EFFICIENT USE OF VIDEO FOR 3D MODELLING OF CULTURAL HERITAGE OBJECTS

B. Alsadik [a,b], M. Gerke [b], G. Vosselman [b]

[a] University of Baghdad, College of engineering, Department of surveying, Baghdad, Iraq.
[b] University of Twente, ITC Faculty, EOS department, Enschede, The Netherlands.
(b.s.a.alsadik, m.gerke, george.vosselman)@utwente.nl

Commission III/1

KEYWORDS: video image sequence, 3D modelling, minimal camera network, blur detection.

ABSTRACT:

Currently, there is a rapid development in the techniques of the automated image based modelling (IBM), especially in advanced structure-from-motion (SFM) and dense image matching methods, and camera technology. One possibility is to use video imaging to create 3D reality based models of cultural heritage architectures and monuments. Practically, video imaging is much easier to apply when compared to still image shooting in IBM techniques because the latter needs a thorough planning and proficiency. However, one is faced with mainly three problems when video image sequences are used for highly detailed modelling and dimensional survey of cultural heritage objects. These problems are: the low resolution of video images, the need to process a large number of short baseline video images and blur effects due to camera shake on a significant number of images.

In this research, the feasibility of using video images for efficient 3D modelling is investigated. A method is developed to find the minimal significant number of video images in terms of object coverage and blur effect. This reduction in video images is convenient to decrease the processing time and to create a reliable textured 3D model compared with models produced by still imaging.

Two experiments for modelling a building and a monument are tested using a video image resolution of 1920×1080 pixels. Internal and external validations of the produced models are applied to find out the final predicted accuracy and the model level of details. Related to the object complexity and video imaging resolution, the tests show an achievable average accuracy between 1 – 5 cm when using video imaging, which is suitable for visualization, virtual museums and low detailed documentation.

1. INTRODUCTION

Nowadays, the generation of a reality based 3D model of architectural objects and monuments is mainly achieved using non-contact measurement methods. The measurements can be applied either by active sensors like laser scanners or passive by cameras. These methods for objects modelling can be distinguished as: image-based modelling (IBM), range-based modelling, or a combination of both techniques (Remondino and El-Hakim, 2006). Image based methods are preferred for limited budget projects beside their practicality and portability in complex sites. Moreover, the advances in the state-of-the-art of image orientation, image dense matching and modelling offer a toolbox for the digital documentation and preservation of cultural heritage (Santagati et al., 2013). Currently, different efficient automated or semi-automated image based modelling software are available in the market like (Acute3D, 2013; EOSsystems, 1994; Photoscan, 2011; Pix4D, 2013) beside the other open source software which offer the same functions like in (Furukawa and Ponce, 2010; Meshlab, 2010; Pierrot-Deseilligny, 2012; Snavely, 2010; Wenzel, 2013; Wu, 2012).

The captured images can be taken either with a static camera (still shots) or a moving camera (video sequence). Usually, high resolution still shot images are used for the 3D modelling and documentation which is captured either from an aerial platform or from the ground. The created 3D models from these images are reliable in the sense of visualization and accuracy. This reliability is based on several factors like: the high resolution of the taken images from either compact or SLR cameras, the low radiometric and geometric distortions of the images and the proper camera network design.

However, the disadvantages of using still image shooting in 3D modelling is the need for proficiency or expertise: the difficulty to capture the needed number of images, the proper pose of the cameras during the capture, and to ensure the required overlap between the images. Consequently, it is difficult for non-professionals to cover the whole object and to avoid an unfavorable wide baseline network configuration (large base/depth ratio) for the 3D modelling (Alsadik et al., 2012, 2013). This wide baseline imaging can represent a difficulty in image based modeling for the image orientation and the subsequent dense matching because of the scale variations and occlusions that may exist. Currently, scale invariant operators like SIFT (Lowe, 2004) and SURF (Bay et al., 2008) represent the state-of-the-art tool for tie points matching . However, these matching operators still have restrictions to match homologous points in wide baseline images (Barazzetti et al., 2010).

After tie point matching a structure-from-motion (SfM) technique is to be used for the computation of the image orientation and the sparse point cloud by bundle adjustment (McGlone et al., 2004). Then, the dense matching is applied for creating a dense point cloud of the object. As recommended in (Haala, 2011; Hullo et al., 2009), it is preferred to keep a reasonable base\depth ratio between (0.15 – 0.30) to have a successful dense matching approach. These aforementioned restrictions put some difficulty on the camera planning stage, which is to be implemented by professionals as mentioned earlier.

On the other hand, the video image sequence represents a short baseline imaging with a high redundancy in the number of images. The image tie points matching can also be done by the mentioned techniques like SIFT or by the so-called feature

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-3-W4-1-2015

tracking like by using (variants of) the KLT method (Tomasi and Kanade, 1992). In the past decade, approaches like in (Nister, 2001; Pollefeys et al., 2008) were used in processing video image sequence to reconstruct 3D scenes which are scaled into reality using a spatial similarity transformation. The major advantage of video imaging is the flexibility of the recording and ease which, in contrast to still image shooting, enables even non-professionals to document sites. However, the resolution of the conventional consumer video cameras and camcorders is not sufficient (less than 3Mp) when compared to the high resolution (HR) still images. Moreover, a significant number of the video image frames are relatively blurry due to the motion of the camera during the capture and this means a loss of some information in the images. Therefore, the created 3D models from video images might be of lower geometric and radiometric quality, compared to HR still imaging.Another drawback of using video sequences for 3D modeling is that a huge amount of data needs to be processed. With a frame rate of 20 images/s a number of 1000 images is easily reached. For still image networks it was demonstrated in (Alsadik et al.,2013) that the required triangulation accuracy and density of point clouds can still be obtained after a systematic and significant reduction of images. In this paper we like to transfer this idea to video sequences: what is the impact a reduction of frames has on the final accuracy?

To this end a method is presented to find the minimum number of video images that guarantees both: a full coverage and a limited amount of blur effect to finally create 3D models. Accordingly, accuracy validations are applied to study the feasibility of using the video image sequence for 3D modelling. Finally, conclusions will be made based on the standards in (Letellier, 2007) to decide whether video imaging represent an alternative to the still imaging in cultural heritage documentation.

2. METHODOLOGY

The key idea of an efficient use of the video image sequence in modelling is by removing blurry video images and in addition, filter out redundant image frames according to some criteria based on coverage. An alternative would be to filter for accuracy to guarantee a pre-defined accuracy (Alsadik et al., 2014b). Fig.1 shows the general pipeline proposed in this research to use the video images for the 3D modelling. Note that in this paper we will not concentrate on the final meshing/texturing step.

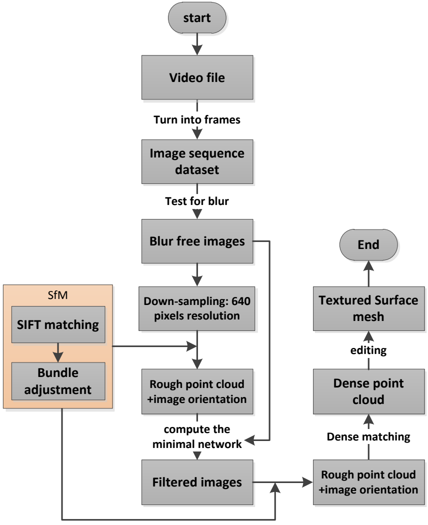


Figure 1. The proposed method for automated 3D modelling from video image sequence

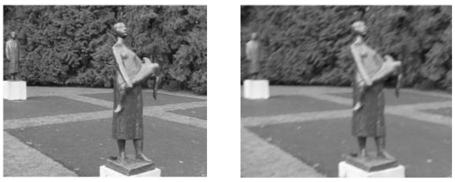
The proposed method is based on having both: rough point cloud of the object and the video images orientation. To this end, we need to apply the SfM technique on the blur free images which is a time consuming approach for large data set. Therefore, we suggest to use the down-sampled images of (640 pixels) with a guided SIFT matching of two consecutive frames to reduce the processing time. All subsequent steps, however, will be done using the full resolution video frames. This will significantly save processing time in dense matching and SfM as will be shown in the two experiments.

2.1 Removal of Blurred Images

Currently, different methods are used to detect the image blur. In this research, the method developed by Crete et al. (2007) will be used for its efficiency and fast implementation. The method is based first on the computation of the intensity differences between neighbouring pixels of the original image. This computation will be repeated, but after intentionally blurring the image with a low-pass filter. These intensity differences before and after the blurring will be compared to evaluate the blur amount. Thus, a metric for sharpness or blurriness is based on either a high or slight variation between the original and the blurred image respectively. Finally, a blur index is computed with a range between 0 to 1 for the best or worst quality respectively.

The software developed by Bao (2009) is used to test the blur on the image dataset which is based on (Crete et al., 2007) paper. Fig. 2 shows a sample test for two consecutive video images for this blur estimation technique.

After testing several datasets an empirical rejection threshold of (0.45) is selected to filter out the blurry images in a dataset and keep the clear sharp images. It should be noticed that the high redundancy in the video images will ensure the sufficient coverage of the images after the filtering.

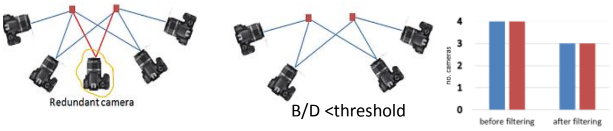


(a) Blur metric= 0.29 (b) Blur metric= 0.46

Figure 2. Sample test for the blur metric computation.

2.2 Minimal camera network

The method of computing the minimal number of images is presented in (Alsadik et al., 2013; Alsadik et al., 2014b) and based on the concept of having at least three cameras viewing simultaneously every object point. Therefore, the cameras are considered redundant if they only result in coverage by more than three cameras as shown in Fig. 3. Another constraint of the B/D ratio is added to ensure a successful dense image matching.



(a) (b) (c)

Figure 3. The concept of filtering redundant cameras. (a) Before filtering. (b) After filtering. (c) Number of covering cameras before and after the filtering.

Accordingly, there is a need to first create a rough point cloud of the object and also to compute the orientation of the images.

This is can be done by applying a SfM on the blur-free full set of the downsampled video images. The rough point cloud is necessary to obtain the shape and size of the object for the subsequent filtering of the redundant video images. In experiments we retrieve this initial information from the down-sampled image sequence to save the processing time. The total time needed for this SfM step and the proposed filtering will significantly reduce the processing time compared to the conventional approach as will be shown.

The methodology of filtering is summarized as follows and shown in Fig. 4:

1- Derive a rough point cloud of the object. The resulted sparse point cloud after the image orientation step with SfM technique is enough for this task.
2- Divide the derived rough point cloud of the object into over-covered and fair-covered. Over-covered points, are the points that appear in more than three cameras while fair-covered points, refer to the points that appear in three cameras.
3- Label the cameras as redundant or significant based on the number of the viewed over-covered points and fair covered points.
4- Arrange the redundant cameras involved in imaging over-covered points according to their coverage (number of points) in an ascending order. The reason for this arrangement is to cancel the redundant cameras that are imaging a fewer number of points and to keep the other cameras.
5- Check the effect of the camera cancelation on the B/D ratio. Accordingly, cancel the camera that is involved in imaging only the over-covered point group and doesn't produce a large B/D ratio in the network configuration.
6- Test the filtering iteratively according to the computed coverage after the camera cancelation and re-label the point cloud in each iteration into over-covered and fair-covered.
7- This procedure is re-iterated starting at step 3. The filtering is repeated until no more redundant cameras involved in imaging.

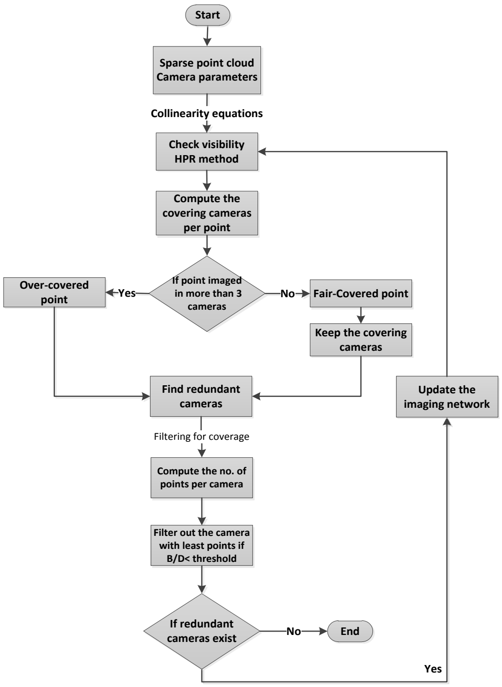

Start
Sparse point cloud
Camera parameters
Collinearity equations
Check visibility
HPR method
Compute the covering cameras per point
Over-covered point
Yes
If point imaged in more than 3 cameras
No
Fair-Covered point
Keep the covering cameras
Find redundant cameras
Update the imaging network
Filtering for coverage
Compute the no. of points per camera
Filter out the camera with least points if B/D< threshold
If redundant cameras exist
No
End
Yes

Figure 4. The workflow diagram of filtering the redundant images.

We make the assumption that the video was constantly in motion, hence we avoid infinite homographys which would render the entire process instable. In addition we request that the length of blurred sequences is small, this means that there is always enough sharp imagery to guarantee a closed image block. Please note that loop closure is implicitly done, since the method will automatically connect for instance images from the start of the sequence with those from the end, because the point reprojection is done on the initial image orientations, not on the initial matching graph.

It is also worth to mention that we used the point cloud visibility with hidden point removal (HPR) method. The concept of this method is applied by assuming the viewpoint C is placed at a sphere origin. The point cloud is projected through the sphere to the opposite outer side in what is named spherical flipping. Spherical flipping reflects a point with respect to the sphere by applying an equation defined by (Katz et al., 2007). The flipped point cloud and the viewpoint will be represented by a convex hull. Then, the transformed points that are located on the convex hull are extracted as visible points.

The major advantages of this method are to determine the visibility without reconstructing a surface compared to other visibility methods beside the simplicity and short implementation time. Moreover, it calculates visibility for dense as well as sparse point clouds, for which reconstruction or other methods, might be failing. However, the disadvantage is realized when a noisy point cloud exists (Mehra et al., 2010). Moreover, it is necessary to set a suitable radius parameter that defines the reflecting sphere (Alsadik et al., 2014a).

3. EXPERIMENTAL TESTS

Two experiments of the presented approach were tested. The first one was for modelling a church building and the second test was for modelling a monument. All the computations were applied on a laptop Dell Latitude E6540 Core i7 and used the state-of-the-art Agisoft photoscan software (Photoscan, 2011). The video imaging is performed by Canon EOS 500D with 1920×1080 pixels in MOV format with a frame rate of 20 fps. Moreover, a still imaging with the same camera is also conducted in a high resolution (HR) of 15MP for details comparison. It is worth to mention that we used the self-calibration approach for all the tested video networks to apply a fully automated SfM. The results are shown in the following sections.

3.1 Church building experiment

The first experiment was applied to the old church building of Enschede in the Netherlands as shown in Fig. 5a which shows the 3D graphical representation taken from google earth (Google, 2010). To verify the video imaging for the 3D modelling, a benchmarking was necessary to have an external validation for the accuracy and reliability of the produced models. Therefore, terrestrial laser scanning (TLS) was conducted around the church building (Fig.6) by using "Trimble CX scanner" where the manufacturer single point accuracy standards were: 4.5 mm @ 30 m. Moreover, five ground control points GCPs were fixed on the church facades as shown in Fig.5b to register the created video based point clouds into the TLS point cloud. The TLS point cloud consists of more than 23 million points as shown in Fig. 6. The GCP target points were marked manually on the corresponding images. The careful zooming and marking was applied by the same observer on the images to decrease the chance of marking errors. Fig.7 illustrates the target design and the target images in the mentioned video and still image resolutions. The imaging scales and ground sample distances GSDs are shown in Table 1.

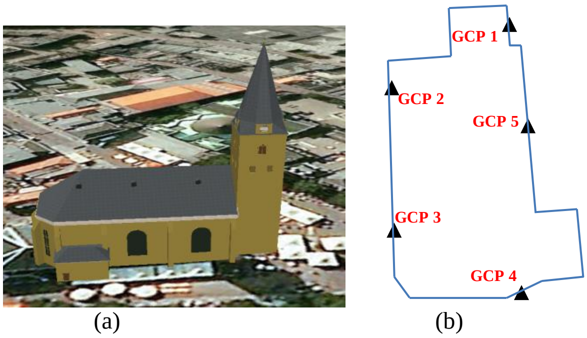


Figure 5. (a) Old Church in Enschede city. (b) The GCPs distribution.

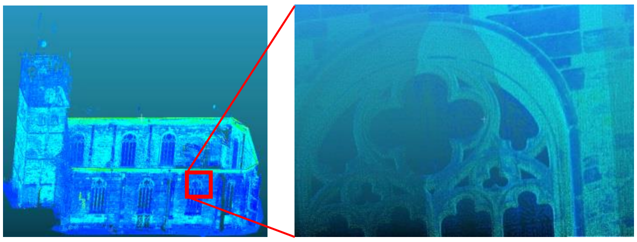


Figure 6. The church point cloud obtained by TLS.

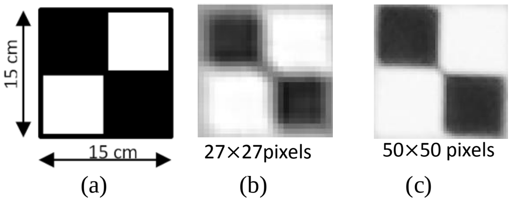


Figure 7. (a) Target design. (b) Target with the video image resolution of 1920×1080 pixels. (c) Target with the still image resolution of 4752×3168 pixels.

Imaging network [pixels]	Av. scale	Pixel size [mm]	GSD [mm]
Video : 1920×1080	1/300	0.020	6
Still images: 4752×3168	1/600	0.005	3

Table 1. The scale and GSD of the video and still imaging networks

To evaluate the reliability and accuracy of using the video imaging for the 3D modelling and documentation, a cloud to cloud distance C2C is computed for a randomly selected four elements of the whole church building. Two windows, one column and a planar façade were tested for this comparison as shown in Fig.8.

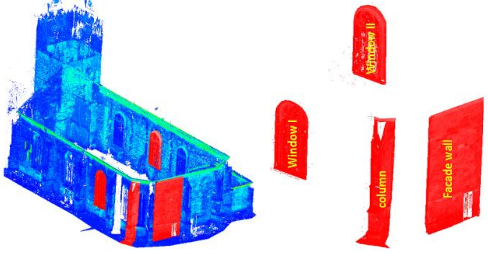


Figure 8. The selected samples (red) of two windows, a column, and a façade for the validation.

The RMSE was computed for the GCPs after image orientation to indicate the quality of the image orientation or SfM computations in the camera network. Due to the lack of additional ground control, we were not able to include independent check points.

The rough point cloud resulting after applying SfM technique was used to compute the minimal number of images in the sequence (section 2.2). This is a reasonable and efficient way to have a fair representation of the study object in an automated faster way as will be shown in Fig. 11. Video images were also down-sampled into 640 pixels before running the SfM technique to reduce the needed time of processing.

The video imaging of 1920×1080 pixels test:

986 image frames are extracted from the video file and the highly blurred images (351 frames) are excluded from the dataset as shown in the bar plot of Fig.9a. The blur-free image network is shown in Fig. 10a. The images were then filtered to a minimum of 347 images (Fig. 10b) where the average number of the detected SIFT points in a single video image was around 10000 points. The number of image pairs in the full pairwise matching decreased from 201295 pairs to 60031 pairs as shown in Fig.9b. The final created dense point cloud before filtering is shown in Fig 10c and the point cloud created from the filtered images is shown in Fig. 10d.

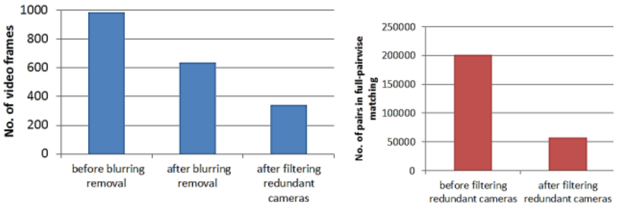


Figure 9. (a) Number of video images before and after filtering. (b) Possible number of stereo pairs within full pairwise matching before and after filtering.

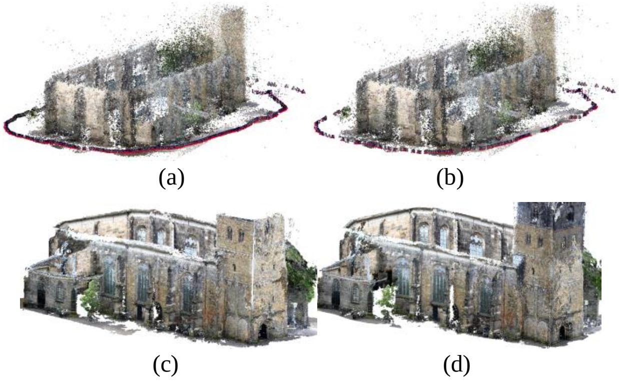


Figure 10. Point cloud from video images of 1920 pixels. (a) SfM output before filtering. (b) SfM output after filtering. (c) Dense point cloud using unfiltered sequence. (d) 3D Dense point cloud after filtering.

Table 2 and Fig. 11 shows the time consumption needed for the SfM and dense matching before and after filtering.

	Before filtering [minutes]	After filtering [minutes]
SfM	210	71
Dense matching	390	86

Table 2. Time consumption for SfM and dense matching for the church experiment

The proposed method as indicated previously is relying on filtering redundant cameras. Our Matlab code consumed 120 minutes for the filtering computation while the SfM consumed 36 minutes to process the full downsampled data set. The total time consumption is illustrated in Fig. 11 where a time reduction of 50% is achieved.

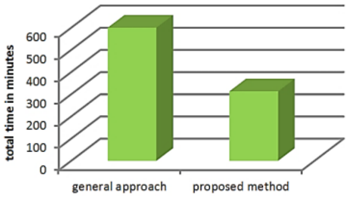


Figure 11. The total time consumption comparison.

The RMSE was computed as shown in Table 3 which shows an error of ≅ 3mm before filtering and ≅ 5mm after filtering. To evaluate the accuracy of the created model before filtering, the C2C distance point based comparison is applied using cloud compare software as shown in Fig. 12 and Table 4. The blue colour refers to near zero shift distances while the red colour refers to larger shift distance.

	$RMSE_X$ [mm]	$RMSE_Y$ [mm]	$RMSE_Z$ [mm]	$RMSE_t$ [mm]
Before filtering	1.9	0.9	1.6	2.6
After filtering	4.8	2.4	1.2	5.4

Table 3. The RMSE of the GCPs before and after filtering of the video network.

Point cloud comparison	No. of points	Mean shift [cm]	Std. deviation [cm]
Window I	221294	2.0	±1.5
Window III	174690	3.6	±2.4
Column	337692	4.9	±2.7
Facade	494347	7.9	±2.0

Table 4. C2C computations for the different parts of the building of the point cloud before filtering

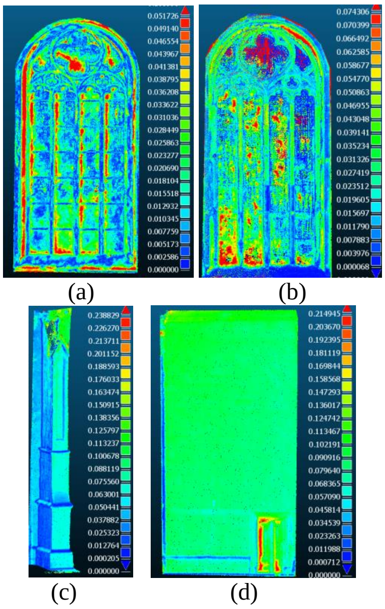


Figure 12. Dense video output before filtering. (a) C2C comparison of window I. (b) C2C comparison of window II. (c) C2C comparison of the column. (d) C2C comparison of the façade.

The C2C distance point based comparison after filtering is also shown in Fig. 13 and Table 5.

Point cloud comparison	No. of points	Mean shift [cm]	Std. deviation [cm]
Window I	205811	2.7	±2.1
Window II	144556	3.0	±2.1
Column	314312	5.3	±3.1
Facade	487977	6.6	±1.9

Table 5. C2C computations for the different parts of the building point cloud after filtering

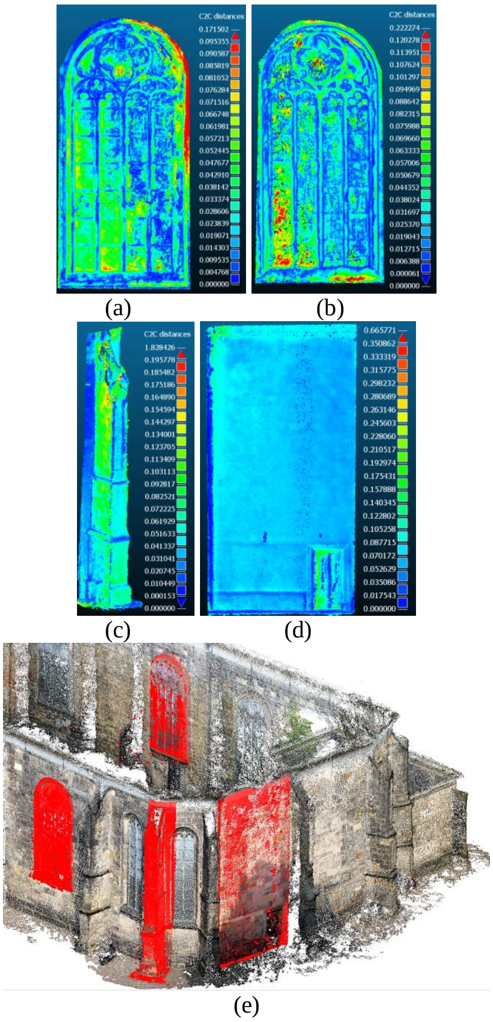


Figure 13. Video based point cloud after filtering (a) C2C comparison of window I. (b) C2C comparison of window II. (c) C2C comparison of the column. (d) C2C comparison of the facade. (e) Video based point cloud and validation parts in red.

The still image shooting of the church:

To evaluate the details represented in the 3D model of the video imaging, a high resolution (4752×3168 pixels) images are taken by an 18 mm Canon camera as mentioned earlier. The expected accuracy in the object space was around 10mm based on a half pixel image measurement accuracy. The complete set of the captured images contained 118 images (Fig.14). The same GCPs are used to reference the camera network. It must be noted that the planning of the image set is applied by a professional user. This proficiency requirement motivates the use of the easy to capture video imaging as investigated in this paper.

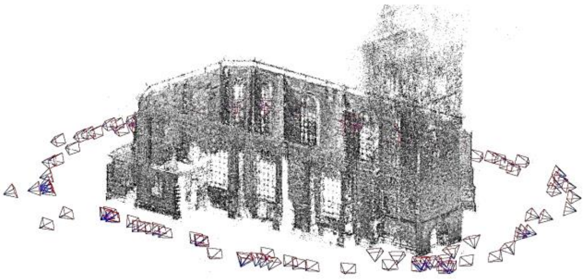
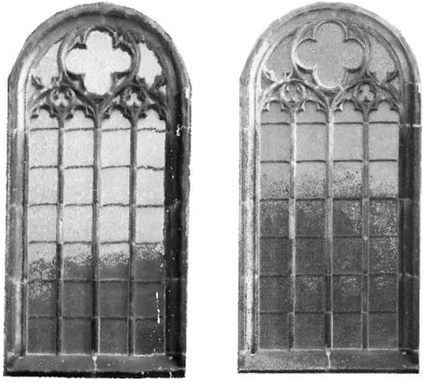


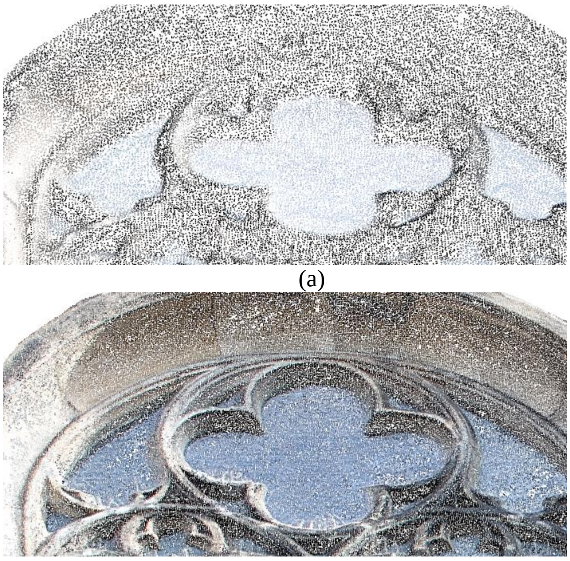
Figure 14. The still imaging camera network

To evaluate the amount of details and visualization acquired from video imaging, a comparison of 3D details with the still imaging is shown in Fig.15.



(a) $\cong 200000\ points$ (b) $\cong 850000\ points$

Figure 15. (a) Video - based point cloud of window I after filtering. (b) Still imaging point cloud of the window I.



(a)

(b)

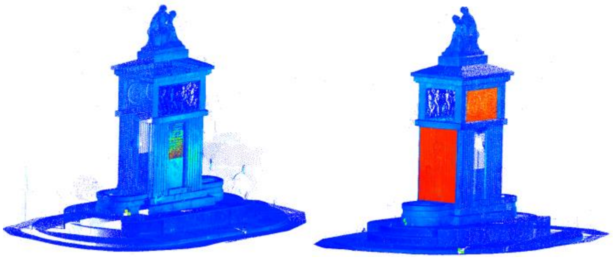
Figure 16. (a) Point cloud from video imaging. (b) Point cloud from HR still imaging.

Fig.16 shows the decoration of window I of the church where the level of details and the number of points is four times higher in the still imaging than the video imaging. This is an expected result because of the high resolution of the still images and their higher geometric and radiometric stability.

From this experiment, two benefits of the developed method are noticed: 1) Filtering didn't significantly reduce the number of points in the final point cloud which is an advantage. 2) Filtering does not have a notable negative impact on the accuracy.

3.2 Monument experiment

The second experiment is applied to a monument in the old city of Enschede of Fig. 17, which is built in 1912 to commemorate the disaster of the city fire in 1863. The point cloud acquired by TLS consisting of 1 million points. For validation computations, two patches (red) are selected as shown in Fig 17b.



(a) (b)

Figure 17. (a) TLS point cloud. (b) Two point patches for validation in red.

A video imaging with a resolution of 1920 pixels was taken around the monument at a scale of 1/250. Three target control points were temporally fixed on the monuments for referencing. The pixel size in the extracted frame was 0.02mm and the ground sample distance GSD was 5mm. A total of 670 video images was acquired from the video stream which were filtered for blur effect to 233 images. The camera network and the rough cloud are shown in Fig.19a. Then, as applied previously in the first experiment, a minimal camera network of 64 images was extracted based on coverage filtering as shown in Fig.19b. The reduction in the number of video images in the filtering steps is shown in Fig. 18. The full pairwise matching was reduced from 224115 pairs to only 1830 pairs.

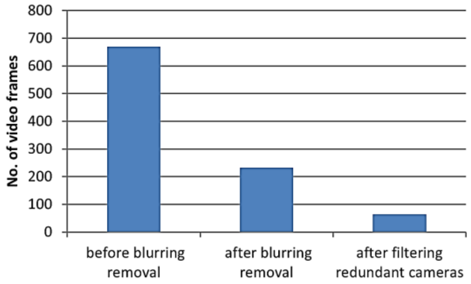
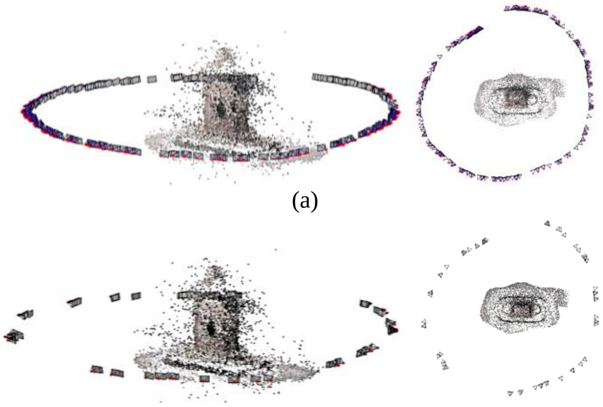


Figure 18. Number of video images of the monument before and after filtering.



(a)

(b)

Figure 19. (a) Video camera network of 233 blur-free images. (b) Minimal video camera network of 64 images.

Accordingly, a dense point cloud after filtering was created and resulted with ≅ 900000 points as shown in Fig.21.
The time consumed for the SfM and dense matching is shown in Fig. 20 before filtering (233 images) and after filtering (64 images).

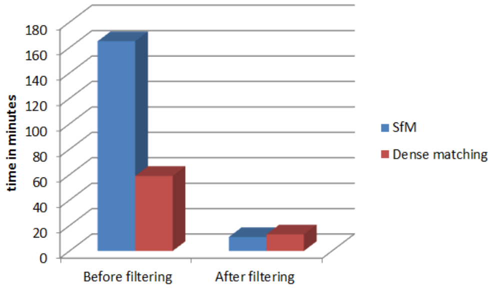


Figure 20. The time consumed for dense matching and SfM before and after filtering for the monument experiment.

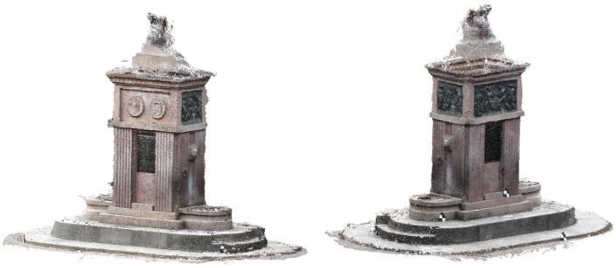
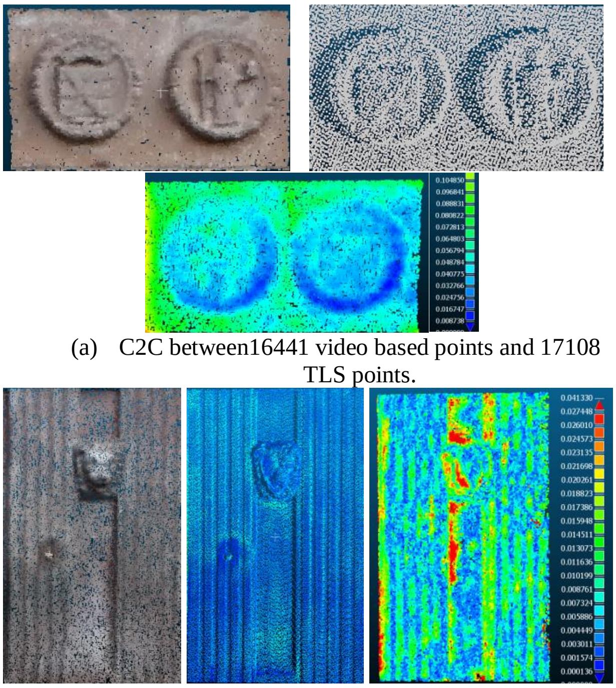


Figure 21. Video based dense point cloud of the monument after filtering.

For validation, two patch clusters of points were selected to check the accuracy of the resulted video based point cloud. The tests shown in Fig.22a and Fig.22b resulted in mean distances of 4.7±1.2cm and 1.0 ± 0.6 cm respectively.



(a) C2C between16441 video based points and 17108 TLS points.

(b) C2C between 38493 video based points and 59073 TLS points.

Figure 22. Validation of the video imaging for the monument. (a) C2C comparison of 1st patch. (b) C2C comparison of 2nd patch.

To validate the extracted details from the video, 40 HR images are captured around the monuments and oriented to produce a 3D model as shown in Fig.23. A black coloured sculpture (1.8×0.8 m^2) on the upper part of the monument is selected to compare the amount of details gathered from the video with respect to the still images (Fig.24). The number of points and consequently the details of the video based point cloud is not adequate for cultural heritage documentation as shown in Fig. 24.

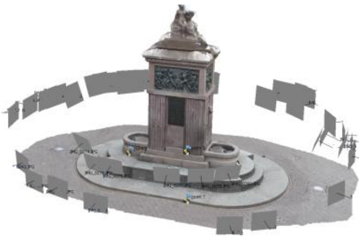
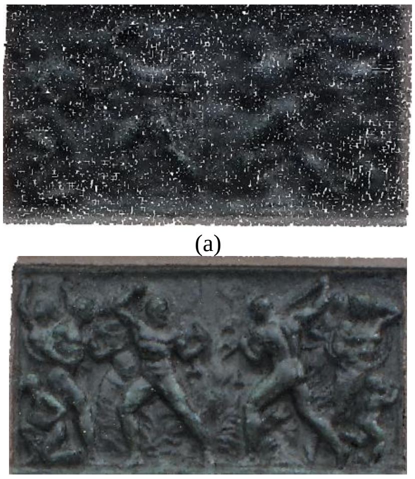


Figure 23. The still imaging network of the monument.



(a)

(b)

Figure 24. (a) Video based point cloud (30000 points) of the sculpture. (b) Still image based point cloud (74000 points) of the sculpture.

4. DISCUSSION AND CONCLUSIONS

In this research a method of using video images (1920×1080 pixels) for 3D modelling was developed and its feasibility for cultural heritage documentation was investigated. Two filtering steps were suggested by firstly removing blurry images from the dataset and secondly to exclude redundant cameras in terms of coverage. Two experimental tests of a church and a monument were presented and the accuracy was evaluated by comparing the created point cloud to a reference TLS point cloud. Four selected sub-point clouds of two windows, a column and a planar façade were used as a reference to evaluate the accuracy of the created video based point cloud. Generally, a significant reduction in the video images was attained from around 1000 video images to an average number of 300 images around the church building. The tests showed an average accuracy of 5cm. Moreover, a high resolution still imaging is applied to clarify and compare the degree of details that can be modelled with this video resolution as shown in Fig.16.
The second test of a monument was implemented with the same video resolution to a scale of 1/250. A reduction from 670 images to only 64 images was obtained by using the proposed technique and resulted with an average accuracy of <5 cm with reference to the TLS point cloud. A comparison to a still imaging is also investigated to conclude about the details

offered by the video (Fig. 24). Accordingly, from both experiments, it is concluded that modelling with video imaging of 1920×1080 pixels is suitable for midrange accuracy applications like planning initial documentation, investigation, small scale visualization and pre-design.

A camera self-calibration proved to be convenient to the video images and to take into account the tangential lens distortion. Strict blur removal was also preferred to have sharper images based on the large number of redundancy offered by video.

Although the GSD has been less than a centimetre in all tests, the final point clouds showed that this accuracy level could not be obtained for the final 3D models.

For future work, it is recommended to investigate the use of the new generation of video cameras with the 4k ability of 8Mp.

REFERENCES

Acute3D, 2013. Smart3DCapture. http://www.acute3d.com/software/.

Alsadik, B.S., Gerke, M. and Vosselman, G. 2012. Optimal camera network design for 3D modeling of cultural heritage. In: ISPRS 2012 Proceedings of the XXII ISPRS Congress: Imaging a Sustainable Future, 25 August - 01 September 2012, Melbourne, Australia. Peer reviewed Annals, Volume I-3, 2012. pp. 7-12.

Alsadik, B., Gerke, M., Vosselman, G., 2013. Automated camera network design for 3D modeling of cultural heritage objects. Journal of Cultural Heritage 14, pp. 515-526.

Alsadik, B., Gerke, M., Vosselman, G., 2014a. Visibility Analysis of Point Cloud in Close Range Photogrammetry, ISPRS. commission V/WG II. ISPRS, Italy, riva Del Garda.

Alsadik, B., Gerke, M., Vosselman, G., Daham, A., Jasim, L., 2014b. Minimal Camera Networks for 3D Image Based Modeling of Cultural Heritage Objects. Sensors 14, pp. 5785-5804.

Bao, D.Q., 2009. Image Blur Metric, in: Inc., M. (Ed.), file exchange Mathworks Inc., USA.

Barazzetti, L., Scaioni, M., Remondino, F., 2010. Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. The Photogrammetric Record 25, pp. 356-381.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110, pp. 346-359.

Crete, F., Dolmiere, T., Ladret, P., Nicolas, M., 2007. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric, SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging, San Jose : États-Unis d'Amérique

EOSsystems, 1994. PhotoModeler www.photomodeler.com.

Furukawa, Y., Ponce, J., 2010. PMVS, 2 ed. http://grail.cs.washington.edu/software/pmvs/.

Google, 2010. Google Earth, 6th ed.

Haala, N., 2011. Multiray Photogrammetry and Dense Image Matching, Photogrammetric Week 2011, Wichmann Verlag, Berlin/Offenbach, pp. 185-195.

Hullo, J.F., Grussenmeyer, P., Fares, S., 2009. Photogrammetry and Dense Stereo Matching Approach Applied to The Documentation of The Cultural Heritage Site of Kilwa (Saudi Arabia), in: CIPA (Ed.), XXII CIPA Symposium ISPRS, Kyoto, Japan. pp. 1-6.

Katz, S., Tal, A., Basri, R., 2007. Direct Visibility of Point Sets. ACM Transactions on Graphics 26, pp.24.

Letellier, R., 2007. Recording, Documentation,and Information Management for the Conservation of Heritage Places, Guiding Principles. Library of Congress Cataloging-in-Publication Data.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision 60, pp. 91-110.

McGlone, J.C., Mikhail, E.M., Bethel, J., 2004. Manual of Photogrammetry, Fifth ed. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, United States of America.

Mehra, R., Tripathi, P., Sheffer, A., Mitra, N.J., 2010. Technical Section: Visibility of noisy point cloud data. Comput. Graph. 34, pp. 219-230.

Meshlab, 2010. Visual Computing Lab - ISTI - CNR. http://meshlab.sourceforge.net/.

Nister, D., 2001. Automatic Dense Reconstruction from Uncalibrated Video Sequence. Stockholm University.

Photoscan, A., 2011. AgiSoft StereoScan, Multi-view 3d reconstruction. http://www.agisoft.ru/.

Pierrot-Deseilligny, M., 2012. MicMac, software for automatic matching in the geographical context. http://www.micmac.ign.fr/index.php?id=6.

Pix4D, 2013. Hands free solutions for mapping and 3d modeling. http://pix4d.com/.

Pollefeys, M., Nister, D., Frahm, J.M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stew, H., Yang, R., Welch, G., Towles, H., 2008. Detailed Real-Time Urban 3D Reconstruction from Video. Int. J. Comput. Vision 78, pp. 143-167.

Remondino, F., El-Hakim, S., 2006. Image-based 3D modelling: A review. The Photogrammetric Record 21, pp. 269-291.

Santagati, C., Inzerillo, L., Di Paola, F., 2013. Image-Based Modeling Techniques for Architectural Heritage 3D Digitalization: Limits And Potentialities. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-5/W2, pp. 555-560.

Snavely, N., 2010. Bundler: Structure from Motion (SfM) for Unordered Image Collections. http://phototour.cs.washington.edu/bundler

Tomasi, C., Kanade, T., 1992. Shape and Motion from Image Streams under Orthography: a Factorization Method. International Journal of Computer Vision 9, pp. 137-154.

Wenzel, M.R.K., 2013. SURE - Photogrammetric Surface Reconstruction from Imagery, in: http://www.ifp.uni-stuttgart.de/publications/software/sure/index.en.html (Ed.).

Wu, C., 2012. VisualSFM : A Visual Structure from Motion System, University of Washington at Seattle. http://ccwu.me/vsfm/