INTEGRATION OF AERIAL, MMS, AND BACKPACK IMAGES FOR SEAMLESS 3D MAPPING IN URBAN AREAS

Zhaojin Li, Bo Wu*, Yuan Li

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong - bo.wu@polyu.edu.hk

KEYWORDS: 3D Mapping, Photogrammetry, Urban, Aerial, MMS, Backpack

ABSTRACT:

Photorealistic three-dimensional (3D) models play an indispensable role in the spatial data infrastructure (SDI) of a smart city. Recent developments in aerial oblique photogrammetry, and the popularity of terrestrial mobile mapping systems (MMSs) offer possibilities for deriving 3D models with centimeter-level accuracy in urban areas. Additionally, advances in image matching and bundle adjustment have allowed 3D models derived from the integration of aerial and ground imagery to overcome typical problems related to 3D mapping in urban areas (e.g., geometric defects, blurred textures on building façades). Nevertheless, this approach may not be suitable for all scenarios owing to innate differences between each platform. Besides, MMS images may not cover regions that cannot be reached by mobile vehicles in urban areas (e.g., narrow alleys, areas far from roads). Meanwhile, backpack systems have garnered attention from the photogrammetry community in recent years due to their flexibility, and regions neglected in previous works can be adequately reconstructed from images collected by backpack systems. This paper presents an approach for effectively integrating multi-source images collected by aerial, MMS, and backpack platforms for seamless 3D mapping in urban areas. The approach includes three main steps: (1) data pre-processing, (2) combined structure-from-motion, and (3) optimal generation of a textured 3D mesh model. The experimental results using aerial, MMS, and backpack datasets collected in a typical urban area in Hong Kong demonstrate the promising performance of the proposed approach. The described work is significant for boosting various types of imagery for integrated 3D mapping in both city scale and street level to facilitate various applications.

1. INTRODUCTION

With more profound recognition of 3D city data, great importance has been attached to 3D photorealistic city models because of their usage in many applications to meet the increasing demand for high geometric accuracy and improved texture (Biljecki et al., 2015; Qiao et al., 2010; Singh et al., 2013). Recent and rapid advances in the development of aerial oblique photogrammetry and unmanned aerial vehicle (UAV) platforms now enable 3D reconstruction of centimeter-level accuracy in large-scale urban areas (Vu et al., 2012; Ye and Wu, 2018).

With the advent of mobile mapping systems (MMSs), close-range photogrammetry based on MMS platforms has been widely used for 3D mapping and modeling applications in urban areas. Assisted by MMSs, recent endeavors to combine oblique aerial images and terrestrial images for improved 3D modeling (Wu et al., 2018) offer 3D building models with optimum geometric accuracy and texture. However, the flexibility of MMS imagery remains insufficient for acquiring information regarding locations that are inaccessible by vehicles. Textural blurs may also occur on MMS images when the vehicle moves at high speed. This problem can be addressed by wearable mapping solutions, such as backpack mapping systems, which have triggered increasing interest because of their flexibility in data collection. In particular, blind zones that cannot be reached by aerial or MMS images (e.g., alleys between tall buildings, regions far from roads) can be correctly reconstructed using imagery collected by backpack systems. Backpack images provide even closer observations of ground objects and enable reconstruction of road furniture or building façades with more detail.

MMSs and backpack systems can generally provide detailed information with flexibility; however, they are unable to offer comprehensive coverage and perspective in the way that aerial images do. An integration of aerial, MMS, and backpack images is therefore an ideal solution for seamless 3D mapping in urban areas, which is desirable in terms of both the scale range and high degree of details. Traditional photogrammetry or multi-view stereo solutions (Furukawa et al., 2010; Magerand and Del Bue, 2020; Schonberger et al., 2016; Westoby et al., 2012) generally include keypoint extraction, image matching, bundle adjustment (BA), dense image matching (DIM), triangulation of mesh models, and texture mapping. However, because of innate differences between the images, these existing solutions cannot be directly used for integrating aerial and terrestrial (including MMS and backpack) images.

Therefore, in this paper, we present an approach to effectively integrate aerial, MMS, and backpack images for optimal 3D reconstruction in large-scale urban areas. An initial data pre-processing step is performed to reduce the amount of computation and impact of moving objects and undesirable illumination conditions. The external orientation (EO) parameters of all of the images are then estimated using combined structure-from-motion (SfM) data. An optimal 3D mesh model is generated by removing geometry conflicts in point clouds generated by DIM (Hirschmuller, 2005; Ye and Wu, 2018) and texture mapping using selected images. Experiment analysis using aerial, MMS, and backpack images collected in Kowloon Bay, Hong Kong, was conducted to evaluate the performance of the proposed approach. Finally, conclusions were drawn and discussed based on the experiment results.

* Corresponding author. Email: bo.wu@polyu.edu.hk

2. INTEGRATION OF UAV, MMS, AND BACKPACK IMAGES FOR SEAMLESS 3D MAPPING

2.1 Overview of the Approach

The workflow of the proposed approach can be divided into three main steps: data pre-processing; combined SfM; and optimal generation of a textured 3D mesh model (Figure 1). The data pre-processing includes optimal image selection, color equalization, and removal of moving objects, which are designed to reduce the computational cost in the subsequent steps and reduce differences between images obtained from different sources. In the combined SfM, the EO parameters of the aerial and terrestrial images are first estimated separately to serve as initial values, which are subsequently refined through a combination of image matching and BA. To generate an optimal 3D mesh model, the DIM point clouds from the aligned images are first improved by detecting and modifying visibility conflicts. The refined point clouds are then used to generate a 3D mesh model and textures are mapped using the images selected in the pre-processed step.

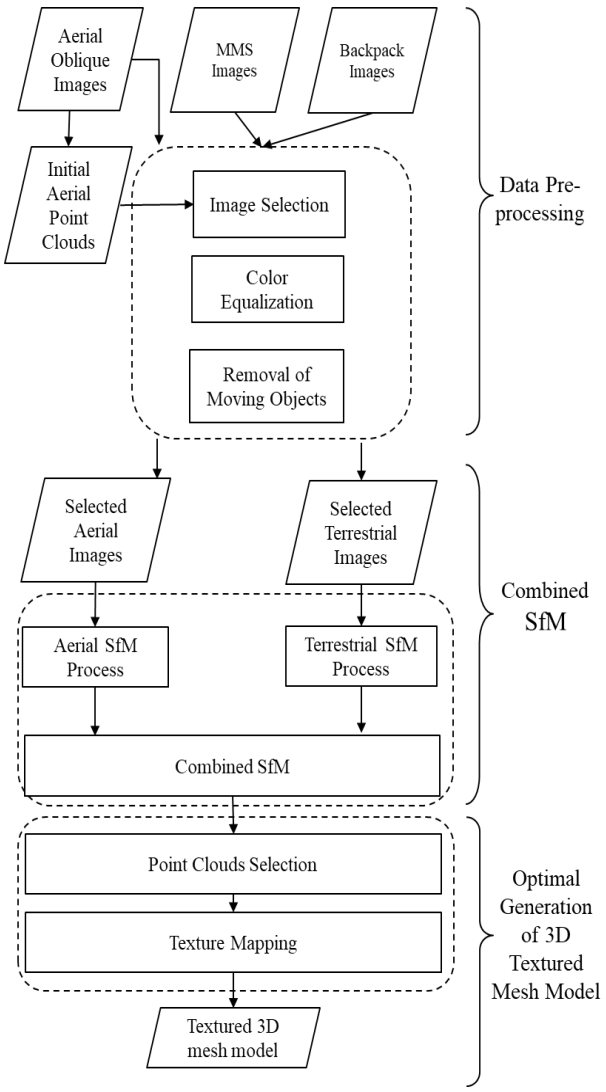

Figure 1. Overview of the proposed approach.

2.2 Data Pre-processing

2.2.1 Image Selection

The images in input datasets are generally unordered; hence, locating all of the image pairs can be exhaustive, especially in consideration of the large volumes of images. Optimal aerial and terrestrial images are therefore selected prior to further processing to minimize computational time and resources. Because aerial images cover a large area, relatively sparse point clouds (e.g., 1-m sample distance) are derived from the aerial images to provide referencing information for the image selection. Building façades and the corners of each plane are further extracted from the sparse aerial point clouds according to the random sample consensus (RANSAC) (Fischler and Bolles, 1981) plane fitting algorithm (Li et al., 2017). Optimal images are then selected for each plane based on the following criteria.

Because the characteristics of aerial and terrestrial images differ, their selection criteria are also not the same. For aerial images, three criteria should be met. The first criterion is to check whether the image was taken at a position where the plane is visible. The second criterion is that the projection of 3 out of 4 corners should be within the range of the image to guarantee that the image contains sufficient area of the façade. The third criterion requires that if a point is visible on an image, it must be the nearest point in its projection direction in the object space, which determines whether the façade is blocked by other buildings.

Three requirements are set for the terrestrial images. First, the distance between the center of the plane and image should be within a certain threshold. Second, the directions between the normal vectors of the plane and image are preferably parallel to each other to ensure that the image possesses a relatively good view. Third, at least a certain ratio (10% used in this study) of the point clouds should be able to be projected within the image range. The abovementioned requirements are progressive and ordered by the computational complexity. If the previous requirements are not met, those following are directly rejected.

2.2.2 Color Equalization

Images obtained from different datasets can differ substantially because illumination conditions vary with time and photosensitive elements of cameras can differ, both of which lead to some extent of mottled texture and difficulty in image matching. Therefore, histogram specification is conducted to reduce this kind of differences in the first place. Because aerial images cover a broad area, leveraging their color histogram to adjust the color of all of the other images usually leads to harmonious visualization effects. A comparison between the textured models without and with color equalization is shown in Figure 2, which reveals the significance of this step. As shown in Figure 2(a), the texture of the building façade displays a noticeable boundary between the aerial and terrestrial images, which is largely improved by applying color equalization.

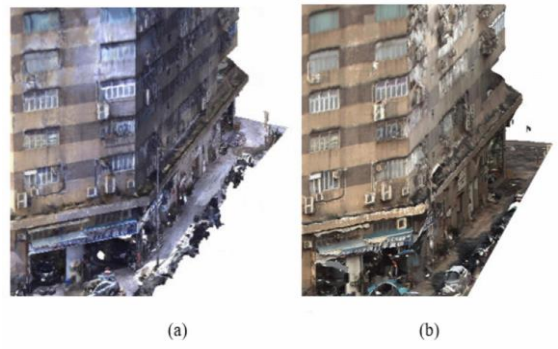

Figure 2. Comparison between mesh models generated without and with color equalization. (a) Mesh model textured with original images. (b) Mesh model textured with processed images.

2.2.3 Removal of Moving Objects

Even though MMS and backpack systems offer complementary views, they still present two main issues. First, the platform of MMS or backpack-fixed cameras must be easily photographed. Second, MMS and backpack systems must be close to the objects. Therefore, typical moving objects in urban scenes (e.g., vehicles, pedestrians) occupy large areas in the MMS and backpack images that cannot be ignored. To address these problems, masks are carefully defined to cover image regions of the mapping platform itself and regions with high possibility of moving objects. The regions within the masks are then ignored in the subsequent image matching step, as shown in Figure 3.

Figure 3. Illustration of the use of image masks. The first row shows original images and the second row shows masks on the images.

2.3 Combined Structure-from-Motion (SfM)

Due to obstruction or multiple reflections of positioning signals caused by densely distributed buildings (Chu and Chiang, 2016; Gruen et al., 2013), the accuracy of Global Navigation Satellite System (GNSS) positioning has been shown to be reduced in urban areas. The initial EO parameters obtained by GPS/IMU on board the MMS or backpack systems may therefore not be sufficiently accurate to generate 3D models. It is therefore imperative to refine their EO parameters by bridging with aerial oblique images through a rigorous mathematical model (Wu et al., 2018), such as SfM.

The first step in the SfM pipeline is image matching, which is based on the tie points between overlapped aerial and terrestrial images. However, due to large variations in view direction and image resolution, many state-of-the-art feature correspondence frameworks, such as affine Hessian and affine Harris (Mikolajczyk and Schmid, 2002), MSER (Matas et al., 2004), and ASIFT (Morel and Yu, 2009; Wang et al., 2018), fail to directly connect aerial and terrestrial images. Hence, in the proposed approach, feature correspondence is first performed on aerial and terrestrial images separately to obtain some initial tie points. Joint feature matching is then carried out to register the aerial images and terrestrial images (Hu and Wu, 2017) to apply constraints on the EO parameter refinement of the terrestrial images.

By capitalizing on the BA algorithm, the estimation of EO and internal orientation (IO) parameters of the images is cast as a non-linear problem to minimize re-projection error by optimally adjusting the positions of images and sparse points corresponding to the tie points (Wu et al., 2015). During this procedure, a small number of ground control points (GCPs) are used to estimate the transformation from the relative positions to the absolute coordinate systems through 3D similarity transformation. After BA, DIM point clouds are generated using dense image matching (Hirschmuller, 2005; Ye and Wu, 2018).

2.4 Optimal Generation of Textured 3D Mesh Models

There will be unavoidable outliers in the DIM point clouds generated from the previous steps due to possible wrong matches or other problems. The filtering of point clouds is therefore crucial to accurately recover geometric information. In our approach, the point clouds are first separated into aerial and terrestrial point clouds for cross-checking. For the terrestrial point clouds, a depth conflict test is implemented. However, this method does not work correctly on aerial point clouds because the rooftops of some low-lying constructions might also be removed. Hence, a different normal-vector-based method is used for the aerial dataset.

Two roughly filtered point clouds are then combined and further operations are conducted to remove redundant points based on the accuracy and spatial smoothness of the point clouds. Once the point clouds are properly selected, 3D mesh models are generated by the Poisson reconstruction algorithm (Bolitho et al., 2009; Kazhdan and Hoppe, 2013) and the textures are mapped using the images selected in the pre-processing stage (Section 2.2.1). Figure 4 illustrates the necessity of this point cloud filtering step.

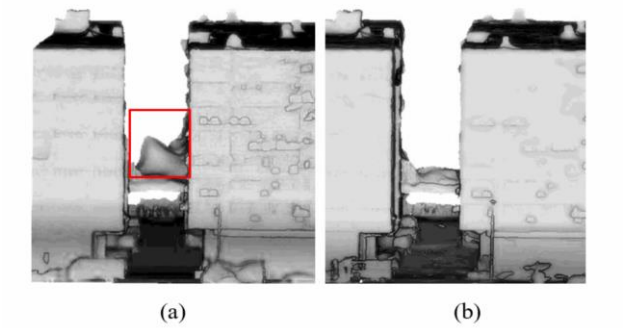

Figure 4. Illustration of point cloud filtering. (a) Mesh model generated without point cloud filtering. (b) Mesh model generated after point cloud filtering, of which the problematic area indicated by the red box in (a) has been removed.

3. EXPERIMENTAL EVALUATION

3.1 Dataset Description

In this paper, three challenging datasets acquired in Kowloon Bay, Hong Kong that cover a built-up urban area of over 40,000 m^2 to evaluate the performance of the proposed approach. As shown in Table 1, the datasets comprise images collected from three different platforms and are divided into three blocks, including a UAV image block, an MMS image block, and a backpack image block. A total of 121 aerial oblique images were obtained using a set of UAV-borne AMC PanOblique cameras, together with 12 Canon EOS 5DS R cameras. The ground sampling distance (GSD) for the aerial image is about 6 cm. In the MMS block, 1895 images were collected by a Leica MMS system, which comprises six cameras. The other 4202 backpack images were collected by the state-of-the-art Leica Pegasus Backpack system, which consists of five cameras. The GSDs of these terrestrial images are about 1 cm. Representative aerial oblique, MMS, and backpack images are shown in Figure 5. The distribution map of these datasets is shown in Figure 6. The backpack dataset contains abundant images of alleys and has the highest usable image ratio.

Together with the images and initial EO parameters, calibrated IO elements are also available. In addition, 70 evenly distributed GCPs in the Hong Kong 1980 (EPSG:2326) spatial reference system are used for control and checking purposes.

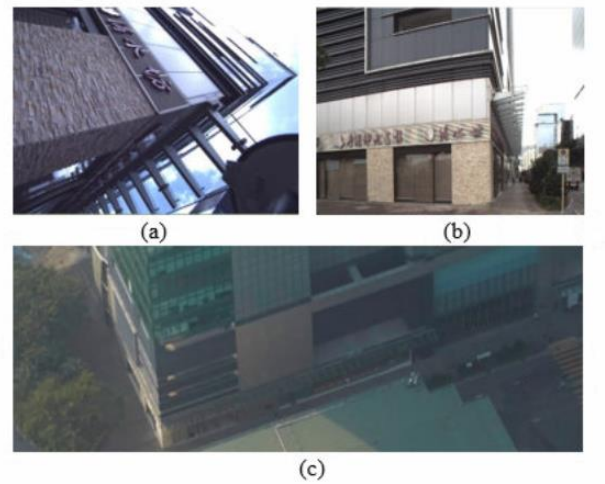

Figure 5. Representative images of the test area in Kowloon Bay, Hong Kong. (a) Backpack image, (b) MMS image, and (c) UAV image.

Figure 6. Experimental area and image distribution. The red and blue lines indicate the trajectories of the MMS and backpack systems, respectively.

3.2 Experimental Results

In this dataset, the MMS laser scanning point clouds georeferenced based on the integrated BA of the aerial and MMS images are used as the ground truth to evaluate the geometrical accuracy. The unsigned cloud-mesh-distance (CMD) is calculated in CloudCompare software and the results are ramped from 0 to 1 m, as shown in Figures 7–9, where blue represents 0 and red represents 1. The average unsigned CMD decreases from >0.5 to 0.1 m in some challenging parts after adding MMS and backpack images. Figures 7–9 show three representative regions, including an alley, a typical building façade, and the bottom part of a tall building.

In the alley region (Figure 7), the unsigned CMD of the mesh model derived from the UAV images approaches 1 m at the end of the alley, which is covered by a rooftop. This area is too deep and narrow to be accessed by the MMS and the perspective range is highly limited, leading to a slight improvement at the beginning of the alley, as shown in the middle part of Figure 7. Owing to the comprehensive views offered by backpack images, all of the alley area is recovered with a low average CMD of about 0.1 m. For the building façade shown in Figure 8, the average CMD decreases from >0.5 to 0.25 m when leveraging the MMS images and is further reduced to 0.1 m after the backpack images are integrally used. However, even with the assistance of backpack images, the bottom part of a tall building shown in Figure 9 remains poorly reconstructed. There are two main reasons for this. First, this building façade is made of glass that strongly reflects and transmits sunlight, which makes it hard to accurately conduct dense image matching (see example in Figure 10). Second,

Dataset	Sensor	Sensor Number	Focal length (mm)	GSD	Image Size (pixels)	Number of Images	Collection Date	Coverage (m^2)
UAV	Canon EOS 5DS R	12	49/35	6 cm	8688×5792	121	07/11/2016 02/12/2016	744,765
MMS	Leica MMS	6	8	1 cm	2048×2048	1895	05/07/2019	142,256
Backpack	Leica Pegasus Backpack	5	6	1 cm	2046×2046	4202	05/07/2019	43,260

Table 1. Information of the dataset used for experimental evaluation.

several backpack images obtained in this area only capture glass surfaces, which results in the failure to match backpack images with MMS and UAV images. Thus, the results in the second and third columns shown in Figure 9 are nearly the same.

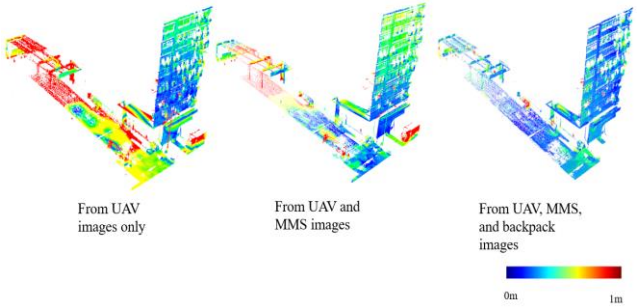

Figure 7. Unsigned CMD of an alley area calculated based on the model from UAV images only, UAV and MMS images, and UAV, MMS, and backpack images.

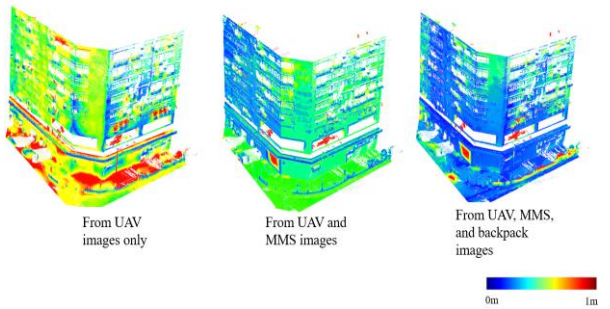

Figure 8. Unsigned CMD of a typical building façade based on the model from UAV images only, UAV and MMS images, and UAV, MMS, and backpack images.

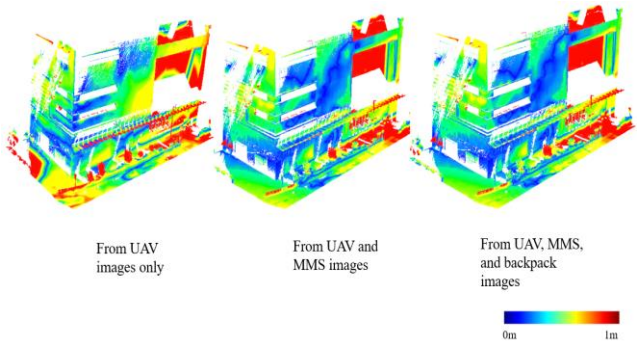

Figure 9. Unsigned CMD of the bottom part of a tall building based on the model from UAV images only, UAV and MMS images, and UAV, MMS, and backpack images.

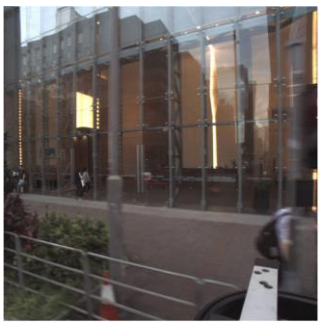

Figure 10. An example of a glassed façade.

Figure 11 shows the overall reconstruction results, which indicate the promising performance of the proposed approach. Figure 12–Figure 14 compare three challenging scenarios: the first columns show the mesh model generated from UAV images only; the second columns show mesh models generated from UAV and MMS images; and the third columns show mesh models generated from UAV, MMS, and backpack images. Even though the UAV images can be leveraged to efficiently reconstruct the scene, detailed street-level information is lost. Additionally, due to the high shooting elevation and occlusions, the textures of regions close to the ground (e.g., bottom parts of buildings) are blurred. When adding the MMS images, the situation of the building bottom is largely improved and the traffic signs are properly mapped. However, information from narrow alleys remains limited without apparent improvement. This problem is solved by including backpack images, which clearly show the boundaries of street objects and plants.

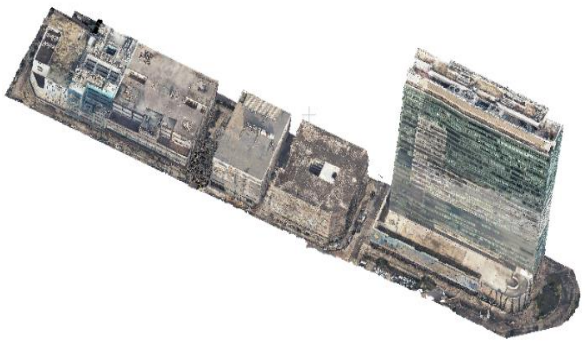

Figure 11. 3D mesh model generated from the integration of aerial, MMS, and backpack images.

Figure 12. Comparison of the 3D models in a narrow alley where MMS images were obtained in the alley.

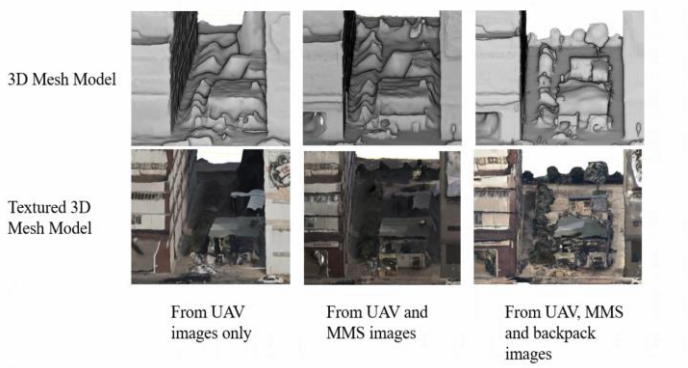

Figure 13. Comparison of the 3D models in another narrow alley where the MMS only obtained images at its two ends.

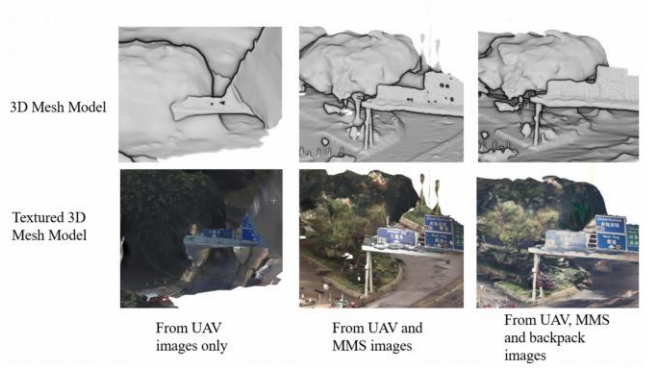

Figure 14. Comparison of 3D models of traffic signs.

4. CONCLUSION AND DISCUSSION

In this paper, we present an effective approach for integrating aerial, MMS, and backpack images for seamless 3D mapping in urban areas. Possible adverse effects are reduced by geometry-based image selection, color equalization, and removal of moving objects. A combined SfM workflow is presented to register the aerial and terrestrial images, followed by point cloud generation and selection, which aims to offer an accurate mesh model for the texture mapping stage.

The experimental results based on actual datasets collected by aerial, MMS, and backpack platforms covering a typical urban area in Hong Kong indicate that the proposed approach can provide improved 3D mapping in large-scale urban areas using integrated multi-resource images. This work is significant to boost various types of imagery for 3D mapping in both city scale and street level to facilitate various applications such as urban planning, urban highways management, and smart city development.

ACKNOWLEDGMENTS

This work was supported by grants from The Hong Kong Polytechnic University (Project No. 1-ZVN6) and the National Natural Science Foundation of China (Project No. 41671426). The authors would also like to thank the Survey and Mapping Office of the Lands Department of the HKSAR government for providing the experimental dataset.

REFERENCES

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Coltekin, A., 2015. Applications of 3D City Models: State of the Art Review. Isprs Int Geo-Inf 4, 2842-2889.

Bolitho, M., Kazhdan, M., Burns, R., Hoppe, H., 2009. Parallel Poisson Surface Reconstruction. Advances in Visual Computing, Pt 1, Proceedings 5875, 678-+.

Chu, C.H., Chiang, K.W., 2016. The Performance of a Tight Ins/Gnss/Photogrammetric Integration Scheme for Land Based Mms Applications in Gnss Denied Environments. Xxiii Isprs Congress, Commission I 41, 551-557.

Fischler, M.A., Bolles, R.C., 1981. Random Sample Consensus - a Paradigm for Model-Fitting with Applications to Image-Analysis and Automated Cartography. Commun Acm 24, 381-395.

Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R., 2010. Towards Internet-scale Multi-view Stereo. 2010 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), 1434-1441.

Gruen, A., Huang, X.F., Qin, R.J., Du, T.W., Fang, W., Boavida, J., Oliveira, A., 2013. Joint Processing of Uav Imagery and Terrestrial Mobile Mapping System Data for Very High Resolution City Modeling. Uav-G2013, 175-182.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. Proc Cvpr Ieee, 807-814.

Hu, H., Wu, B., 2017. Bound-Constrained Multiple-Image Least-Squares Matching for Multiple-Resolution Images. Photogramm Eng Rem S 83, 668-678.

Kazhdan, M., Hoppe, H., 2013. Screened Poisson Surface Reconstruction. Acm T Graphic 32.

Li, L., Yang, F., Zhu, H.H., Li, D.L., Li, Y., Tang, L., 2017. An Improved RANSAC for 3D Point Cloud Plane Segmentation Based on Normal Distribution Transformation Cells. Remote Sens-Basel 9.

Magerand, L., Del Bue, A., 2020. Revisiting Projective Structure from Motion: A Robust and Efficient Incremental Solution. Ieee T Pattern Anal 42, 430-443.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image Vision Comput 22, 761-767.

Mikolajczyk, K., Schmid, C., 2002. An affine invariant interest point detector. Lect Notes Comput Sc 2350, 128-142.

Morel, J.M., Yu, G.S., 2009. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. Siam J Imaging Sci 2, 438-469.

Qiao, G., Wang, W.A., Wu, B., Liu, C., Li, R.X., 2010. Assessment of Geo-positioning Capability of High Resolution Satellite Imagery for Densely Populated High Buildings in Metropolitan Areas. Photogramm Eng Rem S 76, 923-934.

Schonberger, J.L., Zheng, E.L., Frahm, J.M., Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. Computer Vision - Eccv 2016, Pt Iii 9907, 501-518.

Singh, S.P., Jain, K., Mandla, V.R., 2013. Virtual 3d City Modeling: Techniques and Applications. Int Arch Photogramm 40-2-W2, 73-91.

Vu, H.H., Labatut, P., Pons, J.P., Keriven, R., 2012. High Accuracy and Visibility-Consistent Dense Multiview Stereo. Ieee T Pattern Anal 34, 889-901.

Wang, C., Chen, J., Chen, J., Yue, A., He, D., Huang, Q., Zhang, Y., 2018. Unmanned aerial vehicle oblique image registration using an ASIFT-based matching method. J Appl Remote Sens 12.

Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. 'Structure-from-Motion' photogrammetry:

A low-cost, effective tool for geoscience applications. Geomorphology 179, 300-314.

Wu, B., Tang, S.J., Zhu, Q., Tong, K.Y., Hu, H., Li, G.Y., 2015. Geometric integration of high-resolution satellite imagery and airborne LiDAR data for improved geopositioning accuracy in metropolitan areas. ISPRS journal of photogrammetry and remote sensing 109, 139-151.

Wu, B., Xie, L.F., Hu, H., Zhu, Q., Yau, E., 2018. Integration of aerial oblique imagery and terrestrial imagery for optimized 3D modeling in urban areas. ISPRS journal of photogrammetry and remote sensing 139, 119-132.

Ye, L., Wu, B., 2018. Integrated Image Matching and Segmentation for 3D Surface Reconstruction in Urban Areas. Photogramm Eng Rem S 84, 135-148.