

## OPEN-SOURCE IMAGE-BASED 3D RECONSTRUCTION PIPELINES: REVIEW, COMPARISON AND EVALUATION

E.-K. Stathopoulou<sup>1,2</sup>, M. Welpner<sup>1</sup>, F. Remondino<sup>1</sup>

<sup>1</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy  
Email: <estathopoulou><welpner><remondino>@fbk.eu - Web: <http://3dom.fbk.eu>

<sup>2</sup> Laboratory of Photogrammetry, National Technical University of Athens (NTUA), Greece

### Commission II

**KEY WORDS:** photogrammetry, image orientation, SfM, dense image matching, MVS, validation.

#### ABSTRACT:

State-of-the-art automated image orientation (Structure from Motion) and dense image matching (Multiple View Stereo) methods commonly used to produce 3D information from 2D images can generate 3D results – such as point cloud or meshes – of varying geometric and visual quality. Pipelines are generally robust and reliable enough, mostly capable to process even large sets of unordered images, yet the final results often lack completeness and accuracy, especially while dealing with real-world cases where objects are typically characterized by complex geometries and textureless surfaces and obstacles or occluded areas may also occur. In this study we investigate three of the available commonly used open-source solutions, namely *COLMAP*, *OpenMVG+OpenMVS* and *AliceVision*, evaluating their results under diverse large scale scenarios. Comparisons and critical evaluation on the image orientation and dense point cloud generation algorithms is performed with respect to the corresponding ground truth data. The presented FBK-3DOM datasets are available for research purposes.



Figure 1. From UAV image network and sparse point cloud (left and middle) to dense point cloud (right).

### 1. INTRODUCTION

The development of robust computer vision algorithms has facilitated the democratization of the standard photogrammetric pipeline for 3D reconstruction purposes. Towards this end, several software implementations are now available as free, open-source (Table 1) or commercial, providing to users different levels of automatization, parameter tuning and customization. The input is commonly a set of images with extra camera metadata and the output can be, depending on the requirements, a dense coloured point cloud or a triangulated textured mesh. A typical pipeline starts with image orientation that relies on abundant feature matches among the images and the sparse point cloud triangulation (often called Structure from Motion - SfM) by means of incremental and/or global bundle adjustment (BA). Then, a dense 3D reconstruction is performed (normally called Multiple View Stereo - MVS) in order to further densify the sparse point cloud by reconstructing the depth value of almost every pixel correspondence in the 3D space.

The available open-source methods are fairly robust, they offer full access to parameters and they are able to cope even with large sets of unordered and diverse images, but the final 3D data often lack completeness and accuracy, especially while dealing with real-world cases where objects are commonly defined by complex geometries (Figure 1), textureless surfaces, repeated patterns, obstacles or occluded areas. Open-source solutions are normally not designed to support scaled 3D reconstructions with the use of ground control points (GCPs), but with a simple Helmert transformation. On the other hand, in case of closed-source commercial software, there is generally a lack of custom

in-deep parametrization that may often result to misleading output or black-box usage.

#### 1.1 Aim of the paper

The paper considers three of the most common open-source image-based 3D reconstruction pipelines (Table 1):

- *OpenMVG* library combined with *OpenMVS* (Moulon et al., 2016; Moulon et al., 2013; Moisan et al., 2012; Moulon et al., 2012a; Moulon et al., 2012b);
- *COLMAP* pipeline (Schönberger and Frahm, 2016; Schönberger et al., 2016a);
- *AliceVision* computer vision framework (Moulon et al., 2016; Jancosek et al., 2011).

Being open-source, full control of the implemented functions is guaranteed along with parametrization and interchangeability between the pipelines (Table 2). The aim is to check algorithm reliability and performances on large and extensive datasets. Experiments are indeed performed on a series of heterogeneous datasets of large scale scenarios, acquired with high-resolution cameras (Table 3). None of the pipelines allows the usage of GCPs observations for the datum definition, hence all 3D results are scaled and roto-translated with a simple Helmert transformation after the bundle adjustment. Analyses of the results and geometric comparisons are presented and discussed. The rest of the paper is organized as follows: Section 2 reports related work on open-source software for 3D reconstruction, relative benchmark releases and comparison studies. Section 3

describes the respective implementation details for every pipeline that is used in our experiments, as well as the dataset characteristics, while Section 4 presents the experiments and their comparisons, followed by conclusions in Section 5.

## 2. RELATED WORK

Evaluating 3D reconstruction pipelines is a common task in the research community. Remondino et al. (2017) considers large real-world aerial and terrestrial dataset and compares the 3D reconstruction performance using various commercial software and photogrammetric metrics. Similar evaluation studies with commercial software and UAV or underwater images were presented in Georgopoulos et al. (2016), Alidoost and Arefi (2017), Mangeruga et al. (2018) and Vlachos et al. (2019).

### 2.1 Open-source pipelines

Image-based 3D reconstruction has developed immensely during the last decades. Thus, numerous free and open-source solutions became available to the community, with *Photo Tourism* (later known as *Bundler*) being one of the pioneers in the field (Snavely et al., 2006). Few years later algorithms were able to work even with city scale reconstructions (Agarwal et al., 2011). *VisualSfM* was one of the first largely used all-in-one GUI solutions (Wu et al., 2011; Wu, 2013) integrating also the famous PMVS/CMVS (Furukawa and Ponce 2009; Furukawa et al., 2010) dense image matching method. In the last years, many other implementations provide full standalone 3D reconstruction pipelines, such as *COLMAP* (Schönberger and Frahm, 2016; Schönberger et al., 2016a) or a combination of several libraries and algorithms for SfM and/or MVS like *OpenMVG* (Moulon et al., 2016), *MVE* (Fuhrmann et al., 2014), *Theia* (Sweeney, 2015), *OpenMVS*<sup>1</sup> or the *OpenSfM* library<sup>2</sup> of Mapillary. The aforementioned open-source solutions, mostly developed by the computer vision community, target to a broader 3D reconstruction audience and thus, their main purpose is not metric accuracy but rather photorealistic 3D models of arbitrary scale and low geometric quality. *MicMac*<sup>3</sup> (Pierrot-Deseilligny and Paparoditis, 2006; Rupnik et al., 2017) is, on the other hand, a fully photogrammetric open-source pipeline able to handle GCPs and camera constraints (e.g. fixed baselines, etc.).

### 2.2 Accuracy assessment

Comparison and evaluation of algorithm performances have always accompanied algorithm developments. For this, high quality benchmark data are necessary and a plethora of benchmarks were released, usually targeting to one specific subtask of the 3D reconstruction pipeline. The establishment of good reference datasets (ground truth) requires an accuracy level which is commonly two or three times better than the expected results. The choice of the entities be compared and evaluated is also not straightforward (e.g. few single points? an entire surface? small patches?), neither is the procedure (e.g. Euclidean distance? Hausdorff distance?), nor the metrics (e.g. signed distance? sigma MAD? standard deviation? completeness? accuracy? RMS error on plane fitting?). In photogrammetry, important metrics are the standard deviation of unit weight, averaged residuals of image coordinates, the standard deviation of object coordinates, accuracy w.r.t. independent references, relative accuracy and completeness. Less implemented measures

used to judge SfM results are intersection angles, redundancy and number of oriented cameras etc.

### 2.3 Benchmarks

Middlebury University (Seitz et al., 2006) released several datasets of indoor laboratory scenarios focusing either on the evaluation of dense stereo matching<sup>4</sup> algorithms or on multi view stereos. EPFL (Strecha et al., 2008) published real-world outdoor datasets<sup>6</sup> with their corresponding 3D meshes for multi-view stereo and camera calibration evaluation.

Özdemir et al. (2019) recently introduced the 3DOMcity<sup>7</sup> multi-purpose benchmark dataset serving also for the evaluation of the whole photogrammetric 3D urban reconstruction in terms of image orientation and dense image matching. Indoor and outdoor video frame datasets for evaluation of large scale 3D reconstructions are also available in the Tanks and Temples (Knapitsch et al., 2017) and ETH3D<sup>9</sup> (Schöeps et al., 2017) benchmarks.

Other benchmarks in the general field of 3D reconstruction include RGB-D datasets (captured with proper sensors or generated synthetically) to evaluate SLAM, visual odometry and optical flow methods in terms of camera trajectory estimation and 3D reconstruction (Handa et al., 2012; Sturm et al., 2012; Geiger et al., 2013; Menze et al., 2018).

## 3. IMPLEMENTED OPEN-SOURCE PIPELINES

The employed pipelines, their combined approaches and the communication/exchange protocols used in each case are summarized in Table 1. In most of the cases, we used conversion tools already implemented by developers except for the combined *AliceVision+OpenMVS* approach where an in-house converter was developed (it will be soon publicly released). The 3D reconstruction workflows provide a certain level of customization with different algorithms available for each processing step (Table 2) and further customization on demand is achieved through code accessibility.

Pipelines		Communication protocol
SfM	MVS	
COLMAP	COLMAP	native
	OpenMVS	<i>openmvs:InterfaceCOLMAP</i>
OpenMVG	OpenMVS	<i>openmvs:InterfaceOpenMVS</i>
	COLMAP	<i>openmvg:openMVG2COLMAP</i>
AliceVision	AliceVision	native
	OpenMVS	in-house converter
	COLMAP	<i>openmvg:openMVG2COLMAP</i> + in-house converter

Table 1: The employed pipelines and combined approaches.

### 3.1 COLMAP

*COLMAP* (Schönberger and Frahm, 2016; Schönberger et al., 2016a) is a pipeline that implements improved versions of both SfM and MVS which comes also with a graphical user interface facilitating the use by non-experts. Project information is stored in a database structure format. Regarding correspondence search, it implements the well-known SIFT algorithm (Lowe, 2004) providing both CPU and GPU options, followed by an extensive list of feature matching options such as exhaustive matching,

<sup>1</sup> <https://github.com/cdscave/openMVS>

<sup>2</sup> <https://github.com/mapillary/OpenSfM>

<sup>3</sup> <https://micmac.engg.eu/>

<sup>4</sup> <http://vision.middlebury.edu/stereo/>

<sup>5</sup> <http://vision.middlebury.edu/mview/eval/>

<sup>6</sup> <https://www.epfl.ch/labs/cvlab/data/data-strechamvs/>

<sup>7</sup> <https://3dom.fbk.eu/3domcity-benchmark>

<sup>8</sup> <https://www.tanksandtemples.org/>

<sup>9</sup> <http://www.eth3d.net>

	<i>COLMAP</i>	<i>OpenMVG</i>	<i>AliceVision</i>
<b>Keypoint extraction</b>	SIFT (Lowe, 2004)	SIFT (Lowe, 2004), AKAZE (Alcantarilla et al., 2013)	SIFT (Lowe, 2004), AKAZE (Alcantarilla and Solutions, 2011)
<b>Keypoint matching</b>	Exhaustive, Sequential, Vocabulary Tree (Schönberger et al., 2016b), Spatial (Schönberger and Frahm, 2016), Transitive (Schönberger and Frahm, 2016)	Brute Force, ANN (Muja and Lowe, 2009), Cascade Hashing (Cheng et al., 2014)	ANN (Muja and Lowe, 2009) Cascade Hashing (Cheng et al., 2014)
<b>Geometric verification</b>	4 points homography (Hartley and Zissermann, 2003), 5 points relative pose (Stewenius et al., 2006), 7-8 points F-matrix (Hartley and Zissermann, 2003)	4 points homography (Hartley and Zissermann, 2003), 5 points relative pose (Stewenius et al., 2006), 7-8 points F-matrix (Hartley and Zissermann, 2003)	4 points homography (Hartley and Zissermann, 2003), homography growing (Srajer 2016), 7-8 points F-matrix (Hartley and Zissermann, 2003), 5 points relative pose (Stewenius et al., 2006)
<b>Incremental bundle (image resection and triangulation)</b>	P3P (Gao et al., 2003) + DLT, EPnP (Lepetit et al., 2009) + DLT	P3P (Gao et al., 2003) + DLT, EPnP (Lepetit et al., 2009) + DLT	PnP (Gao et al., 2003) + DLT
<b>Global bundle adjustment</b>	CERES <sup>10</sup>	CERES	CERES
<b>Dense point cloud generation</b>	Patch-based stereo (Schönberger et al., 2016a)	Patch-based stereo (OpenMVS – Shen, 2013)	Semi Global Matching (Hirschmüller, 2007)

Table 2: Open-source image-based 3D reconstruction pipelines analysed in this study.

sequential matching, vocabulary tree, spatial matching, transitive matching and custom matching. Image pairs are considered verified if a valid mapping of their geometric relation exists (homography, essential or fundamental matrix) and thus the scene graph is created gradually. 3D reconstruction is done by implementing incremental SfM starting from a carefully selected initial image pair and applying a robust next best view selection algorithm and subsequently multi-view triangulation. The bundle adjustment step uses Ceres solver and global BA every certain steps to improve camera and point estimations and avoid drifting (Schönberger and Frahm, 2016). Multi-view stereo reconstruction is implemented based on the framework of (Zheng et al., 2014) using a probabilistic patch-based stereo approach (Schönberger et al., 2016a).

### 3.2 OpenMVG+OpenMVS

*OpenMVG* provides a complete and neat SfM pipeline based on standard multiple view geometry principles. Feature detection and description are implemented with SIFT (Lowe, 2004) and AKAZE (Alcantarilla et al., 2013), while detection and description of invariant regions can also be used (Xu et al., 2014; Nistér and Stewenius, 2008). Feature matching is implemented by classic brute force, ANN-kD trees (Muja and Lowe, 2009), or cascade hashing (Cheng et al., 2014). Geometric verification of the image pairs is implemented in a similar fashion to *COLMAP*. Sparse reconstruction can be calculated using incremental (Moulon et al., 2012) or global (Moulon et al., 2013) methods followed by bundle adjustment using Ceres solver. For this pipeline combination, we consider dense reconstruction as implemented by the *OpenMVS* library based on patch-based stereo method for large-scale scenes (Shen, 2013).

### 3.3 AliceVision

*AliceVision* is a computer vision framework on which the graphical user interface *Meshroom* is based. Similarly to *OpenMVG*, it implements SIFT (Lowe, 2004) and AKAZE (Alcantarilla et al., 2013) and some variations of them for feature detection and description, while ANN (Muja and Lowe, 2009), cascade hashing (Cheng et al., 2014) and vocabulary trees can be used for matching. The MVS part, based on a typical semi-global

matching method (Hirschmüller, 2007), does not however grant access to the dense point cloud data but rather outputs to a textured mesh model.

## 4. EXPERIMENTS AND EVALUATION

The datasets used in our experiments (Table 3), featuring varying amount of images, acquisition platforms, sensor sizes and ground sampling distance (GSD) values, consist of: a set of aerial images of an ancient temple of complex form (*Nettuno\_aerial*), the respective terrestrial image set of the same temple (*Nettuno\_terrestrial*) as well as the fused one (*Nettuno\_fused*), a set of terrestrial images of a church gate (*Modena*), a dataset (*Barn TaT*) from the Tanks and Temple benchmarks, as well as the newly released 3DOMcity<sup>7</sup> benchmark dataset (Özdemir et al., 2019).

In theory, given the same set of images as input to the several pipelines, the results would be similar and comparable to each other. However, due to the fact that the implementation details of each solution have its own limitations, strengths and weaknesses, the generated results may vary drastically from each other. In this study, we focus on tuning and combining the available parameters, optimizing them according to the needs of each family scenario. At the same time, we implement a further automatization step for each one of the three pipelines to facilitate mass processing of large-scale scenarios that takes into consideration the respective available parameters and lets the user run on demand, organizing the output in a meaningful way. We perform analysis on the image orientation (SfM) results (Tables 4 and 5), as they directly affect the quality of the final dense reconstruction, as well as on the final the 3D dense point cloud (Figure 3, Table 6). The used parameter combinations for each of the employed pipelines were selected carefully after examination and study of all available parameters under each implementation. In more details, for *COLMAP*, as SIFT was the only available descriptor and the sparse reconstruction is achieved with an incremental bundle adjustment (IBA), we choose to experiment with the sequential (S) and exhaustive (E) keypoint matching options. On the other hand, for *OpenMVG* and *AliceVision* we examined both AKAZE and SIFT feature detectors/descriptors, coupled with cascade hashing as criterion for the best match between the descriptors.

<sup>10</sup> <http://ceres-solver.org/>



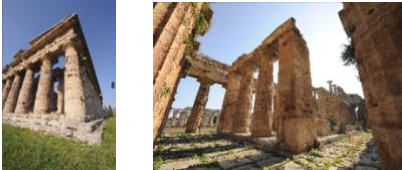
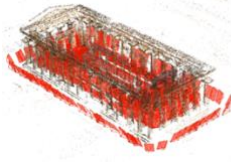
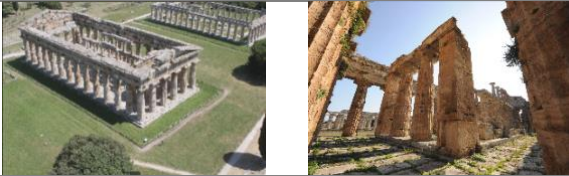
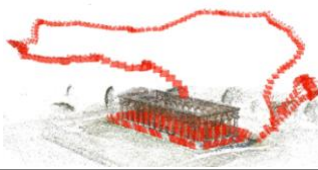





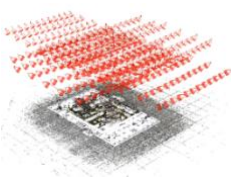
Type	Camera	Provenance	# images	Image resolution & size [px]	GSD [mm/pixel]	Validation / Ground Truth (GT)	# GT points
<b>NETTUNO AERIAL</b>							
Aerial / UAV	Canon EOS 550D, 22.3x14.9 mm, CMOS sensor, 25 mm focal	FBK-3DOM	212	18 MPx, 5184*3456 px (scaled to ½)	ca. 12 (scaled to ½)	C2C vs laser scanner (res. 3 mm)	37.6 M
							
<b>NETTUNO TERRESTRIAL</b>							
terrestrial	Nikon D3X, 36x24mm CMOS sensor, 50mm focal length	FBK-3DOM	404	24 MPx, 6048*4032 px (scaled to ½)	ca. 3 (scaled to ½)	C2C vs laser scanner (res. 3 mm)	37.6 M
							
<b>NETTUNO FUSED</b>							
fused	as in Nettuno_aer & Nettuno_terr	FBK-3DOM	616	as in Nettuno_aer & Nettuno_terr	as in Nettuno_aer & Nettuno_terr	C2C vs laser scanner (res. 3 mm)	37.6 M
							
<b>BARN TaT s</b>							
terrestrial	Sony A7SM2, 23.9 x 35.8 mm CMOS sensor, 21 mm focal length	Benchmark TaT	410	2.1 MPx, 1920*1080 px	ca. 7	C2C vs laser scanner (res. 5 mm)	12.7 M
							
<b>MODENA</b>							
terrestrial	Nikon D750, 35.9 x 24 mm CMOS sensor, 28 mm focal length	FBK-3DOM	55	24 Mpx, 6016*4016 px	ca. 1.5	C2C vs laser scanner (res. 2 mm)	100 M
							
<b>3DOMCITY 7</b>							
aerial (nadir + oblique)	Nikon D750, 35.9 x 24 mm CMOS sensor, 50 mm focal length	Benchmark 3DOMcity	420	24 Mpx, 6016*4016	ca 0.12 (nadir), ca. 0.20 (oblique)	C2C vs photogram. point cloud (res. 0.3 mm)	28.4 M
							

Table 3. Scenarios and datasets considered in the presented evaluation (C2C = Cloud to Cloud comparison; GT = Ground Truth).

Both incremental (IBA) and global bundle adjustments (GBA) for sparse reconstruction were tested. For the processing based on combined pipelines (as defined in Table 1), we considered the best image orientation (SfM) result of every trial and we used it for the successive MVS / dense point cloud generation. As best result is considered the one that satisfies both the number of oriented images criterion (Table 4– more than 90% of the dataset images should be oriented successfully, inspected also visually) as well as the total reprojection error of the BA (Table 5). To be noted that the native *AliceVision* MVS module is providing only a binary triangulated mesh and not an unorganized dense point cloud. Therefore the geometric analyses are performed extracting the vertices of the meshes and *AliveVision* MVS was not combined with any other SfM pipeline.

#### 4.1 Image Orientation / SfM

Table 4 reports the number of oriented cameras per dataset and per pipeline, whereas Table 5 shows the RMS reprojection error of the bundle adjustments for each method.

We can observe that the *Nettuno\_fused* dataset (high resolution aerial + terrestrial images) is the most challenging one, as most of the methods (except *COLMAP*-exhaustive and *OpenMVG*-SIFT-incremental) faced problems in properly orienting all images while achieving a relatively low RMS reprojection error. The AKAZE operator under its *OpenMVG* implementation failed to describe the features because of the high computational cost. In *Nettuno\_aerial* all processing ran quite smoothly over all methods giving enough oriented cameras and relatively low RMS errors, apart from the *AliceVision*-SIFT implementations that resulted relatively high reprojection error values. This behaviour

was expectable given as the UAV flight offered a regular image acquisition with satisfying overlap.

In *Nettuno\_terr* and *Modena* datasets, higher in resolution and acquired by a handheld camera, the AKAZE operator and global adjustment failed to give an adequate number of oriented images, whereas there were almost no problems for sequential or exhaustive matching method and incremental bundle adjustment. For *Barn\_TaT* and *3DOMcity* benchmark datasets most of the pipelines performed well except for *AliceVision*-global that encountered problems orienting the images.

However, after a careful inspection of computed image networks and point clouds (sparse and dense), it was observed that there were few cases where even though the number of oriented images and the reprojection error were considered successful, images were erroneously oriented, resulting to drift effects on the final dense point clouds. Figure 2 shows an example of the *Barn\_TaT* dataset, where all 420 images were oriented by *COLMAP*-sequential with a half-pixel reprojection error (Table 5), yet a visual inspection of the cloud reveals an obvious drift. Thus, for this specific dataset, for the MVS step, the *COLMAP*-exhaustive results were chosen, although the error number was slightly larger (Table 5). Similar errors were observed in *Nettuno\_aerial* and *Nettuno\_fused* dataset, as *AliceVision* orientation was erroneous (397 and 616 oriented cameras, respectively) resulting in not-correct 3D reconstructions (Figure 2).

#### 4.2 Dense 3D reconstruction / MVS

Apart from a qualitative evaluation (Figure 3 – visual inspection on dense clouds and cross sections), the achieved dense point clouds are compared with the available GT data.

SfM - NUMBER OF ORIENTED IMAGES											
	# images	COLMAP		OpenMVG				AliceVision			
		SIFT		AKAZE		SIFT		AKAZE		SIFT	
		S	E	Fast Cascade Hashing				Fast Cascade Hashing			
		IBA	IBA	GBA	IBA	GBA	IBA	GBA	IBA	GBA	
<i>Nettuno_aerial</i>	212	212	212	212	211	212	212	212	<b>210</b>	212	212
<i>Nettuno_terr</i>	404	396	404	404	157	404	403	364	<b>83</b>	404	347
<i>Nettuno_fused</i>	616	212	616	--	--	615	403	212	<b>193</b>	616	544
<i>Barn TaT</i>	410	410	410	410	410	410	410	<b>90</b>	224	397	377
<i>Modena</i>	55	55	55	55	<b>32</b>	55	46	50	39	55	46
<i>3DOMcity</i>	420	420	420	420	418	420	420	407	<b>23</b>	420	219

Table 4. Number of oriented cameras for each pipeline. S: sequential keypoint matching; E: exhaustive keypoint matching; IBA: incremental bundle adjustment; GBA: global bundle adjustment. Worst achieved result per dataset are highlighted in bold.

SfM –RMS REPROJECTION ERROR [pixels]											
	# images	COLMAP		OpenMVG				AliceVision			
		SIFT		AKAZE		SIFT		AKAZE		SIFT	
		S	E	Fast Cascade Hashing				Fast Cascade Hashing			
		IBA	IBA	GBA	IBA	GBA	IBA	GBA	IBA	GBA	
<i>Nettuno_aerial</i>	212	<b>0.45</b>	0.47	<b>0.42</b>	0.46	0.46	0.48	<b>0.73</b>	0.78	0.91	0.95
<i>Nettuno_terr</i>	404	0.45	<b>0.58</b>	0.45	0.49	<b>0.40</b>	0.42	0.38	1.04	<b>0.53</b>	0.57
<i>Nettuno_fused</i>	616	0.44	<b>0.51</b>	--	--	<b>0.43</b>	0.42	0.73	0.76	<b>0.74</b>	0.88
<i>Barn TaT</i>	410	0.49	<b>0.56</b>	<b>0.55</b>	0.62	0.59	0.64	0.45	0.55	<b>0.70</b>	0.52
<i>Modena</i>	55	<b>0.71</b>	<b>0.71</b>	<b>0.49</b>	0.48	0.56	0.55	0.31	0.33	<b>0.82</b>	0.90
<i>3DOMcity</i>	420	<b>0.40</b>	0.48	<b>0.42</b>	1.17	<b>0.42</b>	<b>0.42</b>	0.27	0.39	<b>0.54</b>	0.41

Table 5. RMS reprojection error (pixel) for each pipeline. S: sequential; E: exhaustive; IBA: incremental bundle adjustment; GBA: global bundle adjustment. Best achieved results per dataset and pipeline are highlighted in bold.



Figure 2. Examples of badly reconstructed dense clouds (drift effect) because of erroneously oriented images.  
Up: *Barn\_TaT* dataset oriented with *COLMAP*-sequential (left) and *AliceVision*-SIFT-incremental (right).  
Bottom: *Nettuno\_aerial* dataset, *AliceVision*+*COLMAP* (left), *AliceVision*+*OpenMVS* (right).

Lacking GCPs integration modules, the clouds were co-registered using a 7-parameters ICP method, so introducing possible scale errors among the data. Given the complexity and non-planarity of the analysed scenarios, as metrics we considered the cloud to cloud (C2C) distance as implemented in *CloudCompare*.

Starting from the best result of each SfM method, we further proceed with the dense reconstructions using the native (*COLMAP*, *OpenMVS*, *AliceVision*) and the combined pipelines (Table 1). Although the number of reconstructed points is indicative whether the reconstruction algorithm has produced a complete result or not, yet cannot be used as a robust method to evaluate the quality of the reconstruction.

C2C distances are reported in Table 6. Since *AliceVision* outputs a triangulated mesh model, Cloud to Mesh (C2M) distances are given. The best scores over all pipelines are observed in the *3DOM\_city* dataset, probably due to the high quality of the dataset in terms of resolution and network geometry. On the contrary, the *Nettuno\_aerial* dataset has the highest error values over all pipelines, a fact that can be explained by the camera-object distance and image resolution. The *Nettuno\_terr* dataset resulted to high quality dense point clouds, achieving similar errors to the *Modena* dataset. Error values were generally of the same order of magnitude over all pipelines. *Nettuno\_fused* had higher errors than its respective terrestrial, yet lower than the aerial.

## 5. CONCLUSIONS

The paper presented a review of actual open-source image-based 3D reconstruction pipelines. Experiments were performed with diverse large real-world scenarios, testing the performance of *COLMAP*, *OpenMVG*+*OpenMVS* and *AliceVision*, as well as

their possible combinations. Towards this end, parts of these pipelines were automatized allowing a massive number of tests and observing the output over diverse parameter settings. As metrics for evaluating the SfM part were used the number of oriented images and the RMS reprojection error, while for the dense reconstructions the distance between the resulted cloud and its corresponding ground truth was chosen as evaluation criterion. According to our tests, AKAZE was low-performing with respect to SIFT and it seems that the incremental bundle adjustment grants better results. Assuming correct camera poses, patch-based MVS generally delivers more dense and accurate dense point clouds.

## REFERENCES

- Alidoost, F. and Arefi, H., 2017. Comparison of UAS-based photogrammetry software for 3D point cloud generation: a survey over a historical site, *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. IV-4/W4, pp. 55-61.
- Alcantarilla, P.F., Nuevo, J., Bartoli, A., 2013. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *Proc. BMVC*, Vol. 34(7), pp. 1281–1298.
- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M. and Szeliski, R., 2011. Building Rome in a day. *Communications of the ACM*, Vol., 54(10), pp.105-112.
- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). *J. Computer Vision and Image Understanding*, Vol. 110, pp. 34-359.
- Cheng, J., Leng, C., Wu, J., Cui, H., Lu, H., 2014. Fast and accurate image matching with cascade hashing for 3D reconstruction. In *Proc. CVPR*, pp. 1-8.

	C2C [mm]													
	COLMAP		COLMAP + OpenMVS		OpenMVG + OpenMVS		OpenMVG + COLMAP		Alice		Alice + COLMAP		Alice + OpenMVS	
	mean	STD	mean	STD	mean	STD	mean	STD	mean	STD	mean	STD	mean	STD
<i>Nettuno_aerial</i>	84	130	135	161	109	153	37	128	--	--	--	--	--	--
<i>Nettuno_terr</i>	29	27	33	30	24	24	22	24	41	415	--	--	--	--
<i>Nettuno_fused</i>	50	83	60	94	45	86	34	70	--	--	--	--	--	--
<i>Barn TaT</i>	23	34	17	20	16	20	16	25	--	--	--	--	--	--
<i>Modena</i>	4	6	8	7	6	8	6	11	1	38	6	5	6	5
<i>3DOMcity</i>	0.5	0.9	0.9	1	1	1	0.5	1	2	9	0.5	0.9	1	1

Table 6. C2C analyses (mean and standard deviation – STD) between ground truth (GT) reference and achieved results. Note that all *Nettuno* datasets were scaled down to half resolution for computational efficiency. Dash lines (--) means no data/results available due to bad orientation results.

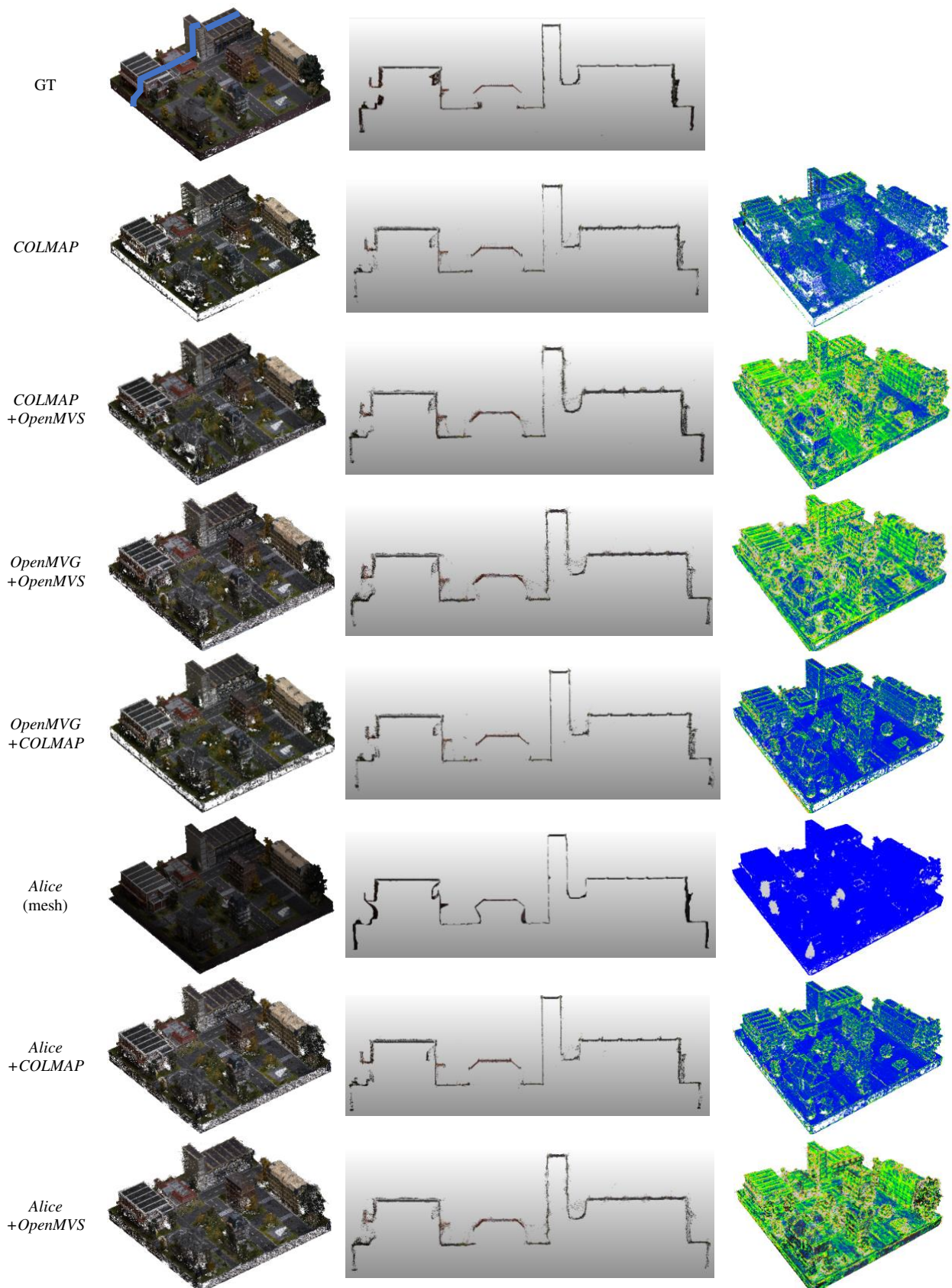


Figure 3. 3DOMcity dataset: derived dense point clouds, corresponding cross-sections and C2C/C2M distances (scale: 0(blue)-5(red) mm) with respect to the GT data for *COLMAP*, *COLMAP+OpenMVS*, *OpenMVG+OpenMVS*, *OpenMVG+COLMAP*, *Alice*, *Alice+COLMAP*, *Alice+OpenMVS*, respectively from top to bottom.

- Furukawa, Y. and Ponce, J., 2009. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 32(8), pp.1362-1376.
- Furukawa, Y., Curless, B., Seitz, S.M. and Szeliski, R., 2010. Towards internet-scale multi-view stereo. In *Proc. CVPR*, pp. 1434-1441.
- Fuhrmann, S., Langguth, F. and Goesele, M., 2014. MVE-a multi-view reconstruction environment. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage*, pp. 11-18.
- Handa, A., Newcombe, R.A., Angeli, A. and Davison, A.J., 2012. Real-time camera tracking: When is high frame-rate best? In *European Conference on Computer Vision*, pp. 222-235.
- Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F., 2003. Complete solution classification for the perspective-three-point problem. *IEEE Trans. PAMI*, Vol. 25, pp. 930-943.
- Geiger, A., Lenz, P., Stiller, C. and Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, Vol. 32(11), pp.1231-1237.
- Georgopoulos, A., Oikonomou, C., Adamopoulos, E. and Stathopoulou, E.K., 2016. Evaluating unmanned aerial platforms for cultural heritage large scale mapping. *Int. Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. XLI-B5, pp. 355-362.
- Hartley, R., Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press: Cambridge, UK.
- Hirschmüller, H., 2007. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. PAMI*, Vol., 30(2), pp.328-341.
- Jancosek, M. and Pajdla, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. In *CVPR 2011*, pp. 3121-3128.
- Knapitsch, A., Park, J., Zhou, Q.Y. and Koltun, V., 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, Vol. 36(4), pp.78.
- Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: An accurate o (n) solution to the PNP problem. *Int. J. Computer Vision*, Vol. 81.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, Vol. 60, pp. 911-10.
- Mangeruga, M., Bruno, F., Cozza, M., Agrafiotis, P., and Skarlatos, D., 2018. Guidelines for Underwater Image Enhancement Based on Benchmarking of Different Methods. *Remote Sensing*, Vol. 10(10).
- Menze, M., Heipke, C. and Geiger, A., 2018. Object scene flow. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 140, pp.60-76.
- Moisan, L., Moulon, P. and Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line*, Vol. 2, pp.56-73.
- Moulon, P., Monasse, P., Perrot, R. and Marlet, R., 2016. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pp. 60-74.
- Moulon, P., Monasse, P. and Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proc. IEEE ICCV*, pp. 3248-3255.
- Moulon, P., Monasse, P. and Marlet, R., 2012. Adaptive structure from motion with a contrario model estimation. In *Asian Conference on Computer Vision*, pp. 257-270, Springer, Berlin, Heidelberg.
- Moulon, P. and Monasse, P., 2012. Unordered feature tracking made fast and easy. In *CVMP 2012*.
- Muja, M. and Lowe, D., 2004. Fast Approximate Nearest Neighbors with automatic algorithm configuration. In *Proc. 4th VISAPP Int. Conference*, Vol. 1, pp. 331-340.
- Nistér, D. and Stewénius, H., 2008. Linear time maximally stable extremal regions. In *European Conference on Computer Vision*, pp. 183-196.
- Özdemir, E., Toschi, I. and Remondino, F., 2019. A multi-purpose benchmark for photogrammetric urban 3D reconstruction in a controlled environment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-1/W2, pp. 53-60.
- Pierrot-Deseilligny, M. and Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery. *Int. Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 36(1/W41).
- Remondino, F., Nocerino, E., Toschi, I. and Menna, F., 2017. A critical review of automated photogrammetric processing of large datasets. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. XLII-2/W5, pp. 591-599.
- Rupnik, E., Daakir, M. and Pierrot-Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. *Open geospatial data, software and standards*. Vol. 2(14).
- Schönberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. In *Proc. CVPR*, pp. 4104-4113.
- Schönberger, J.L., Zheng, E., Frahm, J.M. and Pollefeys, M., 2016a. Pixelwise view selection for unstructured multi-view stereo. In *Proc. ECCV*, pp. 501-518.
- Schönberger, J.L.; Price, T.; Sattler, T.; Frahm, J.M.; Pollefeys, M., 2016b. A vote-and-verify strategy for fast spatial verification in image retrieval. In *Asian Conference on Computer Vision*, pp. 321-337.
- Schöps, T., Schönberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M. and Geiger, A., 2017. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proc. CVPR*, pp. 3260-3269.
- Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. CVPR*, Vol. 1, pp. 519-528.
- Shen, S., 2013. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. *IEEE Transactions on Image Processing*, Vol. 22(5), pp.1901-1914.
- Snavely, N., Seitz, S.M. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. In *ACM transactions on graphics (TOG)*, Vol. 25, No. 3, pp. 835-846. ACM.
- Stewénius, H., Engels, C., Nistér, D., 2006. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 60, pp. 284-294.
- Strecha, C., Von Hansen, W., Van Gool, L., Fua, P. and Thoennessen, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. CVPR*, pp. 1-8.
- Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D., 2012. A benchmark for the evaluation of RGB-D SLAM systems. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 573-580. IEEE.
- Sweeney, C., Hollerer, T. and Turk, M., 2015. Theia: A fast and scalable structure-from-motion library. In *Proceedings of the 23rd ACM Int. Conf. on Multimedia*, pp. 693-696. ACM.
- Vlachos, M., Berger, L., Mathelier, R., Agrafiotis, P., and Skarlatos, D., 2019. Software comparison for underwater archaeological photogrammetric applications, *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol., XLII-2/W15, pp. 1195-1201.
- Wu, C., 2013. Towards linear-time incremental structure from motion. In *Proc. 3DV*, pp. 127-134.
- Wu, C., Agarwal, S., Curless, B. and Seitz, S.M., 2011. Multicore bundle adjustment. In *Proc. CVPR*, pp. 3057-3064.
- Xu, Y., Monasse, P., Géraud, T. and Najman, L., 2014. Tree-based morse regions: A topological approach to local feature detection. *IEEE Transactions on Image Processing*, Vol., 23(12), pp.5612-5625.
- Zheng, E., Dunn, E., Jovic, V. and Frahm, J.M., 2014. Patchmatch based joint view selection and depthmap estimation. In *Proc. CVPR*, pp. 1510-1517.