SIMULTANEOUS REGISTRATION OF GNOMONIC PROJECTIONS AND CENTRAL PERSPECTIVES

LUIGI BARAZZETTI (luigi.barazzetti@polimi.it)
MATTIA PREVITALI (mattia.previtali@polimi.it)
Politecnico di Milano, Milan, Italy

MARCO SCAIONI (marco@tongji.edu.cn)
Tongji University, Shanghai, P.R., China

INTRODUCTION

IN RECENT YEARS different algorithms and procedures have been developed to extract 3D information from *panoramic images* (or, simply, *panoramas*: see, for example, Di et al., 2004; 2008; Schneider and Maas, 2006; Amiri Parian and Gruen, 2010; Fangi, 2010). In fact, when multiple panoramas of the same scene, gathered from different viewpoints, are available, a 3D reconstruction can be carried out from the set of image points and some external information such as ground control points.

A single panorama may combine several images, and the general view of the scene allows for better recognition of the object being investigated. The high-resolution content (gigapixel images) of multiple panoramas can provide for highly detailed 3D reconstructions, and the large field of view may reduce the number of camera stations needed.

The extraction of 3D metric data from panoramas follows the same concept used for standard central perspective imagery (spatial intersection of corresponding rays), although

the equations rely on different geometric models. Some of the most popular camera models for panoramas are based on *spherical* (Fangi and Nardinocchi, 2013) and *cylindrical* (Schneider and Maas, 2004; 2005) projections.

Spherical panoramic images are a particular type of panorama with a large field of view (up to 360° × 180°). From a practical standpoint, it is quite easy to create a spherical panorama. Several photographs taken from the same point are acquired by rotating the camera around its perspective centre. If there is sufficient overlap between the images, feature-based matching algorithms can be used to extract corresponding points. Then, the images are stitched and mapped with the so-called *latitude–longitude* projection (*equi-rectangular* projection). This procedure is also implemented in several commercial software packages. Photogrammetric data processing with this kind of image has proved useful for the reconstruction of indoor environments such as rooms, courtyards or large buildings (Pisa et al., 2010).

Cylindrical panoramic images can be derived from the processing of several central perspective (also referred to as "pinhole") images (Brown and Lowe, 2003), although they are generally acquired using special linear-array rotating cameras. These sensors can obtain very high metric performances. Complete, precise and detailed 3D reconstructions are feasible through space intersection if multiple-oriented panoramas with good distributions in space are available (Luhmann and Tecklenburg, 2002; Schneider and Maas, 2004).

On the other hand, most software packages used in close range photogrammetry do not entail such camera models, resulting in the limited use of panoramic imagery for 3D reconstruction when compared to standard central perspective images captured by digital, consumer-grade cameras. For this reason an alternative automatic approach, able to extend the rigorous photogrammetric processing pipeline to a particular type of panorama, was proposed by Barazzetti et al. (2013). Here, the *gnomonic projection* is adopted as a geometric model for panoramic images, being a particular kind of panorama that provides a superior level of detail (gigapixel images) and still follows the central perspective camera model. The use of such a projection model allows the problem of the external orientation of panoramic images to be reduced to the same mathematical formulation normally adopted in photogrammetry for pinhole images, given a suitable set of inner orientation and additional parameters for the correction of lens distortion. This means that panoramic images and standard images can be registered within the same integrated bundle adjustment.

The aim of this research was the implementation of an improved image matching and orientation procedure, able to simultaneously handle gnomonic projections and central perspective images. The former are exploited for their better geometric resolutions and the larger image sizes. The latter are used instead to strengthen the block geometry and to set up the three-dimensional datum, including external or inner constraints. Image orientation can be carried out by using different methods for the extraction of tie points, namely, interactive (manual) measurements together with target-based and markerless procedures. All observations are then integrated into a unified bundle adjustment.

The main advantage of the proposed methodology concerns the improvement of image resolution. Modern digital single lens reflex (SLR) cameras adopted in close range photogrammetry may feature a geometric resolution of 20 to 30 megapixels. The pixel size usually ranges from 1 to 2 μm in consumer-grade cameras and up to 6 to 9 μm in professional cameras. Gnomonic projections created with the procedure proposed in this paper may feature a much larger resolution, usually 10 to 50 times better than the one achieved with a standard set-up for close range photogrammetric surveying. If the project requires the documentation of large objects with fine details, the use of gnomonic

projections can be a valid alternative to a large block including several standard central perspective images.

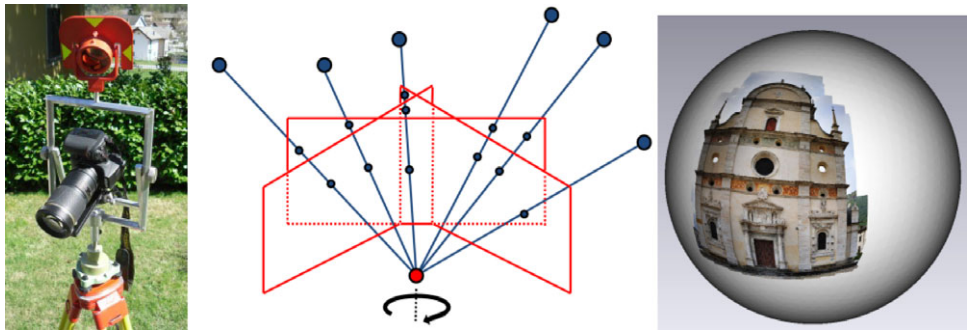
The structure of the paper is organised as follows. First, a geometric model based on *gnomonic projection* is described to map panoramic images into larger central perspective images. These images are then included in an integrated bundle adjustment for computing exterior orientation. Next, techniques to improve the degree of automation of image orientation are illustrated. Some example applications illustrating the procedures discussed in the paper, including some tests for the evaluation of accuracy, are then reported to show the advantages and disadvantages of the methodology.

Generation of the Gnomonic Projection

Gnomonic projections (where the centre of projection is the centre of a sphere) are high-resolution panoramic images based on the central perspective (or pinhole) camera model. Such projections can be generated by using low-cost equipment made up of a special rotating head (Fig. 1) carrying a standard SLR camera equipped with a telephoto lens. The head is intended as a mechanical tool that allows the camera to rotate around the perspective centre.

The description of the head and its mechanical calibration (for example, with a theodolite and vertical wires) are extensively described in Barazzetti et al. (2013). Here, the calibration of the rotating head is carried out by aligning the tripod and two vertical wires with a total station. The ruler of the head (which allows the movement of the camera along the optical axis direction) should be adjusted to find a good alignment between the camera–wire system and the three axes of the head. Finally, different pictures can be acquired by rotating the camera to check whether the wires also remain aligned in the tilted images. Indeed, this is a sufficient condition to prove that the camera rotates around its optical centre.

The automated algorithm for gnomonic projection generation is based on the combined mapping of several standard central perspective (pinhole) images acquired with the rotating camera and the projection of them onto a unary globe (of unit radius) around the perspective centre.



(a) (b) (c)

Fig. 1. (a) The calibrated head adopted for gnomonic projection generation. (b) The principle for generating a panoramic image by using the rotating camera system. (c) The preliminary projection of images on the globe during the mosaicking process. The object is the main façade of the Sanctuary of Tirano, located in Valtellina, Italy (see the example in Fig. 3).

Using the camera calibration parameters, distortion-free images can be generated from the original distorted images. In such a case the collinearity equations, which express the relationship between the image coordinates (x, y) and the corresponding object coordinates (X, Y, Z) of a point, can be written as follows (Mikhail et al., 2001):

$$\begin{bmatrix} x \\ y \\ -c \end{bmatrix} = \lambda \mathbf{R} \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix} \tag{1}$$

where c is the principal distance, $\mathbf{R}$ is the rotation matrix between the object and image coordinate systems, and X_0, Y_0, Z_0 are the coordinates of the perspective centre. Here the offset of the principal point has been removed when distortion-free images are generated ($x_0 = y_0 = 0$).

The collinearity equations can be cast in a compact form with a notation based on homogenous coordinates for both vectors of image $\mathbf{x} = [x, y, 1]^{\mathrm{T}}$ and object coordinates $\mathbf{X} = [X, Y, Z, 1]^{\mathrm{T}}$:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/c \end{bmatrix} \mathbf{R} \begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \lambda \mathbf{K}\mathbf{R}[\mathbf{I}, -\mathbf{X}_0]\,\mathbf{X} \tag{2}$$

where $\mathbf{K}$ is the first matrix of the middle equation, $\mathbf{I}$ is the identity matrix and $\mathbf{X}_0 = [X_0, Y_0, Z_0]^{\mathrm{T}}$ contains the coordinates of the perspective centre.

Two images i and j acquired with a rotating camera (see Fig. 1(b)) give the following equations for a common object point $\mathbf{X}$:

$$\begin{aligned} \mathbf{x}_i &= \lambda_i \mathbf{K}\mathbf{R}_i[\mathbf{I}, -\mathbf{X}_0]\,\mathbf{X} \\ \mathbf{x}_j &= \lambda_j \mathbf{K}\mathbf{R}_j[\mathbf{I}, -\mathbf{X}_0]\,\mathbf{X}. \end{aligned} \tag{3}$$

This means that a direct mapping can be found between the image point coordinates of two overlapping images:

$$\mathbf{x}_i = \lambda_i \mathbf{K}\mathbf{R}_i \frac{1}{\lambda_j} (\mathbf{K}\mathbf{R}_j)^{-1} \lambda_j \mathbf{K}\mathbf{R}_j[\mathbf{I}, -\mathbf{X}_0]\,\mathbf{X} = \lambda_i \mathbf{K}\mathbf{R}_i \mathbf{R}_j^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{x}_j / \lambda_j. \tag{4}$$

As points are expressed by homogeneous coordinates, the final mapping function has the form $\mathbf{x}_i = \mathbf{H}_{ij}\mathbf{x}_j$, where $\mathbf{H}_{ij}$ is a 3×3 matrix with eight degrees of freedom, representing a homography (Hartley and Zisserman, 2004):

$$\mathbf{H}_{ij} = \lambda \mathbf{K}\mathbf{R}_i \mathbf{R}_j^{\mathrm{T}} \mathbf{K}^{-1} \text{ where } \lambda = \lambda_i / \lambda_j. \tag{5}$$

This equation can be extended for a generic set of images gathered with the rotating camera system in order to estimate the unknown parameters for multiple views and generate the final gnomonic projection by using a homography-based bundle adjustment (as the images have the same perspective centre).

The co-registration of all images acquired with the rotating camera requires the extraction of a set of corresponding tie points from overlapping images. The problem is automatically solved by using the scale-invariant feature transform (SIFT) algorithm (Lowe,

2004) where *descriptors* are compared between each pair of images within a *kd-tree* scheme to reduce the search space (Samet, 2006). Finally, the matched image points allow the estimation of the unknown parameters within a bundle adjustment. The effect of lens distortion can be removed beforehand with the set of eight additional parameters obtained from the preliminary photogrammetric calibration of the camera (in other words, a new dataset of distortion-free images is created in order to simplify both the SIFT-based matching and mosaicking phases). The final gnomonic projection is therefore a distortion-free image.

Finally, the extraction of seamless lines for image mosaicking can be easily accomplished as data are related by multiple homographic transformations. Although panoramas could also be generated with hand-held cameras, the head is mandatory for accurate reconstructions because it guarantees that all perspective centres of single images are coincident on the same 3D point. In the case of a calibrated head, there is a unique mapping ($\mathbf{x}_i \leftrightarrow \mathbf{x}_j$) between two (or more) generic images, and the choice of seamless lines is simply performed by selecting the central portion of the images. There is no special need for complex seamless lines that follow objects or their borders as the one-to-one mapping is rigorous for the entire mosaic. More details about this aspect are reported in Barazzetti et al. (2013).

The generation of the final projection is similar to standard cartographic map projection applications where the scene around the camera is the globe and the final 2D map is obtained by projecting a generic point from the unary globe (geographic coordinates φ and λ) onto a tangent plane (coordinates U and V). The plane can be placed in an arbitrary position (oblique aspect in map projection terms), but the implemented algorithm fixes its location by using the centre of the angular field of view (FoV) covered during image acquisition.

The equations mapping the globe coordinates to map coordinates for a gnomonic projection with tangent point φ_0 (central latitude) and λ_0 (central longitude) are

$$\begin{aligned} U &= \frac{\cos\varphi_0 \sin\varphi - \sin\varphi_0 \cos\varphi \cos(\lambda - \lambda_0)}{\sin\varphi_0 \sin\varphi + \cos\varphi_0 \cos\varphi \cos(\lambda - \lambda_0)} \\ V &= \frac{\cos\varphi \sin(\lambda - \lambda_0)}{\sin\varphi_0 \sin\varphi + \cos\varphi_0 \cos\varphi \cos(\lambda - \lambda_0)}. \end{aligned} \tag{6}$$

The intrinsic characteristics of a gnomonic projection mean that distance and shape distortions are pronounced, except very near the tangent point, whereas every geodesic (the shortest distance between two points on the surface of the sphere) is mapped to a straight line. The scale increases from the centre of the projection as $1/\sin\varphi$ for parallels and $1/\sin^2\varphi$ for meridians (scale exaggeration). This means that a gnomonic projection has a limited FoV, similar to that of a normal lens.

At the end of this stage, each mosaic is transformed into one or more gnomonic projections, whose number will depend on the FoV. Then, each gnomonic projection will be treated as a distortion-free central perspective image with a principal distance equal to the calibrated focal length c of the camera installed on the rotating head. Pixel size is also preserved whereas sensor size depends on the FoV covered during image acquisition.

Orientation of Gnomonic Projections and Central Perspectives

As mentioned in the Introduction, the approach for image orientation uses a set of central perspectives to strengthen the block geometry as well as gnomonic projections to

obtain a higher level of detail in terms of ground sampling distance (GSD). The process from image acquisition to the orientation phase is divided into three main phases (Fig. 2):

(1) a block of central perspective images (different to the ones employed in the generation of panoramic images in step (2)) is independently gathered with a calibrated camera;
(2) gnomonic projections are generated from the composition of several pinhole images acquired with a calibrated camera installed on the rotating head; and
(3) a joint bundle adjustment including both types of images is used to estimate the exterior orientation (EO) parameters.

The first phase (image acquisition of central perspectives) can be carried out with a single (or multiple) calibrated camera(s) (Fig. 2 – left). In this case it is assumed that N images are collected to obtain a stable photogrammetric block in terms of network geometry.

The generation of multiple gnomonic projections (phase 2) is, instead, carried out by placing the calibrated head at different stations. For each of M camera stations a variable number of images (M_S) can be collected. The subsequent steps (matching; homography-based bundle adjustment; and image mosaicking) for the generation of the final gnomonic projection are usually carried out *in situ* in order to immediately check the quality of the result (the camera can be linked to a laptop computer for online data processing).

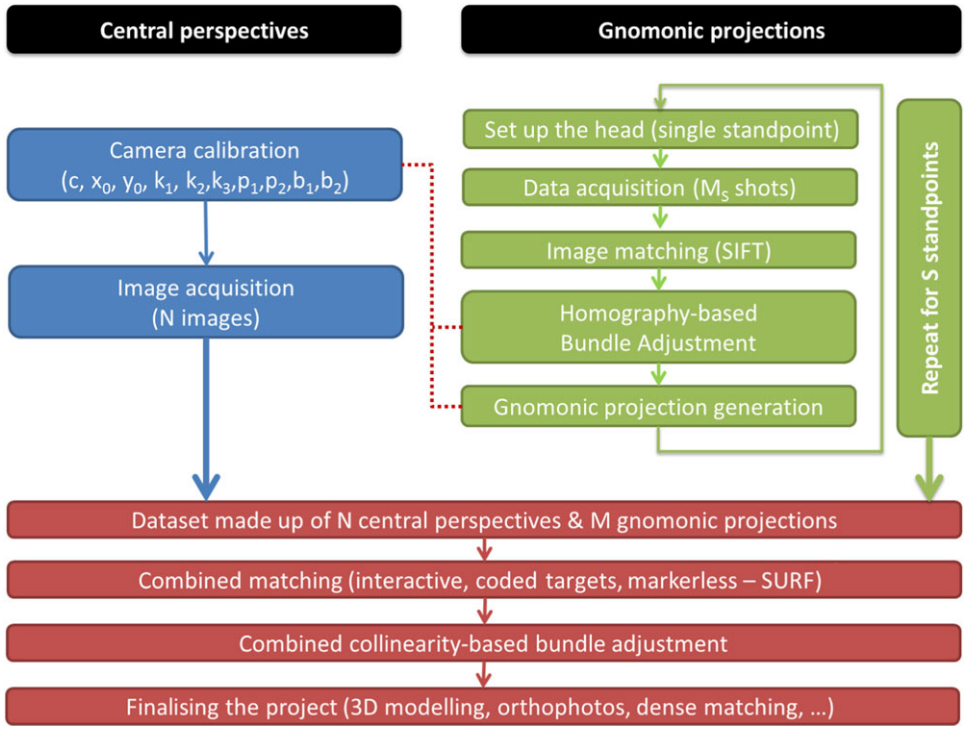


FIG. 2. The flowchart of the procedure for integrating central perspective images (blue), gnomonic projections (green) and their combined uses for photogrammetric purposes (red).

The final gnomonic projection dataset will be composed of M gnomonic projections. These new images will have a variable FoV depending on the angular motion of the head (vertical and horizontal) during data acquisition, but pixel size and principal distance will maintain the same values.

The mathematical model for the combined image orientation (bundle adjustment) is based on the simplified collinearity equations (1) and uses distortion-free images. Although this model is obviously valid for pinhole (central perspective) images, it can also be used for gnomonic projections generated with the implementation discussed here. The algorithm for bundle adjustment is therefore able to consider multiple cameras with their own calibration parameters. For the gnomonic projections, only the principal distance and the metric pixel size are needed as image distortion is removed during the mapping phase onto the projection tangent plane. The whole set of equations (1) can be solved, after linearisation, within a standard Gauss–Markov least squares approach. A stochastic model with unit weights can be used, or, conversely, observations (image point coordinates) can have different weights if their precisions are deemed to differ.

If the dataset is made up of image coordinates alone, the observables do not contain information concerning the datum definition. Therefore, the final 3D reconstruction has an overall ambiguity consisting of an unknown similarity transformation. The rank deficiency in the corresponding least squares normal equation matrix is removed if seven additional linearly independent equations can be found (Luhmann et al., 2014). This ambiguity can be eliminated by introducing a sufficient number of geometrically well-distributed ground control points (GCPs). Otherwise, when GCPs are not available, an inner constraint (that does not involve any external measurements) can be exploited for the combined adjustment to obtain a free network solution (Granshaw, 1980).

The final system is made up of observation and constraint equations, the latter in the form of pseudo-observation equations (Mikhail et al., 2001). Consequently, all variables in the least squares bundle adjustment become weighted observations. The number of unknowns n_u in the joint adjustment can be calculated as follows:

$$n_u = 6(N + M) + 3n_p \tag{7}$$

where n_P is the number of 3D points. Each image results in six EO parameters (three coordinates of the perspective centre and three rotation angles inherent in the rotation matrix $\mathbf{R}$), irrespective of whether it follows the central perspective or the gnomonic projection camera model: they are both parameterised in the same way.

Automated Image Orientation

In the general presentation of the procedure, no assumption has been made on how tie points for the computation of the joint bundle adjustment can be obtained. Solutions based on *manual* (interactive) measurements offer high accuracy but are time consuming, especially in the case of large blocks. *Coded targets* (Cronk et al., 2006), which are automatically recognised, measured and labelled to solve the identification of the image correspondences, are very useful to obtain accurate orientation results, but in many surveys targets cannot be used or applied to the object. The recent trend in close range photogrammetry towards *markerless algorithms* has opened new opportunities and simplified not only the image orientation phase, but also the whole pipeline for 3D modelling, as is witnessed by several commercial and scientific software packages.

Combined datasets made up of markerless gnomonic projections and central perspectives can be automatically processed starting from a set of image points extracted with automated matching techniques. Many operators for automated image matching have been developed in the last decades (Gruen, 2012). However, as these images usually feature a strong scale difference, the image-matching algorithm must be robust enough to cope effectively with such variations (Barazzetti et al., 2011).

For this reason, the scale-invariant SURF (speeded-up robust features) operator was chosen (Bay et al., 2008). The method uses a Hessian matrix-based measure for the *detector* and the distribution of the first-order Haar wavelet responses for the *descriptor*. SURF is claimed to be faster than SIFT-like algorithms and provides similar results in terms of point precision, even though it retrieves a lower number of point correspondences. This is a convenient choice in the case of high-resolution gnomonic projections that require a bigger CPU (central processing unit) time and could have a huge number of interest points.

The implemented algorithm begins by extracting SURF key points on every image. Then a search for corresponding tie points is carried out in a pairwise manner by using the Euclidean distance between descriptors of SURF that are estimated as measurements of the difference. Moreover, a constraint between the first- and the second-best candidates (a "ratio test") is added to be more distinctive (see Barazzetti et al., 2010b). This initial matching phase usually provides a sufficient number of image correspondences but some mismatches are often still present and can be rejected with the robust estimation of the *essential matrix* **E** (Hartley and Zisserman, 2004), as camera calibration parameters are known. The 3×3 **E** matrix has rank 2 and encapsulates the epipolar geometry of a stereopair. It is a homogenous quantity with five degrees of freedom: therefore at least five image correspondences are needed for its estimation (Nister, 2004). This is carried out with robust techniques as they allow the detection of possible outliers in the observations. The implementation uses the maximum a posteriori sample consensus (MAPSAC) estimation technique (Torr, 2002).

The matching procedure based on SURF (for feature extraction) and MAPSAC (for outlier rejection) is applied in a pairwise manner and provides corresponding tie points between pairs of images. The selection of the image combinations to analyse is usually carried out in an exhaustive way if the block is made up of a relatively limited number of images. Otherwise, a knowledge of the block structure or the generation of the so-called "visibility map" (that contains the connections between all image pairs sharing tie points) is exploited (Barazzetti et al., 2010a). It can help reduce the number of image combinations to browse (Snavely et al., 2008).

After completing the pairwise matching phase, tie points are organised into a structure able to initialise the bundle adjustment. From the analysis of image coordinate values, the same tie points found across different pairs can be tracked on multiple images to find manifold corresponding points and run the combined least squares adjustment based on the collinearity equations. Some examples are illustrated in the next section.

Experimental Examples

Two sets of experiments are reported here. The first set contains two applications of the integrated orientation of central perspective and panoramic images after generation of gnomonic projections are described. In the first of these examples, the aim was to derive high-resolution orthophotos of the façade of a historical building based on one panoramic image and a small block of pinhole images. In the second example, more panoramic images are integrated into a larger block of pinhole images for the reconstruction of a small

sculpture. Here, a dense matching stage is also included to demonstrate the potential of this approach for surface reconstruction. In both examples, image orientation is accomplished on the basis of the automatic extraction of tie points as proposed in the previous section.

The second set of experiments was specifically aimed at assessing the accuracy of the methodology using two different laboratory set-ups. In the first one, the theoretical accuracy was checked from the analysis of the covariance matrix of the least squares bundle adjustment. In the second set-up, empirical accuracy was evaluated on the basis of benchmarking data.

A discussion of the results obtained is reported at the end of this section.

High-Resolution Orthophoto of a Building Façade

An example of combined image orientation with scale-invariant features is shown in Fig. 3. A single gnomonic projection (in this case $N \gg M$) was used for the generation of a high-resolution rectified image of the main façade of the Sanctuary of Tirano, located in Valtellina (Northern Italy).

The size of the façade was approximately 14 m × 21 m. A set of 611 images was acquired for the generation of one gnomonic projection with the rotating calibrated head. A large overlap (about 95%) between consecutive horizontal and vertical strips was set up. This overlap can be dramatically reduced in real applications (even less than 50% since homography is a direct image-to-image mapping), limiting the number of images and processing time.

The camera adopted was a Nikon D700 (4256 × 2832 pixels ≈ 12 megapixels; pixel size 8·4 μm) with a 180 mm Tamron telephoto lens. The final gnomonic projection of the façade had a resolution of 23 901 × 25 413 pixels (about 607 megapixels), much larger than the image size achievable with any commercial SLR camera.

(a) (b) (c)

FIG. 3. The main façade of the Sanctuary of Tirano, Valtellina, Italy. (a) The result of the combined orientation of central perspective images (red squares) and one gnomonic projection (blue circle): black dots are the 3D coordinates of tie points on the building façade. (b) The entire rectified projection (1 pixel = 1 mm) generated from the integrated approach. (c) Detail from the rectified projection.

To complete the photogrammetric project, a sequence of 13 pinhole images was acquired with the same camera but using a 20 mm Nikkor lens. The gnomonic projection has a resolution, in this case, 50 times better. Image acquisition (standard central perspectives in this case) required a single strip with short baselines in order to make more efficient the use of automated image-matching techniques for the orientation stage and running the combined bundle block adjustment.

Total data processing took about 8 min, 7 min of which were needed for automated image matching. Sigma naught of the integrated bundle adjustment was 0·55 pixels; 5874 tie points were extracted without any user interaction.

The sparse reconstruction was then scaled with a known distance and a single high-resolution (pixel size 1 mm) view of the front façade was generated by rectifying the oriented gnomonic projection (perspective rectification).

3D Modelling of a Sculpture

The second case study concerns a full 3D object (Fig. 4), namely, the statue of "The Basilisk" in Malesco, Italy. This was surveyed by using 75 pinhole images organised on two different circular strips around the object. A Nikon D700 camera with a 35 mm Nikkor lens was employed. Then, 11 gnomonic projections were generated using the same camera with a 90 mm Tamron lens, increasing the pixel resolution by a factor of three.

The combined matching and adjustment approach provided EO parameters for all central perspective images and gnomonic projections (a posteriori sigma naught of 0·53 pixels; a priori sigma naught of 1 pixel; image point precision ± 1 pixel; 37 320 tie points extracted in object space).

The main problem with this project concerned the orientation of three gnomonic projections which were not automatically oriented as the scene was quite complex. This was mainly due to several points matched on the background. The intersection of rays (collinearity equations) for these points is not favourable because of narrow angles. For this reason, some points were removed during data processing. These points prevented the complete automated orientation of this block, where three gnomonic projections were discarded by the implemented algorithm.

After the image orientation step, the 3D surface of the statue was reconstructed from the oriented images using a development of the multiphoto geometrically constrained matching (MGCM) algorithm (Gruen and Baltsavias, 1988), namely, the dense matching algorithm MGCM+ proposed in Previtali et al. (2011). A mesh of approximately 10 million triangles was obtained. Only pinhole images were used in this phase, whereas gnomonic projections were exploited for texture mapping to obtain a superior level of detail.

Evaluation of Theoretical Accuracy

Evaluations were conducted to demonstrate how the integrated use of both central perspective and gnomonic projections could be a valid alternative to preserve metric accuracy and improve the level of detail. Two experiments were carried out using special sets of coded targets in order to check the accuracy. The first experiment concerned the theoretical accuracy and is reported in this section. The second experiment, detailed in the following section, concerned the empirical accuracy.

In the first experiment, the analysis focused on the variance–covariance matrix from the bundle adjustment in the case of a free network solution, in other words on the assessment of the theoretical accuracy obtainable without using any GCPs. A set of $n_P = 55$ colour

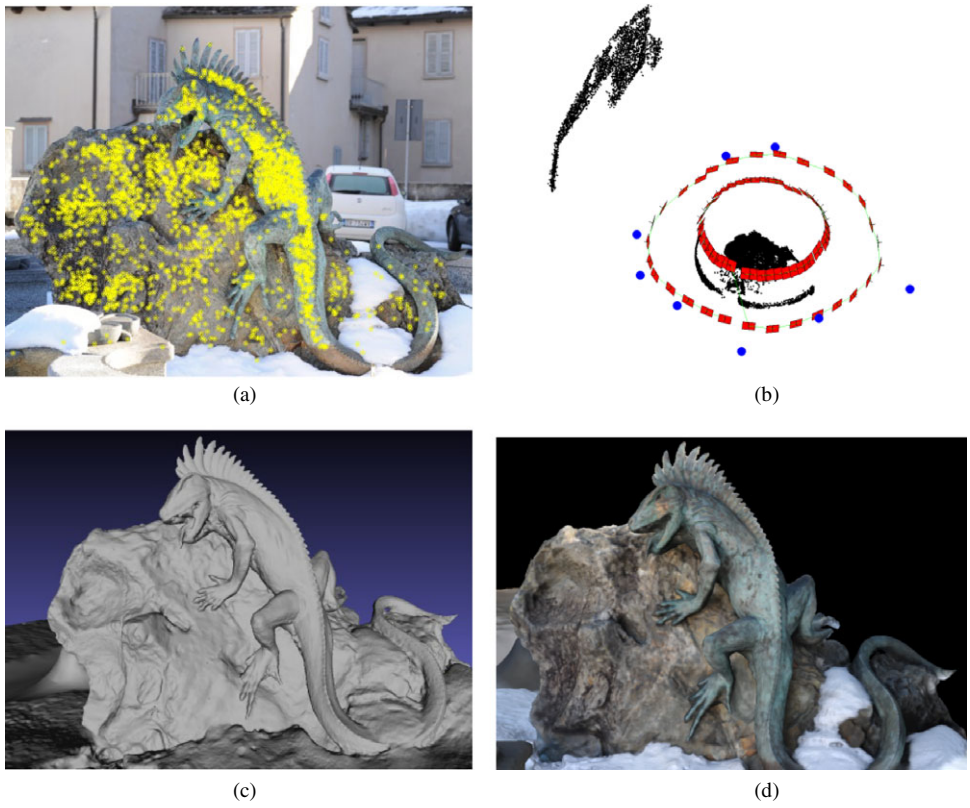
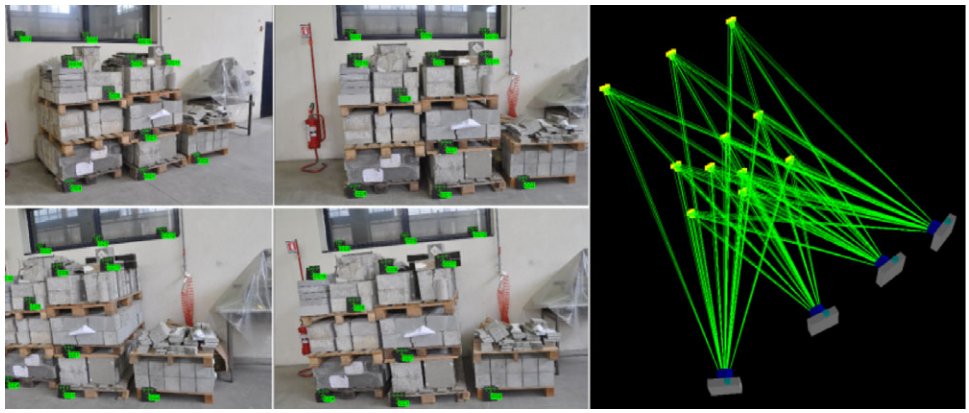


FIG. 4. Some results for the reconstruction of "The Basilisk" statue in Malesco, Italy. (a) Tie points matched on a gnomonic projection of the block. (b) A 3D view with object points and camera positions (red squares are central perspective images; blue circles are gnomonic projections; black dots are tie points). (c) The mesh extracted from dense image matching (MGCM+ algorithm) shown as shaded relief. (d) The same 3D model as (c) but textured by using gnomonic projections.

targets of the iWitnessPRO™ photogrammetric software package (www.photometrix.com) was adopted to extract a precise set of image coordinates (expected precision of about 1/10 pixel).

The block of pinhole images consisted of four convergent views captured with a calibrated Nikon D90 SLR camera (CMOS sensor with size 23·584 mm × 16·664 mm, pixel size 5·5 μm) with a 20 mm Sigma lens (see Fig. 5). The automatic target measurement algorithm of iWitnessPRO™ provided image coordinates and then a free network bundle adjustment was run, completing the orientation phase (total processing time was less than 5 s). Most targets (53 out of 55) were matched in all four images and the minimum intersection angle was 40°. After the bundle adjustment, the project was arbitrarily scaled with a distance of 100 m between two opposite targets at the bottom-left and bottom-right sides. This scale factor did not introduce any deformation in the image block, which was only roto-translated and scaled. The root mean square (RMS) of the image coordinate residuals after orientation turned out to be 0·10 pixels and the theoretical



(a) (b)

FIG. 5. (a) The four pinhole images used in the second experiment with the matched coded targets. (b) 3D view including camera stations and targets. Here multiple targets extracted with the photogrammetric software iWitnessPRO™ were used. Each target is made up of a small plate including five circles.

accuracy of 3D points was 4·12, 4·12 and 10·13 mm along the X, Y and Z axes, respectively.

A set of four gnomonic projections (image size from 63 up to 100 megapixels), generated with a Nikon D700 and a 180 mm Tamron telephoto lens, was then included in this project. Target coordinates were extracted with iWitnessPRO™. The first important result of the combined gnomonic and central perspective bundle adjustment was a global worsening of sigma naught (0·56 pixels) along with a degradation of precision of about a factor of two. The results are illustrated in Table I and a 3D view of camera positions and tie point location is shown in Fig. 6.

Although the combined adjustment featured a larger redundancy (four additional images) and therefore points could be determined from the intersection of a larger number of rays in space, the estimated theoretical accuracies declined. Indeed, the generation of a gnomonic projection required a series of cascaded steps where errors could progressively accumulate in several ways. Firstly, the adopted homography-based bundle adjustment for the generation of panoramic images did not include external constraints and therefore the mosaicking phase could lead to error accumulation if large FoV and long focal lenses are employed. Secondly, the mosaicking phase (transformation from the sphere to the gnomonic projection plane) needed further image resampling and blending algorithms, that is, additional modifications of the original images, and consequently may slightly alter the geometric and radiometric image content. Finally, a residual error during the calibration of the head can become significant in such applications requiring a superior accuracy. All these effects may contribute to the global worsening of accuracy obtained in the combined adjustment.

Evaluation of Empirical Accuracy

The second experiment included a dataset of: (a) 16 standard central perspective images captured with a full-frame Nikon D700 camera (CMOS sensor, size 36 mm × 24 mm, pixel size 8·4 μm) equipped with a 20 mm Nikkor lens; and (b) two sets of gnomonic projections

TABLE I. Estimated theoretical accuracy of 3D point coordinates for the combined adjustment, including four central perspectives and four gnomonic projections generated with a telephoto lens (180 mm focal length).

X	8·522 mm	or	1:20 400
Y	7·913 mm	or	1:22 000
Z	24·629 mm	or	1:12 700

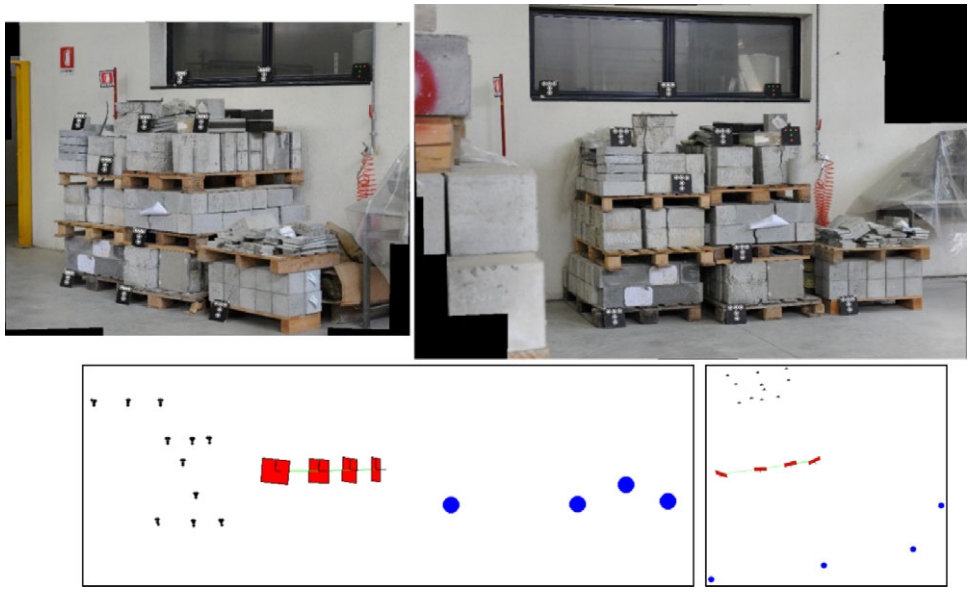


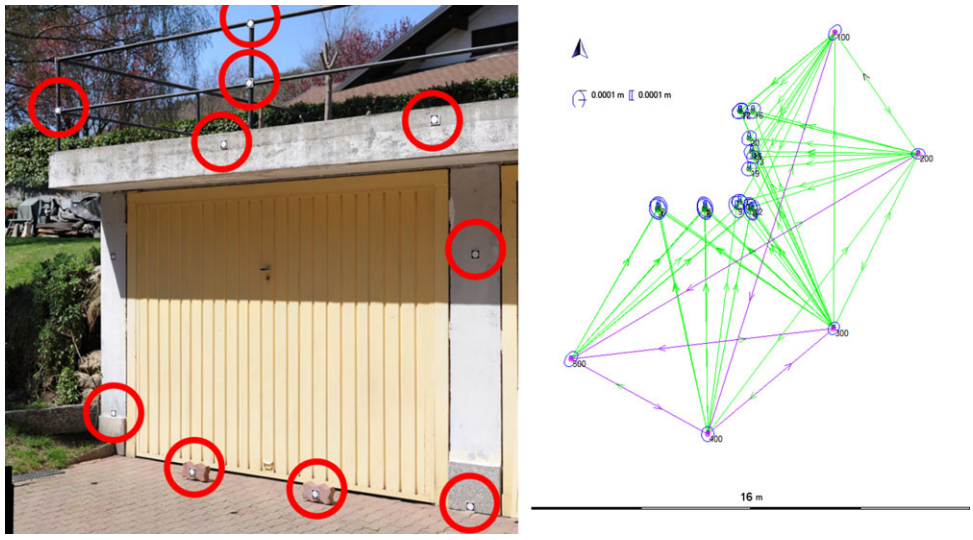
FIG. 6. Top: two gnomonic projections with the colour targets (white marks). In some cases the targets were not correctly measured, for example, the one at the top right. Bottom: 3D visualisation of the combined gnomonic (blue) and central perspective (red) camera positions and target coordinates (black).

created by using the same camera (six projections obtained with a 90 mm Tamron lens and three projections with a 180 mm Tamron lens). The surveyed object was a 3D polygon with coded targets positioned on two orthogonal walls (see a portion of the polygon in Fig. 7(a)).

The targets consist of a white dot with a cross in the middle. A total number of 20 such targets were placed and measured with a Leica TS30 total station (angular precision ±0·15 mgon; range precision ±0·6 mm) to obtain a benchmarking dataset. After least squares adjustment of the geodetic network, 3D point coordinates were computed with a precision better than ±0·2 mm. In Fig. 7(b) the layout of the geodetic network is shown.

Image coordinates were manually measured. It is important to notice that the measurement of the cross (target centre) was quite simple for high-resolution gnomonic projections, although the camera stations of the panoramic images were located far from the object. Indeed, the standard pinhole images provided a smaller GSD than the panoramic images.

The stochastic model for the bundle adjustment was set up as follows: a priori sigma naught of 1 pixel; image point precision ±1 pixel; and GCP precision ±0·2 mm. Eight



(a) (b)

FIG. 7. 3D polygon with coded targets positioned on two orthogonal walls. (a) A portion of the polygon with targets indicated by red circles. (b) The layout of the geodetic network measured by using a total station along with a multiple-intersection scheme from five stations: horizontal error ellipses and vertical error bars (semi-axis 1-sigma) indicate the 3D point precision.

targets were used as GCPs whereas the remaining 12 acted as independent check points. A 3D view showing both sets of camera positions and 3D points after the bundle adjustment is shown in Fig. 8.

The analysis of the covariance matrix $\mathbf{C}_{xx}$ of the bundle adjustment solution gave an average theoretical accuracy for the object points of $\pm 0{\cdot}5$ mm for both the X and Y directions and $\pm 0{\cdot}3$ mm for the vertical Z direction. Estimated sigma naught was 0·97 pixels. The computed difference values of tie point coordinates with respect to the benchmarking data are shown in Table II. These values are quite consistent with the estimated theoretical precisions, except in the case of Z coordinates where the empirical results are larger than the estimated theoretical accuracies.

The same dataset was then processed with the pinhole images alone. In this case the average theoretical accuracy was about $\pm 0{\cdot}2$ mm for all three X, Y and Z components. Accuracy evaluation with total station points provided the results shown in Table II which are slightly better than those of the combined adjustment.

DISCUSSION

The presented examples show that a combined orientation of central perspective images and gnomonic projections, based on the standard collinearity equation camera model including different sets of calibration parameters, allowed the creation of accurate metric reconstructions with a better GSD than central perspectives alone. As the use of gnomonic projections in the integrated bundle adjustment slightly degrades the theoretical accuracy, the block should satisfy the criteria that the number of pinhole images substantially

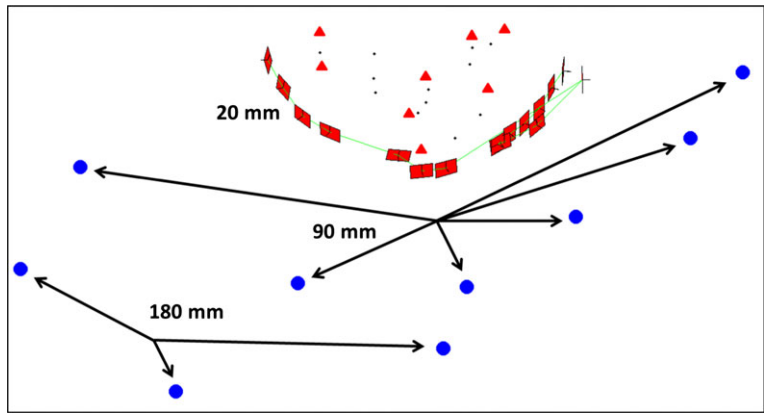


FIG. 8. 3D visualisation of estimated stations of the photogrammetric network including 16 pinhole standard images and nine gnomonic projections acquired with two different focal length cameras (90 and 180 mm). The 12 black dots are check points, whereas the eight red triangles are GCPs. Red squares are central perspective images; blue circles are gnomonic projections.

TABLE II. Statistics on the estimated coordinates of 12 independent check points with respect to benchmarking values measured by a total station.

	Pinhole			*Gnomonic and pinhole*		
	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
Mean (mm)	0·1	−0·5	0·0	−0·1	0·0	0·0
RMS (mm)	0·7	0·7	0·5	0·9	0·6	1·0
Max (mm)	0·7	1·0	0·6	1·1	0·8	1·6
Min (mm)	−1·1	−1·6	−1·1	−1·4	−0·9	−1·9

surpasses the number of gnomonic projections ($N \gg M$) in order to simplify the orientation phase and obtain more accurate metric data. Indeed, the experiments show that an improvement in the global precision can be simply obtained by using more central perspectives than gnomonic projections. The main advantage of the combined bundle adjustment with more pinhole imagery consists in (a) a robust network geometry given by standard imagery and (b) the final superior level of detail of the additional gnomonic projections.

In addition, the number of images for gnomonic projection mapping can be extremely variable. For instance, the use of a huge number of images for the façade dataset (many more than those strictly needed) proved that the current implementation could handle large datasets of single shots fused in one mosaic. The other applications presented in this contribution were, conversely, carried out with a reduced number of images.

It may be argued that a possible alternative to the proposed workflow could be an approach based on a single-step procedure incorporating the generation of panoramas and the co-registration of central perspective images. For instance, in Schneider et al. (2012) an example of such an integrated adjustment is illustrated and discussed. From a theoretical point of view, this approach is more complete because it increases the local redundancy of observations and their controllability (Förstner, 2001). On the other hand, the proposed

multi-step strategy is more suitable from a practical point of view. Indeed, splitting the whole process into several cascaded independent steps has some advantages. For example, it allows for the inspection of intermediate output and their possible applications for other purposes (for instance, a set of independent panoramic images could be used for texture mapping). In addition, image orientation is easier if central perspectives are firstly oriented and then gnomonic projections are progressively concatenated into a unified bundle adjustment.

The examples described in the previous section were carried out with the same camera which was: (a) mounted on the calibrated rotating head; and then (b) used to capture the auxiliary block of central perspective images. This means that the pixel size is constant and the sigma naught of the least squares bundle adjustment can be expressed in pixels. In this case, this parameter is a representative value of the overall theoretical precision. If cameras have a different pixel size, it is important to document sigma naught in metric units and assign proper weights to different kinds of observations. In cases where the precision of different groups of photogrammetric observations cannot be correctly determined before running the bundle adjustment, the use of a unit weight matrix may lead to a biased estimate of the solution. An effective approach to cope with this is to iteratively re-estimate the covariance of groups of homogenous observations (McGlone et al., 2004).

Conclusions

The paper presented an automatic procedure able to integrate central perspective (or "pinhole") images and gnomonic projections into a common photogrammetric bundle block adjustment. The aim was a precise and rigorous photogrammetric image orientation procedure with a superior level of detail in terms of image resolution. The method is particularly useful when: (a) a better level of detail (a better GSD) is needed; (b) long camera-to-object distances have to be employed; and (c) tall objects for which standard blocks made up of different strips cannot be acquired (for example, building façades). The method can be very useful for applications concerning flat objects such as paintings, bas-reliefs, building façades and the like (Remondino et al., 2011; Zhang et al., 2011).

Both image datasets are gathered by using a consumer-grade digital camera following the central perspective model. Gnomonic projections are derived from panoramic images by using stitching techniques that can be applied by installing a pinhole camera on top of a calibrated rotating head. Gnomonic projections can be oriented with the same collinearity model adopted for central perspective imagery. Fusion of single central perspective images into one projection can be operated in a fully automatic way.

The main motivation for integrating pinhole images and gnomonic projections is to exploit the higher resolution of the final output (such as texture mapping and orthophotos). Central perspectives with a lower image resolution can be acquired and processed following typical photogrammetric approaches in order to simplify the combined bundle adjustment within a robust network geometry.

An automatic image-matching procedure able to compute the orientation parameters without interactive measurements in both types of images has been developed and tested. This has allowed the achievement of the same degree of automation inherent in modern photogrammetric procedures that are based on pinhole images, with a higher level of detail in the final product but with far fewer images.

It is also important to mention that in the current implementation, the whole globe is not used in the data processing workflow. This allows the operator to handle gnomonic

projections as standard pinhole images. The use of the whole globe, along with a variable weight matrix, will be taken into consideration in future work in order to handle 360° scenes and will require a combined bundle adjustment based on standard and spherical (Fangi and Nardinocchi, 2013) collinearity equations (this approach follows the same concept proposed by Schneider and Maas (2006) for cylindrical images). A combined camera model will be required to obtain a 360° reconstruction. In this case, particular attention must be also paid to the generation of the global spherical image, especially in the case of long focal length lenses such as those presented in this contribution.

Acknowledgements

The authors would like to thank Fabio Roncoroni for his technical support during the creation of the calibrated head. This research has been partially supported by the National Basic Research Program of China (Project No. 2013CB733204).

References

Amiri Parian, J. and Gruen, A., 2010. Sensor modeling, self-calibration and accuracy testing of panoramic cameras and laser scanners. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1): 60–76.

Barazzetti, L., Remondino, F., Scaioni, M. and Brumana, R., 2010a. Fully automatic UAV image-based sensor orientation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(1). 6 pages.

Barazzetti, L., Scaioni, M. and Remondino, F., 2010b. Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. *Photogrammetric Record*, 25(132): 356–381.

Barazzetti, L., Forlani, G., Remondino, F., Roncella, R. and Scaioni, M., 2011. Experiences and achievements in automated image sequence orientation for close-range photogrammetric projects. *SPIE* 8085. 13 pages.

Barazzetti, L., Previtali, M. and Scaioni, M., 2013. Stitching and processing gnomonic projections for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 79(6): 573–582.

Bay, H., Ess, A., Tuytelaars, T. and van Gool, L., 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3): 346–359.

Brown, M. and Lowe, D. G., 2003. Recognizing panoramas. *International Conference on Computer Vision*, 2: 1218–1225.

Cronk, S., Fraser, C. and Hanley, H., 2006. Automated metric calibration of colour digital cameras. *Photogrammetric Record*, 21(116): 355–372.

Di, K., Xu, F. and Li, R., 2004. Constrained bundle adjustment of panoramic stereo images for Mars landing-site mapping. *4th International Symposium on Mobile Mapping Technology*, Kunming, China, 29th to 31st March. 6 pages (e-doc).

Di, K., Xu, F., Wang, J., Agarwal, S., Brodyagina, E., Li, R. and Matthies, L., 2008. Photogrammetric processing of rover imagery of the 2003 Mars Exploration Rover mission. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2): 181–201.

Fangi, G., 2010. Multiscale multiresolution spherical photogrammetry with long focal lenses for architectural surveys. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(5): 228–233.

Fangi, G. and Nardinocchi, C., 2013. Photogrammetric processing of spherical panoramas. *Photogrammetric Record*, 28(143): 293–311.

Förstner, W., 2001. Generic estimation procedures for orientation with minimum and redundant information. *Calibration and Orientation of Cameras in Computer Vision* (Eds. A. Gruen & T. S. Huang). *Springer, Heidelberg, Germany*, 34: 63–94.

Granshaw, S. I., 1980. Bundle adjustment methods in engineering photogrammetry. *Photogrammetric Record*, 10(56): 181–207.

Gruen, A., 2012. Development and status of image matching in photogrammetry. *Photogrammetric Record*, 27(137): 36–57.

Gruen, A. and Baltsavias, E. P., 1988. Geometrically constrained multiphoto matching. *Photogrammetric Engineering & Remote Sensing*, 54(5): 633–641.

HARTLEY, R. I. and ZISSERMAN, A., 2004. *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press, Cambridge, UK. 672 pages.

LOWE, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91–110.

LUHMANN, T. and TECKLENBURG, W., 2002. Bundle orientation and 3-D object reconstruction from multiple-station panoramic imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(5): 181–186.

LUHMANN, T., ROBSON, S., KYLE, S. and BOEHM, J., 2014. *Close-Range Photogrammetry and 3D Imaging*. Second edition. De Gruyter, Berlin, Germany. 684 pages.

MCGLONE, J. C., MIKHAIL, E. M., BETHEL, J. S. and MULLEN, R., 2004. *Manual of Photogrammetry*. Fifth edition. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA. 1151 pages.

MIKHAIL, E. M., BETHEL, J. and MCGLONE, J. C., 2001. *Introduction to Modern Photogrammetry*. Wiley, New York, USA. 479 pages.

NISTER, D., 2004. Automatic passive recovery of 3D from images and video. *IEEE Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*. 438–445.

PISA, C., ZEPPA, F. and FANGI, G., 2010. The spherical photogrammetry as metric recording instrument for cultural heritage. The San Galgano Abbey, Siena, Italy and the Roman Theatre, Sabratha, Libya. *Second ACM Workshop on eHeritage and Digital Art Preservation*, Florence, Italy. 6 pages.

PREVITALI, M., BARAZZETTI, L., SCAIONI, M. and TIAN, Y., 2011. An automatic multi-image procedure for accurate 3D object reconstruction. *4th International Congress on Image and Signal Processing*, IEEE Conference Record Number 18205, Shanghai. 5 pages.

REMONDINO, F., RIZZI, A., BARAZZETTI, L., SCAIONI, M., FASSI, F., BRUMANA, R. and PELAGOTTI, A., 2011. Review of geometric and radiometric analyses of paintings. *Photogrammetric Record*, 26(136): 439–461.

SAMET, H., 2006. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, Burlington, Massachusetts, USA. 1024 pages.

SCHNEIDER, D. and MAAS, H.-G., 2004. Application and accuracy potential of a strict geometric model for rotating lines cameras. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(5/W16). 5 pages (on CD-ROM).

SCHNEIDER, D. and MAAS, H.-G., 2005. Combined bundle adjustment of panoramic and central perspective images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5/W18). 4 pages (on CD-ROM).

SCHNEIDER, D. and MAAS, H.-G., 2006. A geometric model for linear-array-based terrestrial panoramic cameras. *Photogrammetric Record*, 21(115): 198–210.

SCHNEIDER, J., SCHINDLER, F., LÄBE, T. and FÖRSTNER, W., 2012. Bundle adjustment for multi-camera systems with points at infinity. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3: 75–80.

SNAVELY, N., SEITZ, S. M. and SZELISKI, R., 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2): 189–210.

TORR, P. H. S., 2002. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1): 35–61.

ZHANG, Y., ZANG, Z., SUN, M. and KE, T., 2011. Precise orthoimage generation of Dunhuang wall painting. *Photogrammetric Engineering & Remote Sensing*, 77(6): 631–640.