AUTOMATIC ORIENTATION OF IMAGE SEQUENCES FOR 3D OBJECT RECONSTRUCTION: FIRST RESULTS OF A METHOD INTEGRATING PHOTOGRAMMETRIC AND COMPUTER VISION ALGORITHMS

L. Barazzetti , M. Scaioni

Politecnico di Milano - Lecco Campus, via M. d'Oggiono 18/a 23900 Lecco, Italy
(luigi.barazzetti, marco.scaioni)@polimi.it

KEY WORDS: Automation, Matching, Modelling, Orientation, Processing, Sequences

ABSTRACT:

Image sequence analysis for 3-D modeling is an aim of primary importance in Photogrammetry and Computer Vision. Both disciplines approach this issue, but with different procedures. The result is the reconstruction of a 3-D model of the captured object featuring different characteristics in terms of accuracy and time needed to built it up. The goal of this work is the implementation of an algorithm which is capable of computing the orientation parameters of an image sequence in an automatic way. Input elements are the calibration parameters of the camera and the images. Particular attention must be paid during the sequence acquisition: each image must present a good overlap with the next images. The algorithm makes use of computer vision and photogrammetric techniques in an integrated approach for the sequence processing, in order to obtain the exterior orientation parameters and the 3-D coordinates of the automatically matched points. The main advantage of a combined use of the methods of both these disciplines is the possibility to solve some typical problems of one with the other, and vice versa. The process starts with the automatic matching between image triplets by using a feature based matching followed by blunder removal through robust estimation methods of the camera poses. Then a progressive resection alternated to triangulation is carried out to join to the first triplet some further images belonging to other triplets. This task is performed by applying a "Structure & Motion" technique based on collinearity equations. The developed algorithm has been successfully tested with synthetic and real data.

1. INTRODUCTION

Automatic reconstruction of 3-D objects with image sequences is an issue still open in Computer Vision (CV) and Photogrammetry. The demand of 3-D models is increasing in several fields, such as Cultural Heritage, Computer Graphics, Robotics and many others. The features of a 3-D model are highly dependent on the use of the model, and can be very variable in term of accuracy and time for their creation.

In this last years both CV and photogrammetric communities afforded the reconstruction problem by using different methods to solve for the same problems, such as camera calibration, orientation, object reconstruction and modelling. The same terminology used for addressing the same task in both disciplines is sometimes diverse. On the other hand, the integration of methods and algorithms coming from them can be used to improve both. Indeed, in photogrammetry the accuracy is a goal of primary relevance and implicates particular attention since the image acquisition phase. Calibration parameters are always used to model the distortion, especially when non-metric cameras are used, fact that is commonplace in modern digital Photogrammetry. The resolution of the images for close-range applications is often larger than 6-8 mega-pixels and seems increasing. In some cases the adopted images can feature a resolution of some tens of mega-pixels. The orientation is based on a photogrammetric bundle adjustment by using collinearity equations and tie points are manually measured. The use of coded targets placed on the object allows one to improve the process automation.

Although many photogrammetric digital stations have tools for automatic orientation of the image block based on *markerless* tie points, they work especially with images in normal configuration (typical of an aerial block, where today the problem can be considered solved). Applications of close-range Photogrammetry often require convergent images and these packages rarely can be used with success (Remondino & Ressl, 2006). Some photogrammetric low-cost software (e.g. iWitness, PhotoModeler) can automatically orient a close-range image block by using coded targets. On the other hand, the analysis of an image sequence with markerless tie points appears as a still open issue, especially when wide baseline images are used. First results on automatic relative orientation can be found in Forlani *et al.* (2005), Hao & Mayer (2003), Läbe & Förstner (2006), and Remondino & Ressl (2006).

In the field of Computer Vision the automatic orientation approach is based on the use of low resolution sensors (generally less than 1 mega-pixels) and a huge number of images can be analyzed. In this sense it is interesting somehow the effective number of pixels can be similar in both disciplines, but with a complete different point a view. The literature is vast and examples can be found in Beardsley *et al.* (1996), Dick *et al.* (2004), Pollefeys *et al.* (1999). A significant difference with Photogrammetry is the lack of calibration, that means that distortion parameters but also inner orientation parameters (principal distance and principal point coordinates) are not necessary. This choice is motivated by the need to analyse image sequences where the knowledge of the calibration is rarely achievable, thus without any limitation in the image acquisition. For this reason the algorithms used in CV are termed *uncalibrated* (Hartley & Zisserman, 2004). The reconstruction with uncalibrated methods makes use of multiple-view geometry and presents a projective ambiguity which can be removed with self-calibration techniques (the term self-calibration has a different meaning in Photogrammetry) to obtain a *metric* reconstruction where the only ambiguity is represented by a scale factor. A fundamental advantage is the possibility to use different sensors or a zoom lens without knowledge of them in the analysis phase.

Another important aspect is the necessity to automatically create a model in order to speed up the process. In some cases real time (or quasi real time) reconstruction is necessary.

A typical reconstruction in CV (but not limited to) starts with the analysis of *image pairs* (e.g. Pollefeys *et al.*, 2004) or *triplets* (e.g. Hao, 2003). Then a "Structure and Motion" (S&M) approach is performed to contemporarily determine camera poses and coordinates of object points. The reconstruction is completed with a *dense matching* process to build a 3-D surface model, that can be then texturized for a photorealistic visualization. The number of frames that it is possible to analyse can be really huge.

Some commercial packages that use an approach based on automatic image tracking are available (e.g. Bonjou). These software allow the estimation of the path of a camera in case images present a very short baseline. For this reason video-cameras are generally used. With a S&M approach the trajectory of the camera is automatically calculated from some tens of points (or hundreds) matched with a tracking algorithm. Most applications of these software regard the virtual reality and entertainment industry.

Although the method used in CV appear really interesting, their applications in Photogrammetry is quite limited. Many differences separate CV and photogrammetric communities, starting from the size of the used sensors, the presence or absence of calibration and the adopted orientation procedures. A description of the main difference between both disciplines is presented in Förstner (2002), where common interests and a possible interaction between both fields are presented.

2. THE DEVELOPED ALGORITHM FOR IMAGE SEQUENCE ORIENTATION

2.1 Overview

The algorithm presented in this paper allows the automatic orientation of an image sequence. Input elements are the calibration parameters of the adopted camera and the images. A basic pre-requisite for the correct working of the method is the order of the images; in fact, each image (*i*) must present an overlap with the next two images (*i*+1 and *i*+2). The automatic matching procedure is carried out between image triplets, in a way that a point appears on 3 or more images. This condition ensures a good redundancy and makes the method more robust.

The algorithm makes use of typical methods of Photogrammetry and CV in an integrated approach. On the other hand, the use of an image sequence is not strictly mandatory, being the algorithm capable to work with any block made up of convergent shots. The other functions can be used for the analysis of generic image block and the order of the image can be unknown. In this case a matching between pairs is carried out to determine tie points across all combinations of images in the block. The main difference between both algorithms is the time necessary during the F.B. Matching stage, that rapidly increases with the total image number. In figure 1 the computational time of the algorithm presented in this paper (in blue) is compared to that of the other algorithm (in red). The image used for this test have a resolution of 800×535 pixels. As it is possible to see, the sequence algorithm has a linear computational time, while the time of the second algorithm becomes of hours if the image number increases.

The method allows one to determine the orientation parameters of a sequence composed of some tens of images. Calibration parameters must be known for every image. The orientation procedure is based on collinearity equations without the use of ground control points, and can be assumed as a relative multi-image orientation. The image points are detected with SIFT (Lowe, 2004) followed by a robust outlier detection based on CV camera poses estimation techniques. The choice of this matcher is motivated by its good invariance to changes of scale, translation, rotation (around optical axis) and illumination. It provides good results also with moderate affine transformations and probably is the best matcher today available (Mikolajzyk & Schmid, 2003). The survey is then completed with dense matching algorithms to create a point cloud or with manual editing.

Globally, the process can be divided in four main parts which are described in the following sections.

The algorithm is implemented in standard C code (for the part concerning Feature-Based matching), Fortran 77 (the L.S. Matching algorithm; for details about its implementation see Gianinetto & Scaioni, 2008), and MATLAB for other computations.

The user interface of the algorithm is presented in figure 2, where only the functions indicated with a red rectangle refer to image sequences.

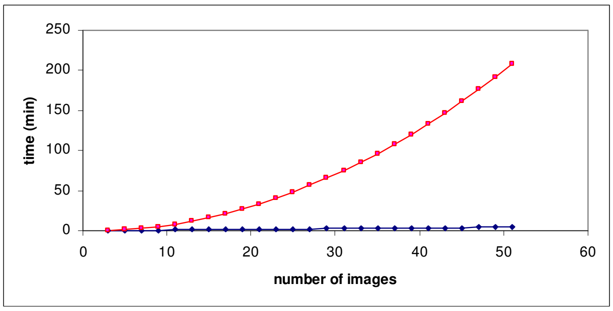


Figure 1. Computational time for the matching of an image sequence with the algorithm here presented (in blue) and with the generic image block algorithm (in red).

2.2 Feature-Based Matching

As the sequence composed of images where each image presents an overlap with the next two images, the whole sequence can be split in *n*-2 triplets (1,2,3; 2,3,4; …). For each triplet a matching with SIFT operator is carried out, in a way to determine homologous points. The use of this feature extractor is motivated by the optimal invariance of the method to scale variation and rotation around the optical axis of the cameras, which are typical of terrestrial shots. The method is also robust enough to changes in illumination. In any case, after matching some outliers could be erroneously found.

The SIFT operator is applied to all images of each triplet. Then, one image is selected as "master" and each point found on these is looked for its homologous point on the other two images ("slaves"). At the end of point transfer inside the triplet, only threefold points are stored.

In order to remove outliers from the dataset, a robust estimation of the fundamental matrix *F* by using Least Median of Squares (LMedS) method is carried out on each image pair (Scaioni, 2000). The fundamental matrix *F* is the geometrical representation of the epipolar geometry in CV, and is a 3×3 rank 2 homogenous matrix which satisfies the condition:

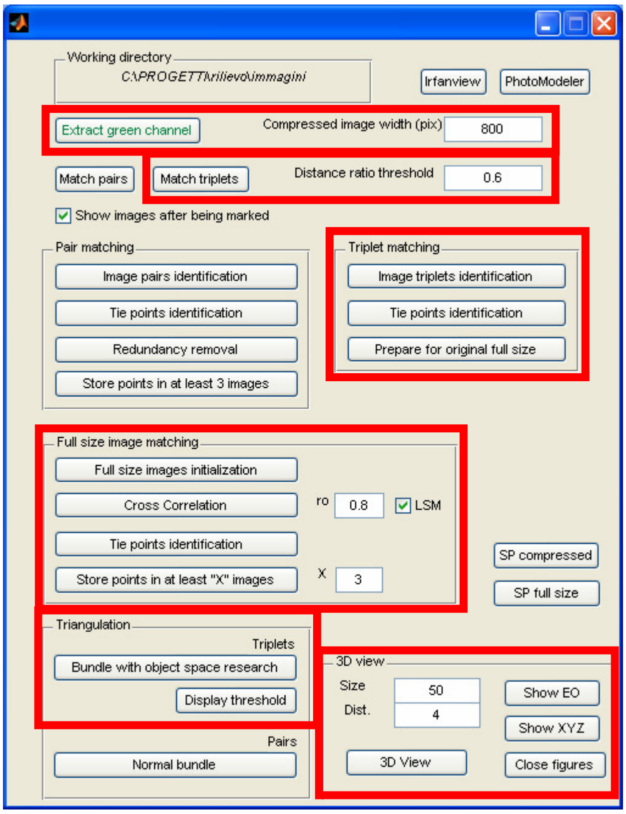
$$\mathbf{x}'^{T}_{i}\, F\mathbf{x}_i = (x'_i, y'_i, 1)F\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = 0 \qquad (1)$$

where $\mathbf{x}_i$ and $\mathbf{x'}_i$ are the homogenous image coordinates of an homologous point on the two images. The epipolar lines of an image pair can be calculated through *F* by using the following relations:

$$\mathbf{l}'_i = F\mathbf{x}_i \quad , \quad \mathbf{l}_i = F^T\mathbf{x}'_i \qquad (2)$$

Eq. (1) can be used to estimate *F*. In fact with a set of matches (at least 8 points) the elements of *F* can be written in terms of image coordinates as follows:

$$\begin{aligned} & x_i' x_i f_{11} + x_i' y_i f_{12} + x_i' f_{13} + y_i' x_i f_{21} + y_i' y_i f_{22} + \\ & + y_i' f_{23} + x_i f_{31} + y_i f_{32} + f_{33} = 0 \end{aligned} \qquad (3)$$


Working directory: C:\PROGETTI\rilievo\immagini

Compressed image width (pix): 800

Distance ratio threshold: 0.6

☑ Show images after being marked

ro: 0.8 ☑ LSM

X: 3

Size: 50

Dist.: 4

Figure 2. The graphical user interface of the algorithm. In red the functions used for a sequence analysis, while the others are used for a free-shape image block.

Eq. (3) corresponds to the following homogenous linear system:

$$A\mathbf{f} = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_i'x_i & x_i'y_i & x_i' & y_i'x_i & y_i'y_i & y_i' & x_i & y_i & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}\mathbf{f} = \mathbf{0} \qquad (4)$$

where the vector **f** contains the elements of *F*. The system can be solved by means of Singular Value Decomposition (SVD), which minimize ǁ *A***f** ǁ under condition ǁ **f** ǁ =1.

System (4) presents numerical stability problems if pixel coordinates are used, thus a preliminary normalization of the coordinates is suggested. The normalization consists in a translation and scaling of the image coordinates and leads to an improvement of the conditioning number of *A*.

After the solution of vector equation (3), the matrix *F* should be substituted with an opportune matrix *F'* which satisfies the singular constraint of *F*. This can be done by minimizing the *Frobenius norm* || *F'-F* || with the condition det(*F*) = 0, via SVD again.

Eq.s (1) and (2) allow the robust estimation of *F* and the outlier removal (Scaioni, 2000). The method here used to estimate *F* is LMedS, which is based on the estimation of the median of square of residual errors (Rousseeuw & Leroy, 1987). Its application to outlier detection between image pairs is performed by considering the distance between an image point and the corresponding epipolar line. A closed solution for this problem is not known and a random sampling technique is implemented.

The application of this method to an image triplet is based on the estimation of two fundamental matrices. Although the best elements for an image triplet analysis is the *trifocal tensor* and not the fundamental matrix (which encapsulates the geometry of an image pair only), many tests demonstrated that the fundamental matrix is more than sufficient for a correct outlier rejection in many real applications. In any case, the trifocal tensor will be implemented in the algorithm to increase the stability soon.

In figure 3 an example of matching between triplets is shown. The entire image sequence here used is composed of 61 images acquired in order to outline two concentric rings. The imaged object is a statue located in the municipality of Malesco (Italy).

Figure 3. Three consecutive triplets of an image sequence and the extracted tie points by F.-B. Matching.

Before starting F.-B. Matching, all images were automatically compressed into a resolution of 640×480 pixels and only the green channel was stored. The matching process with SIFT and the outlier removal step on the compressed image required 300 seconds roughly.

2.3 Reconstruction initialization

Once all image triplets were matched the reconstruction of the object can be initialized. So far, a set of triplets which internally share a set of tie points among is available. However, each triplet is independent from the others; only three-fold tie points exist in the block. The next step of the procedure to achieve the external orientation of the whole block is a triplet concatenation. To achieve this goal, the first triplet is taken as starting point for a progressive concatenation of all the others.

First of all, the algorithm computes a relative orientation between the three images, in order to determine the object coordinates (with a scale ambiguity) and the external orientation parameters.

The main problem during this step is due to the linearization of the collinearity equations and the need of good approximate orientation parameters. A first test was carried out by using the *essential matrix E*, which is the specialization of the fundamental matrix *F* when calibration is known. It allows a direct metric reconstruction. Inner orientation parameters and *F* allow the estimation of *E* as follow:

$$E = K^T F K = \begin{bmatrix} f & 0 & x_P \\ 0 & f & y_P \\ 0 & 0 & 1 \end{bmatrix}^T F \begin{bmatrix} f & 0 & x_P \\ 0 & f & y_P \\ 0 & 0 & 1 \end{bmatrix} \qquad (5)$$

where *K* is called *calibration matrix*. In CV community the essential matrix was attributed to Longuet-Higgins (1981), but Thompson (1959) had introduced this method more than twenty years before. After some test (Barazzetti & Scaioni, 2008) we verified that a reconstruction from an image pair with the essential matrix was not accurate enough and the possibility to use it for determining approximate values (instead of a more traditional photogrammetric approach) is not an optimal choice when accurate measurements are necessary. The words of Horn (1990), presented also in Fraser (2005), describe the limits of an orientation procedure with the essential matrix: "Overall, it seems that the two-step approach to relative orientation, where one first determines an essential matrix, is the source of both limitations and confusion." In any case the essential matrix results useful for the automatic removal of points which lie on the background, resulting in weaken the block geometry; this new function will be soon included in the algorithm.

The orientation of the first image triplet is carried out in two step by using a photogrammetric approach. At the beginning only two images are used. They form a stereo pair with a low convergence (in case of a sequence) and allow the creation of a non-linear system of equation based on the co-planarity model (Mikhail *et al.*, 2001). The hypothesis of quasi parallel optical axis makes the relative orientation simple, avoiding the challenge to find good approximate values for highly convergent images. At the end of relative orientation, coordinates of tie points in an arbitrary "model" reference system are computed by intersection.

In the second step the orientation of the third image is recovered by using a *space resection* on the "model" coordinates of tie points. Lastly a bundle adjustment is carried out to improve the estimation of parameters and coordinates for the whole triplet.

2.4 Concatenation of the triplets

The reconstruction obtained from the first triplet is the starting point of a progressive resection alternated to triangulation in order to join each image along a "Structure & Motion" approach based on collinearity equations.

The first triplet (image 1,2,3) shares with the second triplet two images (2,3), which are now already oriented. For this reason the object coordinates of the image points on the second triplet can be determined by using a triangulation. Starting from these points, the orientation parameters of the image 4 can be determined by using space resection. In this step the algorithm recognizes also points which lie in both triplets; by this approach, the number of rays-per-point might become higher than 3, with consequent increment of local redundancies. Lastly a bundle adjustment is carried out to improve the estimation of all parameters. The procedure is repeated as far as the complete concatenation of all image triplets. In figure 4 camera poses and the object coordinates of tie points for the statue of figure 3 are shown.

For a closed sequence the first image is matched with the second and the last one, in order to strengthen the geometry.

This approach is quite similar to that used in CV, but the orientation procedure is based on collinearity equations.

Before analysing the full resolution images, it is convenient to orient a block made up of images at lower resolution. In fact the F.B. matching with SIFT is applied only on the compressed images because of computational cost of the algorithm itself.

As it is possible to see in figure 4, images can be distributed on two concentric rings with a different radius and height from ground. This means that images present a sidelap and tie points can be found also between images which are not consecutive. In order to link together tie points on not adjacent images, a research in the object space was developed. This solution is described in the section 3.2.

In order to transfer tie points from the low resolution to the full resolution images, a two-steps area based matching technique is applied. The image coordinates on the original image are whose detected by SIFT multiplied by the compression factor. Firstly a preliminary *normalized cross-correlation* is computed, then the Adaptive L.S. Matching (Gruen, 1985) allows one to define the final coordinates of tie points.

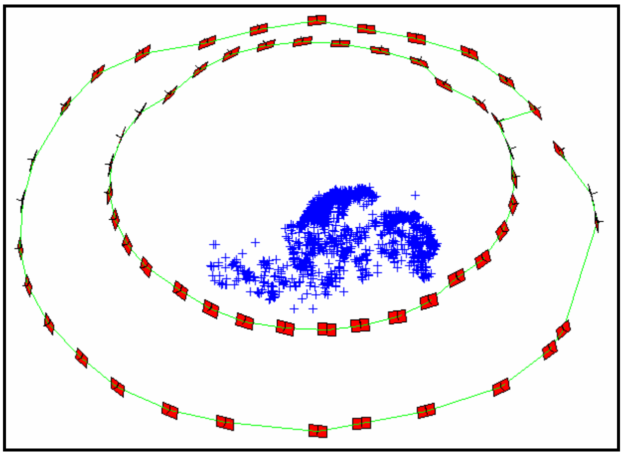


Figure 4. Camera poses and 3-D coordinates of tie points.

Particular attention must be paid in this phase. In fact, this procedures is carried out after the ordering and storage of the data into a tree structure, in a way that the matching is performed by fixing the point on the first image where it appears (in a progressive order in the sequence) and it is moved only on the remaining images.

After the image coordinates refining, the original images can be definitely oriented in an arbitrary reference system, considering that no GCPs nor external constraints were used.

2.5 Completion of the survey

After the determination of the orientation parameters and the object coordinates of the automatically matched tie points, the object reconstruction can be completed with manual or automatic procedures.

At the present time the algorithm has not a user friendly interface. For these reasons a convenient choice is to export the results in a format which is compatible with a low-cost photogrammetric software. The output of the algorithm is then automatically exported into PhotoModeler, which has all the basic functions to carry out this task. Moreover, the version "Scanner" of this package has a module for *dense surface matching* and allows the creation of point clouds from images; in figure 5 is shown the result obtained for the block in figures 3 and 4.

In addition, PhotoModeler offers also the possibility to perform the absolute orientation of the model, i.e. to compute a 3-D conformal transformation. Another advantage of the use of a commercial package to complete the survey is the possibility to export the data in several formats.

Figure 5. The extracted point cloud (2.264.000 points) for the statue of the example adopted in these sections.

3. TESTS WITH DIFFERENT BLOCK GEOMETRY

In this section some applications with different image block geometry are presented. The main difference between them regards the possibility to find tie points not only in consecutive images (like in a ring sequence) but also between images which present a sidelap (like an aerial block).

3.1 Block with a ring shape

The example proposed in this section is the popular "dinosaur sequence" of the University of Hannover, which is well-known in the CV community because it is used to test and to compare different algorithms. It is a turntable sequence of 36 images taken under a circular motion, so that images are equally spaced of 10° (Fitzgibbon & Zisserman, 1998). The resolution of the image is 720×576 pixels and no calibration is provided.

For the application with the proposed method only 18 images were used in order to improve the geometry of the block and an approximate calibration was fixed (no distortion parameters, principal point in the image centre, focal length a priori estimated). An automatic matching was carried out by including the first image two times (at the beginning and at the end) in order to close the sequence (Fig. 6).

The elaboration of this sequence required no more than 3 minutes, and camera poses are presented in figure 7.

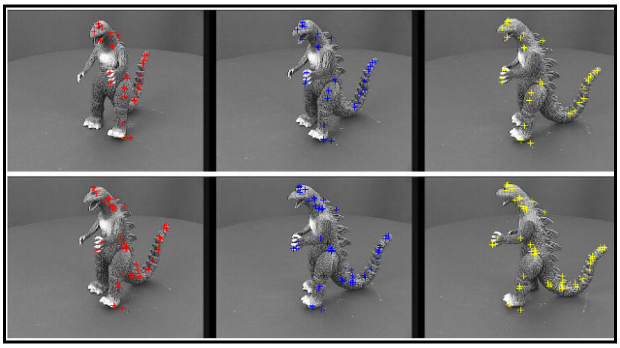


Figure 6. Some triplets of the "dinosaur sequence".

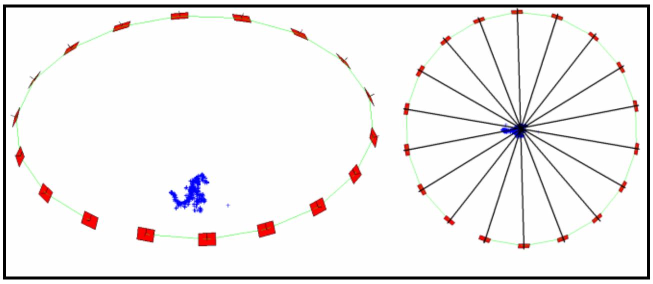


Figure 7. Camera poses and 3-D points. Even without a good calibration images result equally spaced.

3.2 Block with several strips

A block made up of several strips contains tie points also between images on different strips. This is the typical case of an aerial block, but also terrestrial applications could require this kind of block configuration.

The block used in this test is composed of 49 images on 6 strips. The object is a bas-relief located on Alexander Manzoni's monument in Lecco (Italy).

The main problem during the analysis of a sequence with strips is the automatic detection of points between non-consecutive images. The procedure here developed to solve this task is based on a research in the object space. In particular, after the connection of a new triplet, the distance between each ray back projected from the 3-D point to the image and any other ray is calculated. Rays closer than a prefixed threshold are assumed to be homologous. To avoid the identification of image points which lie on only 2-3 images and to improve the robustness of the method, a minimum limit of 4 rays was fixed. The threshold value depends on the precision of the object points and varies after the addition of a new triplet. This dynamic threshold t is estimated by considering the precision of the points along the axis of the object reference system by using an index similar to the GPS *Position Dilution of Precision* (PDOP):

$$t = \max\left[\sqrt{\sigma_{Px}^{2} + \sigma_{Py}^{2} + \sigma_{Pz}^{2}}\right] \qquad \forall P \qquad (6)$$

To remove possible errors in this step, the linear correlation coefficient between the matched points is then estimated.

The average theoretical precision of the objects coordinates was $\sigma_x = \pm 1.3$ mm, $\sigma_y = \pm 1.5$ mm and $\sigma_z = \pm 1.9$ mm (z represents the depth) and the object is approximately 1500×1200 mm.

In figure 8 the camera poses of this survey are shown.

The main advantage of a research in the object space is related to the speed of the procedure. Another possibility could be a more complete image matching, where each image is matched with all the others. This solution was implemented in another algorithm which can analyse an image block acquired in a random manner by using only image pairs (Fassi *et al.*, 2009). It demonstrated a little better accuracy but is also much more time consuming and does not appear convenient when the images form a sequence.

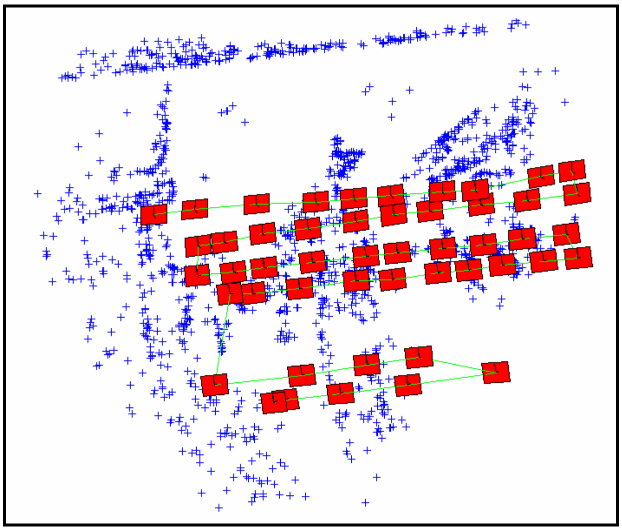


Figure 8. Block geometry of a survey of a bas-relief.

4. APPLICATION TO CULTURAL HERITAGE AND ACCURACY ANALYSIS

4.1 The survey of the temple G1 in My Son

The method was used to perform the reconstruction of the temple "G1" in My Son (Vietnam, see Fig. 9). The temple was named by UNESCO a World Cultural Heritage and a 3-D model is necessary to prove its current condition with a numerical analysis on its stability. As it is possible to see in figure 9, the temple is very damaged and restoration works are necessary.

Figure 9. The temple G1 in My Son.

The camera used is a Nikon D80 with a 18 mm lens, that was calibrated just before the acquisition of the image of the temple. The block is composed of 41 images taken under an approximately circular motion around the temple (Fig. 11). No terrestrial laser scanning instrument was available for this survey, and the 3-D model was created by using only digital photogrammetry.

On the facades of the temple some targets were applied to execute a survey with a semi-automatic approach, in order to compare the results of the implemented automatic algorithm. These targets were measured also by a Total Station; unfortunately, GCPs on each side were measured into an independent reference system w.r.t. the other sides.

The complete survey and the model are described in Barazzetti *et al.* (2009), while in this paper only the results with the implemented algorithm are presented.

4.2 The survey with the implemented algorithm

The original images were compressed to a resolution of 1500×1004 pixels and only the green channel was used. The F.-B. Matching procedure with SIFT needed 4 hours roughly and more than 20000 points were automatically identified. Points on the background were then removed by using a mask.

The analysis was performed by using only the F.-B. Matching without a following L.S.M. refining. Starting from the measured image points the first triplet was oriented, and the remaining ones were connected by the S&M approach based on collinearity equations and progressive bundle adjustments. The first image was joined with the last one to close the whole sequence. This operation required no more than 1 hour and became progressively slow with the introduction of each new triplet.

At the end of the orientation process, a scale factor was fixed by using the distance between two points on the South façade. Camera poses are shown in figure 11.

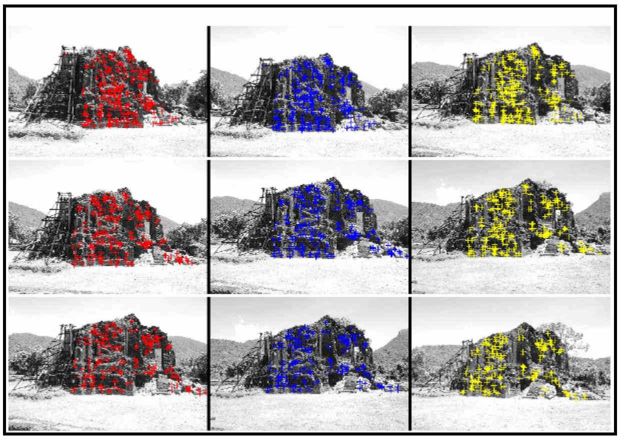


Figure 10. Three matched triplets of the temple sequence.

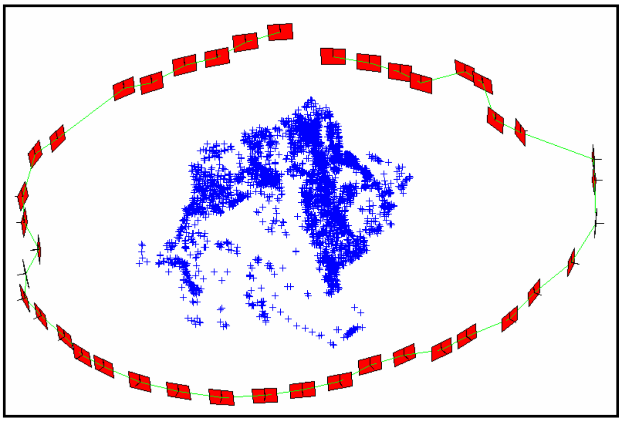


Figure 11. Camera poses of the temple sequence.

4.3 Accuracy analysis of the algorithm

The coordinates of the targets on each facades were compared with the Total Station measurements. As the measurements of the Total Station are not connected through a geodetic network, the comparison with the photogrammetric measurement was performed by analyzing the four façades independently. First of all, the interdistances between all targets were computed obtaining the result shown in table 12. As it is possible to see, only the South façade is bias-free. This is probably due to the constraint fixed only on this facade. In any case, the standard deviation is always lower than 8 mm.

To test the coordinate accuracy of the photogrammetric model a transformation was necessary. This is due to the independent measurements of the Total Station, which are independent for the four facades.

Facade	**North**	**South**	**East**	**West**
mean (mm)	5	0	10	14
std.dev (mm)	3	5	7	8
max (mm)	13	14	27	34
min (mm)	-1	-26	-4	-17

Table 12. Interdistance comparison between Total Station and photogrammetric results with the implemented algorithm.

For this reason, a comparison on the coordinates was performed by estimating a rototranslation with variation in scale to transform the coordinates of the model to be compared with those of the four independent stations. The results are shown in table 13. In this case, the use of a scale factor in the transformation allows a bias removal and demonstrates an accuracy of few millimetres, surely more than sufficient for the purpose of the survey.

The precision of the coordinates results better than 10 mm and more than sufficient for the purpose of the survey. A 3D model was then created by using the dense surface matching of PhotoModeler Scanner by exporting the calculated orientation parameters and 3-D points. The point cloud is shown in figure 14.

Facade	**North**			**South**		
	x	**y**	**z**	**x**	**y**	**z**
mean (mm)	0	0	0	0	0	0
st.dev (mm)	2	3	1	6	5	4
max (mm)	2	5	2	15	14	5
min (mm)	-4	-6	-2	-12	-6	-8
Facade	**East**			**West**		
	x	**y**	**z**	**x**	**y**	**z**
mean (mm)	0	0	0	0	0	0
st.dev (mm)	4	6	3	2	6	2
max (mm)	6	10	8	3	8	3
min (mm)	-7	-8	-3	-4	-13	-2

Table 13. Comparison with Total Station coordinates

Figure 14. The point cloud extracted with PhotoModeler Scanner by using orientation parameters computed with the proposed S&M algorithm.

5. CONCLUSIONS AND OUTLOOK

An algorithm for automatic orientation with markerless images sequences through a Structure and Motion strategy has been presented here. It allows one to carry out the object reconstruction with a precision similar to that achievable with manual measurements. The algorithm makes use of photogrammetric and computer vision algorithms in an integrated approach to obtain an accurate 3-D model.

New functions will be implemented soon into the procedure. These include a new *local descriptor* implementation, a *multi-image L.S. Matching*, an automatic removal of the points on background (by using the *essential matrix*), the robust *trifocal tensor* estimation and a strategy for automatic point decimation. This last function consists in an intelligent identification of a

minimum dataset of tie point in order to perform the orientation of the images. In this case, a minimal set of image points is extracted from all matched points considering the position of each point on the image, the number of image where the point lies and the distance to each other point on the same image.
The algorithm is not implemented to be fast but a real-time analysis is not a goal for this work. In any case, a computational time of few hours is good for many photogrammetric applications.
Lastly, this algorithm will be associated with another one which allows the elaboration of images taken under a random geometry which was only mentioned in this paper.

ACKNOWLEDGEMENTS

Authors would like to thank Prof. Luigia Binda and Marco Cucchi (Dept. of Structural Engineering, Politecnico di Milano), for the availability of the images of the temple G1 in My Son.

REFERENCES

Barazzetti, L., Binda, L., Cucchi, M., Scaioni, M., Taranto, P., 2009. Photogrammetric reconstruction of the My Son G1 temple in Vietnam. *IAPRSSIS*, 37 P5(W1)

Barazzetti, L., Scaioni, M., 2008. SIFT and multiple view geometry for automatic 3D modelling with image sequences. *Proceedings of the National Conference ASITA,* pp.269-275

Beardsley, P, Torr, P. and Zisserman, A., 1996: 3D model acquisition from extended image sequences. Proceedings of ECCV'96, Lecture Notes in Computer Sciences, 1065: 683-695

Dick, A. R., Torr, P.H.S, Cipolla, R., 2004. Modelling and Interpretation of Architecture from Several Images. *IJCV*, 60(2): 111-134

Fassi, F., Barazzetti, L., Branka, C., Brumana, R., Scaioni, M., 2009. RC-Heli and Structure from Motion techniques for the 3D reconstruction of a Milan Dome spire. *IAPRSSIS*, 37 P5(W1)

Fitzgibbon, A.W., Zisserman A., 1998. Automatic camera recovery for closed and open image sequences. *Proceedings of the 5th European Conference on Computer Vision,* 1: 331-326

Fraser, C., 2005. Network Orientation Models for Image-Based 3D Measurement. *IAPRSSIS*, 36 P5(W17).

Forlani, G., Roncella, R., Remondino, F., 2005. Structure and motion reconstruction of short mobile image sequences. VII Conference on Optical 3D Measurement Techniques

Förstner, W., 2002. Computer Vision and Photogrammetry – mutual questions: geometry, statistics and cognitions. *Bildteknik/Image Science, Swedish Society for Photogrammetry and Remote Sensing*, 151-164

Gianinetto, M., Scaioni, M., 2008. Automated geometric correction of high-resolution pushbroom satellite data. *PE&RS*, 74(1): 107-116

Gruen, A., 1985: Adaptive least square correlation: a powerful image matching technique. South African Journal of PRS and Cartography, 14(3), 175-187

Hartley R., Zisserman A., 2004. *Multiple View Geometry in computer vision*. Cambridge University Press

Hao, X., Mayer, H., 2003. Orientation and Auto-Calibration of Image Triplets and Sequences *IAPRSSIS*, 34(3/W8): 73-78

Horn, B.K.P., 1990. Recovering baseline and orientation from essential matrix. MIT AI Memo

Läbe, T., Förstner,W., 2006. Automatic relative orientation of images. *Proceedings of the 5th Turkish-German Joint Geodetic Days*, Berlin, Germany

Lounguet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections, *Nature,* 292: 133-135

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *IJCV,* 60(2): 91-110

Mikhail, E.M., Bethel, J., McGlone, J.C., 2001. *Introduction to Modern Photogrammetry*. John Wiley & Sons, Inc.

Mikolajczyk, K., Schmid, C., 2003. A performance evaluation of local descriptors. *Transaction of Pattern Analysis and Machine Intelligence*, 27(10) 1615-1630

Pollefeys, M., Koch, R., Van Gool, L., 1999: Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. *IJCV,* 32(1): 7-25.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R., 2004. Visual modeling with a hand-held camera. *IJCV,* 59(3): 207-232

Remondino, F., Ressl, C., 2006. Overview and experience in automated markerless image orientation. *IAPRSSIS*, 36(3): 248-254

Rousseeuw, P.J., Leroy, A.M. (1987). *Robust Regression and Outliers Detection*. John Wiley, New York

Scaioni, M., (2000). Robust estimation of relative orientation through the fundamental matrix. *IAPRSSIS*, 32(W8/2): 253-259

Thompson, E.H., 1959. A rational algebraic formulation of the problem of relative orientation. *Photogrammetric Record*, 3(13): 55-59