www.czasopisma.pan.pl PAN www.journals.pan.pl
POLSKA AKADEMIA NAUK

GEODESY AND CARTOGRAPHY
Vol. 62, No 2, 2013, pp. 123-138

DOI: 10.2478/geocart-2013-0007

3D modeling of architectural objects from video data obtained with the fixed focal length lens geometry

Paulina Deliś, Michał Kędzierski, Anna Fryśkowska, Michalina Wilińska

Military University of Technology
Faculty of Civil Engineering and Geodesy
Institute of Geodesy
Department of Photogrammetry and Remote Sensing
2 Kaliskiego St., 00-907 Warsaw, Poland
e-mail: pdelis@wat.edu.pl; mkedzierski@wat.edu.pl,
afryskowska@wat.edu.pl, mwilinska@wat.edu.pl

Received: 13 March 2013 / Accepted: 24 June 2013

Abstract: The article describes the process of creating 3D models of architectural objects on the basis of video images, which had been acquired by a Sony NEX-VG10E fixed focal length video camera. It was assumed, that based on video and Terrestrial Laser Scanning data it is possible to develop 3D models of architectural objects. The acquisition of video data was preceded by the calibration of video camera. The process of creating 3D models from video data involves the following steps: video frames selection for the orientation process, orientation of video frames using points with known coordinates from Terrestrial Laser Scanning (TLS), generating a TIN model using automatic matching methods. The above objects have been measured with an impulse laser scanner, Leica ScanStation 2. Created 3D models of architectural objects were compared with 3D models of the same objects for which the self-calibration bundle adjustment process was performed. In this order a PhotoModeler Software was used. In order to assess the accuracy of the developed 3D models of architectural objects, points with known coordinates from Terrestrial Laser Scanning were used. To assess the accuracy a shortest distance method was used. Analysis of the accuracy showed that 3D models generated from video images differ by about 0.06 ÷ 0.13 m compared to TLS data.

Keywords: video image, image orientation, bundle adjustment, Terrestrial Laser Scanning, 3D modeling

1. Introduction

Worldwide, the problem of modeling 3D architectural objects using a video camera had often been discussed (Frahm et al., 2008; Hansen et al., 2004; Tian et al., 2010). Detailed 3D models are used in digital preservation and restoration, physical replicas, virtual tourism, research and education. What distinguishes this technique from other data acquisition methods for 3D modeling are low costs, rapid data acquisition and the possibility of working from an unstable instrument platform, the possibility of obtaining a whole sequence of images in a short time (not as in case of a digital camera – only single images), and last but not least, there is common access to this popular type of image data.

In literature, the most frequent restrictions of data acquisition with a video camera include (Frahm et al., 2008): instability of intrinsic camera parameters, significant lens distortions, low resolution of video images, instability of the video images radiometry (the changes in appearance between subsequent scenes could potentially be non-linear).

The appearance of a commercial fixed focal length video camera, for which it is possible to determine the intrinsic camera parameters, enabled the achievement of high accuracies of the 3D models in comparison with commonly known methods of 3D modeling using video techniques.

2. Photogrammetric and Computer Vision 3D modeling methods based on video data

Methods for 3D modeling based on video data can be divided into two groups. The first group includes methods which are used in Computer Vision (CV), in which emphasis is placed on the speed and automation of developing a 3D model of a building, usually at the expense of accuracy. The second group includes photogrammetric methods, in which accuracy is a priority factor.

In CV, techniques for developing 3D models of architectural objects from video data are most commonly based on the "Structure From Motion" (SFM) approach. The SFM relies on determining of the 3D point coordinates and the camera projection matrix simultaneously, based on homologous points measured on a large number of images. There are SFM methods which start with an initial reconstruction from two views (Hansen et al., 2004) or from three views (triplet-based approach) (Hao and Mayer, 2003). In case of the initial reconstruction from two views, having identified corresponding points in two views, it is possible to compute their epipolar geometry. In order to apply the SFM approach for more than two views, sequential and factorization algorithms have been developed (Robertson and Cipolla, 2009).

The reconstruction ambiguity could be reduced by changing the geometry from projective to metric through self-calibration. A large majority of self-calibration algorithms used in CV impose intrinsic camera parameters as constant but unknown (Faugeras et al., 1992). Self-calibration in CV methods is described by (Zhang, 2000).

Alternative approaches developed for calibrated Structure From Motion algorithms include Sum-of-Sum of Square-Differences (SSSD) (Okutomi and Kanade, 1993) and space-carving (Kutulakos and Seitz, 2000).

Photogrammetric 3D modeling methods require knowledge of the intrinsic camera parameters which are determined during the camera calibration process. In order to process data from the variable focal length video camera, a self-calibration process was performed.

Among self-calibration methods, linear and non-linear techniques are distinguished. Self-calibration methods which are a combination of linear and non-linear techniques are also proposed (Heikkilä and Silven, 1997).

An example of a linear technique of self-calibration is the Direct Linear Transformation (DLT). The DLT method is based on the collinearity equations which are extended by an affine transformation of the image coordinates (Abdel-Aziz and Karara, 1971). However, this method has some drawbacks. One of the disadvantages is liability to ambiguity if all control points are in a common plane.

Among non-linear techniques, the most popular method is the extended colinearity equation model, which forms the basis of the self-calibration bundle adjustment (Brown, 1971). The basic mathematical model is provided by the non-linear collinearity equations, usually extended by additional parameters, including radial and decentering lens distortions. The bundle adjustment is a very flexible method which provides a simultaneous determining of all system parameters along with estimates of the precision and reliability of the extracted calibration parameters. This method requires a favorable network geometry, convergent and rotated images, with well distributed points throughout the image. In Bauer et al. (2008) self-calibration methods such as DLT, space resection, the Newton method, adjustment using Gauss Markov and point features or lines features are described.

3. Developing 3D models of object using automatic matching methods

In photogrammetry, in order to develop 3D models of architectural objects, it is common to use automatic matching methods. Image matching is defined as the establishing of similarities between two or more images to reconstruct surfaces in 3D. The result of matching is a disparity map or 3D point cloud. Among image matching, two main methods are distinguished: Area Based Matching and Feature Based Matching. While the ABM method is based on image intensity patterns, the FBM method uses features (Remondino et al., 2008).

The Area Based Matching (ABM) can be performed with cross-correlation using the correlation coefficient as a similarity measure. Cross-correlation works well if the patches contain enough signal without noise and if geometrical and radiometric distortions are minimal. To overcome these problems, image reshaping parameters and radiometric corrections were considered, leading to a nonlinear Least Squares Matching (LSM) estimation algorithm. Disadvantages of ABM include: the need for a small searching range for successful matching, a large data volume which must be handled and the requirement of good initial values for the unknown parameters (in the case of LSM). Problems occur in areas with occlusions, areas with a lack of or repetitive texture or if the surface does not correspond to the assumed model (for example, planarity of the matched local surface patch).

The Feature Based Matching (FBM) determines the image correspondence using image features. There are three types of features: interest points, edges, and regions.

The most common strategy to match features is the computation of the Euclidean or Mahalanobis distance between the descriptor elements. Matching with larger features is also called Relational Matching. Compared to ABM, FBM techniques are more flexible with respect to surface discontinuities, less sensitive to image noise and require less approximate values. The accuracy of FBM is limited by the accuracy of the feature extraction process.

The photogrammetric method of developing detailed 3D models of architectural objects is described in (Remondino, et al., 2008). Barazetti and Scaioni (2009) proposed a method based on integration algorithms used both in photogrammetry and CV.

4. Development of 3D models of architectural objects from data acquired with a calibrated video camera

Development of a 3D model of architectural objects using a classic photogrammetric method consists of the following steps: video data acquisition, selection of video frames for orientation, image orientation by bundle block adjustment, TIN models generating using an image matching method and 3D model filtering and texturing.

The main problem associated with video data is the instability of the intrinsic camera parameters. In our research, a Sony NEX-VG10E fixed focal length video camera (f = 16 mm) was used (Table 1).

Table 1. Parameters of Sony Handycam NEX-VG10E video camera

sensor size	CMOS 23.4 mm × 15.6 mm
video resolution	1920 × 1080
pixel size	10.8 µm
number of frames per second	25 fps
video format	AVCHD (MPEG-4 AVC (H.264))

The acquisition of video data was preceded by calibration of the fixed focal length video camera. For this purpose, a 2D calibration test delivered from the Topcon Image Master Calib software was used (Figure 1). A mathematical model of camera was based on perspective projection.

Based on the averaged calibration results which were performed for 10 sets of video frames, intrinsic camera parameters were obtained. Distortion of the lens has a characteristic shape that combines features of barrel and pincushion distortions (Figure 2).

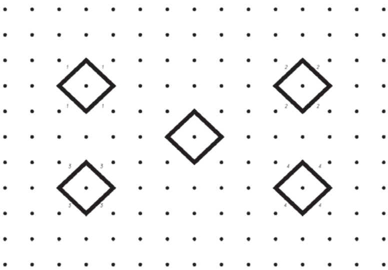


Fig. 1. Calibration target field (Topcon Image Master Calib software)

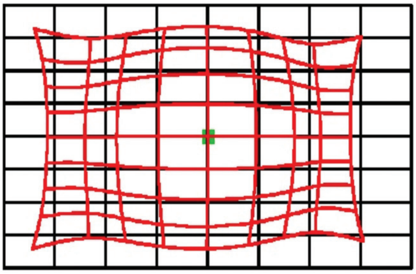


Fig. 2. Distribution of the objective distortion. The red lines are the distortion before correction, the black line are the ideal values. Scale: 20:1

4.1. The selection of video frames for orientation

Developing 3D models of architectural objects had to be preceded by video frames selection to be used in the orientation process.

In literature, there are a lot of keyframe selection methods. (Torr et al., 1998) proposed the Geometric Robust Information Criterion (GRIC). Depending on the type of data, the GRIC method allows the use of one of two models: epipolar geometry (F) or homography (H). It is assumed that for very small baselines the homography model is always selected. Pollefeys's approach (Pollefeys, et al., 2002) relied on setting the first keyframe of a keyframe pairing at the first view, while the second keyframe is the last view for which the number of tracked feature points is above 90% of the number of feature points tracked at the view where F-GRIC becomes smaller than H-GRIC. (Ahmed et al., 2010) proposed a Point to Epipolar Line Criterion (PLEC) algorithm as an additional criterion for keyframe selection to the GRIC method. (Seo et al., 2003) used a cost function which takes into consideration: the ratio of the number of correspondent points to the total number of features, the homography error and spatial distribution of corresponding points. Thormählen's method (Thormählen et al., 2004) relied on selecting the keyframe pairing with the lowest expected estimation error of initial camera motion and object structure. A new formula for computing a n-th video frame used in the orientation is proposed (Deliś, 2012):

Assuming that the terrain size of the image scene S in the horizontal direction is equal to:

$$S = P_x \cdot L_H \tag{1}$$

where: P_x – terrain pixel size, L_H – number of pixels of the video camera matrix in the horizontal direction.

Assuming that the coverage between successive video frames equals to 60% S, the distance between successive perspective centers should be equal to 40% S (Figure 3).

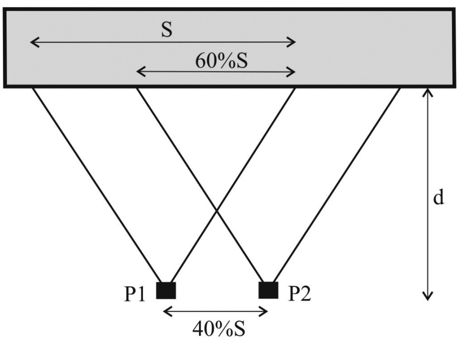


Fig. 3. Relationship between successive image scenes' S coverage and distance between successive perspective centres P1, P2

Based on the general formula for the velocity, the time t in which the video camera travels the distance equal to 40% S could be expressed:

$$t = \frac{40\%S}{v} \tag{2}$$

where: v – velocity of video camera.
Signifying for I the number of video frames Q per second:

$$Q = t \cdot I \tag{3}$$

The n-th video frame used for the orientation is expressed by the formula:

$$Q = \frac{2}{5}\frac{p_x \cdot Y}{f \cdot v} \cdot L_H \cdot I \tag{4}$$

where: p_x – image pixel size, Y – distance between video camera and object, f – focal length, v – velocity of video camera, L_H – number of pixels of the video camera matrix in the horizontal direction, I – the number of video frames Q per second.

The proposed formula (4) can be used with the assumption that the distance from the video camera to the object and velocity were constant during the entire time of video data acquisition. Angular external orientation parameters should also be unchangeable. Therefore, the model proposed by us is ideal for the development of 3D models of building facades.

Furthermore, the quality of video frames which can be used in the orientation processes should also be taken into consideration. The important factor is image blur. As the image blur can degrade orientation and matching results. (Fulton and Fraser, 2009) proposed a method for measuring the blur on the video frames in the frequency domain, in order to eliminate those frames for which the degree of blur is greatest.

4.2. Orientation of video images and creating 3D models of architectural objects

It was assumed, that based on video and Terrestrial Laser Scanning data it is possible to develop 3D models of architectural objects.

In order to create a 3D model of an object using photogrammetric methods, external orientation parameters should be known. For orientation of the video frames (which had been selected based on the formula described above) the bundle adjustment method was used. This method allows for the accurate determining of the external orientation parameters and the 3D coordinates of all tie points at the same time along with a full assessment of the accuracy of alignment.

3D models of two architectural objects were created: elements of St. Anne's Church and Nożyk Synagogue. Both of these architectural objects are located in Warsaw. The above objects have been measured with an impulse laser scanner, Leica ScanStation 2.

The orientation of selected video frames was based on measurements of points with known coordinates. For this purpose, point coordinates from a Terrestrial Laser Scanning (TLS) were used. Using point clouds from TLS is advantageous for two reasons. Using points with known coordinates will provide more accurate orientation. Moreover, point clouds from Terrestrial Laser Scanner could be used to analyze the accuracy of 3D models generated from video frames. The theoretical accuracy of a single position measurement using the Leica ScanStation 2 at a 1 ÷ 50 m range is m_{TLS} = 6 mm.

In addition, there are 4 groups of factors, which could have an impact on the accuracy of TLS data (Hejbudzka et al., 2010). The first group concerns the instrumental properties, the second group deals with the object properties, the third one refers to the scanning geometry and the last one relates to atmospheric conditions. The influence of some of these factors on the accuracy of our measurements can be considered negligible, due to the fact that the instrument had been rectificated, weather conditions remained unchanged (scanning had been conducted during windless, clear weather, at 20°C) and a large majority of the object's surfaces are smooth.

In order to ensure a high degree of accuracy it was important to ensure an incidence angle of approximately 0°, by property designing the distribution of instrument positions. For St. Anne's Church's facade and for Nożyk Synagogue's side entrance, data was acquired from two laser scanner positions. Point cloud resolution p_{res}, relative orientation error m_{ori} and mean error of point position m_{pos} are in the Table 2.

Table 2. The main parameters of TLS data for St. Anne's Church's facade and Nożyk Synagogue' side entrance

	St. Anne's Church's facade	Nożyk Synagogue's side entrance
p_{res}	0.015 m	0.011 m
m_{ori}	0.036 m	0.027 m
m_{pos}	0.039 m	0.030 m

Based on previous studies (Kędzierski et al., 2010), it was assumed that for this type of data, the accuracy of point position m_{pos} is mainly affected by the three parameters: the theoretical accuracy of a single position measurement using the Leica ScanStation 2 m_{TLS}, point cloud resolution p_{res} and relative orientation error m_{ori}.

The mean error of point position m_{pos} was computed using formula (5):

$$m_{pos} = \sqrt{\mathrm{m_{TLS}}^2 + \mathrm{p_{res}}^2 + m_{ori}^2} \tag{5}$$

It was assumed that the control points accuracy is equal point position m_{pos}. In the orientation process for St. Anne's Church's facade 5 video frames, 11 control points and 12 tie points were used. For Nożyk Synagogue's side entrance 5 video frames, 14 control points and 11 tie points were used. Differences of computed coordinates values for control points for St. Anne's Church's facade did not exceed $dX = dY = dZ = 0.04$ m and for Nożyk Synagogue's side entrance $dX = dY = dZ = 0.02$ m (Tables 3 and 4).

Table 3. Control points discrepancies for St. Anne's Church's façade

	dX [m]	dZ [m]	dY [m]
max discrepancies	0.031	0.039	-0.016

Table 4. Control points discrepancies for Nożyk Synagogue's side entrance

	dX [m]	dZ [m]	dY [m]
max discrepancies	-0.012	0.013	-0.016

Topcon Image Master software is designed for an orientation accuracy of 1 pixel or less so that the accuracy of stereo image measurement is about the same as the planar resolution and depth resolution. When the orientation accuracy is below 1 pixel, the depth resolution ΔY can be calculated by the following formula (6) (Topcon Image Master Operation Manual, 2007).

$$\Delta Y = \frac{Y}{B} \cdot GSD \tag{6}$$

where: Y – distance between video camera and object, GSD – planar resolution, B – base length.

Table 5. The parameters, resolution and accuracy of stereo image measurement for Nożyk Synagogue

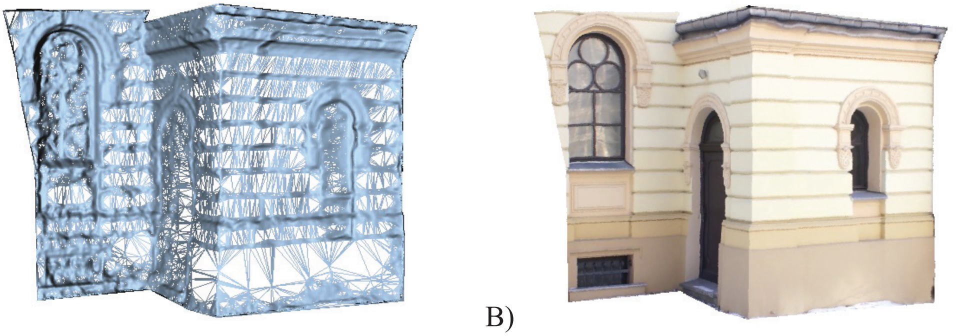
pair name	B [m]	Y [m]	B/Y	GSD [m]	ΔY [m]
1-2	4.54	12.06	0.38	0.008	0.022
2-3	4.36	12.25	0.36	0.008	0.023
3-4	4.04	12.41	0.33	0.008	0.026
4-5	3.73	12.60	0.30	0.008	0.029
			mean	0.008	0.025
			st. dev. S	0.0002	0.0032

Table 6. The parameters, resolution and accuracy of stereo image measurement for Nożyk Synagogue

pair name	B [m]	Y [m]	B/Y	GSD [m]	ΔY [m]
1-2	0.81	7.09	0.11	0.005	0.042
2-3	2.11	6.98	0.30	0.005	0.016
3-4	1.82	7.96	0.23	0.005	0.023
4-5	2.19	7.16	0.31	0.005	0.016
			mean	0.005	0.024
			st. dev. S	0.0003	0.0123

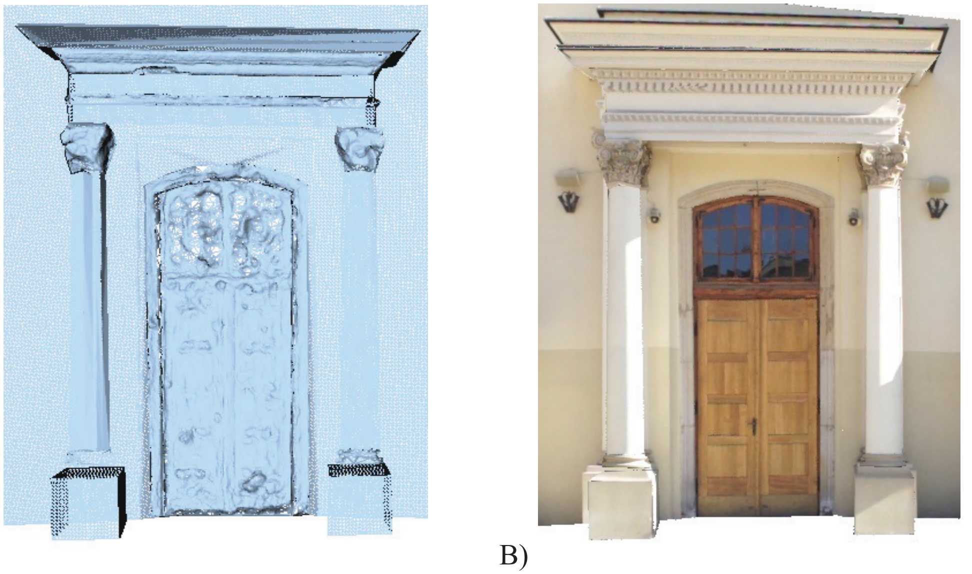
The geometric model of the Nożyk Synagogue's side entrance is characterized by a better planar resolution GSD = 5 mm compared to the geometric model of St. Anne's Church's facade GSD = 8 mm. However, in terms of depth ΔY, the geometric accuracy of both models is similar and amounts to ΔY = 25 mm (Tables 5 and 6).

In order to create 3D models of architectural objects, the automatic matching method (LSM) was used. As a result, 3D models of architectural objects in the TIN form with an 0.02 m interval mesh were generated (Figures 4, 5). In order to create the 3D models, the Topcon Image Master Software was used.



A) B)

Fig. 4. TIN models of Nożyk Synagogue generated from video images with calibrated parameters of inner orientation
(A) without a texture (B) with a texture



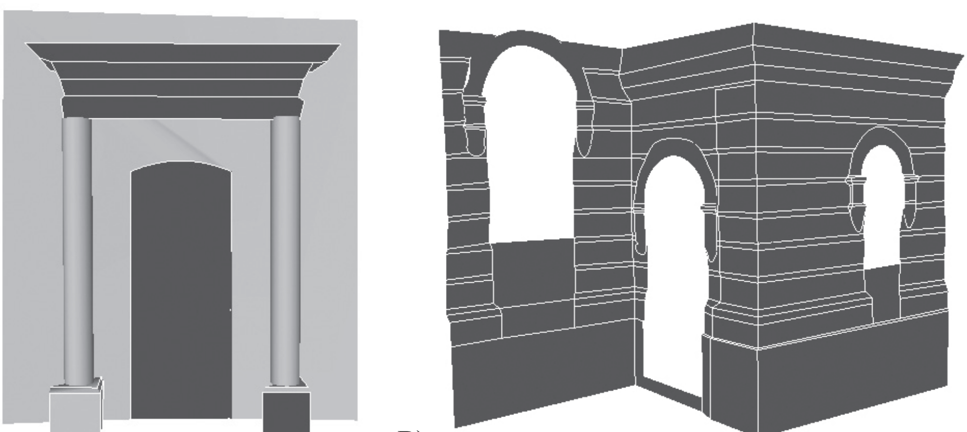
A) B)

Fig. 5. TIN models of St. Anne's Church facade generated from video images with calibrated parameters of inner orientation
(A) without a texture (B) with a texture

5. Development of 3D models of architectural objects from video data based on a self-calibration bundle adjustment

Created 3D models of architectural objects were compared with 3D models of the same objects for which the self-calibration bundle adjustment process was performed. The self-calibration and 3D models creation processes were performed using the Photo Modeler Scanner software. In order to create 3D models with a self-calibration bundle adjustment method the same video frames were used. The orientation of selected video frames was based on using points from Terrestrial Laser Scanning too. In this case, 3D models were created through the process of manual edge measurements without using automated matching methods. For St. Anne's Church's facade 30 control points and 44 tie points were used. For Nożyk Synagogue's side entrance 14 control points and about 200 tie points were used.

The RMS errors of the self-calibration bundle adjustment for the Nożyk Synagogue, and for St. Anne's Church's are respectively: 1.683 and 1.968 pixels. After measuring tie points and edges on the video images, 3D surface models were generated (Figure 6).



A) B)

Fig. 6. 3D surface models extracted from video data based on a self-calibration bundle adjustment
A) St. Anne's Church, B) Nożyk Synagogue

6. Accuracy analysis of 3D architectural models generated from video data

In order to assess the accuracy of the developed 3D models of architectural objects, points with known coordinates from Terrestrial Laser Scanning were used. In order to compare 3D models the IMInspect module of Polyworks software was used. To assess the accuracy a shortest distance method was used. The method relied on computing the minimum point-to-surface distance d between TLS data and the developed 3D models (Figures 7, 8).

Table 7. Accuracy analysis of the 3D models extracted from video data compared to TLS data

3D model		St. Anne's church		*Nożyk* Synagogue	
method		with calibrated parameters of inner orientation	self-calibration bundle adjustment	with calibrated parameters of inner orientation	self-calibration bundle adjustment
number of points		1488515		430842	
mean distance between TLS data and 3D model d_{mean}		0.028 m	0.041 m	0.011 m	0.003 m
RMSE		0.13 m	0.12 m	0.08 m	0.06 m
max error S_{max}		1.31 m	1.28 m	0.67 m	0.37 m
st. dev. S		0.13 m	0.11 m	0.08 m	0.06 m
percent of points within	$\pm 1\sigma$	82%	75%	91%	80%
	$\pm 2\sigma$	95%	93%	97%	94%
	$\pm 3\sigma$	98%	99%	98%	97%

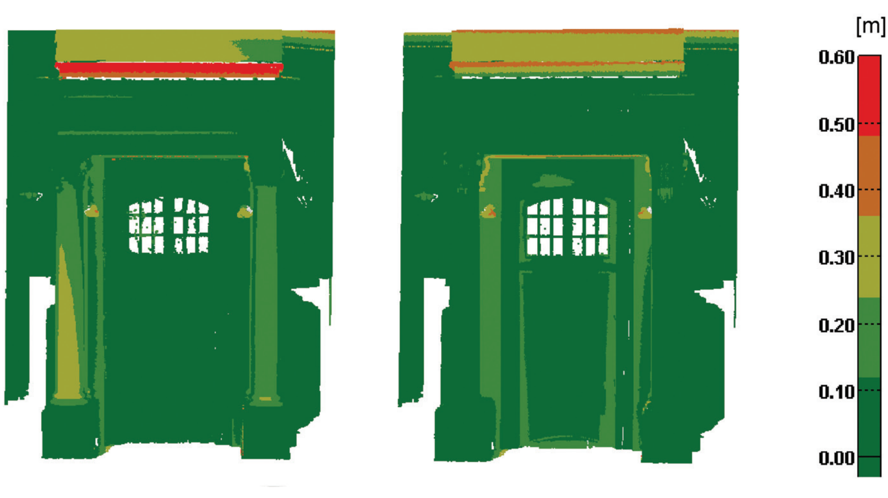


Fig. 7. Difference maps of 3D models of St. Anne's church facade generated from video data and TLS data. Left image: 3D model generated with from video images for which the calibration process was carried out. Right Image: 3D model generated from video images for which the self-calibration bundle adjustment process was performed

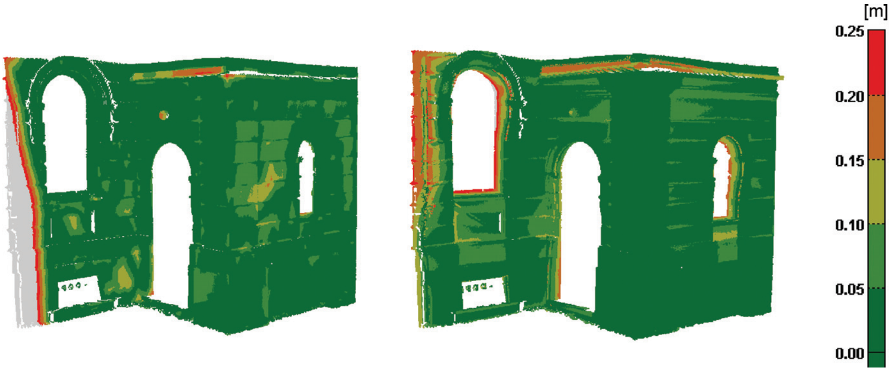


Fig. 8. Difference maps of 3D models of the Nożyk Synagogue generated from video data and TLS data. Left image: 3D model generated from video images for which the calibration process was carried out. Right Image: 3D model generated from video images for which the self-calibration bundle adjustment process was performed

Results for 3D models generated from video data with parameters of inner orientation and based on self-calibration technique are similar (Table 7). For St. Anne's church, the mean distance between points from TLS data and 3D models generated from video data for which the calibration process was carried out is d_{mean} = 0.028 m and for which the self-calibration bundle adjustment was used is d_{mean} = 0.041 m. For Nożyk Synagogue, the mean distance between points of 3D models generated from TLS data and from video data for which the calibration process was carried out is d_{mean} = 0.011m and for which the self-calibration bundle adjustment was used is d_{mean} = 0.003m.

For both examples slightly better results were achieved for 3D surface models generated from video data for which the self-calibration bundle adjustment was used in PhotoModeler Software. It may be due to the form of the created 3D models. 3D models created using an automated matching method in a TIN form could include some points, which, in reality, do not belong to architectural structures.

7. Conclusion

The paper contains an overview of the methods of 3D modeling of architectural objects from video data using both photogrammetric and Computer Vision. The process of creating 3D models of architectural objects on the basis of video images, which had been acquired by a Sony NEX-VG10E fixed focal length video camera.

In the orientation of selected video frames point coordinates from a Terrestrial Laser Scanning (TLS) were used. The acquisition of video data was preceded by the calibration of the fixed focal length video camera. It allowed to determine the intrinsic camera parameters. After the orientation process, using automatic matching methods, 3D models of architectural objects from video data in the TIN form were created.

The accuracy of control points for St. Anne's church was m_{pos} = 0.039 m and for Nożyk Synagogue m_{pos} = 0.030 m. Control points discrepancies for St. Anne's Church's facade did not exceed $dX = dY = dZ = 0.04$ m and for Nożyk Synagogue $dX = dY = dZ = 0.02$ m. It was assumed, that the accuracy of stereo image measurement is equal to the planar and depth resolutions. For St. Anne's Church facade, the planar resolution was $GSD = 8$ mm and the depth resolution was $\Delta Y = 25$ mm. The geometric model of the Nożyk Synagogue was characterized by a slightly better resolution. The planar resolution was $GSD = 5$ mm and the depth resolution was $\Delta Y = 24$ mm.

Developed 3D models of architectural objects were compared with 3D models of the same objects for which the self-calibration bundle adjustment process was performed. In this case 3D models were created in the process of manual edge measurements.

In order to assess the accuracy of the developed 3D models of architectural objects, points with known coordinates from Terrestrial Laser Scanning were used. To assess the accuracy a shortest distance method was used. Analysis of the accuracy showed that the accuracy of 3D models generated from video images differs by about 0.06 ÷ 0.13m compared to TLS data.

Results of 3D modeling of buildings using video imagery can be used by companies or institutions which deal with the production of architectural documentation. What is more, the proposed measurement method based on a video camera is ideally suited to complement data obtained by other methods.

Acknowledgements

The paper is the result of research within statutory research in the Faculty of Civil Engineering and Geodesy, Institute of Geodesy, Department of Photogrammetry and Remote Sensing Military University of Technology, Warsaw.

References

Abdel-Aziz, Y. I. & Karara, H. M. (1971). Direct Linear Transform from comparator coordinates into object space coordinates. In Proceedings of the Symposium on Close-Range Photogrammetry, Vol. 1, January 1971, (pp: 1-18), Falls Church, Virginia, USA: American Society of Photogrammetry.

Ahmed, M., Dailey, M., Landabaso, J. & Herrero, N. (2010). Robust key frame extraction for 3d reconstruction from video streams. International Conference on Computer Vision Theory and Applications (VISAPP), 17-21 May 2010 (pp: 231-236). Angers, France: Springer-Verlag.

Barazzetti, L. & Scaioni M. (2009). Automatic orientation of image sequences for 3d object reconstruction: first results of a method integrating photogrammetric and computer vision algorithms. In Proceedings of 3D-ARCH 2009, 25-28 February 2009 (on CD-ROM). Trento, Italy: International Society of Photogrammetry and Remote Sensing (ISPRS).

Bauer, S., Luber, A., & Reulke, R. (2008). Evaluation of camera calibration approaches for video image detection systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B1), 5-11. DOI: 10.1.1.184.3536.

Brown, D.C., (1971). Close-range camera calibration. *Photogrammetric Engineering*, 37(8), 855-866. DOI: 10.1.1.14.6358.

Deliś, P. (2012). Integracja danych z Naziemnego Skaningu Laserowego i danych obrazowych pozyskanych kamerą wideo. [Integration of Terrestrial Laser Scanning data with image data acquired by a video camera]. *Biuletyn WAT*, 4/2012, 39-54.

Eos Systems Inc. (1992-2008). PhotoModeler Scanner 6 [computer software]. Vancouver: Eos Systems.

Faugeras, O., Luong, Q.-T. & Maybank, S., (1992). Camera self-calibration: theory and experiments. *Computer Vision – ECCV, Lecture Notes in Computer Science*, 588, (321-334). DOI: 10.1007/3-540-55426-2_37.

Frahm, J.M., Pollefeys, M., Clipp, B. Gallup D., Raguram R., Wu, C. & Zach., C. (2008). 3D Reconstruction of architectural scenes from uncalibrated video sequences. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 38(5), DOI: 10.1.1.156.3312.

Fulton, J.R., & Fraser, C.S. (2009). Automated reconstruction of buildings using a hand held video camera. *Innovations in Remote Sensing and Photogrammetry*, *Lecture Notes in Geoinformation and Cartography,* 393-404. DOI: 10.1007/978-3-540-93962-7_30.

Hansen, von W., Thonnessen, U., & Stilla, U. (2004). Detailed relief modeling of building facades from video sequences. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(3), 967–972.

Hao, X. & Mayer, H. (2003). Orientation and Auto-Calibration of Image Triplets and Sequences. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3), 73–78. DOI: 10.1.1.118.1201.

Heikkilä, J. & Silven, O., (1997). A four-step camera calibration procedure with implicit image correction. In Proceedings of IEEE Computer Society Conference Computer Vision and Pattern Recognition, 17-19 June 1997 (pp. 1106–1112). San Juan, Puerto Rico: Institute of Electrical and Electronics Engineers (IEEE).

Hejbudzka K., Lindenbergh R., Soudarissanane S., & Humme A. (2010), Influence of atmospheric conditions of the range distance and number of returned points in Leica ScanStation 2 point clouds. In Proceedings of the ISPRS Commission V Mid-Term Symposium ‘Close Range Image Measurement Techniques, 21-24 June 2010 (pp. 282–287), Newcastle upon Tyne, United Kingdom: International Society of Photogrammetry and Remote Sensing (ISPRS).

Kędzierski, M., Wilińska, M., & Fryśkowska A. (2010). Opracowanie ortofoto obiektu zabytkowego metodami fotogrametrii bliskiego zasięgu z wykorzystaniem naziemnego skaningu laserowego. [Generation of orthophotos of historic structures using close range photogrammetry and terrestrial laser scanning]. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, 21/2010, 149–158.

Kutulakos, K. N. & Seitz, S. M. (2000). A Theory of Shape by Space Carving. *International Journal of Computer Vision*, 38(3), 307-314. DOI: 10.1.1.27.3066.

Okutomi, M. & Kanade, T., (1991). A multiple-baseline stereo. In Proceedings of IEEE Computer Society Conference Computer Vision and Pattern Recognition 15(4), 3-6 June 1991 (pp: 63–69), Lahaina, Maui, Hawaii: Institute of Electrical and Electronics Engineers (IEEE).

Pollefeys, M., Gool, L.V., Vergauwen, M., Cornelis, K., Verbiest, F. & Tops, J. (2002). Video-to-3d, In Proceedings of Photogrammetric Computer Vision, 34(3), 9-13 September 2002 (pp. 252–258). Graz, Austria: International Archive of Photogrammetry and Remote Sensing

Remondino, F., El-Hakim, S.F., Gruen, A., & Zhang L. 2008. Turning images into 3-D models. *Signal Processing Magazine IEEE*, 25(4), 55–65. DOI: 10.1109/MSP.2008.923093.

Robertson, D.P. & Cipolla R. (2009). Structure from Motion. In Varga. M. *Practical Image Processing and Computer Vision* (pp. 1-49). Wiley 1 edition.

Seo, J., Kim, S., Jho, C., & Hong, H. (2003). 3D estimation and keyframe selection for machmove. In Proceedings International Technical Conference on Circuits/Systems, 7-9 July 2003, Bokwang Phoenix Park, Korea: ITC-CSCC.

Thormählen, T., Broszio, H. & Weissenfeld, A., (2004). Keyframe selection for camera motion and structure estimation from multiple views. In Proceedings of the European Conference on Computer

Vision, Lecture Notes in Computer Science, Vol. 3021, 11-14 May 2004 (pp: 523–535). Prague, Czech Republic: Springer.

Tian, Y., Gerke, M., Vosselman, G. & Zhu, Q. (2010). Knowledge-based building reconstruction from terrestrial video sequences, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(4), 395–408. DOI: 10.1016/j.isprsjprs.2010.05.001.

Topcon Corporation (2007-2008), Topcon Image Master [computer software]. Livermore: Topcon.

Topcon Corporation (2007-2008), Topcon Image Master Operation Manual, 2007.

Torr, P., Fitzgibbon, A., & Zisserman, A. (1998). Maintaining multiple motion model hypotheses through many views to recover matching and structure. In Proceedings Sixth International Conference on Computer Vision, 4-7 January 1998 (pp:485-491). Bombay, India: Computer Vision.

Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22(11), 1330-1334. DOI: 10.1.1.145.7481.

Modelowanie 3D obiektów architektonicznych na podstawie danych wideo pozyskanych z wykorzystaniem obiektywu stałoogniskowego

Paulina Deliś, Michał Kędzierski, Anna Fryśkowska, Michalina Wilińska

Wojskowa Akademia Techniczna
Wydział Inżynierii Lądowej i Geodezji
Instytut Geodezji
Zakład Teledetekcji i Fotogrametrii
ul. gen. S. Kaliskiego 2, 00-907 Warszawa
e-mail: pdelis@wat.edu.pl; mkedzierski@wat.edu.pl,
afryskowska@wat.edu.pl, mwilinska@wat.edu.pl

Streszczenie

Artykuł zawiera opis procesu opracowania modeli 3D obiektów architektonicznych na podstawie obrazów wideo pozyskanych kamerą wideo Sony NEX-VG10E ze stałoogniskowym obiektywem. Przyjęto założenie, że na podstawie danych wideo i danych z naziemnego skaningu laserowego (NSL) możliwe jest opracowanie modeli 3D obiektów architektonicznych. Pozyskanie danych wideo zostało poprzedzone kalibracją kamery wideo. Model matematyczny kamery był oparty na rzucie perspektywicznym. Proces opracowania modeli 3D na podstawie danych wideo składał się z następujących etapów: wybór klatek wideo do procesu orientacji, orientacja klatek wideo na podstawie współrzędnych odczytanych z chmury punktów NSL, wygenerowanie modelu 3D w strukturze TIN z wykorzystaniem metod automatycznej korelacji obrazów. Opracowane modele 3D zostały porównane z modelami 3D tych samych obiektów, dla których została przeprowadzona samokalibracja metodą wiązek. W celu oceny dokładności opracowanych modeli 3D obiektów architektonicznych wykorzystano punkty naziemnego skaningu laserowego. Do oceny dokładności wykorzystano metodę najkrótszej odległości. Analiza dokładności wykazała, że dokładność modeli 3D generowanych na podstawie danych wideo wynosi około 0.06 ÷ 0.13m względem danych NSL.