

COMPARISON OF SPATIAL RECONSTRUCTION SOFTWARE PACKAGES USING DSLR IMAGES

¹Árpád SOMOGYI, ²Tamás LOVAS, ³Árpád BARSÍ

Department of Photogrammetry and Geoinformatics, Faculty of Civil Engineering
Budapest University of Technology and Economics
Műegyetem rkp. 3, H-1111 Budapest, Hungary
e-mail: ¹somogyi.arpad@epito.bme.hu, ²lovas.tamas@epito.bme.hu
³barsi.arpad@epito.bme.hu

Received 15 November 2016; accepted 6 February 2017

Abstract: Point cloud based modeling has become widely used in recent years in the engineering practice. This paper presents several point clouds produced by different software packages using pixel based multi view spatial reconstruction. The purpose of the investigation is to give statistical quality measures beyond the apparent differences between the point clouds. Reference measurements have been carried out by terrestrial laser scanning. There are proprietary and free software solutions among the tested programs.

Keywords: Point cloud, Reconstruction, Photogrammetry, Terrestrial laser scanner

1. 3D reconstruction with photogrammetry

The definition of photogrammetry is given by the International Society for Photogrammetry and Remote Sensing (ISPRS) in 2016 as follows: 'Photogrammetry is the science and technology of extracting reliable three-dimensional geometric and thematic information, often over time, of objects and scenes from image and range data' [1]. Photogrammetry can be divided into four groups based on the number of images used to produce three-dimensional geometry: single-, two- three and/or n-view geometry [2], [3]. In case of n-view photogrammetry it is possible to produce very detailed surface information despite relatively a small amount of measured points. This kind of procedures is presented in this paper.

2. Algorithms and methods for the reconstruction

Automatic image orientation is a very active field in the photogrammetric community. Image orientation aims reconstructing the coordinates of the perspective center and the viewing direction by evaluating large set of captured images. Different approaches have been proposed and developed over the years; one of them is the Structure-from-Motion (SfM) from the Computer Vision (CV). SfM is a multistage method; the first task is the extraction of feature points in each image that is followed by their matching. Feature extractions can be carried out by various methods, for example Förstner-operator [4], Harris corner detector [5], Scale-Invariant Feature Transform (SIFT) [6], Speeded Up Robust Features (SURF) [7]. New key-point matching strategies and closed-form techniques has been developed to estimate an initial relative position between pairs of images [8], [9], [10]. Incorrect matching is filtered out using a RANdom SAMple Consensus (RANSAC)-type algorithms [11]. The spatial reconstruction is carried out based on the inlier points. The last step of the process is the Bundle Adjustment (BA) [12] of which purpose is to minimize the remaining model errors.

A sparse point cloud is available after the adjustment. The calculated camera positions and orientations can be used for the pixel-wise surface reconstruction (also called ‘dense’ reconstruction). Energy-based methods provide the best results [13], [14], [15], [16].

3. Examined software

The images were processed in the following software: 3DF Samantha; Agisoft Photoscan; Autodesk Remake; CMPMVS; COLMAP; Meshrecon; The Multi-View Environment (MVE); Sure; Visual SfM.

3DF Samantha is an SfM pipeline, which is able to automatically recover the relative camera positions and orientation along with a sparse reconstruction of the scene directly from a set of casual images [17]. The software can be used for research and non-commercial purposes. Used version: 3DF Samantha v.1.7.

Agisoft Photoscan is a proprietary stand-alone software product that performs photogrammetric processing of digital image sets and generates various 3D spatial data from point clouds to textured mesh models. After loading the images, it is possible to enter the camera calibration data that are considered as fix values or initial parameters. To achieve good reconstruction results it is necessary to mask all irrelevant elements on the source images. The accuracy of the camera position estimation, the pair pre-selection, the key and tie point limits can be defined at the photo align processing stage. After this step the sparse point cloud is derived and based on the estimated camera positions the software can calculate depth information for each camera to be combined into a single dense point cloud. After dense point cloud has been reconstructed it is possible to generate polygonal mesh model, which can be textured in a further step. Used version: Agisoft Photoscan 1.2.4.

Autodesk ReMake is proprietary software for converting scenes captured with photos or scans into high-definition 3D meshes. These meshes then can be modified in

different ways also in this environment. ReMake simplifies complex processes since it was designed for users who require top-quality digital models of real-life objects but have little or no 3D modeling expertise. The program is allowed to run in the cloud, which is highly advantageous, since the pixel based spatial reconstruction can require enormous computational capacity. Used version: Autodesk ReMake 17.24.1.4.

CMPMVS is multi-view reconstruction software from the Center for Machine Perception (CMP), which is a research center at the Czech Technical University in Prague [18]. The input is a set of perspective images and (intrinsic and extrinsic) camera parameters, while the output is a textured surface model. Used version: CMPMVS 0.6.0.

COLMAP is a general-purpose SfM and Multi-View Stereo (MVS) pipeline with a graphical and command-line interface [19], [20]. The software is licensed under the GNU General Public License. The full reconstruction workflow can be separated into four steps: feature detection and extraction; feature matching and geometric verification; structure and motion reconstruction; Multi-View Stereo. Used version: COLMAP 2.0.

MeshRecon is free software to create 3D mesh models. It uses image sets and optionally an orientation file (.nvm) as input; however, it is complete reconstruction software. Used version: Meshrecon 2016/03/30.

The Multi-View Environment (MVE), is developed at TU-Darmstadt by Michael Goesele's research group [21]. It is an implementation of a complete end-to-end pipeline for pixel based spatial reconstruction. It contains SfM multi-view stereo, surface reconstruction and texturing. The individual steps are available as command line applications. Used version: MVE 2016/05/17.

SURE free is a software solution for multi-view stereo, which enables the derivation of dense point clouds from a given set of images and their orientations [22]. It supports several images (jpg, png, tif) and orientation (out, nvm, prj) formats. The software can be used for research and non-commercial purposes; however, the number of images is limited to 1000, while the sum of the resolution for all input images is limited to 5000 megapixels in total. Used version: SURE 1.4.

VisualSFM is an application for spatial reconstruction using structure from motion, which is free for personal, non-profit, or academic use [23], [24], [25]. As the name of the software package suggests, it has a graphical user interface. The reconstruction pipeline consists of five steps from the image loading to Furukawa's dense reconstruction tool (PMVS/CMVS) [15]. Used version: VisualSFM 0.5.

4. The case studies

Three case studies were carried out to investigate the software packages: one outdoor and two indoor measurements. The images and videos were captured by a Canon EOS 760 DSLR camera, and the reference measurements were captured by a Faro Focus 120 S terrestrial laser scanner (TLS). The scanner's ranging error is ± 2 mm at 10 m.

Boxes were chosen as test objects for the indoor cases (*Fig. 1*). Still images were captured around the items with uniform density and appropriate overlap in the first arrangement. The image parameters were as follows: image count: 27; resolution:

6000×4000; focal length: 28 mm; exposure time: 1/50 s; f/stop: 3.5. There was laser scanning from three positions, the transformation parameters were calculated with Iterative Closest Points (ICP) method [26], [27]. Point spacing for each of the positions were selected 6 mm at 10 meters. Prior to the second indoor measurement, the boxes were packed in matte newspapers, then three videos (format: MPEG-4, resolution 1920×1080 pixel, frame-rate: 25 fps, bitrate: 28k kb/s) were captured around the objects. The measurement of the reference point cloud was similar to the previous ones.

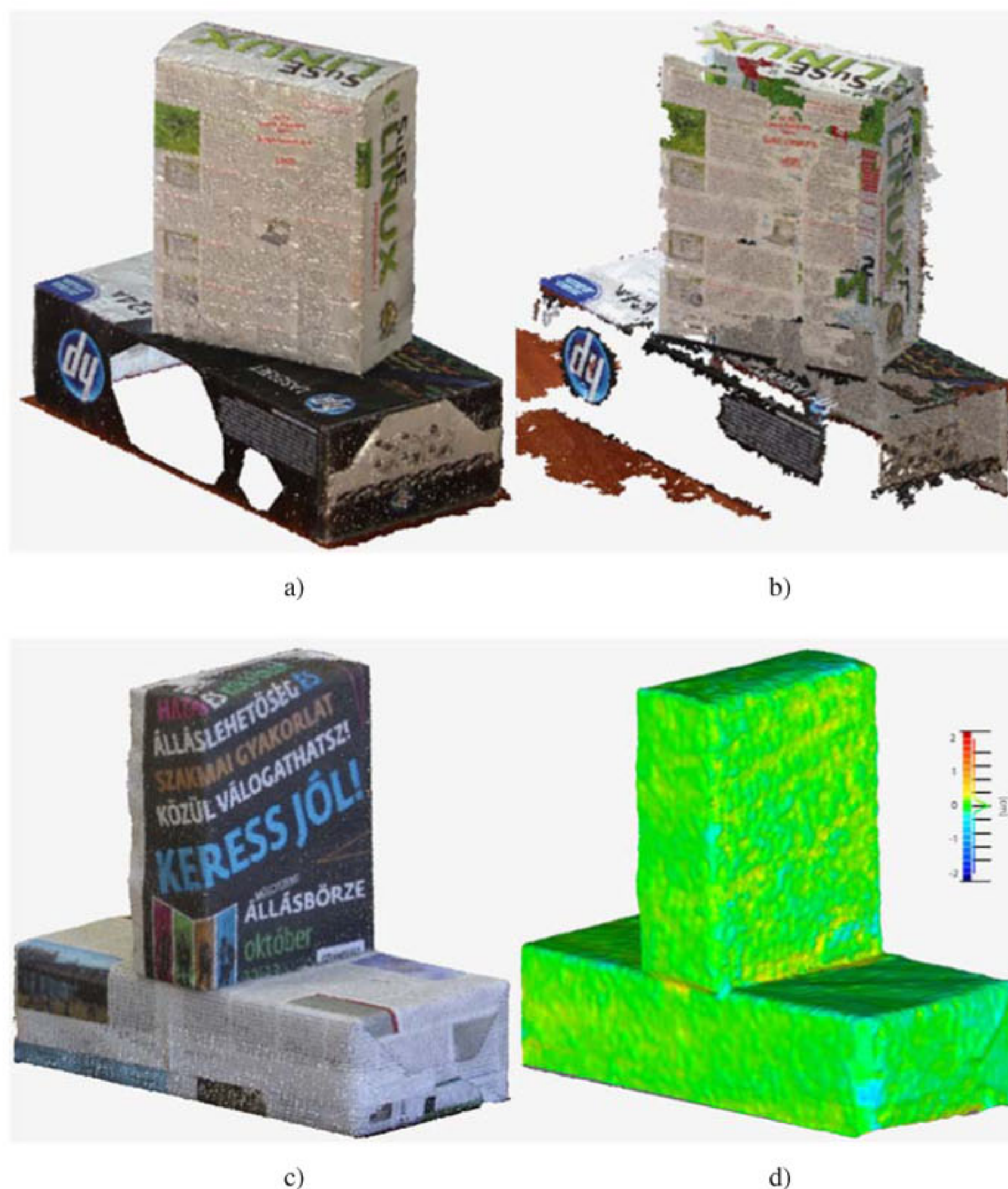


Fig. 1. Indoor test I: a) Reconstructed scene by CMPMVS; b) Reconstructed scene by Photoscan (ultra-high quality setting); Indoor test II: c) Reconstructed scene by Photoscan (ultra-high quality); d) Differences for Photoscan (ultra-high quality) compared to TLS measurements

For the outdoor measurements one of the lions of the Chain Bridge at Budapest was selected (*Fig. 2*), which is located on the northern part of the Pest side. The survey was carried out in March 2016. The Faro TLS was used to produce the reference point cloud; to ensure the appropriate coverage; two stations were established on either side of the road. Point spacing for each of the positions were selected 3 mm at 10 meters. ICP method was used to calculate the alignment between the point clouds. Still images were captured from roughly 180° around the sculpture. Image count: 25; resolution: 6000×4000 pixels.

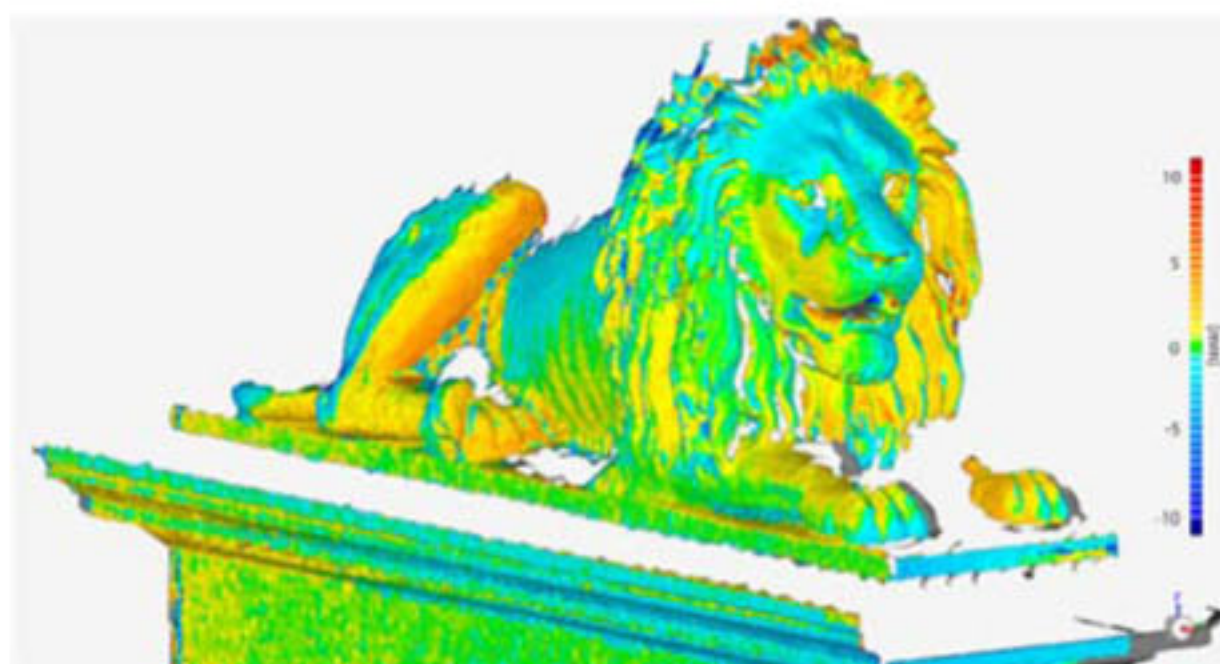


Fig. 2. Outdoor test: Color-coded difference image for Agisoft Photoscan (Ultra High quality setting) compared to the TLS reference

5. The comparison of the derived point clouds

The point clouds produced by different software packages have different scale and orientation. If they are aimed to be analyzed properly, they have to be transformed into a common reference system. The transformation parameters were calculated in CloudCompare. The rough alignment was carried out by 4-point-alignment [28] then the fine registration was performed with scalable ICP method.

The comparison of the registered point clouds was made in Geomagic Control 2014. A surface model was created based on the laser-scanned data then the difference of each individual point clouds was calculated. The statistical quality measures (the mean of all differences (full range), the mean of only negative and the mean of only positive differences as well as and the Standard Deviation (SD)) were collected in the tables (*Table I-III*). Some point cloud regions were excluded from the analysis due to their considerable distance from the reference surface (*Fig. 1d*).

There are large variations in the number of the points of the outdoor measurements (*Table I*), from 2.5 thousands to almost 28 million. The sparse point cloud produced by Photoscan has higher standard deviation. There are significantly larger values compared to other versions, however, those decrease in the dense cloud reconstruction step (*Fig. 2*). Comparing the sparse and dense point clouds for the same methods, it is noticeable that the former ones are more affected by errors. There is one exception; due to the extreme computational requirements, the input image resolution for the COLMAP's densification had to be reduced significantly [29], [30].

Table I

Differences between the point clouds of the software packages

Outdoor test						
Software package	Type of point cloud	Number of points	Mean difference [mm]			SD [mm]
			-	+	Full range	
3DF Samantha	Sparse	13 344	-15	14	-1.4	23
	Sparse*	2 562	-32	17	-7.6	34
Agisoft Photoscan	Dense (Medium)	1 865 985	-14	9	-4.1	19
	Dense (High)	7 356 294	-11	7	-2.7	18
	Dense (Ultra high)	27 979 914	-13	7	-3.7	19
Autodesk Remake	From TIN model**	1 409 656	-12	8	-2.6	20
CMPMVS	From TIN model***	2 437 068	-9	5	-1.8	16
COLMAP	Sparse	12 166	-6	5	-0.3	9
	Dense****	62 084	-12	14	0.7	21
Meshrecon	From TIN model	264 113	-10	7	-0.7	18
MVE	Dense	11 569 941	-8	8	-0.3	14
Sure	From TIN model	5 905 579	-14	11	-1.9	20
Visual SfM	Sparse	16 436	-12	9	-1.2	18
	Dense (PMVS/CMVS)	1 221 237	-6	5	-0.6	12
TLS reference	-	216 464				
*19%, **13%, ***11%, ****10% out of investigated range						

The point clouds of the different methods resulted in very similar statistics (*Table I-III*). However, it does not mean that they could be used in the same way. During the first indoor measurement, the surfaces of the measured boxes were glossy, and the lower box was black, which was very challenging for all the software packages (*Fig. 1a*). Photoscan created a very incomplete point cloud with the ultra-high quality setting; typically the surroundings of the texts were reconstructed. However, the lower settings (medium and high) resulted larger point clouds, they achieve better globally coverage. Three software packages (Remake CMPMVS, Meshrecon) could reconstruct almost the entire scene.

In the case of the second indoor measurement all the reconstructed point clouds were nearly complete due to the applied matte newspaper cover. From the recorded video streams every 25th images were extracted and saved as still images. In some cases, not every trajectory could be processed due to the enormous computation requirements or the total amount of pixels (of all images) is limited in the software. This negative impact can be experienced with COLMAP and Sure. The differences between the sparse and dense point clouds can be seen in *Table III*.

Mesh models were created from the point clouds, and then their surface and volume were calculated. Almost all software packages created smaller model than the reference, except MVE. Significant volume and surface differences cannot be observed; it remained below 10 percent in all cases. The less accurate models (sparse point clouds from Remake and Meshrecon) come from the noisy point clouds.

Table II

Differences between the point clouds of the software packages

Indoor test I						
Software package	Type of point cloud	Number of points	Mean difference [mm]			SD [mm]
			-	+	Full range	
3DF Samantha	Sparse	6 478	-2	2	0.6	3
	Sparse	4 850	-1	2	0.4	2,1
Agisoft Photoscan	Dense (Medium)	880 921	-2	2	0.3	3
	Dense (High)	1 031 918	-1	2	0.8	3
	Dense (Ultra high)	428 307	-1	2	0.3	2
Autodesk Remake	From TIN model	97 933	-1	1	0.3	2
CMPMVS	From TIN model	400 282	-2	1	0.4	3
COLMAP	Sparse	3 509	-1	5	2.8	6
	Dense	27 127	-4	2	-1.8	5
Meshrecon	From TIN model	322 711	-2	2	0.3	3
Sure	From TIN model	143 443	-1	1	0.1	2
	Sparse	6 843	-1	1	0.3	2
Visual SfM	Dense (PMVS/CMVS)	265 819	-1	1	-0.1	2
TLS reference	-	1 236 750				

Table IV shows the dimensions of the boxes derived by the parallel plane alignments, and the volume information. Since the planes were fitted with error filtering, the volumes are more accurate. Generally the obtained difference for the dense models is less than 1%, although Colmap and Meshrecon are outliers having 5% and 4%, respectively (Fig. 3).

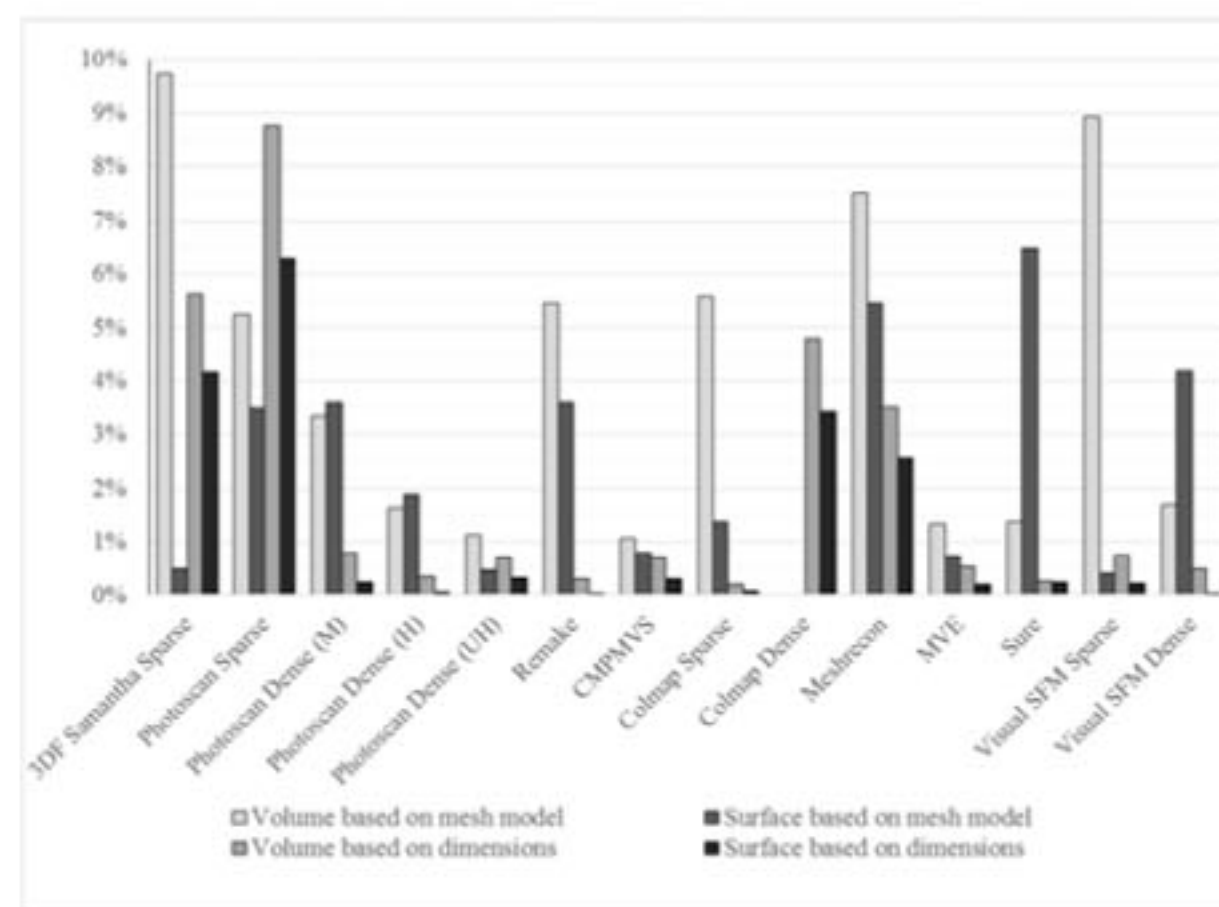


Fig. 3. Volume/surface absolute difference from the reference model

Table III

Differences between the point clouds of the software packages

Indoor test II						
Program	Type of point cloud	Number of points	Mean difference [mm]			SD [mm]
			-	+	Full range	
3DF Samantha	Sparse	20 082	-4	3	-1.0	5
	Sparse	6 028	-4	4	-0.9	5
Agisoft	Dense (Medium)	39 278	-1	1	-0.1	1
Photoscan	Dense (High)	172 750	-1	1	0.0	1
	Dense (Ultra high)	424 597	-1	1	-0.1	1
Autodesk Remake	From TIN model	9 686	-1	1	0.0	1
CMPMVS	From TIN model	297110	-1	1	-0.1	1
COLMAP	Sparse	11 535	-2	1	-0.2	3
	Dense	44 080	-5	4	-3.1	5
Meshrecon	From TIN model	39 454	-2	0	-1.4	1
MVE	Dense	44 606	-1	1	0.1	1
Sure	From TIN model	174 607	-2	2	-0.1	3
	Sparse*	14 228	-3	3	-0.8	5
Visual SfM	Dense (PMVS/CMVS)	114 633	-1	1	-0.2	2
TLS reference	-	162 305				

*11% out of investigated range

6. Conclusions

The analyzed spatial reconstruction software packages can fulfill the expectations in a similar research approach. The packages were tested with indoor and outdoor test sets, with still image and sampled video stream image sets. Almost all commercial and freely available software have supported the whole reconstruction workflow, starting with the image alignment and resulting textured surface model. Two main measures were in the focus: accuracy and completeness (degree of covering). The reference was obtained by terrestrial laser scanning, which resulted adequate amount of object points as well as color information. The analyzed packages are very similar regarding accuracy, but show significant variance in completeness. The obtained number of points in the dense model is independent from the accuracy or from the completeness. In all cases the sparse point clouds have radically lower quality measures, than that of the dense versions; therefore the densification phase is to be considered essential. The follow-up texturizing option supports visualization.

Table IV

The differences between the resulted point clouds of the software packages for indoor test II

Program	Type of point cloud	Volume [dm ³]*	Surface [dm ²]*	Upper box [mm]			Volume [dm ³]**
				Lower box [mm]			
				a	b	c	
3DF Samantha	Sparse	12.908	44.17	253	200	96	13.421
				109	405	194	
Agisoft Photoscan	Sparse	13.549	42.83	249	198	93	12.972
				109	403	192	
	Dense (Medium)	13.820	42.78	258	204	98	14.331
				111	415	199	
	Dense (High)	14.066	43.55	259	204	99	14.269
				110	415	199	
Dense (Ultra high)	14.137	44.17	259	205	99	14.319	
			110	417	199		
Autodesk Remake 2017	From TIN model	13.519	42.79	259	204	98	14.263
				110	416	199	
CMPMVS	From TIN model	14.148	44.04	259	204	99	14.318
				110	416	199	
COLMAP	Sparse	13.501	45.00	259	204	99	14.248
				110	415	198	
	Dense	NA	NA	257	200	98	13.535
				108	411	191	
Meshrecon	From TIN model	13.226	41.96	255	202	97	13.721
				109	410	195	
MVE	Dense	14.489	44.70	258	205	99	14.294
				110	416	199	
Sure	From TIN model	14.102	47.26	258	204	98	14.183
				109	416	199	
Visual SfM	Sparse*	13.022	44.20	259	204	99	14.324
				110	415	198	
	Dense (PMVS/CMVS)	14.058	46.24	258	204	99	14.290
TLS reference	-	14.299	44.38	111	416	198	14.219
				259	204	99	
				109	417	198	
* based on surface model; ** based on dimensions							

* based on surface model; ** based on dimensions

References

- [1] Chen J., Dowman I., Li S., Li Z., Madden M., Mills J., Paparoditis N., Rottensteiner F., Sester M., Toth C., Trinder J., Heipke C. Information from imagery: ISPRS scientific vision and research agenda, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 115, 2016, pp. 3–21.
- [2] Hartley R., Zisserman A. *Multiple view geometry in computer vision*, (2nd Ed) Cambridge University Press, New York, NY, USA, 2003.
- [3] McGlone C., Mikhail E., Bethel J. *Manual of photogrammetry*, (5th Ed.) American Society for Photogrammetry and Remote Sensing, 2004.

- [4] Förstner W., Gülch E. A fast operator for detection and precise location of distinct points, corners and centers of circular features, *Proceedings of the ISPRS Intercom Mission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, 2-4 June 1987, pp. 281–305.
- [5] Harris C., Stephens M. A combined corner and edge detector, *Alvey Conference*, Manchester, United Kingdom, 31 August - 2 September 1988, pp. 147–152.
- [6] Lowe D. G. Object recognition from local scale-invariant features, *Proceedings of the International Conference on Computer Vision*, Corfu, Greece, 20-27 September 1999, pp. 1150–1157.
- [7] Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features, *9th European Conference on Computer Vision*, Graz, Austria, 7-13 May 2006, pp. 404–417.
- [8] Nistér D. An efficient solution to the five-point relative pose problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 6, 2004, pp. 756–770.
- [9] Frahm J. M., Gallup D., Johnson T., Raguram R., Wu C., Jen Y.H., Dunn E., Clipp B., Lazebnik S., Pollefeys M. Building Rome on a cloudless day, *European Conference on Computer Vision*, Heraklion, Greece, 5-11 September 2010, pp. 368–381.
- [10] Mayer H. Efficient hierarchical triplet merging for camera pose estimation, *German Conference on Pattern Recognition, GCPR*, Münster, Germany, 3-5 September 2014. Springer, Berlin, pp. 99–409.
- [11] Fischler M. A., Bolles R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM*, Vol. 24, No. 6, 1981, pp. 381–395.
- [12] Triggs B., McLauchlan P., Hartley R., Fitzgibbon A. Bundle adjustment - A modern synthesis, *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice (ICCV'99)*, Springer-Verlag, London, UK, 1999, pp. 298–372.
- [13] Hirschmüller H. Stereo processing by semi-global matching and mutual information *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, 2008, pp. 328–341.
- [14] Pierrot-Deseilligny M., Paparoditis N. A multi-resolution and optimization based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI-1/W41, 2006, pp. 1–5.
- [15] Furukawa Y., Ponce J. Accurate, dense and robust multi-view stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, 2010, pp. 1362–1376.
- [16] Bulatov D., Wernerus P., Heipke C. Multi view dense matching supported by triangular meshes, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 66, No. 6, 2011, pp. 907–918.
- [17] Gherardi R., Farenzena M., Fusiello A. Improving the efficiency of hierarchical structure-and-motion, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, 13-18 June 2010, pp. 1594–1600.
- [18] Jancosek M., Pajdla T. Multi-view reconstruction preserving weakly-supported surfaces, *CVPR 2011, IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, Colorado, USA, 20-25 June 2011, pp. 3121–3128.
- [19] Schönberger J. L., Frahm J. M. Structure-from-motion revisited, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 26 June - 1 July, 2016, pp. 4104–4113.
- [20] Schönberger J. L., Zheng E., Pollefeys M. Pixelwise view selection for unstructured multi-view stereo, *Proceedings of 14th European Conference on Computer Vision, ECCV 2016, Part III*, Amsterdam, The Netherlands, 11-14 October 2016, pp. 501–518.

- [21] Fuhrmann S., Langguth F., Goesele M. MVE, A multi-view reconstruction environment, *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage*, Darmstadt, Germany, 6-8 October, 2014, pp. 1-8.
- [22] Rothmel M., Wenzel K., Fritsch D., Haala N. SURE: Photogrammetric surface reconstruction from imagery, *Proceedings LC3D Workshop*, Berlin, Germany, 4-5 December 2012, pp. 1-9.
- [23] Wu C. Towards linear-time incremental structure from motion, *Proceedings of the 2013 International Conference on 3D Vision (3DV'13)*, Washington, DC, USA, June 29 - July 1 2013, pp. 127-134.
- [24] Wu C. SiftGPU: A GPU implementation of scale invariant feature transform, (SIFT) 2007, <http://www.cs.unc.edu/~ccwu/siftgpu/> (last visited 21 November 2016).
- [25] Wu C., Agarwal S., Curless B., Seitz S. M. Multicore bundle adjustment, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, United States, 20-25 June 2011, pp. 3057-3064.
- [26] Besl P. J., McKay N. D. A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, 1992, pp. 239-256.
- [27] Chen Y., Medioni G. Object modeling by registration of multiple range images, *Image Vision Computing*, Vol. 10, No. 3, 1992, pp. 145-155.
- [28] Aiger D., Mitra N. J., Cohen-Or D. 4-points congruent sets for robust pairwise surface registration, *ACM Transactions Graph.* Vol. 27, No. 3, Article 85, 2008, pp. 1-10.
- [29] Molnár B. Developing a web based photogrammetry software using DLT, *Pollack Periodica*, Vol. 5, No. 2, 2010, pp. 49-56.
- [30] Berényi A. Laser scanning in engineering survey, - An application study, *Pollack Periodica*, Vol. 5: No. 2, 2010, pp. 39-48.