

Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach

*Original*

Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach / Aicardi, Irene; Chiabrando, Filiberto; Lingua, Andrea Maria; Noardo, Francesca. - In: JOURNAL OF CULTURAL HERITAGE. - ISSN 1296-2074. - ELETTRONICO. - 32:(2018), pp. 257-266. [10.1016/j.culher.2017.11.006]

*Availability:*

This version is available at: 11583/2705081 since: 2018-07-10T12:21:12Z

*Publisher:*

Elsevier

*Published*

DOI:10.1016/j.culher.2017.11.006

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



Available online at  
**ScienceDirect**  
[www.sciencedirect.com](http://www.sciencedirect.com)

Elsevier Masson France  
**EM|consulte**  
[www.em-consulte.com](http://www.em-consulte.com)



## Review

# Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach



Irene Aicardi<sup>a</sup>, Filiberto Chiabrando<sup>b</sup>, Andrea Lingua<sup>a,\*</sup>, Francesca Noardo<sup>a</sup>

<sup>a</sup> Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, c.so Duca degli Abruzzi, 24 10129 Torino, Italy

<sup>b</sup> Department of Architecture and Design (DAD), Politecnico di Torino, v.le Mattioli, 39 10125 Torino, Italy

## INFO ARTICLE

### Historique de l'article :

Reçu le 6 décembre 2016

Accepté le 14 novembre 2017

Disponible sur Internet le 21 February 2018

### Keywords :

Photogrammetry  
 Computer vision  
 Structure-from-motion  
 Algorithms  
 3D modelling  
 Orthophoto

## ABSTRACT

The techniques of measuring and 3D modelling based on images, as is typical in photogrammetry, grew in interest again in recent years, since a new generation of software tools has spread. These ones implement in different measure the algorithms developed by computer vision, increasing the automation of the standard photogrammetric process. This made the use of image-based approaches for 3D models reconstruction enormously increase, which is an essential part of the Cultural Heritage documentation and analysis processes. Starting from these assumptions, the aim of the paper is to evaluate what and where it is possible nowadays to find the main differences between photogrammetry and computer vision approaches and how these have to be considered in the choice of the processing technique. The analysis has been performed starting from a theoretical point of view in order to trace the main characteristics of the two methods. Moreover, in order to complete the investigation, an experimental part is reported on two particular cases study, considered as representative of two types of usually surveyed objects. The results allow to enlighten some differences between the two image processing approaches, in terms of accuracy and achieved products.

© 2017 Les Auteurs. Publié par Elsevier Masson SAS. Cet article est publié en Open Access sous licence CC BY (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The great back in vogue, in the last years, of the use of images for measuring 3D features, made many software have a wide success in many application fields, including cultural heritage documentation and analysis. They offer the possibility of a cheap acquisition, fast and automatic processing and high accuracies in the results. After the skepticism and the enthusiasm are past, the necessity arise for photogrammetrists and geomatics users to know more about the behaviour of the recently spread software tools. These implement techniques and algorithms mainly developed by the computer vision (CV) community, a different but more and more familiar discipline. For researchers or professional actors working with geomatics, it is a strategic topic to know when and how to prefer a photogrammetry-based software or a computer vision-derived technique.

In all the branches of cultural heritage field (mobile museum objects, architecture, archaeological sites, and so on), the 3D survey is an essential support for a number of activities: the object

documentation, different kinds of analysis (statistical analysis, historical reconstructions, etc.), the communication and promotion of the sites, and so on. The possibility to generate very accurate and detailed 3D models from imagery is a great opportunity, since a number of tools exist for the method to be employed by a wide community of users, with limited costs. However, the results need to be controlled and critically considered to be reliable. Moreover, it is necessary to contemplate different kinds of processing tools or acquisition procedures based on the kind of object to be surveyed (evaluating the geometry, the texture, the distance, the complexity of its shape, and so on). It is therefore necessary to be conscious of the characteristics of the main two methods for measuring objects and generating 3D models from images (photogrammetry and computer vision derived techniques), in order to choose, each time, the most suitable one.

If we analyze the definition of Photogrammetry (PH) and Computer Vision we can find that:

- “Photogrammetry is the art and science of determining the position and shape of objects from photographs” [1];
- “Computer Vision is a mathematical technique for recovering the three-dimensional shape and appearance of objects in imagery” [2].

\* Corresponding author. Politecnico di Torino, c.so Duca degli Abruzzi, 24 10129 Torino, Italy.

Adresse e-mail : [andrea.lingua@polito.it](mailto:andrea.lingua@polito.it) (A. Lingua).

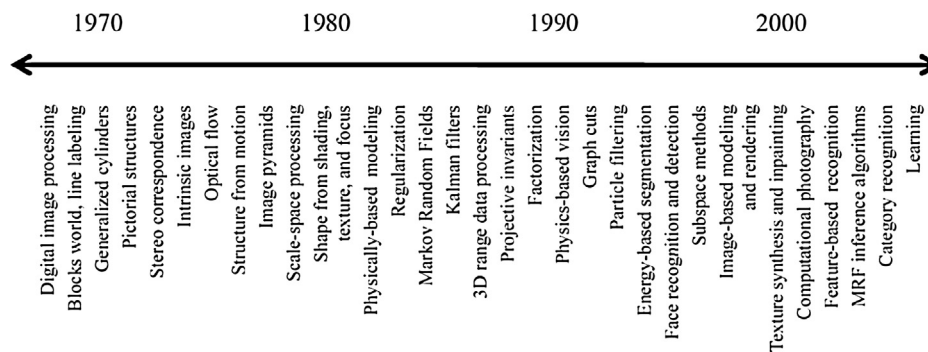


Fig. 1. Timeline of some of the most significant topic in the field of Computer Vision [2].

The main feature is that both the techniques start from the analysis of 2D images to discover 3D shape information, even if the employed approach is sometimes different: originally, the goal of photogrammetry was the measurement of the position of a set of 3D points, while the computer vision aimed at the final appearance of the model.

The main goal of PH was for mapping purposes; consequently, the technique was linked to the achievement of the best reachable metric accuracy. Instead, the CV idea was related to the automation of the process, in particular regarding the possibility to make the data understandable to computers for automatically extracting information from images (image segmentation, individuation of lines with similar value or depth, individuation of the steepest gradient, pattern or face recognition, semantic network and so on) [3].

The birth of PH could be traced back to the 1480, starting from the studies of Leonardo da Vinci [4]. It was then improved through the years, using as base the central projection concept and developing new techniques to reconstruct the 3D model (interior and external parameters problem). The PH had a revival in the 1990s thanks to the digital images [5].

On the other hand, the CV started from the 1970s with the aim to endow robots with the human visual perception, in order to improve their intelligent behaviour. The history of this science can be easily summarized in the Fig. 1 below [2].

By the time, in the field of 3D model reconstruction, PH and CV came to have the same goal and purpose, approaching the problem from two different points of view.

First of all, they start from the same mathematical model (the central projection), but the first CV algorithms use a linear approach to solve the problem, while the PH theory generally considers a non-linear solution that must be linearized according to the approximation of the initial parameters. For this reason, in many cases, for the photogrammetric processing is preferable the interior orientation parameters of the used camera to be known, or, at least, stable, while the CV approach starts from the concept of not calibrated cameras (unknown interior orientation parameters) to reconstruct the 3D shape of the area of interest.

As it is mentioned before, the processing steps for the two techniques could be considered different. Simplifying the basic workflow of the two techniques, the PH starts from the knowledge of the camera parameters (focal length and principal point) and then moves to the extraction of homologous points. Instead the CV starts from the analysis of the images to discover common points (homologous points) between the images and then it reconstructs the geometry of the acquisition.

However, even if sometimes they seem to be two very different approaches, nowadays it is difficult and sometimes impossible to define the border between the two applications in the definition of the used algorithms and the obtained results. The frontier

is even fainter when considering that some CV techniques are currently implemented in up-to-date photogrammetric software. For example, feature-extraction algorithms, dense reconstruction techniques and so on. It is thus important to investigate the tools that regulate these processes, in order to reach a higher consciousness of the methods used by the software for suitably choosing one of them for the specific application needs.

The purpose of this paper is to analyze and compare the digital photogrammetry technique with the computer vision method in the field of 3D metric reconstruction from images [6–8]. As a first step, the mathematical structure must be understood as a basis for the following phases. Moreover, the goal of each approach is defined and the procedures characterizing the techniques are described. In order to show the differences of the achieved results, finally, an experimental part reports two representative case studies in the Cultural Heritage field.

## 2. A brief foreword about the algorithms

In PH and CV, images (stereo pairs or sequential images) are used in order to extract further information. The aim of photogrammetry can be pursued by means of well-established algorithms to extract the geometry and the physical model information. In computer vision the great advantage is the implementation of a human-like capability to recognize 3D information in the image data in non-human entities (computers). However, a main difference rests.

The photogrammetric concept starts from a physical model of the image creation

Instead, the CV has a strong connection with the mathematics and the computer science application and it somehow loses its connection with the physical models

Especially in CV, the 3D reconstruction from the images starts from the image matching techniques and the recognition of interest points on 2D images. Digital image matching can be defined as an establishment of automatic correspondences among the primitives extracted from two or more digital images depicting, at least partly, the same scene (Heipke, 1996).

Over the years, sophisticated algorithms have been implemented by researchers, but a fully automatic and always effective image matching method can still be improved. This is due to the information lost during the image acquisition. Indeed, for a given point in one image, its correspondences on other images may not exist due to occlusion. Moreover, it is possible to have more than one match, due, for example, to repetitive texture patterns or even no matches, according to image noise or lack of textures; semi-transparent object surfaces could also give similar problems. For these reasons image matching techniques can be ranked in the ill-posed problem class (there is no guarantee that a solution exist and that it is unique and stable respect to small variations of the input data, Terzopoulos, 1986).

Geometrical constraints exploit the mathematical models, which describe the physical problem to reduce the research space.

In both cases, the starting model is a perspective camera (pin-hole camera, as reported in [9]), even if in the Computer Vision sometimes it is possible to simplify the Structure from Motion (SfM) problem using the orthographic approach [10], or considering the para-perspective case. In this case, the perspective rays project the points in an image plane from the projection centre. The perspective geometry can be expressed starting from the epipolar constrain. It reduces the 2-dimensional search space to 1 dimension: if the images orientation is known, the candidates to match belong to a line, called “epipolar line”; instead, if the orientation parameters are only approximated, the search space is reduced to a band centered at the epipolar line.

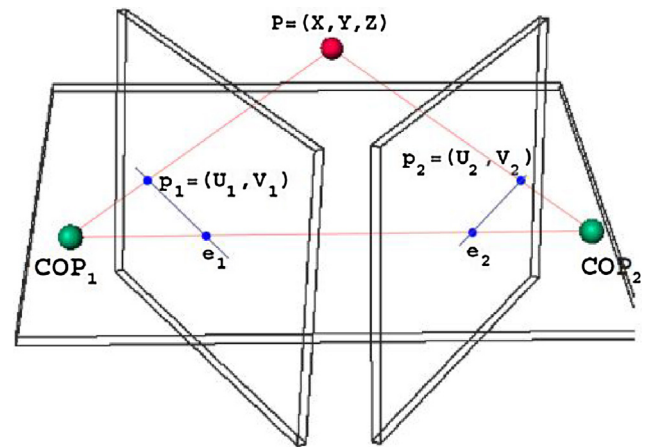


Fig. 2. Epipolar geometry schematization [12].

The photogrammetric approach expresses the relationship between image and world points in terms of colinearity equations	To define homogeneous coordinates, the computer vision generally uses a linear approach
--	---

### 2.1. The photogrammetric approach

In photogrammetry the most used model is the central projection.

It is a geometric procedure that transforms a 3D entity into a 2D reality and it occurs when you have an object, a projection centre and a projection plane oriented in any way with respect to the projected object. This projection is then produced because a series of straight lines (or projecting rays) connect the points of the object with the projection centre and, intersecting the projection plane, generate the image points, that are the points projections.

When we acquire a frame, the object point P, the projection centre O and the image point P' lie on a same straight line. The mathematical representation that indicates the colinearity conditions representing the alignment between the points in the image and in the object system can be expressed in the following way:

$$\frac{\xi - \xi_0}{c} = \frac{X' - X'_0}{Z'_0 - Z'}$$

$$\frac{\eta - \eta_0}{c} = \frac{Y' - Y'_0}{Z'_0 - Z'}$$

where  $\xi, \eta, \zeta$  ( $\zeta = 0$  for the image points and  $\zeta = c$  for the projection centre) represent the image system and  $X, Y, Z$  the object system.

The coordinates  $X', Y', Z'$  of the point P and the coordinates  $X'_0, Y'_0, Z'_0$  of the projection centre may be processed in the system  $X, Y, Z$  by the spatial rotation matrix R:

$$\begin{pmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} X' - X'_0 \\ Y' - Y'_0 \\ Z' - Z'_0 \end{pmatrix}$$

So, if we multiply the matrix above for the matrix  $R^T = R^{-1}$ , and replace in the equations, explicit the image coordinates, we will get the relationship between the image coordinates and the ground ones, called colinearity equations:

$$\xi = \xi_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)}$$

$$\eta + \eta_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)}$$

The physical meaning of the photogrammetric approach is described by the colinearity equations, which demonstrate that each

point is projected into a unique image point, if it is not occluded by other object points [11].

The described relations are strictly connected to the image generation inside a camera using the concept of the central projection geometry.

The main problem is that the colinearity condition is expressed in terms of non-linear equations and it is solved with a least square estimation. For this reason, the expressions can be rewritten in terms of differential relations, and approximated initial parameters must be used to solve the system. The main theoretical problem is the definition of the initial parameters to be used and the final solution must be solved with an iterative procedure in order to refine the solution derived from the approximated estimation. From a practical point of view, this issue was solved with the introduction of the fully automatic pre-bundle adjustment initial parameter determination, which was introduced in most of the available photogrammetric software.

### 2.2. The computer vision strategy

In the CV the 3D reconstruction from images has been developed considering the possibility to use mathematical techniques. The CV algorithms start from the same relation of co-planarity that in this case is described in homogeneous coordinates to represent the transformations through matrix operations. In the case of stereovision with two un-calibrated cameras the relationship between the two images can be described by the epipolar constraint that is widely used in automated matching process because it reduces the matching ambiguity problems and computational costs. In the epipolar geometry (Fig. 2), the epipolar plane, made by the object point and the two centres of projection, defines the epipolar lines as its intersections with the two image planes [12].

The mathematical ideal model (since it does not include non-linear and second order effects, such as lens distortions) can be expressed using the following formula:

$$Z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & s & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = M_1 M_2 \bar{X} = M \bar{X}$$

where:

- ( $u, v, 1$ ) are the homogeneous image coordinates;
- $Z_c$  is the space points depth under camera system;

$(\bar{X}) = (X, Y, Z, 1)^T$  is space points homogeneous coordinates under object-space coordinate system;

$M = (M_1 \ M_2)$  is a  $3 \times 4$  projection matrix composed by:

$M_1$  that represents the camera calibration matrix containing:  $f_x, f_y$ : the focal length along the two axis;  $u_0, v_0$ : coordinates of the principal point in the image plane;  $s$ : the skew value between the  $x, y$  axis.

$M_2$ : the exterior orientation matrix that contains:

$R$ : the rotation;  $t$ : the translation of the system.

The expression used in CV has a more compact mathematical expression and it has no physical meaning as it describes the relationship between the space points coordinates and its image point coordinates.

The main significant advantage in this relation is that it can be expressed by a linear equation.

In this geometry, the epipoles ( $e_1$  and  $e_2$ ) are defined by the intersections between the line that connects the two projection centres and the image plane. For this reason, the 3D point  $P$  can be described as its projection in the two image planes with the  $P_1(u_1, v_1, 1)$  and  $P_2(u_2, v_2, 1)$  coordinates and can be modelled by a linear equation:

$$P_1^t \ F \ P_2 = 0$$

where  $F$  is a  $3 \times 3$  matrix, called “fundamental matrix”, used to map from a point in one image to a line (its epipolar line) in the other one. It is a 9-parameters matrix with only 7 degrees of freedom due to the fact that the scale factor is not significant and  $\text{Det}(F) = 0$  (rank 2).

This expression is valid for any pair of corresponding points in the two images and it describes the geometry of the correspondences in the two images.

The definition of  $F$  is independent from the knowledge of interior camera parameters, but not from the relative pose of the cameras.

Over the time, different techniques to estimate the fundamental matrix parameters have been defined using different numbers of points and algorithms. The most popular one was proposed by Longuet-Higgins [13] and then refined by Hartley (1985). It is an 8-point algorithm that uses linear expression and can be fast and easily implemented in the software analysis.

If we expand the previous equation we can set a linear constraint of  $F$  for a pair of observed point:

$$u_1 u_2 f_{11} + u_1 v_2 f_{12} + u_1 f_{13} + v_1 u_2 f_{21} + v_1 u_2 f_{22} + v_1 f_{23} + u_2 f_{31} + v_2 f_{32} + f_{33} = 0$$

Using this equation for the  $N$  corresponded features, it is possible to set a linear system of the form  $Af = 0$ . Obviously, the system can be solved with 8 points, but it is preferable to use additional points and to apply a least squares approach.

However, the presented epipolar linear approach is the first model used in the CV analysis and it has been demonstrated that it presents some instabilities due to the fact that some parameters are considered as independent. The technique is sensitive to noise and it can generate numerical ill-conditioning in the case of images presenting only rotation (no translation).

For this reason, some non-linear approaches have been presented. They started from the perspective projection camera model, but they do not use the epipolar geometry. In this case there is a direct correlation between the real 3D coordinates of the points and their 2D projections, starting from a given focal length value. Using this approach the solution can be reached through an iterative non-linear estimation with a least-square evaluation, in order to improve the accuracy.

The stereovision application can be expanded introducing a third image in the estimation of the fundamental matrix. This approach uses the so-called trifocal tensor proposed and developed by Hartley [14], Shashua and Werman [15], Faugeras and Papadopoulou [16] among others.

Also in this case, there is a trifocal plane formed by the intersection of the three centres of projection with the three image planes that form the trifocal lines  $t_1, t_2, t_3$ .

In this case there are two epipoles for each image ( $e_{i,j}$ ) and it is possible to directly use the epipolar geometry considering the three fundamental matrices. However, in order to avoid some limitation in their use, it is preferable to use the trifocal tensor to estimate the coordinates of the third point starting from the other two points. Typically, one uses this tensor ( $t$ ) to map a line in Image 1 ( $l_1$ ) and a line in Image 2 ( $l_2$ ) to a line in Image 3 ( $l_3$ ). Also in this case the relation can be expressed by a linear equation that can be used to map points simply considering the intersections of the mapped lines:

$$l_3 = t(l_1 \cdot l_2)$$

This tensor is defined by 27 scalars (i.e. it can be considered as a  $3 \times 3 \times 3$  cube operator, 18 degrees of freedom) and it can be estimated knowing some corresponding points in each of the three images.

Finally, it is possible to consider the more general case of the multi-view geometry in which more than two images have to be processed with the presence of multifocal acquisitions.

The techniques, according to [17], can be divided into two groups:

- global matching: it reconstructs the entire model using all the points and the cameras involved. It uses a factorization approach or a bundle block adjustment (BBA);
- partial matching: it starts from the separate reconstruction of the camera and points space through only some points and then reconstruct the global geometry.

This is only a brief overview about the general approaches that have been used in the born of photogrammetry and computer vision in the field of 3D reconstruction.

In order to define what are the main features they present, it is important to analyze the approach used in the two techniques during the entire 3D reconstruction workflow.

### 2.3. Two ways, one direction. Comparing algorithms

Starting from the computer vision linear system, we want to try to investigate if there is a relationship between the CV and the photogrammetric equations.

In particular, the analysis starts from these specific hypotheses:

- the CV uses a linear system equation:  $P = k[R^T \ | \ t]$ , where  $k$  is the calibration matrix,  $R^T$  is the transposed rotation matrix and the translation matrix is  $t = R^T \cdot C$ ;
- the rotation matrix  $R$ , according to the CV strategy, is composed by 9 independent parameters;
- the focal length has the same value in the two directions:  $f_x = f_y$ ;
- the pixel is considered as underformed:  $s = 0$ .

Under these constraints, we can assume that  $u = \xi$  and  $v = \eta$  and it is possible to consider the following relation:

$$\begin{pmatrix} \xi \\ \eta \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & \xi_0 \\ 0 & f & \eta_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_{11} & r_{21} & r_{31}[-r]_{11}x_0 + r_{21}y_0 + r_{31}y_0 \\ r_{12} & r_{22} & r_{32}[-r]_{12}x_0 + r_{22}y_0 + r_{32}y_0 \\ r_{13} & r_{23} & r_{33}[-r]_{13}x_0 + r_{23}y_0 + r_{33}y_0 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Multiplying the matrices, the new three-equations system can be easily defined as:

$$x - x_0$$

substituting the third equation, that means, dividing everything by the side of the third equation, we get:

$$\begin{cases} \xi = \xi_0 + f \frac{r_{11}(x - x_0) + r_{21}(y - y_0) + r_{31}(z - z_0)}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33}(z - z_0)} \\ \eta = \eta_0 + f \frac{r_{12}(x - x_0) + r_{22}(y - y_0) + r_{32}(z - z_0)}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33}(z - z_0)} \end{cases}$$

The obtained system contains the same equations used in photogrammetry (colinearity equations), except for a sign that is due to the used coordinates system.

Thanks to this simple comparison, is it possible to state that the two techniques start with a different approach (linear equation and independent parameters), but can be lead back to the same concept of co-planarity.

### 3. The processing phases, from images to 3D model

Using both the analyzed strategies, the workflow consists in different steps that are summarized in the following part.

#### 3.1. The interior orientation and camera calibration procedures

The calibration of the camera determines the interior orientation of the images and therefore the possibility to use them for measuring purposes.

Well-affirmed photogrammetry manuals [18] describe different strategies for traditional camera calibration. These include laboratory calibration, in which, for example, a goniometer was used for determining the centre of the bundle of rays yielded in the photograph; this was a good practice in traditional photogrammetry. Another possibility is the test-field calibration: in this case, a bundle adjustment is used in order to obtain interior and exterior orientation of an image, once the coordinates of spatially well-distributed and redundant points are known in the three coordinate directions. Test-field calibration also includes methods that use orthogonality, planarity, lines verticality of the image features in order to estimate the interior orientation. Finally, a self-calibration practice was used in photogrammetry, with the advantage of taking into account the systematic errors of the whole photographic system (for example, due to the environmental conditions). For effectively exploiting such technique, it is necessary to take particular care on the imaging configuration and control points distribution, in order to adapt in the best way to any kind of geometry of the object. In the past, the unavailability of well-distributed control points or excessively plane objects could limit such procedure. Following the development of CV algorithms for feature extraction, millions of points are available for each image, and they are generally distributed on the entire area and, consequently, in all the object directions. This offers the ideal conditions for performing a self-calibration, and the bundle block adjustment. A

further processing for photogrammetric camera calibration uses points with unknown coordinates for performing a self-calibration [18]. Several images with the same focus setting and different camera position and direction are used. The condition to be cared is the visibility of a subset of points in at least three images. A bundle block adjustment or other algorithms are applied, together with a free network adjustment to solve the orientation and the calibration as part of it. This procedure is nowadays developed and usually employed in software using CV algorithms. These can take advantage, again, from the high number of automatically extracted Tie Points (TPs), which provide high redundancy of reference points and high number of rays for performing the bundle block adjustment. For this reason, the use of self-calibration increased in recent years in digital photogrammetry (Luhmann, 2013).

The same procedures used for photogrammetric camera calibration are employed by CV derived methods, which though can exploit the different algorithms (as described in section 1.1.2) [2]. The major flexibility of the CV approach permits to compute intrinsic parameters, and use the images acquired from more irregular cameras, with less stable lenses, such as some low-cost action cameras video frames [19], which are hard to be estimated through a common photogrammetric process. At the same time, common software using CV derived methods can easily use images acquired from different cameras in the usual processing [20], which was not trivial for photogrammetry, even if the present software can do that.

The CV based software often have expressed procedures for camera calibration, which use predisposed planar pattern with known well-individuated points (they are frequently configured like chessboards), better if filling the acquired images [3]. These calibration systems exploit the computing of vanishing points by CV algorithms [21,22], so that they can be able to estimate the camera calibration parameters [3]. A similar system is known as “Zhang calibration” and exploits the homography of the photographed well-known object (the planar pattern with orthogonal lines at known distances) and its representation in the image [17].

#### 3.2. The exterior orientation: from stereoscopic models to SfM

Nowadays, most of the applications require managing high number of images, often not acquired from metric cameras. These blocks need high performance computers to solve the orientation problem of wide block at the same time. This problem is investigated in this paragraph.

The usual generation of the 3D models from images, which are oriented in a reference system using few Ground Control Points (GCPs) in a second step, uses photogrammetric BBA techniques. This computes unknown coordinates in the models from GCPs known coordinates starting from stereoscopic bundles [18].

Photogrammetry uses the bundle block adjustment (BBA) technique for directly computing the relations between the image and the object coordinates, avoiding the intermediate models and minimizing the reprojection errors.

In this case, the single image is the elementary unit [18]. This was developed in the 1950s and in the 1970s it was extended to include the additional parameters in order to auto-calibrate the camera or to re-estimate the parameters for adjusting the possible systematic errors [18,23]. The BBA phase is now usually implemented in common photogrammetric software.

The orientation phases improve thanks to the recently introduced CV algorithms.

For performing the process, some corresponding points must be used. It is possible to use the GCPs and TPs, which allow to estimate the external orientation parameters. Updated photogrammetric software often implement SIFT [24], SURF and similar algorithms [17] to extract the TPs for helping and improve the orientation process

Computer vision introduced the automatic feature extraction using algorithms such as Harris-Stephen operator, Forstner, Smallest Univalued Segment Assimilating Nucleus (SUSAN), Features from Accelerated Segment Test (FAST) operators [25] Scale-Invariant Feature Transform (SIFT) [24], Speeded Up Robust Feature (SURF) and similar [17]. Similar points in CV techniques are matched for the process of SfM [2]

### 3.2.1. Structure from motion

A technique similar to the BBA is used in computer vision derived software: structure from motion is the simultaneous recovery of 3D structure of the object and pose of the cameras from image correspondence [2]. SfM uses the principles of epipolar geometry for reconstructing both external orientation of images and the 3D structure of the object. Different approaches can be used: the epipolar geometry can be estimated starting from the features matching, eliminating the outlier matchings and evaluating the more reliable ones [17]. Others use an incremental approach to reconstruct the geometries, starting from two images with characteristics of high number of matches and a suitable position of the cameras, so that they can offer a correct starting point from the epipolar geometry point of view. The structure is re-estimated and adjusted at each newly introduced image. A hierarchic approach could also be used, dividing the set of images in clusters of similar represented objects [17].

For minimizing the re-projection (and thus measurement) errors, a bundle adjustment borrowed from photogrammetry is often applied [17]. Since the number of images and tie points is extremely increased with respect to the original photogrammetric applications [18], the relations used to compute them evolved in order to reduce the computation cost of the operation (Hartley and Zisserman, 2003).

### 3.3. The geometry reconstruction

Even if the present integrated software of PH and CV give similar outputs, originally PH software produced, as a priority, object points measurements, while CV tools aimed at a dense point cloud supporting a captivating 3D model.

It is possible to notice the same difference that exists between the traditional surveying workflow (object interpretation – interesting limited number of feature individuation – feature measurement and plotting), as, for example, the topographic method requires, and the new methods, which rapidly measure a high number of superabundant data which are then interpreted and chosen (such as laser scanning data). This is due to the cheapness and the quickness of the acquisition and measurement in such new devices, which permit to quickly have lots of data to be managed.

In photogrammetry, the remarkable features are measured points with high metric integrity, which has always been the main aim of the discipline. Indeed, a point cloud from a photogrammetric processing should be near free of noise

On the other hand, with CV software, a dense cloud is produced, which can be noisy and which does not include proper breaklines. The classification and the interpretation of the point cloud describing the surveyed object is therefore a following phase of the processing

The point cloud produced by the CV algorithms, similarly to the laser scanner point clouds may present noise, or irregularities, which must be eliminated for a synthetic representation of the object, useful for its employment (reading, interpretation, analysis)

by various users (accustomed, for example, to traditional orthogonal projection representation). Therefore, a post-processing phase should be always considered. In a future, computer vision technique of image interpretation could be further used in order to analyze and classify the produced raster for automatically having a sort of plotting, may be also with some more semantic value already associated [3,11].

### 3.4. Some user-side considerations

It is undeniable that the introduction of automatic operators, often deriving from CV techniques permits the achievement of more user-friendly digital photogrammetry software tools. In CV-based software, it is possible to immediately have the 3D model of an object, following simple instruction for the taking of images, so that some of them are now common widespread software or smartphone apps that anyone can use (e.g. 123DCatch Autodesk <http://www.123dapp.com/catch>, ReCap Autodesk <http://www.autodesk.com/products/recap-360/overview>). The risk is to rely blindly on these tools and obtaining incorrect models.

What is always advisable is to exploit the quickness given by the automation and the potentiality of the new techniques, but not to avoid the last careful control on the results with some accurately measured checkpoints and some controlled profiles.

## 4. Photogrammetry and photogrammetric computer vision applications

Considering the practical side, such theoretical premises result in different kinds of applications, exploiting one method or the other one.

The photogrammetric algorithms can reach very good accuracies when the geometry of the surveyed object is comparable to the objects seen from an aerial view; that is, generally, where there is a dominant planar dimension. A number of examples can be found in literature; [26–30] are only few of them, regarding cultural heritage issues.

On the other hand, the introduction of computer vision techniques and procedures permitted to add a major automation (even to photogrammetric processes), but, in particular, they allow to model more complex objects, only exploiting the algorithms implemented in the common software tools. Some example of such processes, selected from a huge literature of case studies are: [31–42]. Moreover, the computer vision-derived dense matching methods allow obtaining major performances in the presentation of the results. Since they permit the automatic extraction of the positions of millions of points, the geometry of the acquired objects is reconstructed in greater detail.

As a test of the analyzed approaches, in this section two different case studies are reported. The two objects are characterized by a similar distance from the object (approximately 20 m), but they have a very different geometry, which influence the algorithms performances in their reconstruction from images.

The first example consists in an archaeological site, with images acquired from an Unmanned Aerial Vehicle (UAV) flying at a low altitude. Although being a close range imagery, the archaeological rests had a quite horizontal development, similarly to the landscapes seen by the aerial views.

The second example here described regards some baroque vaults. In this case the distance of the camera from the object is similar (15 m), but the geometry is very complex, and the surfaces are developed in very different directions.

It is important to notice that the images acquisition schema was chosen with similar criteria, maintaining the images parallel to the same plane.



**Fig. 3.** A general view of the archaeological area of Via Gemina.

The datasets were processed using three software implementing SfM and photogrammetric techniques. The first employed software is the well-known and widespread commercial package Agisoft Photoscan (PS) (Professional version, 0.9.1) that, in this paper, is considered as totally SfM oriented. A lot of other tools employed by the users community (such as VisualSfM, 3DZephyr, Pix 4D, Context Capture, OpenSfM, etc.) have a similar approach.

As representative of traditional digital photogrammetric software (including e.g. Inpho, Strabo, etc.), the software tool “Intergraph Imagine photogrammetry” (also known as Leica Photogrammetric Suite – LPS, version 11.0) is used.

Nowadays, the trend of most part of software tools is the integration of both algorithms groups (deriving from photogrammetry and computer vision procedures). They often use structure-from-motion methods to reach first parameters estimation, in case these are not known or perfectly calibrated; then, they continue with a photogrammetric refinement of the results, and finally produce dense 3D products employing computer vision methods for dense matching. Therefore, it is true that the difference between the two methods is more and more fainter, and almost disappeared. However, it is useful to be aware of the characteristics of the two methodologies in order to choose the most suitable solution for every application case.

For this reason, a third software was employed, MicMac (MM) the open source software developed by IGN France, which is expressly a middle ground between SfM and Photogrammetry, employing both kind of algorithms in the different parts of the processing [43]. A number of similar software also exist, such as Menci APS suite, Correlator 3D, Aspect3D, and many more.

The products that were realized using the SfM approach are strictly connected to the Ground Sample Distance (GSD) and to employed parameters for the generation of the dense cloud. In the following test site the employed parameters were set up at a high level that means that a 3D point was extracted each two pixels. The results showed that the employed strategy allows obtaining an accuracy that, as is well-known, is directly connected to the pixel size.

In the next sections, after a short description of the main characteristic of the surveyed objects, the followed strategy (image acquisition and orientation) and results (products, processing time and achieved accuracy) are reported. For the accuracy evaluation, as usual, the RMSE (Root Mean Square Errors) were evaluated. In photogrammetry the RMSE means the distance between the input (source) location of a ground control point or check point (GCP or

CP) and the retransformed location for the same GCP after the BBA. It is a measure of how closely the retransformed location matches the output location of a point. These values were employed in order to have a statistical index for each test site. Further tests were realized as well comparing the 3D results like the Digital Surface Models (DSM) but since the results confirm the accuracy obtained on the analyzed points in the present paper in order to provide a synthetic index only the results connected to the RMSE are reported.

We do not expose the single software workflows, since they are very similar in every package, and their differences often depend on the processing settings. We report here only the results, to be compared for a rough evaluation of the exposed techniques, which can be further investigated considering the number of cases described in literature.

#### 4.1. Test site 1: the archaeological area of Aquileia (UD) Italy

The first case study is the area containing the remains of the roman “Domus dei putti danzanti” in Aquileia (UD, Italy) and a portion of an ancient roman street (cardo) made up of a stone pavement, which shows the typical humpback section for the reflux water drainage (Fig. 3). It is located in a very important area of the ancient city, right between the Forum and the river port [44].

In 2011 and following years, the structures remains of the “domus dei putti danzanti” and the cardo were object of a complete laser scanning and UAV photogrammetric survey realized as a combined educational and research project [45,46]. Since a high accuracy were required for the documentation purposes, a flight at a very low height (lens axis pointing at nadir) was performed with a multicopter UAV (HexaKopter produced by MikroKopter) at 20 m above the ground, in order to obtain an average Ground Sample Distance (GSD) of 58 mm. The UAV was equipped with a mirror less camera a NEX-5 with a pixel size of 5.22  $\mu\text{m}$  and a 16 mm focal length, over the archaeological area (approximately 6000  $\text{m}^2$ ) 130 images were acquired and employed with a large overlapping (80% longitudinal and 65% lateral).

##### 4.1.1. The results of the processing

In the processing of the Domus, it is interesting to underline that using the SfM approach the extracted and employed TPs for the orientation were close to 1 million, while in the case of LPS were close to 1 thousand (1200) but the accuracy results were absolutely comparable or better than the one obtained using MM or PS. The number of extracted points is clearly related to employed algorithm

**Table 1**  
Accuracy assessment of the achieved orthophotos and DSMs.

Horizontal residuals	Min Res. (mm)	Max Res. (mm)	RMSE (mm)
Photoscan	4.4	16	11
MicMac	7.1	17	19
LPS	3.2	15	10
Vertical residuals	Min Res Z (mm)	Max Res Z (mm)	RMSE ( $10^{-4}H$ )
Photoscan	4.1	24	62.5
MicMac	6.8	26	117
LPS	3.1	23	24

In this test site the best results, in terms of RMSE are given by the photogrammetric processing, performed with LPS.

for TPs extraction. In LPS it is impossible to set up the operator that probably work at the lower level of the pyramidal subsample of the images (1/32). On the other hand, with the SfM tools this parameter could be set up and usually for a good result the images are subsample at 1/2 in order to obtain million of TPs that are filtered as well in order to eliminate the blunders (this operation is performed manually in LPS). The SfM procedure is faster and sometimes easier (especially using PS), but the control on the processing steps is quite impossible. On the other hand, using MM and LPS the manual interaction during the workflow need an expert user, but improve the control on the accuracy of each part of the process. In the case of Aquileia site, this methodology delivered a better accuracy on the final orthophoto as well (Table 1). The produced orthophotos and DSMs were all similar, since they do not present noticeable noise or incongruence in their parts.

#### 4.2. Test site 2: a baroque vault, in the Stupinigi royal residence

A more complex surface was surveyed on the vault of the hall of honour of the Stupinigi royal residence (TO, Italy) (Fig. 4). The oval hall is closed by a composed vault: a rib vault in the centre and four bowl-shaped vaults linked together by plane surfaces and arcs. The whole hall is decorated with frescoes of the hunting goddess Diana who triumphs in a trompe-l'oeil technique painted architectural frame. Moreover, most of the architectural elements in the hall (columns, capitals, friezes, and so on) are not sculpted but the relief is painted onto a smooth, plastered surface.

A calibrated photogrammetric camera (Canon EOS-1Ds Mark II) was used for the image acquisition. The camera has the following characteristic: pixel size  $7.2 \times 7.2 \mu\text{m}$ , sensor size  $24 \times 36 \text{ mm}$ , image size:  $4992 \times 3328$  pixels, equipped with a 20 mm focal lens.



**Fig. 4.** An interior view of the hall of honor in Stupinigi residence.

The vaults system was acquired by 19 nadir images from a scaffolds at about 8 meters from the ground floor, disposed as a cross along the two axis of the hall, with a reciprocal overlap around 80–90%.

##### 4.2.1. The results of the processing

The detailed textured surface of the vault allows to successfully complete the orientation process with the SfM approach, as well as the other steps in order to deliver the final orthophoto, useful for restoration and conservation purposes.

The adjustment was obtained using some natural GCPs (particulars of the decoration drawings of the vault) measured using a Total Station. In the following Fig. 5 the orthophoto integrated in a 2D drawing with contours is reported. It is possible to see how the geometry automatically generated has little difference compared to the real geometry of the vaults, as it was also verified through the comparison to some laser scanner data [38].

In this case, the produced orthophotos and DSMs have important differences, as it is possible to notice in Fig. 4.

In this specific case, using a traditional photogrammetric approach, the problems were highlighted especially in the realization of the DSM and orthophoto (Fig. 6 right). The first one, automatically extracted by using the LPS process was quite noisy, as it is also possible to deduce from the height map of the generated DSM (Fig. 6 left). As a consequence, the final orthophoto was not immediately suitable for architectural documentation, mainly due to the presence of some disturbing elements (e.g. the lateral pillars and some erroneous texture projections). It needs to be accurately edited in order to realize a representation comparable with the ones achieved using the SfM approach.

As expected, in this case, the complex surface of the vault is better reconstructed by a SfM software, which indeed presents the minor RMSE on the residuals (Table 2).

## 5. Conclusions

In the paper, a discussion about the typical photogrammetric approach compared to a computer vision one for the realization of metric products was realized. How it is possible to deduce from the first part, it is quite difficult nowadays to underline the real break point between the two methodologies. Certainly, the photogrammetric approach is more controlled: the measurements, accuracy, final products etc. are typically analyzed from a metric point of view. On the other hand, the algorithms introduced by CV increased the workflow automation. Moreover, CV algorithms are useful for easily obtain a 3D model from various images, also employing less stable lenses.

Naturally, starting from these assumptions, the two methods today converge towards a comprehensive methodology that could be considered the new era of the digital photogrammetry.

In the experimental section, two different data sets were processed in order to show the differences of the results given by the two techniques in two situations: a case of a near planar surface, similar to a terrain portion, usually surveyed in the cartographic measurements, and the complex surface of a composite vault. As also deducible from the experiences described in the wide available literature, the performed tests confirm what are the most suitable techniques to be used in the different cases.

The results of the SfM approach, especially in close range and complex objects, are more accurate and complete.

On the other hand, with common nadir aerial cameras data, surveying a near-plane surface, actually a traditional digital photogrammetric approach is preferred.

In conclusion, it is possible to state that the best solution obviously should be the complete integration of the two approaches, but until this complete achievement, it will be



Fig. 5. Orthophoto produced by means of the SfM approach4 integrated in a 2D representation with contours (II).

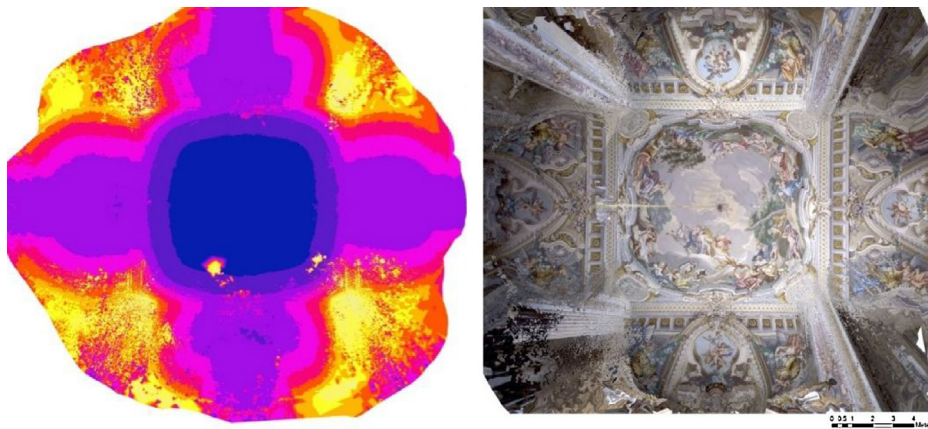


Fig. 6. DSM (left) and Orthophoto (right) generated by LPS (without editing).

**Table 2**

Accuracy assessment of the achieved orthophotos and DSMs.

Horizontal residuals	Min Res. (mm)	Max Res. (mm)	RMSE (mm)
Photoscan	2.4	8.9	7.8
MicMac	2.1	31	12
LPS	60	104	30
Vertical residuals	Min Res. Z (mm)	Max Res. Z (mm)	RMSE ( $10^{-4}H$ )
Photoscan	12	21	3.7
MicMac	11	35	5.7
LPS	1.2	109	10.7

In the afore described test area the best results, in terms of RMSE are given by the SfM processing due to the more complex shape of the acquired object.

still useful to evaluate each used software in relation to the predominant implemented approach and the surveyed objects characteristics.

From our point of view, some further step needs to be done for the processing improvement. First of all, at present it is not possible to control the TPs automatic computation that would have advantages from a better interaction (manually inserting some of them or eliminating outliers). This is a common problem in SfM software, and needs to be improved since sometimes with an only

automatic based approach it is not possible to obtain a correct orientation.

Another aspect is related to the typical 3D drawing (digital plotting extracted from the photogrammetric process). This part is under development in this last period in order to complete the workflow. The faster and more accurate approach for image orientation will be integrated with the possibility to extract or plot the traditional drawings, for documenting the surveyed object in a synthetic, understandable and codified form, and producing maps and cartography as well.

Concluding, what is always advisable is to be conscious of the operating principles of the different methods, in order to choose the processing technique and the acquisition workflow and modalities in a suitable way, considering the surveyed object and the survey aims. If a unique frame will include the different possibilities, then it will be easy to adapt the processing to whatever exigency.

Thanks to these last developments it is possible to state that actually we are arriving to realize a photogrammetric workflow for everyone. We do not know if this was the objective of the pioneer researcher in the photogrammetric field, but we are sure that our past professors have this aspiration when in the early 1970s introduced the photogrammetric course in the area of the architecture and the engineering science.

## Références

- [1] K. Kraus, *Photogrammetry*, 1, Dümmler, 1994.
- [2] R. Szeliski, *Computer vision: algorithms and applications*, Springer Science & Business Media, 2010.
- [3] R. Klette, *Concise computer vision*, Springer, London, 2014.
- [4] F. Doyle, The historical development of analytical photogrammetry, *Photogramm. Eng.* 30 (2) (1964) 259–265.
- [5] J.C. Shears, J.W. Allan, Softcopy photogrammetry and its uses in GIS, In: *Proceedings of International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI, Part B4, Vienna, 1996.
- [6] F. Remondino, M.G. Spera, E. Nocerino, F. Menna, F. Nex, State of the art in high density image matching, *Photogramm. Rec.* 29 (146) (2014) 144–166.
- [7] R.I. Hartley, J.L. Mundy, Relationship between photogrammetry and computer vision, in: *Optical engineering and photonics in aerospace sensing*, International Society for Optics and Photonics, 1993, pp. 92–105.
- [8] J.I. Kim, T. Kim, Comparison of computer vision and photogrammetric approaches for epipolar resampling of image sequence, *Sensors* 16 (3) (2016) 412.
- [9] R. Mohr, B. Triggs, *Projective geometry for image analysis*, Technical report, International Society for Photogrammetry and Remote Sensing, Vienna Congress, WG III/2 Tutorial, 1996.
- [10] C. Tomasi, T. Kanade, Shape and motion from image streams under orthography: a factorization method, *Int. J. Comput. Vis.* 9 (2) (1992) 137–154.
- [11] C. McClone, E. Mikhail, J. Bethel, *Manual of Photogrammetry*, American Society for Photogrammetry and Remote Sensing, 2004 [Bethesda, MD].
- [12] T. Jebara, A. Azarbayejani, A. Pentland, 3D structure from 2D motion IEEE Signal Processing Magazine, 3D And Stereoscopic Visual Communication, 16, 1999.
- [13] H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* 293 (1981) 133–135.
- [14] R. Hartley, Lines and points in three views – an integrated approach, In: *Proceedings of the ARPA IU Workshop*. DARPA, Morgan Kaufmann, 1994.
- [15] A. Shashua, M. Werman, On the trilinear tensor of three perspective views and its underlying geometry, In: *Proceedings of IEEE International Conference on Computer Vision (ICCV'95)*, 1995.
- [16] O. Faugeras, T. Papadopoulos, A nonlinear method for estimating the projective geometry of 3 views, In: *Sixth International Conference on Computer Vision*, 477–484, 1998.
- [17] A. Fusiello, *Visione Computazionale: appunti delle lezioni*, Informatic Department, University of Verona, 2005.
- [18] K. Kraus, *Photogrammetry*, 2, Dümmler, 1997.
- [19] C. Balletti, F. Guerra, V. Tsioukas, P. Vernier, Calibration of action cameras for photogrammetric purposes, *Sensors* 14 (9) (2014) 17471–17490.
- [20] N. Snavely, S.M. Seitz, R. Szeliski, Photo tourism: exploring photo collections in 3D, *ACM Trans. Gr. (TOG)* 25 (3) (2006) 835–846.
- [21] R. Hartley, A. Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.
- [22] L.L. Wang, W.H. Tsai, Camera calibration by vanishing lines for 3-D computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (4) (1991) 370–376.
- [23] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment – a modern synthesis, in: *Vision algorithms: theory and practice*, Springer Berlin Heidelberg, 2000, pp. 298–372.
- [24] D.G. Lowe, Distinctive image features from Scale-Invariant Keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110, <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [25] T. Luhmann, S. Robson, S. Kyle, J. Boehm, Close-range photogrammetry and 3D imaging, 2013 [Walter de Gruyter].
- [26] E. Gülch, Photogrammetric measurements in fixed wing UAV imagery, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* (2012) 381–386 [Melbourne, Australia].
- [27] T. McCarthy, A.S. Fotheringham, G. O'Riain, Compact Airborne Image Mapping System (CAIMS), *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* 36 (5) (2007) C55 [Padua, Italy].
- [28] M. Hendrickx, W. Gheyle, J. Bonne, J. Bourgeois, A. De Wulf, R. Goossens, The use of stereoscopic images taken from a microdrone for the documentation of heritage – an example from the Tuetka burial mounds in the Russian Altay, *J. Archaeol. Sci.* 38 (11) (2011) 2968–2978.
- [29] A. Bujakiewicz, M. Kowalczyk, P. Podlasiak, D. Zawieska, 3d reconstruction and modelling of the contact surfaces for the archaeological small museum pieces, *Int. Arch. Photogramm. Remote Sensing* 36 (5) (2006) 56–61.
- [30] F. Chiabrando, A. Spanò, Digital wide scale orthoprojections and mapping from low-height aerial images, *J. Cult. Herit.* 10 (2009) e49–e58.
- [31] F. Lafarge, F. Keriven, M. Brédif, V.H. Hiep, Hybrid multi-view reconstruction by jump-diffusion, *IEEE Comput. Vis. Pattern Recogn.* (2010) 350–357.
- [32] R. Roncella, C. Re, G. Forlani, Performance evaluation of a structure and motion strategy in architecture and cultural heritage, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* 38 (2011) 285–292.
- [33] C. Balletti, F. D'Agnano, F. Guerra, P. Vernier, From point cloud to digital fabrication: a tangible reconstruction of ca' venier dei leoni, the Guggenheim museum in Venice, *Ann. Photogramm. Remote Sensing Spat. Inf. Sci.* (2016) 43–49.
- [34] F. Chiabrando, E. Donadio, F. Rinaudo, SfM for orthophoto to generation: a winning approach for cultural heritage knowledge, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* 40 (5) (2015) 91.
- [35] J. García-Gago, D. González-Aguilera, J. Gómez-Lahoz, J.I. San José-Alonso, A photogrammetric and computer vision-based approach for automated 3D architectural modeling and its typological analysis, *Remote Sensing* 6 (6) (2014) 5671–5691.
- [36] M. Capone, M. Campi, R. Catuogno, Gothic churches in Paris St-Gervais et St-Protais image matching 3d reconstruction to understand the vaults system geometry, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* 40 (5) (2015) 423.
- [37] I. Aicardi, F. Chiabrando, N. Grasso, A.M. Lingua, F. Noardo, A. Spanò, UAV photogrammetry with oblique images: first analysis on data acquisition and processing, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* (2016) 835–842.
- [38] F. Chiabrando, A. Lingua, F. Noardo, A. Spanò, 3D Modelling of trompe l'oeil decorated vaults using dense matching techniques, *Ann. Photogramm. Remote Sensing Spat. Inf. Sci.* 2 (5) (2014) 97.
- [39] L. Piermattei, W. Karel, A. Vettore, N. Pfeifer, Panorama image sets for terrestrial photogrammetric surveys, *ISPRS Ann. Arch. Photogramm. Remote Sensing Spat. Information Sci.* 3 (5) (2016).
- [40] F. Nex, F. Remondino, UAV for 3D mapping applications: a review, *Appl. Geomatics* 6 (1) (2014) 1–15.
- [41] N. Grasso, Immagini sferiche per il rilievo di ambienti indoor, *Boll. Soc. Ital. Fotogramm. Topografia* 2 (2016) 1721–1791.
- [42] V. Di Pietra, Fotogrammetria terrestre e tecnologia tablet in scenari post sisma: il caso studio della chiesa di sant'Agostino in Amatrice, *Boll. Soc. Ital. Fotogramm. Topografia* 2017 (2) (2017) 1721–1791.
- [43] M.P. Deseilligny, I. Clery, Apero an open source bundle adjustment software for automatic calibration and orientation of set of images, In: *Proceedings of the ISPRS Symposium, 3DARCH11* (Vol. 269277, 2011, March), 2011.
- [44] F. Fontana, E. Murgia, La domus dei "Putti danzanti" lungo la via Gemina: alcuni elementi dell'apparato decorativo, in: J. Bonetto, J.M. Salvadori (Eds.), *L'architettura privata ad Aquileia in età romana*, Atti del convegno di studio, Padova, Italy, 2012, pp. 297–308.
- [45] F. Chiabrando, A. Lingua, M. Piras, Direct photogrammetry using UAV: tests and first results, *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.* (2013) 81–86.
- [46] F. Chiabrando, F. Fontana, A. Spanò, E. Murgia, 3D models from reality based survey techniques: the domus of "Putti danzanti" in Aquileia, In: *Attidella XII International Conference of AIPMA, Context and Meaning*, Athens, Greece, Vol. XVI-XX, 2013.
- [47] R. Hartley, In defence of the 8-point algorithm, In: *Proceedings of the 5th International Conference on Computer Vision*, 1064–1070, 1995.
- [48] T. Kanade, M. Okutomi, A stereo matching algorithm with an adaptive window: theory and experiment, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (9) (1994) 920–932.
- [49] W.H. Lee, K. Yu, Bundle block adjustment with 3D natural cubic splines, *Sensors* 9 (12) (2009) 9629–9665.
- [50] T. Kanade, O. Amidi, Q. Ke, Real-time and 3D vision for autonomous small and micro air vehicles Decision and control, 2, CDC, 2004, pp. 1655–1662 [43rd IEEE conference].
- [51] Z.F. Wang, Z.G. Zheng, A region based stereo matching algorithm using cooperative optimization, *Comp. Vis. Pattern Recogn.* (2008).
- [52] Z. Zhang, A flexible new technique for camera calibration, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11) (2000) 1330–1334.