 2. 2004. Standardization and related activities – General vocabulary. www.iso.org

ISO/TC 172/SC 6 (ISO 17123), www.iso.org

ISO/TC 213 (ISO 10360), www.iso.org

iWitness, iWitness, www.iwitnessphoto.com

JRC 3D Reconstructor, Gexcel, www.gexcel.it/it/software

make, www.gnu.org/software/make

MicroMap, Geoin, www.geoin.it/Micromap.jsp

NRC – MSS. www.nrc-cnrc.gc.ca/eng/rd/mss/index.html

PhotoModeler Scanner, EOS Systems, www.photomodeler.com

PhotoScan, Agisoft , www.agisoft.ru

Photosynth, Microsoft, www.photosynth.net

Planetek Italia s.r.l. www.planetek.it

PMVS2, www.di.ens.fr/pmvs

PolyWorks, InnovMetric, www.innovmetric.com

proj4, trac.osgeo.org/proj

Project PhotoFly, Autodesk, forums.autodesk.com/t5/Project-Photofly/bd-p/507

ReCap, Autodesk, www.autodesk.com/products/recap/overview

SAL Engineering, www.salengineering.it

Scape Capture / ShapeScan, ShapeQuest Inc., www.shapecapture.com

SCENE, FARO Laser Scanner Software, www.faro.com/en-us/products/faro-software/scene/overview

SOCET SET, BAE Systems, www.geospatialexploitationproducts.com

STAR*NET, MicroSurvey, www.microsurvey.com/

SURE, www.ifp.uni-stuttgart.de/publications/software

Surfer, Golden Software Inc., www.goldensoftware.com/products/surfer

TAPEnADe project, www.map.archi.fr/?p=170

The London Charter for the Computer-Based Visualization of Cultural Heritage, www.londoncharter.org/

UMR 3495 CNRS/MCC MAP-Gamsau. www.map.archi.fr/?page_id=381

UNESCO, en.unesco.org

Vedaldi, 2010, www.robots.ox.ac.uk/~vedaldi/code/siftpp.html

VIM3, JCGM 200:2012 International vocabulary of metrology
www.bipm.org/en/publications/guides/vim.html

VisualSFM, ccwu.me/vsfm

ACKNOWLEDGMENTS

I would like to express my special gratitude to my supervisor Alessandro Capra, Full Professor in Geomatics at the University of Modena and Reggio Emilia. From the beginning of my research studies, he supported my efforts and encouraged me to deepen my personal scientific interests.
He taught me that measurement is responsibility.

I also like to give my special thanks to Marco Dubbini, who always believed in me.
He taught me that measurement is enthusiasm.

I thank all my colleagues, Eleonora Bertacchini, Cristina Castagnetti, Riccardo Rivola and Paolo Rossi, who worked with me from the beginning of my research studies.
They taught me that measurement is fun.

I thank Livio De Luca, who hosted me at UMR CNRS/MCC MAP-Gamsau Laboratory (Marseille, France) and supported my research studies.
He taught me that measurement is multidisciplinarity.

I thank Jean-Angelo Beraldin, who hosted me at NRC-MMS (Ottawa, Canada), and all his family, who made me feel at home. Angelo showed me what real metrology is.
He taught me that measurement is beauty.

Furthermore, I would like to thank all researchers who, through their papers and books, encouraged me to love photogrammetry. As shown by the Reference section, they are numerous. Even though most of them don't know me, they all played a paramount role in my personal research growth.
They taught me that measurement is knowledge.

I finally want to give my special thanks to all people who love me. If I'm here, it's only because of them.
They taught me that measurement is not everything in life.