The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2010)
A. Artusi, M. Joly-Parvex, G. Lucet, A. Ribes, and D. Pitzalis (Editors)

Automation in Multi-Image Spherical Photogrammetry for 3D Architectural Reconstructions

Luigi Barazzetti*, Gabriele Fangi**, Fabio Remondino***, Marco Scaioni*

* Department Building Engineering Science and Technology, Politecnico di Milano, Italy
E-mail: (luigi.barazzetti, marco.scaioni)@polimi.it, Web: http://www.polimi.it
** Università Politecnica delle Marche, Ancona, Italy, Email: g.fangi@univpm.it
*** 3D Optical Metrology – B. Kessler Foundation (FBK), Trento, Italy
E-mail: remondino@fbk.eu, Web: http://3dom.fbk.eu

Abstract

The derivation of 3D metric information from spherical images for interactive exploration and realistic 3D modeling is receiving great attention due to their high-resolution content, large field-of-view, low-cost, easiness, rapidity and completeness. We present a methodology for accurate 3D reconstruction from spherical (panoramic) images acquired by mosaicking separated frame images captured with a rotating head and a consumer grade or SLR digital camera. In particular we focus the attention on the orientation of the panoramas which is achieved by extracting the necessary tie points with a new fully automated procedure, based on feature matching and robust estimators. Results of the automated panorama orientations and 3D reconstructions of architectural scenes are presented and discussed.

1. Introduction

The documentation with spherical or panoramic photography is getting a very common practice for many kinds of visual applications (e.g. Google Street View, 1001 Wonders, etc.). In addition, the derivation of metric results from spherical images for interactive exploration, accurate documentation and realistic 3D modeling is receiving great attention due to high-resolution content, large field-of-view, low-cost, easiness, rapidity and completeness of the method. Cylindrical or spherical panoramic images are generally acquired with expensive linear array and rotating panoramic cameras which have very high metric performances ([LT02], [PG04], [SM04]). 3D reconstructions are then feasible, e.g. via space intersection [LT04], if multiple oriented panoramas are available. Another common method to produce a panoramic image [SS97] is to acquire a set of partly overlapped images from a unique point of view with a consumer or SLR digital camera which is rotated around its perspective centre. Images are then stitched together and eventually the derived panorama is projected on a virtual sphere. Several commercial packages are also available to perform this task (Realviz Stitcher, PTgui, Autopano, etc.) Generally, the radial image distortion can be also compensated during the stitching process ([BL03], [BL07]). The projection sphere of these panoramic images is mapped in the cartographic plane with the so-called longitude-latitude representation (or equi-rectangular projection) [Sny87] from which the angular directions can be drawn. If multiple panoramas of the same scene, acquired from different point of view, are available, the 3D reconstruction of the scene can be achieved. Therefore this easy and low-cost solution allows to acquire almost Gigapixel images with great potential not only for visual needs, but also for metric applications and 3D modeling purposes. A single panorama can replace a lot of images and the sight of the whole environment can allow a better understanding of the scene. The high resolution content of the panoramas allows for highly detailed 3D reconstructions while their large field-of-view limits the number of acquisitions.

The article presents a pipeline developed to perform 3D reconstructions from multiple spherical (panoramic) images generated by stitching together frame images. The experiments are performed on architectural scenes, for their 3D metric documentation and modeling. Particular attention will be given to the procedure developed to automatically orient the spherical images. The procedure is quite fast and demonstrated its robustness also in the case of wide-baseline panoramas. No commercial solution is available for this task and a classical SfM (Structure from Motion) approach on the single frames is not usable. After the automated orientation step, the interactive 3D reconstructions of the architectural structures and features (planar façades, roofs and edges) can be performed, in order to derive sparse point clouds and the required architectural drawings and plans (Figure **1**). Indeed no automated image correlation procedure is available for spherical images. Furthermore the 3D reconstruction of large architectures is still a manual-driven procedure as dense point clouds are not sufficient to correctly describe all the important features required for drawings, sections and plans.

2. Multi-image spherical photogrammetry

After the stitching of the single images acquired from a unique stand-point (e.g. with a dedicated tripod and a motorized head), the generated mosaic (with barrel and pincushion distortion corrections) is projected onto a sphere

DOI: 10.2312/PE/VAST/VAST10S/075-081

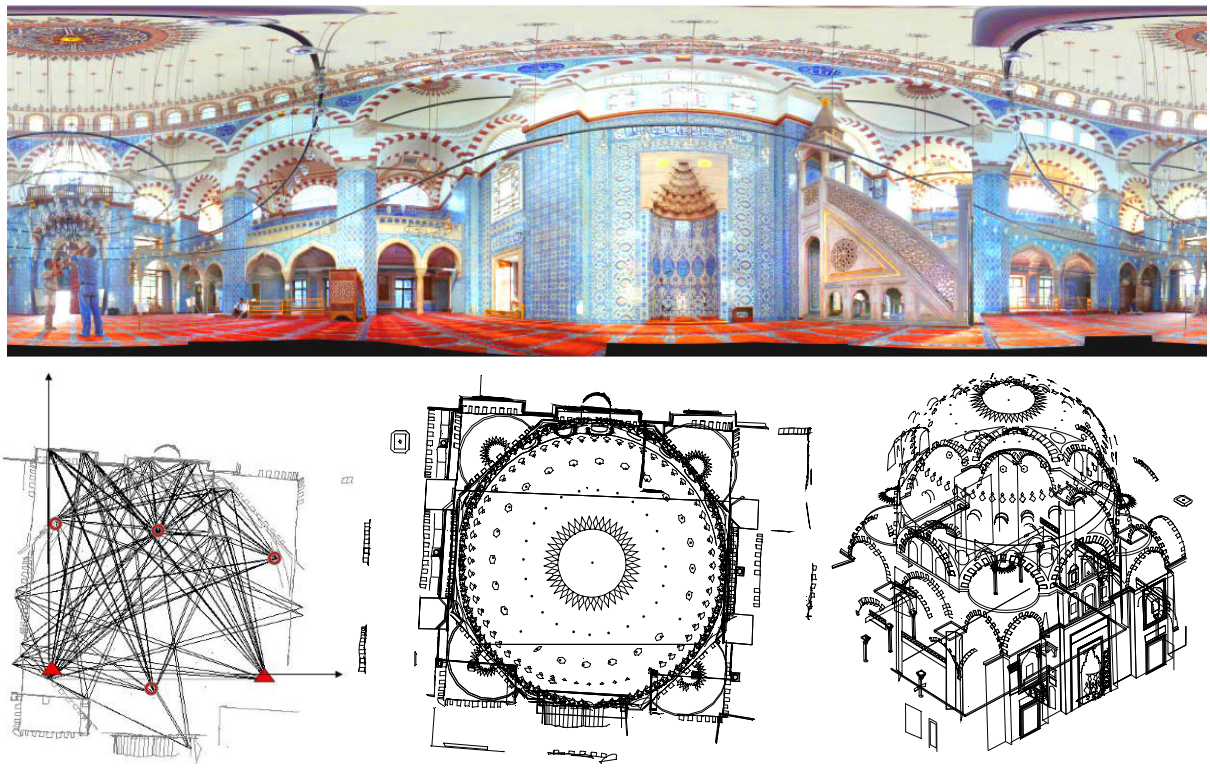


Figure 1: *A typical spherical (or panoramic) image of the Rustem Pascià (Istanbul, Turkey) mosque's interior created mosaicking together different frame images (top). The geometric resolution of each panoramic image can vary from 50 up to 150-200 Megapixels. Using a network of 6 panoramas, the interior of the mosque was digitally reconstructed in 3D (bottom).*

whose radius R is arbitrary but equal to the focal length of the camera in order to keep the original image resolution. Afterwards the sphere is mapped with the so-called longitude-latitude representation (or equi-rectangular projection) onto a plane with the image coordinate given by:

$$= R\theta, \; y = R\varphi \qquad (1)$$

with the horizontal (θ) and vertical (φ) directions as shown in (Figure 2). Such representation is neither conform nor equivalent.

The poles of the sphere are represented by two segments of length equal to the circumference of the sphere and therefore equator and poles have the same length. On the other hand, the height of the panorama is equal to the development of a meridian. From such representation the angles of direction of the projective line can be drawn and, knowing the extension a of the panorama, the radius of the generating sphere is derived as $R = a/2\pi$.

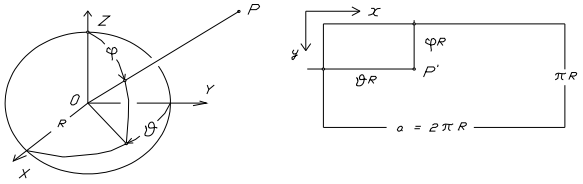


Figure 2: *Relationship between image and object coordinates in case of spherical images (i.e. the latitude-longitude projection).*

Compared to a classical theodolite, where setting the principal axis along the local plumb-line is a simple task, in the case of spherical images it is not possible to do the same with enough accuracy. Therefore two correction angles (α_x, α_y) around the two horizontal axes should be estimated and applied (Figure 3).

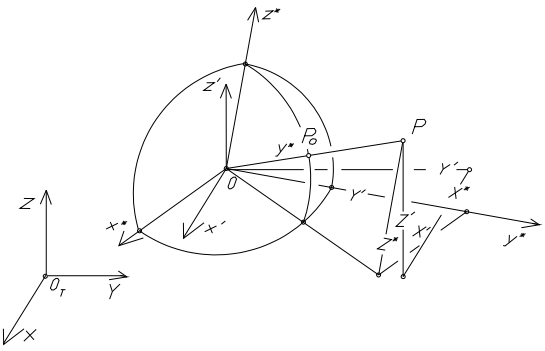


Figure 3: *Relationship between image and object coordinates before and after the two corrections applied to compensate for the missing verticality.*

After this correction [Fan06], if multiple panoramas are available, they can be photogrammetrically processed with a reformulation of the standard perspective camera model from Cartesian to spherical coordinates [Fan07], given a sufficient number of corresponding points between the panoramas (Section 2.1). Once the camera (sphere) poses are estimated along with the 3D object

coordinates of the image correspondences (Section 2.2), the 3D reconstruction of the scene can take place (Section 2.3) with the normal surveying methods of intersection and resection.

2.1 Feature correspondences extraction

Automation in the orientation and reconstruction phases demonstrated impressive results especially in the field of Computer Vision ([PV02], [VV06], [FFG09], [ASS09], [FCSS10]), but also in the photogrammetric community with the implementation of strategies to derive accurate orientation results ([RFR05], [LF06], [RR06], [BRS09], [BRS10]). However, all these approaches use pinhole images and cannot be employed in the case of spherical images. From this point of view, a solution for the automated and accurate analysis of this particular kind of data is still missing. For the orientation of large panoramas (typically in the order of 50-150 Megapixels), the interactive identification of image correspondences might require long elaboration time, thus automation becomes necessary. The presence of illumination changes, occlusions, moving objects and large baselines, requires a powerful and reliable procedure.

The developed method is based on the extraction and matching of scale invariant features with the SIFT [Low04] and SURF [BETV08] operators, combined with some robust estimators for the detection of wrong correspondences ([RL87], [Tor02]) based on the estimation of the fundamental matrix [HZ04]. The methodology is similar to some traditional approaches for pinhole images, however it was extended in order to accommodate also spherical images (or panoramas) and deliver the required correspondences for the camera pose estimation. When spherical images are unwrapped onto a plane, the image content presents different resolutions (width and height), scale changes and the impossibility to use any camera model in Cartesian coordinates. Indeed while a pinhole image is described by its camera calibration parameters, a generic spherical image is only described with its circumference a, which corresponds to the image width (in pixels) under an angle of 2π. In fact, a spherical image can be intended as a unitary sphere S around the perspective centre and 3D point coordinates $\mathbf{u}$ can be expressed in terms of longitude θ and co-latitude φ as:

$$\mathbf{u} = \begin{bmatrix} u & v & w \end{bmatrix}^T = \begin{bmatrix} \sin\varphi\cos\theta & \sin\varphi\sin\theta & \cos\varphi \end{bmatrix}^T$$
$$\|\mathbf{u}\| = 1 \qquad (2)$$

The relation between a point onto the sphere and its corresponding 3D coordinates is $\mathbf{u} = \mathbf{X} / \| \mathbf{X} \|$. Homogenous image coordinates $\mathbf{m}$ can be mapped onto the sphere using the equi-rectangular projection:

$$\mathbf{m} = \begin{bmatrix} m_1 & m_2 & 1 \end{bmatrix}^T = \begin{bmatrix} R\theta & R\varphi & 1 \end{bmatrix}^T \qquad (3)$$

where $R = a/(2\pi)$ is the radius of the sphere.

Once the invariant features are extracted on the unwrapping spherical images, the median of the longitudes $\mu(\theta)$ is subtracted from θ, obtaining new longitudes $\theta^* = \theta - \mu(\theta)$. This allows the projection of a generic points $\mathbf{p}$ of the sphere onto the plane $u = 1$ as:

$$\mathbf{p} = \begin{bmatrix} 1 & p_2 & p_3 \end{bmatrix}^T = \begin{bmatrix} 1 & tg\theta^* & \cot\varphi/\cos\theta^* \end{bmatrix}^T \qquad (4)$$

Here, p_2 and p_3 can be intended as the inhomogenous image coordinates of a new pinhole image. Moreover, the centre of the spherical images is also the projection centre of the new pinhole image, with the advantage that given two spherical images S and S', an outlier can be removed by robustly estimating a fundamental matrix.

Obviously, this procedure cannot cope with large longitude variations. However, the partitioning of the spherical images into 4 zones ($k\pi/2 \leq \theta < (k+1)\pi/2$, $k = \{0,1,2,3\}$) produces 4 local pinhole images that can be independently processed.

Given the high resolution content of the panoramic images, the employed feature operators and matching strategy deliver a great number of image correspondences. To speed up the processing and bundle adjustment procedures, a reduction strategy is applied (Figure 5b). Each image is divided into rectangular cells and for each cell only the point with the best multiplicity (i.e. the number of images in which the point was matched) is stored.

The extracted image correspondences are then used to derive, with a bundle solution in spherical coordinates, the camera poses and a sparse 3D geometry of the analyzed scene.

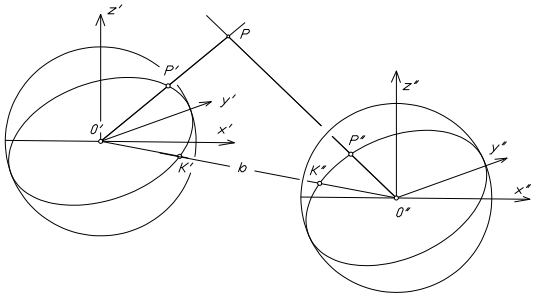


Figure 4: *The coplanarity condition between two spherical images.*

2.2 Geometric estimation

An image point P' and its corresponding 3D object point P must satisfy the collinearity principle expressed as:

$$x = R \cdot \left(-\theta_0 + atg \frac{(X - X_0) - \alpha_y (Y - Y_0)}{(Y - Y_0) + \alpha_x \cdot (Z - Z_0)} \right)$$
$$y = R \cdot a\cos \frac{\alpha_y (X - X_0) - \alpha_x (Y - Y_0) + (Z - Z_0)}{\sqrt{(X - X_0)^2 + (Y - Y_0)^2 + (Z - Z_0)^2}} \qquad (5)$$

where (see Figure 2 and Figure 3):

x, y … image coordinates of the panoramic image;
X, Y, Z … object 3D coordinates;
R … radius of the sphere;
α_x, α_y … correction angles (roll and pitch);
θ_0 … heading;
X_0, Y_0, Z_0 ... object coordinates of the centre of the sphere.

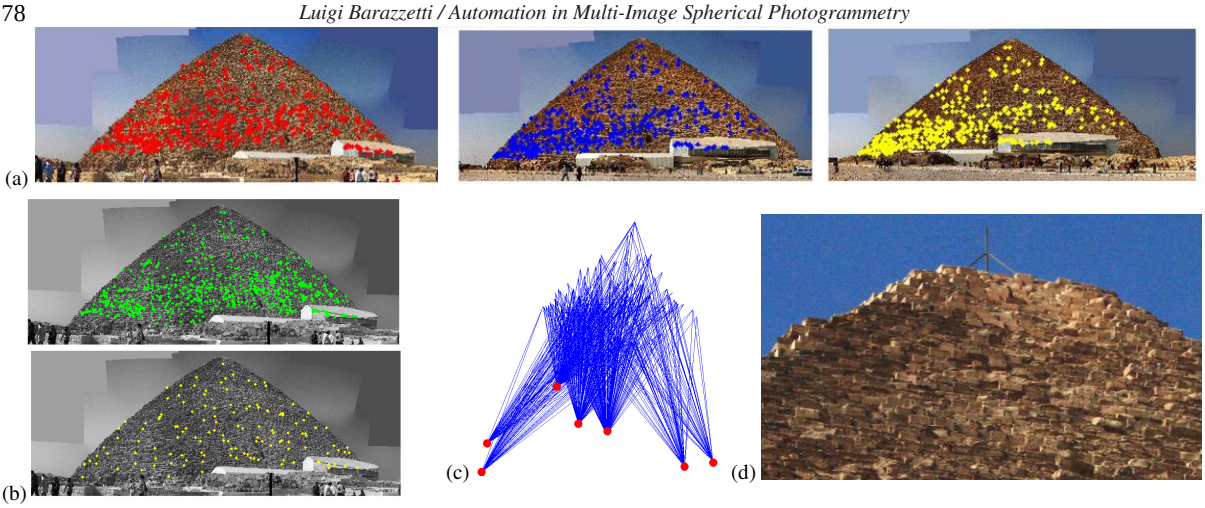


Figure 5: *Example of the automated orientation of panoramic images of the Keope's Pyramid (Egypt). (a) the extracted correspondences between an image triplet. (b) the results of the reduction procedure on the extracted feature correspondences. (c) the derived camera poses of 6 panoramas using a bundle solution in spherical coordinates. (d) the high geometric content of a panorama acquired from a distance of ca 200 m with a lens of 12 mm focal length (50 mm equivalent).*

Given a set of panoramas and image correspondences, the poses of the spheres and the object coordinates of the feature correspondences can be computed using Eq. (5) with a least squares bundle adjustment solution providing some 3D control points information to solve for the datum deficiency and derive metric results.
If no control information is available, one panorama can be relatively oriented with respect to any other one using the coplanarity condition:

$$\begin{bmatrix} X' & Y' & Z' \end{bmatrix} M'^{T} \begin{bmatrix} 0 & -bz & by \\ bz & 0 & -bx \\ -by & bx & 0 \end{bmatrix} M'' \begin{bmatrix} X'' \\ Y'' \\ Z'' \end{bmatrix} = 0 \qquad (6)$$

with (see Figure 4):
bx, *by*, *bz* … components of the baseline *b* between the two panoramas;
M' and *M"* … rotation matrices of the panoramas;
X',Y',Z' … object coordinates of point *P'*.
X",Y",Z" … object coordinates of point *P"*.
Some visual results of the extracted and reduced correspondences and the derived camera poses are shown in Figure 5.

2.3 Interactive 3D reconstruction

Once the camera poses have been recovered, the successive goal is the 3D reconstruction of the scene. In case of free form objects (statues, bass-relieves, etc.) automated matching procedure would be necessary to derive all the geometric details and discontinuities of the scene. At the moment no commercial solution is available to automatically match panoramic images. On the other hand, in case of architectural objects, the digital reconstruction is primarily carried out with interactive measurements (Figure 6) as fully automated procedures for 3D reconstructions of architectures lead to poor geometric results.

Starting from the estimated poses, the user selects homologous features in order to derive the 3D structures of the scene. Geometric elements like lines, curves and surfaces can be drawn and afterwards textured for photo-realistic rendering. In addition, external constraints (e.g. coplanarity of some particular points or parallelism among a set of lines) can be applied to improve the geometry of the 3D reconstruction.
The developed 3D modeling system can exchange information with several modeling and visualization packages (AutoCAD, 3D Studio Max, Maya, Rhinoceros, etc.) in order to perform some post-processing or to create animations, virtual tours and special effects. Figure 7 shows the interaction with 3D Studio Max for the editing and 3D rendering of the main façade of the Monastery of Petra (ca 70×40 m), modelled with only 3 spherical images (image resolution ca 2 cm).

Figure 6: *Some typical 3D modeling results from panoramic images based on interactive measurements.*

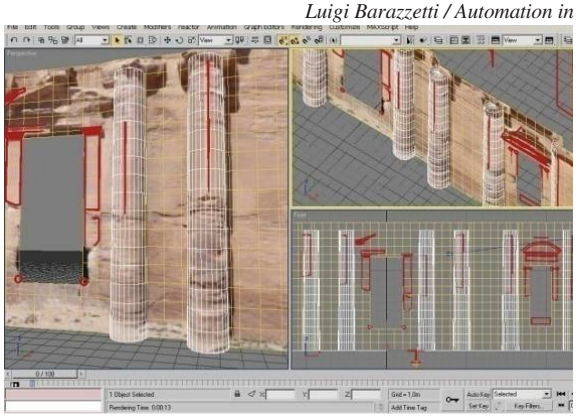


Figure 7: *The interactive 3D reconstruction of the scene starting from the oriented panoramas.*

3. Experiments and considerations

The developed 3D modeling procedure was used to automatically orient a set of panoramic images and then reconstruct the analysed scene with interactive measurements. Three datasets featuring different characteristics in terms of number of images, image resolution, field-of-view, object size, shape and texture are presented in detail (Figure 8).

Block 1 comprehends 16 spherical images of the interior and exterior of a church in Ascoli Piceno, Italy. The field of view of most images is larger than 270° (Figure 8a). The automated matching procedure to extract the image correspondences demonstrated to be robust even in the case of topological variations (e.g. object on the left in one image appears on the right in another one and vice versa). The white arrows of Figure 8b facilitate the interpretation of the matching results and demonstrate how the method can work with large disparities between the images. These large disparities and the strong geometric distortion of the images underline the importance of using reliable feature matching strategies.
As clearly visible from the results, the method provides for a great number of image correspondences, which are generally reduced to speed up the adjustment phase. The large number of image correspondences is primarily due to the high resolution of the images which are always processed at their original size allowing the derivation of the fine geometric details of the scene (Figure 8d).
The orientation phase is carried out with a bundle adjustment solution in spherical coordinates. An interesting issue is related to the presence of outliers in the set of image correspondences. These mismatches are due to the use of only two images in the outlier rejection step. Indeed the robust estimation of the fundamental matrix is performed on image pairs and gross errors lying on the epipolar line cannot be removed with the analysis of the epipolar geometry. However, these mismatches can be removed using a photogrammetric bundle adjustment coupled with a powerful statistical model for the analysis and detection of blunders, which becomes essential for these automated applications. The recovered camera poses (Figure 8c) allow then the scene 3D reconstruction, which is achieved with interactive measurements on stereo pairs. For this specific example, the accuracy and quality of the obtained 3D results was so good that the model could be used for restoration works. The derived geometric details and some architectural cross-sections (at scale 1:50) are shown in Figure 8d.

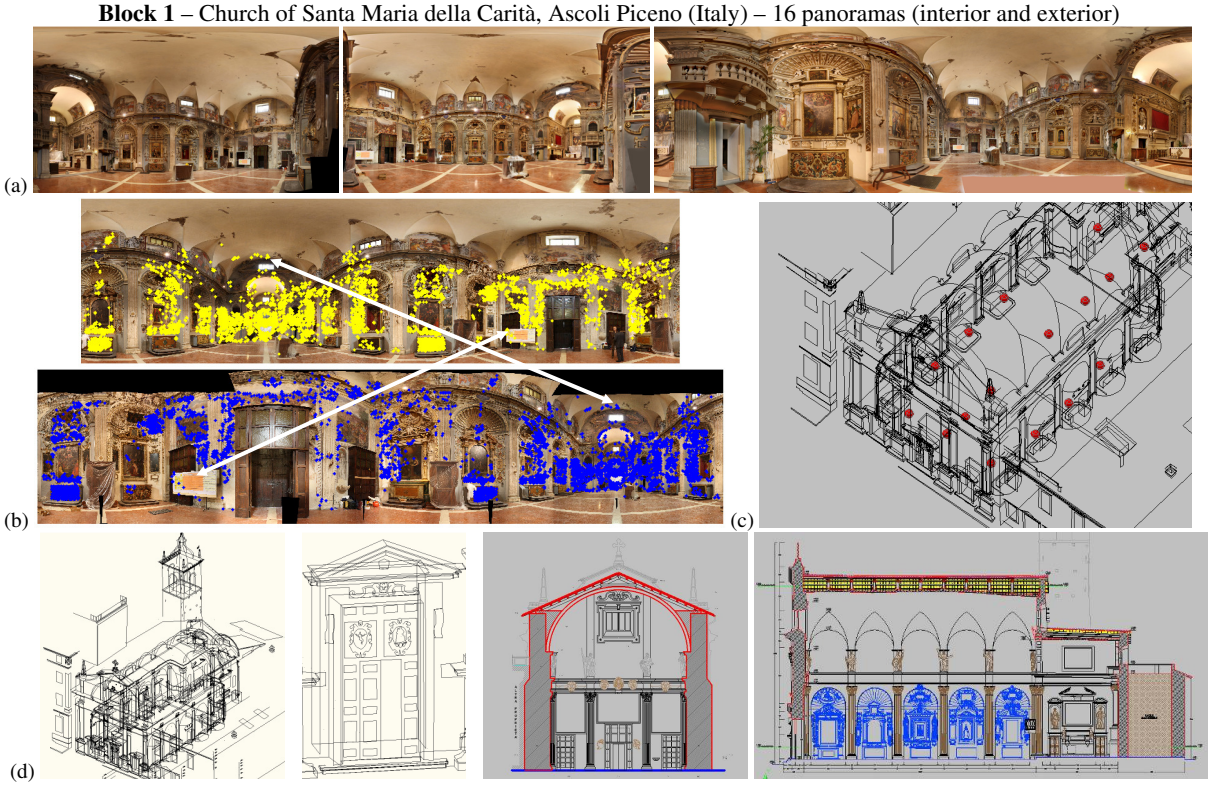
Block 2 consists of 3 spherical images of the main façade of the Monastery of Petra, Jordan. Each panorama is made up of ca 12-15 frame images and contains up to 95 Megapixels. The detected correspondences and the scheme of the image network are respectively shown in Figure 8e and Figure 8f. The statistics results of the bundle solution reported a precision of the computed camera poses of 4-5 cm, similar to those achievable with manual measurements. The derived orientation results allowed the successive scene reconstruction and the creation of a textured 3D model, which is shown in Figure 8g with the employed spherical images.

Block 3 shows the 3D modeling results for the church of San Pietro in Spoleto (Italy). The use of 3 spherical images allowed the accurate reconstruction of even small details on the top of the church (Figure 8l).

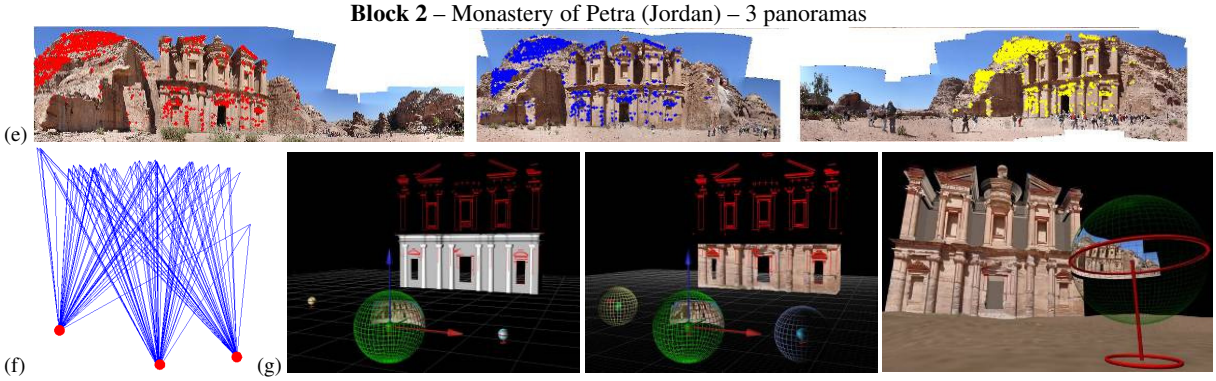
4. Conclusions

The metric use of panoramic images is nowadays feasible and with great perspective in many application fields. Panoramic images obtained by stitching together single frame imagery allow the generation of almost Gigapixel images with very high resolution contents and large fields of view. These are two requirements for detailed and photo-realistic 3D modeling, in particular for architectural applications.
The presented methodology is based on the automated orientation a set of panoramas using a camera model in spherical coordinates and the successive interactive scene reconstruction. Interaction is indeed not neglected as automation is still in the future for the 3D reconstruction of complex man-made structures.
The automated orientation phase demonstrated a great reliability even in case of scene's topological variations. The SURF operator gave the best results in terms of speed, number and distribution of extracted features as well as precision of the final estimated camera poses. Regarding the 3D reconstruction of the architectural scenes, the achieved 3D models are very detailed thanks to the high resolution of the panoramas. Similar geometric and radiometric results could be achieved with traditional pinhole images (frame), but with a larger number of images and processing time. Given the specific requirements and goals of project, the automation in the 3D restitution phase is not of primary importance and user-assisted reconstructions are mandatory.
Further improvements are related to the development of some ad-hoc procedures for image matching and dense 3D reconstruction for detailed areas, with particular attention to computational issues. Indeed, in some cases, the network geometry could allow the elaboration using only particular image combinations, with a consequent reduction of the processing time.

Block 1 – Church of Santa Maria della Carità, Ascoli Piceno (Italy) – 16 panoramas (interior and exterior)



Block 2 – Monastery of Petra (Jordan) – 3 panoramas



Block 3 – Church of San Pietro, Spoleto (Italy) – 3 panoramas

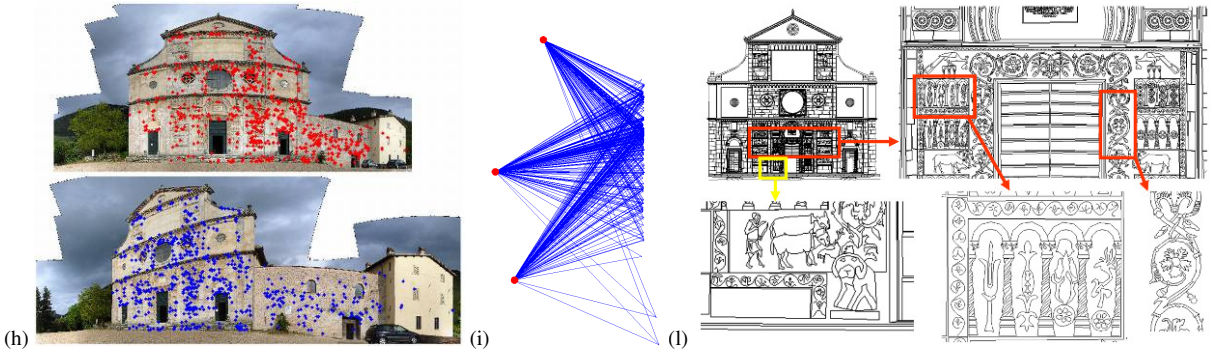


Figure 8: *Results for several datasets of panoramic images, showing the automatically matched points (b, e, h), orientation results (c, f, i) and 3D reconstructions of the scene (d, g, l). More details of the figures are given in the text.*

Acknowledgements

The authors would like to thank Emanule Ministri, Enzo d'Annibale, Sabino Massa, Elisa Cingolani, Giacomo Mariotti and Massimiliano Villani for their valuable contributes during the digital restitution of the presented 3D models.

References

[ASS09] S. AGARWAL, N. SNAVELY, I. SIMON, S.M. SEITZ, R. SZELISKI. Building Rome in a day. *Int. Conference on Computer Vision*, Kyoto, Japan, 2009.

[BETV08] H. BAY, A. ESS, T. TUYTELAARS, L. VAN GOOL. SURF: Speeded up Robust Features. *Computer Vision and Image Understanding*, 110(3), pp. 346-359, 2008.

[BRS09] L. BARAZZETTI, F. REMONDINO, M. SCAIONI. Combined use of Photogrammetric and Computer Vision techniques for fully automated and accurate 3D modeling of terrestrial objects. Proc. of "Videometrics, Range Imaging, and Applications X", *SPIE*, Vol. 7447, 12 pp, San Diego (CA, USA), 2009.

[BRS10] L. BARAZZETTI, F. REMONDINO, M. SCAIONI. Automation in 3D reconstruction: results on different kinds of close-range blocks. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences,* Vol. 38(5), 2010.

[BL03] M. BROWN, D.G. LOWE. Recognising panoramas. *In Proc. of 9th ICCV,* Vol.2, 2003.

[BL07] M. BROWN, D.G. LOWE, D. G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision* 74(1), 59-73, 2007.

[Fan06] G. FANGI. Investigation On The Suitability Of The Spherical Panoramas By Realviz Stitcher For Metric Purposes. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences* Vol. 36(5), 2006.

[Fan07] G. FANGI. The Multi-Image spherical panoramas as a tool for architectural survey. Proc. of XXI Int. *CIPA Symposium*, Athens, Greece, 2007.

[FFG09] M. FARENZENA, A. FUSIELLO, R. GHERARDI. Structure and Motion pipeline on a hierchical cluster tree. *In: proc. of IEEE Int. Work. on "3D Digital Imaging and Modeling" (3DIM)*, Kyoto, Japan, 2009.

[FCSS10] Y. FURUKAWA, B. CURLESS, S.M. SEITZ, R. SZELISKI. Toward Internet-scale multiview stereo. *IEEE Computer Vision and Pattern Recognition*, San Francisco, California, June 13-18, 2010.

[HZ04] R.I. HARTLEY, A. ZISSERMAN. *Multiple view geometry in Computer Vision*. Cambridge, Cambridge University Press, ISBN 0521540518, 672 pp., 2004.

[LF06] T. LABE, W. FÖRSTNER. Automatic relative orientation of images. *Proc. of the 5th "Turkish-German Joint Geodetic Days"*, 6 pp., 2006.

[Low04] D. LOWE. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), pp. 91-110, 2004.

[LT02] T. LUHMANN, W. TECKLENBUR G. Bundle orientation and 3-D object reconstruction from multiple-station panoramic imagery. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 34(5), pp. 181-186, 2002.

[LT04] T. LUHMANN, W. TECKLENBURG. 3-D Object reconstruction from multiple-station panorama imagery. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 34(5/W16), on CD-ROM, Dresden, Germany, 2004.

[PG04] J.A. PARIAN, A. GRUEN. An advanced sensor model for panoramic cameras. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol 35(5), pp. 24-29, 2004.

[PV02] M. POLLEFEYS, L. VAN GOOL. From images to 3D models. *ACM Communications*, 45(7), 50-55, 2002.

[RR06] F. REMONDINO, C. RESSL. Overview and experience in automated markerless image orientation. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 36(3), pp. 248-254, 2006.

[RFR05] R. RONCELLA, G. FORLANI, F. REMONDINO. Photogrammetry for geological applications: automatic retrieval of discontinuities in rock slopes. Proc. of "Videometrics VIII", *SPIE IS&T Electronic Imaging*, pp. 17-27, 2005.

[RL87] P.J. ROUSSEEUW, A.M. LEROY. *Robust Regression and Outlier Detection*. New York, John Wiley, ISBN 978-0471852339, 352 pp., 1987.

[SM04] D. SCHNEIDER, H.-G. MAAS. Application and Accuracy Potential of a Strict Geometric Model for rotating Lines Cameras. *Int. Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 34(5/W16), on CD-ROM, Dresden, Germany, 2004.

[SS97] R. SZELISKI, H. SHUM. Creating full view panoramic image mosaics and environment maps. *Proc. of SIGGRAPH*, pp. 251-258, 1997.

[Tor02] P. TORR. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *Int. Journal of Computer Vision*, 50(1): 35-61, 2002.

[VV06] M. VERGAUWEN, L.J VAN GOOL. Web-based reconstruction service. *Machine Vision and Applications*, 17(6), pp. 411-426, 2006.