2019 EC³
2019 European Conference on Computing in Construction
Chania, Crete, Greece
July 10-12, 2019

PRECISE RECONSTRUCTION OF GEOMETRIC PRIMITIVES IN BUILT ENVIRONMENTS

Lingling Wang, Hanbin Luo, Ying Zhou, Cheng Zhou
Huazhong Univ. of Science and Technology, Wuhan, Hubei, China

Abstract

Precise reconstruction of the built environment is very useful for the management of the construction site. As far as the reconstruction of large-scale built environments is concerned, the reconstruction effect still needs to be further improved. Considering that most of the structures are piece-wise planar/linear in the built environment, this paper proposes a method for reconstructing the geometric structure of the scene that display its appearance precisely. This method focuses on reconstructing objects with plane and edge structures, such as buildings, to achieve the reproduction of their geometry. The paper introduces a new dense reconstruction algorithm, the patch based stereo matching algorithm to refine a sparse point cloud to produce a dense point cloud. This method further merges three-dimensional (3D) line into the dense point cloud to optimize the geometric line of the model. The experiment demonstrates that the improved method has a flawless reconstruction effect on the geometric primitives of buildings.

Introduction

The technique of three-dimensional (3D) reconstruction has become an important tool and is being widely used to develop infrastructure and landscape models to better manage cities and assets. This technique can be used for a variety purposes such as providing the spatial information needed to operate and maintain infrastructure (Kim et al., 2005), recognising structural components and monitoring their progress (Son et al., 2010), inspecting bridges (Lattanzi et al., 2014) and creating retrospective 'as-builts' of historic buildings (Yang et al., 2010; Arayici, 2007).

Currently, the precision of the reconstructed model based on the state-of-the-art technology has attained cm-level for the image-based modeling technique or mm-level laser scanning technology for simple scene or single object (Ma et al., 2018). Reconstructed scenes have been typically simple objects such as single-span bridge and a set of items on the desktop. For a large range of scenes with multiple building entities, the effect of model reconstruction needs to be improved. These models can well realize the visual maintenance and management of the construction site (Kim et al., 2005). In order to better reconstruct the built environment, a method for reconstructing its important geometric structure is proposed. Most of the structures of the built environment are piece-wise planar/linear (Raposo et al., 2014). If the edge and surface reconstruction precision can be improved, this can better display important elements such as buildings and roads in the built environment, which will be very useful for the application of 3D models.

Therefore, this paper proposes an improved algorithm for generating a point cloud to achieve a fast and precise reconstruction of the 3D structure of the built environment. Specifically, an improved dense reconstruction method, the patch based stereo matching algorithm, is introduced to quickly process images (especially high-resolution images) to generate a precise point cloud (Shen, 2013). Further, for built environment, this paper incorporates a 3D line extraction algorithm for dense point clouds to optimize the geometric structures of the model (Hofer et al., 2015).

The paper commences with a review of point cloud reconstruction method (Section 2). Newly developed algorithms for reconstructing 3D structure are then presented (Section 3). The algorithms are then tested and discussed using experiments (Section 4). Finally, the contributions and implications for future research are identified (Section 5).

Reconstruction of point cloud

The image-based 3D reconstruction method is a relatively popular technique for reconstructing the built environment and involves two key steps: (1) Restoring the structure from the motion (SFM) (Wu, 2013); and (2) Multi-view Stereo (MVS) (Jensen et al., 2014). The SFM technology restores the camera motion parameters and scene structure using low-level features (point features) from multi-view images to reconstruct sparse 3D scenes. Common feature extraction algorithms include Scale-invariant Feature transform (SIFT) (Chang et al., 2008), Speeded Up Robust Features (SURF) (Bay et al., 2008), and Fast Retina Keypoint (FREAK) (Alahi et al., 2012).

The sparse point cloud that is composed of seed points in areas with rich textures can only reflect the discrete state of the scene and therefore it must be densely reconstructed to improve its integrity and precision. This has been typically undertaken using three MVS based methods identified in Table 1.

Table 1: The methods for dense reconstruction

Method	Specific algorithm	Characteristics
Voxel-based	Voxel coloring framework (Seitz et al., 1999) Space Carving (Kutulakos et al., 2000) Graph-cut optimization (Cipolla et al., 2005)	A series of regularly stacked cubes or polygon surfaces
Feature point growing-based	Region growing (Otto et al., 1989) PMVS (Furukawa et al., 2009)	To expand sparse seed points to weakly textured areas
Depth-map merging-based	Window-based (Goesele et al., 2006) Markov Random Field optimization (Campbell et al., 2008) DAISY feature matching (Tola et al., 2012)	To find a depth map of each image, and merge them

Voxel-Based Methods include Voxel coloring algorithm, Space Carving algorithm and son on. These methods need to provide a better initial volume, such as Visual hull, otherwise the optimization may converge to a local minimum. Voxel resolution affects the accuracy of reconstruction. However, increasing the resolution of voxels will cause the storage space to grow at lot, which consumes a lot of computing resources. Nowadays, the most popular feature point growing-based methods is PMVS algorithm, which ranks in the forefront of dense reconstruction method in terms of model integrity and precision. However, for multi-planar scenes, images often exhibit a single, repeating texture so that the reconstructed model by PMVS algorithm will have many holes. The DAISY feature descriptor is an efficient tool for calculating depth maps in depth-map merging-based methods. It can quickly realize the reconstruction of high-resolution images. The depth-map merging-based methods is particularly suitable for the reconstruction of large scenes. The higher the resolution of the image is, the denser the reconstructed point cloud will be.

Although MVS-based reconstruction has been greatly developed, it still needs to be further improved. In the DTU standard database, a large number of clips, holes, etc. can be found in the point cloud of different scenes reconstructed by several popular algorithms (Furukawa et al., 2009; Tola et al., 2012).

Precise Reconstruction of 3D Structure

Akin with previous studies, pictures of the scene were collected from different angles to create a 3D model (Zheng et al., 2016). An SfM pipeline is performed to obtain the corresponding camera poses and a sparse point cloud. Based on the sparse point cloud, the patch based stereo matching algorithm is introduced to recover the details of the scene by generating additional points and eliminate noise. Then perform 3D line extraction from the sparse point cloud. Finally, the 3D line is merged with the dense point cloud to generate a complete point cloud of the scene for built environment. Figure 1 shows the generation process of the point cloud.

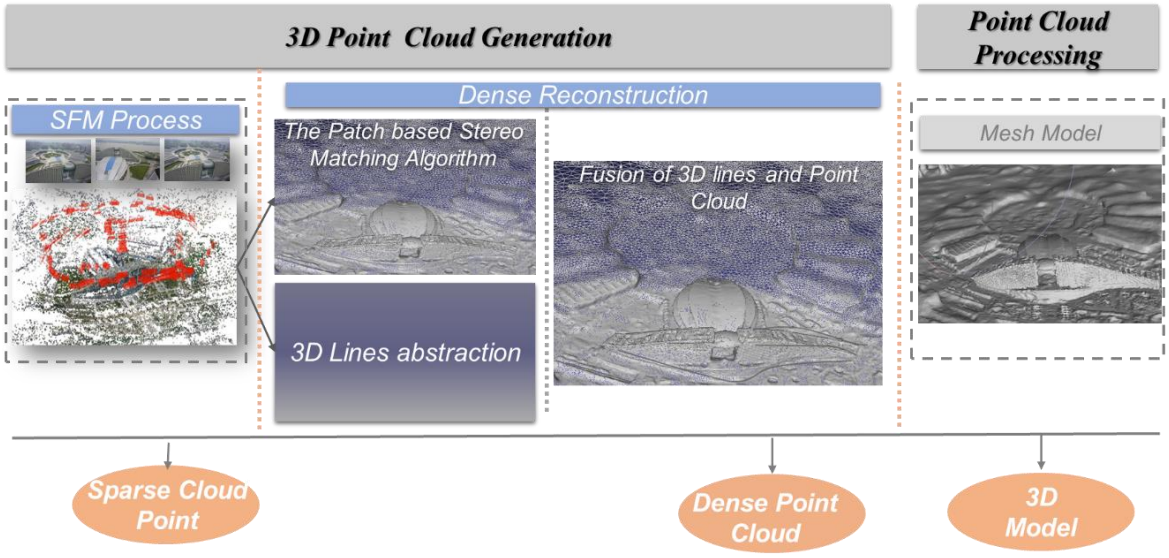


Figure 1: The improved point cloud generation process

For SfM pipeline, specifically, feature point extraction is performed using the SIFT algorithm. Then, match these feature points with its nearest k neighbours by using the k-d tree algorithm (Zhiliang et al., 2018). Then calculate the parameters of the camera for sparse reconstruction and finally adjust the camera's positional parameters to get a sparse point cloud.

Dense Reconstruction

After obtaining the internal and external parameters of the camera, the reconstructed data of the sparse 3D scene can be obtained, along with generating a sparse point cloud. The patch based stereo matching algorithm is introduced (Shen, 2013). This algorithm is based on the existence of a satisfactory spatial neighbourhood near the projection plane. Unlike the conventional method of obtaining decentralised, discrete depth maps (e.g., stereoscopic vision), a continuous depth map of a precision subpixel is reconstructed. The patch based stereo matching algorithm consists of four steps: stereo pair selection, depth-map computation, depth-map refinement, and depth-map merging. The patch based stereo matching algorithm spatial depth-map propagation algorithm could be easily parallelized at image level.

We first find its reference picture for each image. A good candidate reference image should have a similar viewing direction as the image, and have a suitable baseline. The baseline should not be too short, otherwise it will generate a lot of information redundancy. The baseline should not be too long to have less common coverage of the scene. After forming its image pair for each image, we calculate the depth map for the image, which is to calculate the depth for each pixel in the image.

Each pixel in the input image needs to find a good support plane which can minimize aggregated matching costs between the pixel of the image with the corresponding pixel in its reference image. The support plane is essentially a local tangent plane of the scene surface. A square window n*n centered on the pixel in the given image is used to calculate its support plane. The other pixels in the window on the image is projected onto the reference image R by homography. The photometric consistency of pixels between the image and its reference image can be compared by the Normalized Cross Correlation (NCC) (Yoo et al., 2009). We can find a patch plane with the lowest matching cost. Each pixel in the picture corresponds to a 3D plane. The initial 3D planes are then refined by slightly adjusting the parameters of the 3D planes, including its angle and position, to achieve a smaller matching cost. Finally, all depth maps are merged to obtain a dense point cloud.

Extraction of 3D Lines

The rich information features of line segments are often lost or unclear when models are re-built using point features. For example, straight line segments at the edges of a model are sometimes uneven and poorly represented (e.g. roof plane). Thus, the 3D line segments need to be extracted from the model.

In line with previous studies 3D lines are generated based on the 3D sparse model, containing a sparse set of 3D points $P=\{P_1, \ldots P_K\}$ and camera poses. A set of 2D line segments is found on each image, and each 2D line consists of two endpoints. Specifically, we use the line segmentation operator LSD to get the 2D line segment on each image (Gioi, 2012).

The 2D lines segments extracted in different images are matched to associate 2D line segments corresponding to the same 3D line on the object. We use epipolar matching constraints to establish a set of potential correspondences for each line segment individually (Hofer et al., 2014). To simplify the calculation of the match, we only match the 2D line segments on the image to the 2D line of its nearest neighbours. After these matches, we built a series of line segment correspondences.

For all matches of 2D line segments, we need to further rule out the wrong matches. We use a novel similarity measure based on positional- and angular reprojection errors between a 3D hypothesis and 2D segments. Specifically, for each line segment correspondence, find its corresponding 3D line. The 3D line is projected on other images to obtain projected line segments. The more similar the projected line is to the 2D line extracted on other images, the more likely the 3D line is to be seen by the image. If the 3D line can be seen on more images, the more reliable the 3D line is. Retain 2D line segment correspondences that form these reliable 3D lines. According to this rule, we can clarify whether the matching of a 2D line segment is credible.

A 2D line segment does not correspond to a 3D line hypothesis, but the 2D line on each image can only be a projection of a 3D structure. Therefore, we have to calculate the most likely 3D position for each 2D line. Based on the above calculations, we determine the position of each 2D line, which is the one that can be seen by most images and has the highest similarity among all the assumptions of the 3D line. We can select the most plausible correspondence for each 2D segment as its 3D position hypothesis. Finally, we perform clustering 2D segments based on their spatial proximity in 3D to obtain the final correspondence set and 3D model.

Fusion of 3D Lines and Point Cloud

After the 3D line is extracted, the reconstruction of the entire point cloud is achieved by fusing the 3D line with the dense point cloud. The method of fusion is to extract the points on the 3D line and draw a set of points. The number of points to be extracted needs to take into account the point density of the point cloud. The greater the density of dots, the more points are extracted from the 3D line.

Reconstructing the 3D point cloud after combining the 3D line reconstruction method and depth propagation algorithm significantly improves the quality of the model. On the one hand, 3D line reconstruction can show sharp straight-line features of some edges of the model. Dense reconstruction of point cloud by depth propagation algorithm, on the other hand, can often contain more modelling information that can correct inaccurate 3D lines.

The following steps are processing the point cloud, including the elimination of interference, surface mesh reconstruction and texture synthesis. The mesh reconstruction method initially performs a Delaunay triangulation on the dense point cloud (Fang et al., 1995). After constructing the Delaunay tetrahedron, the model surface is determined. Then use the plane-sweeping principle to remove hallucinated surfaces that are often related to missing cameras. That leads to cleaning the space and the hallucinated surfaces are further generated from sparsely distributed false positive points. To enhance the visualization, texture synthesis arises when a picture of an object is projected back on to the corresponding surface of a model.

Experiments and Validation

The proposed method in this paper could reconstruct the built environment. In order to test the reconstruction effect of the method on the geometric primitives of a built environment model, a set of residential buildings was selected as the experimental scene. Then the box girder scene was reconstructed to quantitatively evaluate the effects of the proposed method.

Visual effect

This experiment qualitatively shows that the proposed method can reconstruct large and complex scenes. Figure 2 shows a model of a group of housing construction scene. The proposed method can faithfully reproduce the building and its surroundings. Whether it is a building, a road, a car, or a tree, they can maintain a good shape and texture in the model. Except for some trees, their structure is too irregular to be fully expressed.

(a) The entire model (b) Small objects in the model, which are house, car, road & meadow, tree

Figure 2: 3D model of a group of housing construction scene

Quantitative Experiment

The proposed method is compared with some mature software packages or products, including Visual SFM + CMVS (Teeravech, 2013) and Autodesk Remake (the original 123d Catch) (https://www.autodesk.com/products/remake/overview) and Kinect and Laser scanning technology (Leica Scan Station P30 / P40) by a box girder scene. Figure 3 shows the models for all methods.

We can see the overall appearance of all models in Figure 3, and the models of the proposed method, Kinect and laser scanners are quite perfect. We further evaluate the quality of the point clouds using the following criteria: (1) point density (ρ); (2) average error (ε). The point density refers to the number of points in the unit cube. The density distribution of the points is counted by calculating the percentage of regions with the same number of point density. Average error refers to the difference between the point cloud generated by several methods and the

ground truth.

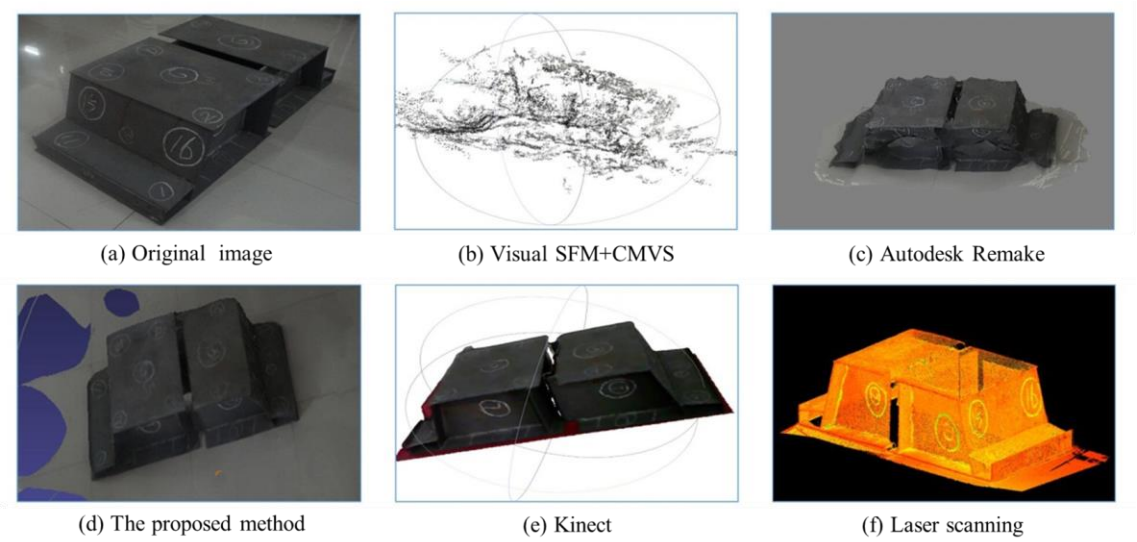


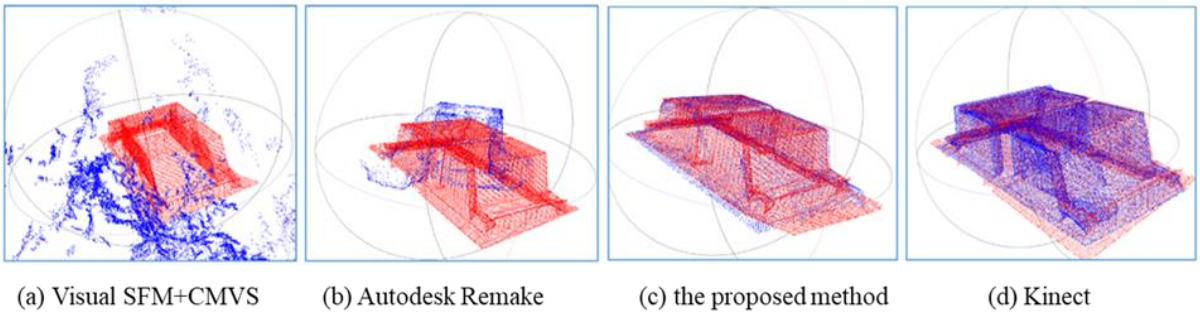
Figure 3: The reconstructed model

Laser scanning is currently a high-stability and high-accurate technology. The laser scanner selected in this experiment is FARO Focus3D X130 and its ranging error is ±2 mm with a resolution of up to 70 million pixels. The point cloud generated by laser scanning are taken as a ground truth. The average error ε are defined as (1).

$$\varepsilon=\sqrt{\frac{\sum_{i=1}^{n}(\Delta x_i^{\,2}+\Delta y_i^{\,2}+\Delta z_i^{\,2})}{n}} \quad (1)$$

Where n is the number of matching points. $\Delta x_i = x_i^* - x_i$, $\Delta y_i = y_i^* - y_i$, $\Delta z_i = z_i^* - z_i$ represents the spatial coordinates (x_i, y_i, z_i) of the laser point cloud minus the spatial coordinates (x_i^*, y_i^*, z_i^*) of the corresponding points in Visual SFM+CMVS/ Autodesk Remake/ Kinect and the proposed method.

Then the all point clouds are converted to the same coordinate system by registration alignment. The alignment results are shown in Figure 4. We calculate the point-to-point errors by comparing every other point cloud with the Laser point cloud.



Figures 4: The point cloud aligned to the laser scanning point cloud

Figure 5 shows the density distribution of their point cloud. Then we can calculate the average error, point density. Table 2 represents the average error, the point density reconstruction time and model type for the four methods (Visual SFM+CMVS, Autodesk Remake, and the proposed method and Kinect).

Table 2: Average error (ε), Point density (ρ), Reconstruction time and Model type for four methods

Method	ε (cm)	ρ (pt/cm^3)	Time (min)	Model type
Visual SFM+CMVS	Match failed	Invalid	3.5	Point cloud
Autodesk Remake	Match failed	Invalid	6.3	Mesh
The proposed method	6.3931	1.1515	8.5	Mesh
Kinect	6.5592	6.5169	6.1	Point cloud

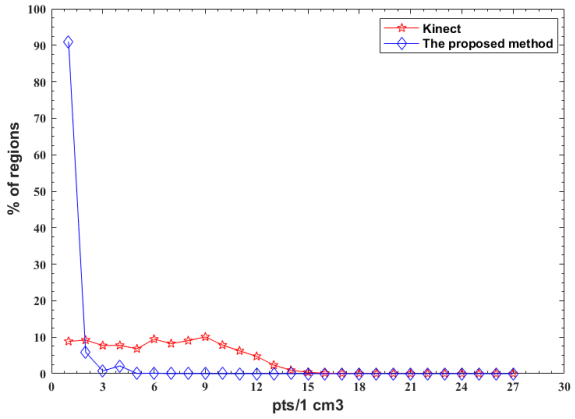


Figure 5: The point distribution of the point cloud for Kinect and the proposed method

As shown in Figure 4, the proposed method and the Kinect align well with the laser point cloud, while the others fail to do so. Except for this reason that there are not enough points for the point cloud generated by Visual SFM+CMVS, Autodesk Remake, many points in the point cloud are very different from the ground truth and have a big error.

In Figure 5, it can be seen that the point density of the point cloud generated by the proposed method is relatively sparse, concentrated at 1 pts/cm3. The number of points for point cloud generated by the Kinect is large and the density distribution of the points is quite different, and the point density varies from 1 pts/cm3 to 12 pts/cm3.

For these four models, the models generated by Autodesk Remake and the proposed method are mesh model. Generating mesh models (Autodesk Remake and the proposed method) takes longer than generating point cloud models (Visual SFM+CMVS and Kinect), because point clouds need to perform grid reconstruction and texture synthesis. Mesh model is usually able to meet more visualization purposes than the point cloud model. Of four methods, the proposed method shows the best precision.

In general, the proposed method outperforms the Visual SFM+CMVS, Autodesk Remake, and Kinect. The model generated by the proposed method has a small error and have a great visual effect.

Conclusions

Based on the existing image-based 3D reconstruction method, this paper improves the process of dense reconstruction, and focuses on improving the accuracy of geometric primitives (plane and line) reconstruction in built environments, which helps enlarge the applications of 3D reconstruction techniques in civil engineering. In the experiment, the approach presented in this paper is very good for achieving precise reconstruction of planes and lines. Compared with several existing methods, the reconstruction performance of the proposed method is also very good.

Regarding future work, we plan to use more means to test the performance of the proposed method. In this paper, the quantitative experiment was not carried out using the built environment scene, but was replaced by a box girder scene composed of lines and planes, which may lead to the advantages of the proposed method not shown. In addition, the comparison methods selected in this experiment are several easy-to-obtain methods, and the comparison between the proposed methods and the more advanced methods needs further study. We also research further to improve the reconstruction performance of geometrically irregular objects.

References

Alahi, A., Ortiz, R., & Vandergheynst, P. (2012) Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 510-517). Ieee.

Arayici, Y. (2007) An approach for real world data modelling with the 3D terrestrial laser scanner for built environment. *Automation in Construction*, 16(6), 816-829.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008) Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), 346-359.

Campbell, N. D., Vogiatzis, G., Hernández, C., & Cipolla, R. (2008) Using multiple hypotheses to improve depth-maps for multi-view stereo. In *European Conference on Computer Vision* (pp. 766-779). Springer, Berlin, Heidelberg.

Chang, Y., Lee, D. J., Hong, Y., & Archibald, J. (2007) Unsupervised video shot detection using clustering ensemble with a color global scale-invariant feature transform descriptor. *EURASIP Journal on Image and Video Processing*, *2008*(1), 860743.

Cipolla, G. V. P. T. R., Vogiatzis, G., & Torr, P. H. S. (2005) Multi-view stereo via volumetric graph-cuts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA (Vol. 2, pp. 391-398).

Fang, T. P., & Piegl, L. A. (1995) Delaunay triangulation in three dimensions. *IEEE Computer Graphics and Applications*, *15*(5), 62-69.

Furukawa, Y., & Ponce, J. (2009) Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8), 1362-1376.

Goesele, M., Curless, B., & Seitz, S. M. (2006) Multi-view stereo revisited. *In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 2, pp. 2402-2409). IEEE.

Hofer, M., Donoser, M., & Bischof, H. (2014) Semi-Global 3D Line Modeling for Incremental Structure-from-Motion. In *BMVC*.

Hofer, M., Maurer, M., & Bischof, H. (2015) Line3d: Efficient 3d scene abstraction for the built environment. In *German Conference on Pattern Recognition* (pp. 237-248). Springer, Cham.

Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., & Aanæs, H. (2014) Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 406-413).

Kim, C., Haas, C. T., & Liapi, K. A. (2005) Rapid, on-site spatial information acquisition and its use for infrastructure operation and maintenance. *Automation in Construction*, 14(5), 666-684.

Kutulakos, K. N., & Seitz, S. M. (2000) A theory of shape by space carving. *International journal of computer vision*, 38(3), 199-218.

Lattanzi, D., & Miller, G. R. (2014) 3D scene reconstruction for robotic bridge inspection. *Journal of Infrastructure Systems*, 21(2), 04014041.

Otto, G. P., & Chau, T. K. (1989) 'Region-growing'algorithm for matching of terrain images. *Image and vision computing*, 7(2), 83-94.

Raposo, C., Antunes, M., & Barreto, J. P. (2014) Piecewise-planar stereoscan: structure and motion from plane primitives. In European Conference on Computer Vision (pp. 48-63). Springer, Cham.

Seitz, S. M., & Dyer, C. R. (1999) Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2), 151-173.

Shen, S. (2013) Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. *IEEE transactions on image processing*, 22(5), 1901-1914.

Son, H., & Kim, C. (2010). 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. *Automation in Construction*, 19(7), 844-854.

Teeravech K. 2013. An introduction to 3D reconstruction using VisualSFM and PMVS2/CMVS. Remote Sensing and Geographic Information Systems, School of Engineering and Technology, Asian Institute of Technology.

Tola, E., Strecha, C., & Fua, P. (2012) Efficient large-scale multi-view stereo for ultra high-resolution image sets. *Machine Vision and Applications*, 23(5), 903-920.

Von Gioi, R. G., Jakubowicz, J., Morel, J. M., & Randall, G. (2012) LSD: a line segment detector. *Image Processing On Line*, 2, 35-55.

Wu, C. (2013) Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision-3DV 2013* (pp. 127-134). IEEE.

Yang, W. B., Chen, M. B., & Yen, Y. N. (2011) An application of digital point cloud to historic architecture in digital archives. *Advances in Engineering Software*, 42(9), 690-699.

Yoo, J. C., & Han, T. H. (2009). Fast normalized cross-correlation. *Circuits, systems and signal processing*, 28(6), 819.

Ma, Z., & Liu, S. (2018) A review of 3D reconstruction techniques in civil engineering and their applications. *Advanced Engineering Informatics*, 37, 163-174.

Zheng, M., Zhu, J., Xiong, X., Zhou, S., & Zhang, Y. (2016) 3D model reconstruction with common hand-held cameras. *Virtual Reality*, 20(4), 1-15.