FISHEYE LENSES FOR 3D MODELING: EVALUATIONS AND CONSIDERATIONS

L. Barazzetti[1], M. Previtali[1], F. Roncoroni[2]

[1]Dept. of Architecture, Built environment and Construction engineering (ABC)
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, Italy
[2]Polo Territoriale di Lecco, via Previati 1/c, Lecco
(luigi.barazzetti, mattia.previtali, fabio.roncoroni)@polimi.it
http://www.gicarus.polimi.it

Commission II

ABSTRACT:

Fisheye lenses are becoming more popular in complete image-based modelling projects of small and narrow spaces. The growing interest in fisheye lenses is confirmed by the availability of different commercial software incorporating a fisheye camera model. Such software are now able to carry out the steps of the image processing pipeline in a fully automated way, from camera calibration and orientation to dense matching, surface generation, and orthophoto production. This paper highlights the advantages (and disadvantages) of fisheye lenses when used for 3D modelling projects through different commercial software. The goal is not only a comparison of commercial software, but also an analysis of the additional issues that arise when a fisheye lens is used for 3D modelling. Results confirm that a fisheye lens is suitable for accurate metric documentation, especially when limited space is available. On the other hand, additional issues where found during the camera calibration/image orientation step as well as the texture generation and orthophoto production phases, for which particular attention is required.

1. INTRODUCTION

The fisheye camera model for photogrammetric applications has been extensively studied, tested and validated in the first decade of 2000s. Calibration procedures were presented by Abraham and Förstner (2005), Schwalbe (2005), Van den Heuvel et al. (2006) and Schneider et al. (2009), among the others.

The recent introduction of the fisheye camera model in some commercial packages for automated image-based 3D modelling (such as PhotoScan, Pix4D and ContextCapture) has allowed both professional and "less expert" users to generate 3D models in a fully automated way, starting from a set of digital images. Results presented in Strecha et al. (2015) confirm the new level of automation achievable for the different steps of the image modelling workflow: camera calibration, dense matching and surface extraction.

Such level of automation for fisheye cameras is quite similar to the automation already achievable in projects based on central perspective cameras (pinhole cameras). However, the risk of unreliable and "crude" digital reconstructions because of the lack of expertise in basic surveying concepts has already been described in Nocerino et al. (2014), in which the authors presented inaccurate reconstructions obtained from pinhole (central perspective / frame) images.

In the case of a fisheye lens, the short focal length coupled with an extreme distortion makes automated 3D modelling more complex. This could provide inaccurate 3D models without metric integrity. Indeed, an unfavourable network geometry for object reconstruction coupled with a process integrating also camera calibration parameters, can easily result in deformed reconstruction.

The variable ground sampling distance is also important to plan an appropriate scheme for image acquisition, since different parts of the object are captured with variable resolution.

The incorporation of the fisheye camera model in commercial software is a clear indicator about how users are becoming more familiar with such distorted projections, not only for photographic purposes but also for metric applications. Nowadays, automated fisheye image processing is possible without turning them into pinhole images, exploring the full potential of their wide field of view. Fisheye can significantly reduce the typical number of images required for indoor applications, simplifying both the acquisition phase (limiting the size of the image dataset) and the orientation step with a more reliable bundle block adjustment.

As there is no unique camera model for fisheye lenses, different mathematical formulations were incorporated in commercial software. This means that a direct software comparison based on interior/exterior image parameters is not possible and the same set of calibration parameters cannot be used in different packages. Results are therefore very software-depended (say technology-driven). Only an evaluation of the final model (in terms of metric accuracy, completeness, resolution, level of detail, etc.) can define the quality of the reconstruction.

In this paper, we carried out some experiments that analyse all the steps of the image processing pipeline. Different software ware tested to highlight the main advantages and disadvantages in a 3D modelling project carried out with such distorted projections.

2. FISHEYE CAMERA CALIBRATION

As mentioned, different mathematical models for camera calibration and orientation are available for fisheye lenses. A comprehensive review and accuracy evaluation is presented in Schneider et al. (2009), in which targets were used to simplify the tie point extraction phase, obtaining more precise image coordinates. In this work, we decided to try out three commercial software incorporating a fully automated workflow for 3D modelling: from camera calibration to surface reconstruction. The experiments presented in the next sections (accuracy of image orientation, point cloud generation and surface extraction) were always carried out with a calibrated

Nikon D700 with a 16 mm Nikkor fisheye lens, which means that calibration parameters (interior orientation and distortion coefficients) were estimated beforehand, eliminating the unknowns for camera calibration from a project for object reconstruction (fixed calibration).

This choice is motivated by the need of a particular network geometry for camera calibration, as illustrated in Remondino and Fraser (2006) for the case of central perspective images (pinhole camera model). Such image blocks require convergent images with roll variations and variable camera object distances.

The camera was set at infinity during image acquisition to remove errors caused by auto-focus. As the software employed cannot use only coded targets for image orientation, we decided to exploit the targetless camera calibration principle presented in Barazzetti et al. (2011) and Stamatopoulos and Fraser (2014). 74 images of a wall with a good texture were acquired and processed with the three software, obtaining an overall RMS of image coordinates of about 0.2 – 0.4 pix, which is worse than a typical calibration with coded targets, for which ±0.1 pix is expected. On the other hand, the targetless project has a larger redundancy and demonstrated to be equivalent to the traditional approach based on targets.

As mentioned, different software could exploit different mathematical models for image orientation. ContextCapture and Pix4D use an equidistant model in which the angle θ between an incident ray and the camera direction is estimated as:

$$\theta = \frac{2}{\pi} arctan\left(\frac{\sqrt{X^2+Y^2}}{Z}\right) \qquad (1)$$

where (X, Y, Z) are 3D coordinates in a camera centred reference system. The relationship between image coordinates (x, y) and 3D points is then written as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} C & D \\ E & F \end{bmatrix} \begin{bmatrix} \rho X/\sqrt{X^2+Y^2} \\ \rho Y/\sqrt{X^2+Y^2} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \qquad (2)$$

where C, D, E, F, c_x, c_y form an affine transformation, and ρ can be estimated as:

$$\rho = p_0 + \theta + p_2\theta^2 + p_3\theta^3 \qquad (3)$$

PhotoScan is based on the equidistand projection with the generic form:

$$x = \frac{f}{\sqrt{\left(\frac{x}{Y}\right)^2+1}} arctan\left(\frac{\sqrt{X^2+Y^2}}{Z}\right) + c_x + \Delta x$$

$$y = \frac{f}{\sqrt{\left(\frac{x}{Y}\right)^2+1}} arctan\left(\frac{\sqrt{X^2+Y^2}}{Z}\right) c_y + \Delta y \qquad (4)$$

where f is the focal length, and $\Delta x, \Delta y$ are additional terms to compensate for systematic error. Additional parameters are described by the radial symmetric distortion and decentring distortion proposed by Brown (1971) as well as parameters to model affinity and shear (El-Hakim, 1986).

A visualization of camera poses after bundle adjustment is shown in Fig. 1. An evaluation of calibration parameter quality is not simple for the lack of complete statistics (variance-covariance matrix) to check (at least) parameter precisions and correlations. For this reason, the quality of calibration parameters will be checked during the next phases of the reconstruction (see next sections), in which calibration parameters will be assumed as fixed values.
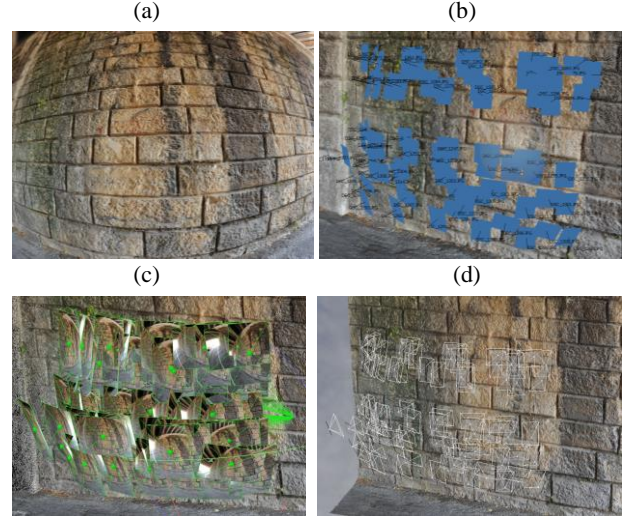
(a) (b)

(c) (d)

Figure 1. (a) An image of the block used for camera calibration and camera poses in the calibration project with PhotoScan (b), Pix4D (c) and ContextCapture (d).

3. METRIC ACCURACY OF LONG FISHEYE IMAGE SEQUENCES

Previous work carried out by different authors (e.g. Nocerino et al., 2014) demonstrated the lack of accuracy in the case of long image sequences, especially when few control points are used or a reliable mathematical model for absolute orientation is not taken into consideration. This is the case of free-network solutions that are then rigidly rotated, translated and scaled with a 7-parameter transformation.

The short focal length coupled with a wide field of view and a strong visual distortion of fisheye lenses makes the problem of network deformations even more important. The aim of this section is to analyse network distortion for the case of calibrated and uncalibrated cameras. A long sequence made up of 93 images was acquired with the Nikon D700 equipped with the 16 mm fisheye lens. A set of 3D points (targets) was measured with a total station Leica TS30, obtaining 22 points with a precision better than ±1 mm to be used as ground control points and check points.
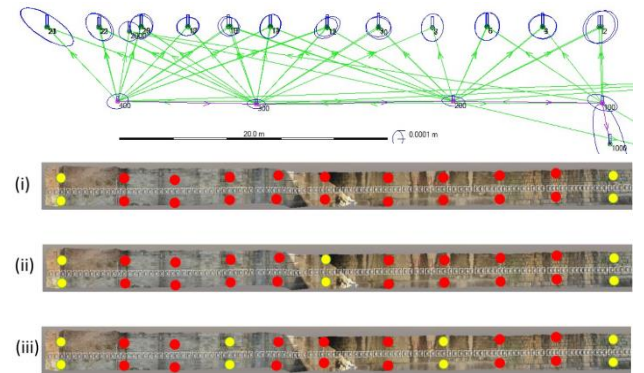

Fig. 2. The straight wall used to try out the accuracy of image orientation. Top: 3D points measured with a total station; bottom: the different ground control points (yellow) and check points (red) used in the different image-based projects.

Image orientation was carried out with different software and their calibration parameters, which were estimated in section 2. Different ground control point / check point configurations were tested to check metric accuracy. Figure 2 shows the configurations used in this work: (i) 4 GCPs and 18 check points, (ii) 6 GCPs and 16 check points, and (iii) 8 GCPs and 14 check points.

Figure 3 shows orientation results (camera poses and 3D points) for the different software: the image sequence is about 40 m and the average baseline between consecutive images is 0.43 m. One of the problems is an overall bending effect in the sequence. Ground control points and a reliable mathematical model for image orientation (in which GCPs are rigorously incorporated to reduce network deformations) are mandatory for accurate image-based projects.
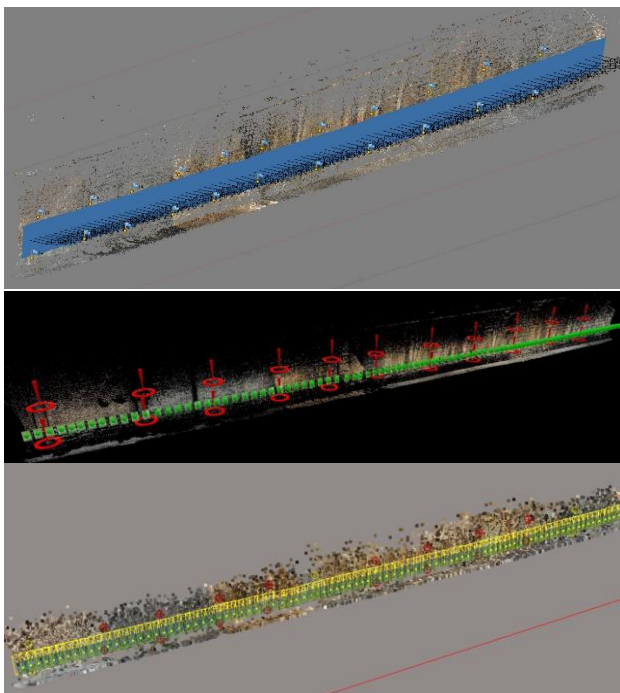

Figure 3. Camera poses and 3D points estimated with the different software.

Figure 4 shows check point errors (in terms of RMS values) for the used software. Results for a configuration with 4 ground control points and 18 check points demonstrate that large errors were found for all software (top). The reconstruction is always affected by a significant geometric deformation, which is a bending effect for ContextCapture and Pix4D (main errors for *Y* coordinates, i.e. the depth), whereas PhotoScan has an additional relevant error along *X*.

The number of control points was then increased by adding two points in the middle of the sequence, obtaining an improvement of metric accuracy for Pix4D and ContextCapture, which use GCPs to remove network deformations. PhotoScan absolute orientation is instead based on a rigid 7-parameter transformation, which cannot modify the geometry obtained with a free-network adjustment. Results with the last point configuration (bottom, 8 ground control points and 16 check points) highlight an error of few millimetres for Pix4D and ContextCapture. The overall metric accuracy with PhotoScan (some centimetres) is much worse, also when compared to the average ground sampling distance (some millimetres).

Errors with 4 GCPs and 18 check points (meters)	RMS X	RMS Y	RMS Z
PIX4D	0.0100	0.1246	0.0054
PHOTOSCAN	0.1203	0.0371	0.0456
CONTEXTCAPTURE	0.0107	0.0476	0.0064

Errors with 6 GCPs and 16 check points (meters)	RMS X	RMS Y	RMS Z
PIX4D	0.0025	0.0137	0.0012
PHOTOSCAN	0.0687	0.0230	0.0259
CONTEXTCAPTURE	0.0033	0.0141	0.0028

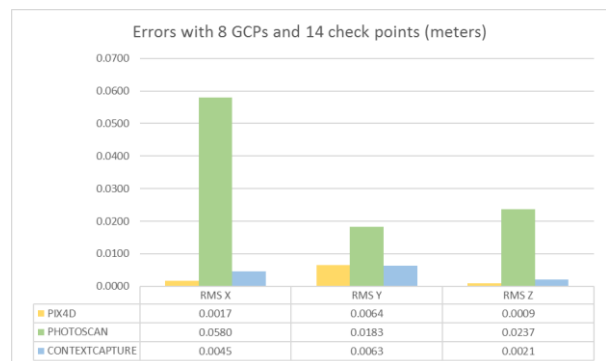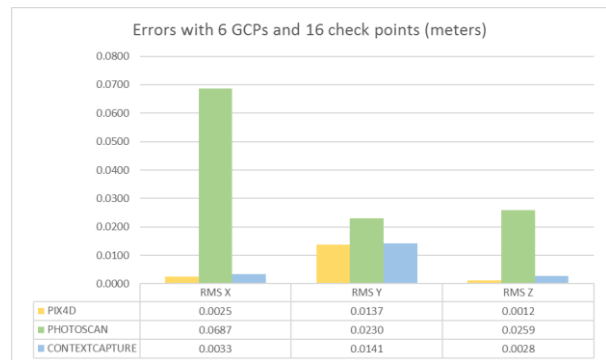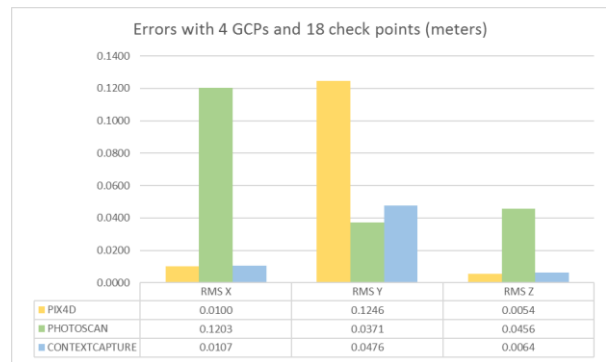| Errors with 8 GCPs and 14 check points (meters)	RMS X	RMS Y	RMS Z
PIX4D	0.0017	0.0064	0.0009
PHOTOSCAN	0.0580	0.0183	0.0237
CONTEXTCAPTURE	0.0045	0.0063	0.0021

Figure 4. Metric accuracy achieved by different software with different control point / check point configurations.

Results demonstrate that a good metric accuracy can be obtained with fisheye lenses. On the other hand, a geometric model for absolute orientation able to incorporate GCPs and remove network deformations is needed. At the same time, check points remain mandatory to check the real metric accuracy, whereas statistics on image points (e.g. RMS of image coordinates) are not sufficient to understand the quality of a project.

Finally, the same sequence was oriented without using the calibration parameters estimated in section 2. The deformed shape of the sequence (with ContextCapture) is shown in Fig. 5. A long sequence of images does not provide a reliable network geometry able to incorporate calibration parameters as additional unknowns. The overall error is much larger than the error achieved with the same camera and a pinhole lens (20 mm Nikkor), for which a deformation was also quite evident. On the other hand, the use of a 16 mm fisheye gave a curvature of about 45°, much larger than the deformation achieved with the pinhole camera.

For this reason, a generic reconstruction project should be always carried out with a calibrated camera, especially in the case of fisheye lenses.

Figure 5. Results with calibration parameters estimated in the orientation step (a straight wall becomes a circular sector). The geometry of a long sequence is not reliable and calibration parameters cannot be incorporated as additional unknowns.

4. POINT CLOUD CREATION AND SURFACE RECONSTRUCTION WITH FISHEYE LENSES

The surface of a room with a vault was reconstructed with a set of 65 images. The shape of the room required a set of images in front of the walls coupled with some "normal photographs" that captured the vault and its frescoes. The connection between "vertical" and "horizontal" images was guaranteed by some convergent images with an angle of 45°.

Image acquisition required only 3 minutes. The scale ambiguity of the image-based reconstruction was removed with a known distance between two targets.

Data processing was carried out with PhotoScan, Pix4D and ContextCapture. Image orientation was fully automated and calibration parameters were assumed as fixed quantities. Shown in Fig. 6 is a 3D view with the different software.
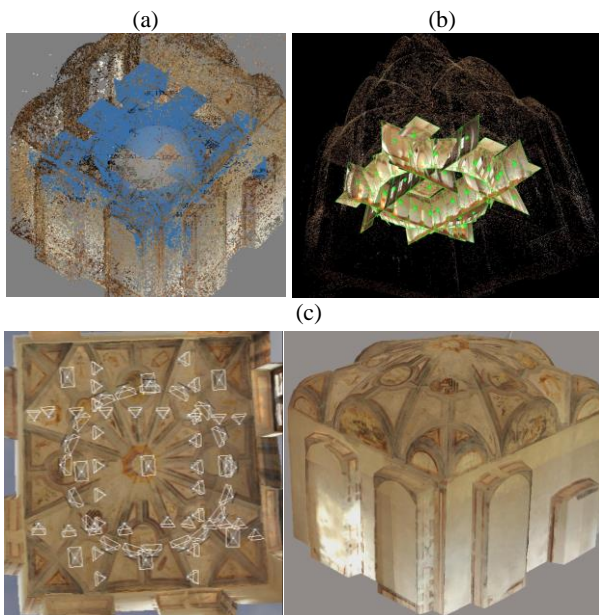
(a)   (b)

(c)

Figure 6. Results with different software for a room with a vault reconstructed with a fisheye lens.

The accuracy evaluation was carried out by comparing the point clouds generated with the three software and a laser scanning point cloud collected with a Faro Focus 3D (precision ±2 mm). Image-based results were registered in the same reference system of laser scanning data with the ICP algorithm of Geomagic Studio. Accuracy was then evaluated with CloudCompare, obtaining the error maps shown in Fig. 7.

Statistics revealed very similar results for the different software, which were 3.4mm ± 2.1mm for PhotoScan (average and standard deviation), 3.5mm ± 2.2mm for Pix4D and 3.3mm ± 2.2mm for ContextCapture. This means that results achieved with different software are comparable in terms of metric accuracy.
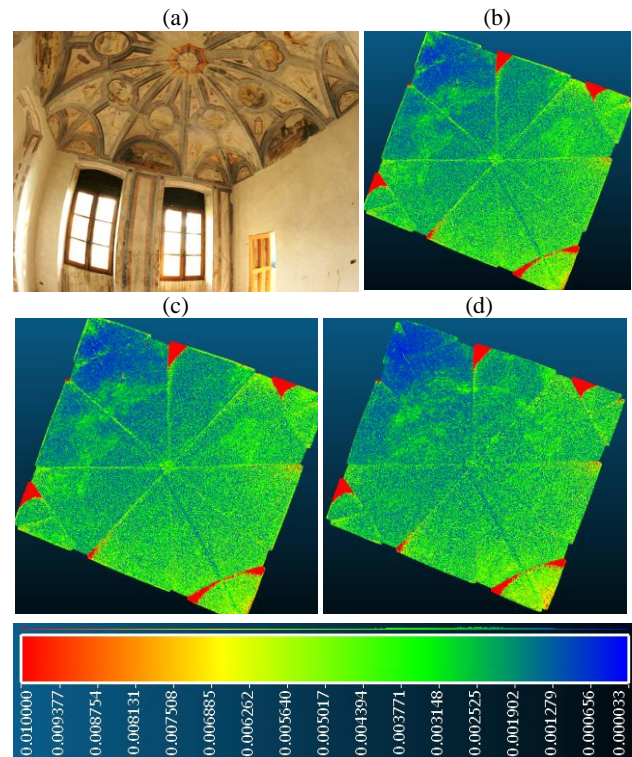
(a)   (b)

(c)   (d)

Figure 7. An image of the vault (a) and comparison between laser and image-based point clouds for the different software: (b) PhotoScan, (c) Pix4D, (d) ContextCapure.

5. TEXTURE MAPPING AND ORTHOPHOTO GENERATION WITH FISHEYE LENSES

Texture mapping is an aspect that plays an important role in 3D modelling. Here, the characteristics of fisheye lenses should be taken into consideration to create sharp photorealistic models. In fact, the achieved results revealed an important limitation.

The creations of photorealistic products (textured mesh and orthophotos) with a high metric accuracy is one of the reasons that made automatic software for 3D modelling for image processing very popular. The opportunity to capture object texture with consumer or professional cameras is also more attractive than reconstruction based on laser scanning technology, in which the incorporated camera does not provide the quality of results achievable with photogrammetric solutions.

The first consideration is that fisheye lenses allow one to capture images with a wide field of view. The number of images can be strongly reduced, especially in small and narrow spaces.

On the other hand, the opportunity to generate true-orthophotos requires images with a complete coverage of the object. Although images can be decimated when compared to a more traditional approach based on standard frames, a complete reconstruction still requires more images than those strictly necessary in terms of image overlap, especially for 3D objects with geometric irregularities and occlusions.

However, this was not the main issues, which is instead related to an unwanted "blur effect" in the final texture. An example is shown in Fig. 8. The orthophoto of the wall was generated with a fisheye lens because of the limited space available for image acquisition. The narrow corridor is about 1 m wide the use of a fisheye lens was a good solution for 3D modelling.

The reconstruction was initially carried out with a single strip, which was sufficient to capture the whole object. The first phases of the photogrammetric process were carried out in a fully automatic way, obtaining an accurate mesh. The texture mapping step revealed instead a significant drawback close to the edges of the model.
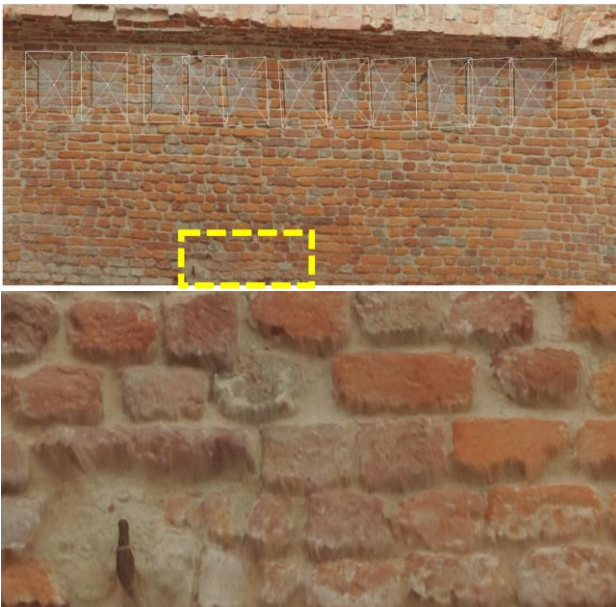

Figure 8. The unwanted effect resulting from the use of the whole image.

The variable GSD of a fisheye photograph is not the only reason behind this unwanted effect. The wide field of view allows one to capture a large portion of the object, but areas close to the image edges are imaged with a very narrow angle. The effect becomes extremely evident when the image is orthorectified.

The same wall was therefore reconstructed with an additional strip (Fig. 9), so that only the central part of the images was used during the orthorectification process. This allowed one to overcome the previous limitation, which however doubled the number of images in the project.

Similar results were found for other projects aimed at generating textured 3D models and orthophotos. The general consideration is that the wide field of view of a fisheye lens cannot be fully used for textured model and orthophoto production when the final texture must be sharp and homogenous. Additional images are needed to guarantee a uniform ground sampling distance and a smaller viewing angle, that could be intended as the opposite requirement of image-based projects with fisheye lenses.
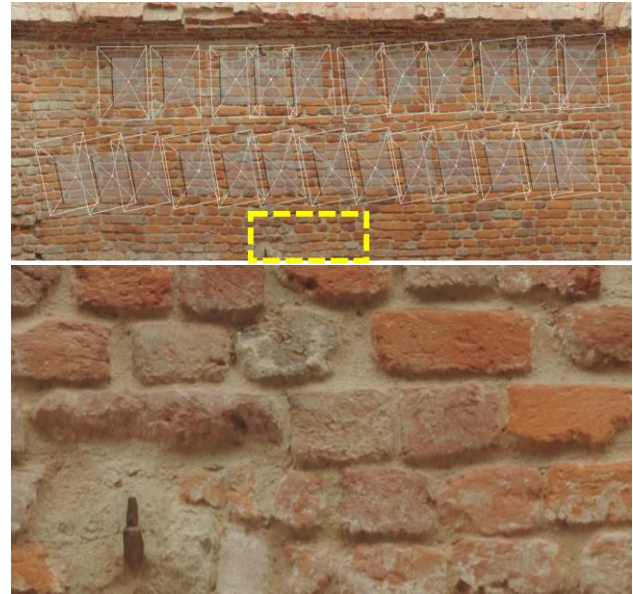

Figure 9. Results with an additional strip allow one to generate a better textured model / orthophoto.

6. 3D MODELLING EXAMPLES AND CONCLUSIONS: WHY USE A FISHEYE LENS

The previous sections highlighted that a fisheye lens is suitable for accurate 3D modelling, notwithstanding particular attention has to be paid for the use of the whole field of view, especially for the case of textured models and orthophotos.

A camera equipped with a fisheye seems a convenient choice for small and narrow spaces, in which a huge number of traditional pinhole image would be necessary.

An example is the very narrow space shown in Fig. 10, which was automatically reconstructed with 222 images, i.e. are more than those strictly needed. The scene is a 360° narrow corridor with a very irregular shape. The large number of images allowed one to capture the same areas from multiple viewpoints, increasing metric accuracy and obtaining a good coverage for orthophoto generation. Data processing was carried out with ContextCapture and a calibrated fisheye camera.

Figure 10. A narrow scene reconstructed with 222 images.

Another example is shown in Fig. 11, which required only 19 images. The barrel vault and the vertical walls form a small 3D space for which the use of a fisheye allowed a very rapid a simple acquisition phase. Automatic image processing allowed one to reconstruct a textured 3D model (top) from which very detailed orthophotos where extracted. In the case of the barrel vault, the reconstructed surface was unrolled by fitting a cylinder. The final orthophoto (middle) follows the curvature, in which *x*-coordinates correspond to the effective length measured along a circumference. The vertical wall was instead orthorectified with a traditional planar projection (bottom).

Figure 11. Textured 3D models (top), orthophotos of the unrolled vault (middle) and vertical wall (bottom) generated with a calibrated fisheye camera.

Such results demonstrate the potential of fisheye lenses for accurate 3D modelling. On the other hand, some issues have to be taken into consideration, among which the importance of camera calibration. The cameras should be calibrated beforehand by using an image block with suitable geometry. Network deformations for wrong camera calibration can be much larger than typical deformations with pinhole images. In addition, sharp textures require to limit the wide field of view of fisheye lenses, which could be intended as the paradox of a reconstruction based on fisheye images.

REFERENCES

Abraham, S., Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. ISPRS Journal of Photogrammetry and Remote Sensing 59 (5), 278-288.

Barazzetti, L., Mussio, L., Remondino, F., Scaioni, M. (2011). Targetless Camera Calibration. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-5/W16, pp.335-342.

Brown, D., 1971. Close-Range Camera Calibration. Photogrammetric Engineering 37(8), 855-866.

El-Hakim, S.F., 1986. Real-time image meteorology with ccd cameras. Photogrammetric Engineering and Remote Sensing 52 (11), 1757-1766

Fraser, C.S., Edmundson, K.L., 2000. Design and implementation of a Computational Processing System for Off-Line Digital Close-Range Photogrammetry. ISPRS Journal of Photogrammetry & Remote Sensing, 55(2): 94-104.

James, M.R., Robson, S., 2014. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. Earth Surf. Proc. Landf.; 39:1413–1420.

Nocerino, E., Menna, F., Remondino F., 2014. Accuracy of typical photogrammetric networks in cultural heritage 3D modelling projects. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, Volume XL-5, 465-472.

Remondino, F., Fraser, C,. 2006. Digital camera calibration methods: Considerations and comparisons. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.; 36: 266–272.

Schneider, D., Schwalbe, E., Maas, H.-G., 2009. Validation of geometric models for fisheye lenses. ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 259–266

Schwalbe, E., 2005. Geometric Modelling and calibration of fisheye lens camera systems. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 5/W8), (on CD-ROM).

Stamatopoulos, C., Fraser, C.S, 2014. Automated Target-Free Network Orientation and Camera Calibration. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2014;1: 339–346.

van den Heuvel, F., Verwaal, R., Beers, B., 2006. Calibration of fisheye camera systems and the reduction of chromatic aberration. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 5), (on CD-ROM).

Agisoft PhotoScan: http://www.agisoft.com

ContextCapture: https://www.acute3d.com/

Pix4D: https://pix4d.com/

Geomagic: http://www.geomagic.com

CloudCompare: http://www.danielgm.net/cc/