

THE FEASIBILITY OF 3D POINT CLOUD GENERATION FROM SMARTPHONES

N. Alsubaie ^{a,*}, N. El-Sheimy ^a

^a Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, Canada - (nalsubai, elsheimy)@ucalgary.ca

Commission V, WG V/4

KEY WORDS: Smartphones, Dense Matching, Mobile Mapping System, Closed Range Photogrammetry, Linear Features Constraints

ABSTRACT:

This paper proposes a new technique for increasing the accuracy of direct geo-referenced image-based 3D point cloud generated from low-cost sensors in smartphones. The smartphone's motion sensors are used to directly acquire the Exterior Orientation Parameters (EOPs) of the captured images. These EOPs, along with the Interior Orientation Parameters (IOPs) of the camera/ phone, are used to reconstruct the image-based 3D point cloud. However, because smartphone motion sensors suffer from poor GPS accuracy, accumulated drift and high signal noise, inaccurate 3D mapping solutions often result. Therefore, horizontal and vertical linear features, visible in each image, are extracted and used as constraints in the bundle adjustment procedure. These constraints correct the relative position and orientation of the 3D mapping solution. Once the enhanced EOPs are estimated, the semi-global matching algorithm (SGM) is used to generate the image-based dense 3D point cloud. Statistical analysis and assessment are implemented herein, in order to demonstrate the feasibility of 3D point cloud generation from the consumer-grade sensors in smartphones.

1. INTRODUCTION

The demand for dense 3D point clouds has been increasing over the past decade. This demand is to satisfy a variety of applications, including 3D object reconstruction and mapping. In general, 3D point clouds can be acquired through two different remote sensing systems; active and passive. An active system has the ability to acquire a precise and reliable 3D point cloud of an object directly (i.e., with a laser scanner). However, this system is expensive when compared to a passive system. Passive systems have the ability to acquire and reconstruct a 3D of an object from a set of overlapping images using digital cameras along with the knowledge of the EOPs of these captured images, the IOPs of the involved camera, and the corresponding points in the overlapping images. The IOPs can be obtained from the camera calibration process, while EOPs can be obtained using one of two geo-referencing approaches in photogrammetry; *indirect* and *direct*. The difference between the two depends on how EOPs are determined, which involves the *position* and *orientation* of the involved images. The indirect approach uses a set of control points to determine EOPs, while the direct approach uses on-board GPS/INS position and orientation systems for calculating EOPs at the time of exposure, using Mobile Mapping System (MMS) (El-Sheimy, 2008).

Today's smartphones are getting ever more sophisticated and smarter reaching to close the gap between computers and portable tablet devices (such as iPad). Current generation of smartphones are equipped with micro-electro mechanical systems (MEMS) based navigation sensors (such as gyroscopes, accelerometers, magnetic compass, and barometers), offering the potential for integrating these sensors with GPS for outdoor applications (e.g. the *iPhone6* integrates 3-accelerometers, 3-gyroscopes, pedometer, compass, barometer, step detector, and step counter). These sensors are needed for a direct geo-referenced system, as illustrated in Figure 2. Furthermore, nowadays smartphones are all equipped with high resolution digital cameras at the same resolution of current land-based MMS and thus allow mapping at large operational range (e.g. 100m). As a result, smartphones are now getting quite popular for data collection projects including point data mapping tasks and it is not far before we see smartphones are adopted as mobile

mappers and thus the importance of this paper. The objective of this paper is to explore the feasibility of using consumer-grade smartphones for direct geo-referenced 3D point cloud generation from overlapping imagery.

2. RELATED WORK

2.1 MMS and Smartphone

MMS started in 1991 as land-based systems with the GPSVanTM developed by the Center for Mapping at Ohio State University (Ellum, 2001). This system integrated a code-only GPS receiver, two digital CCD cameras, two colour video cameras, two gyroscopes, and an odometer. The sensors were mounted on a van, and coordinated to calculate the position of an object relative to the vehicle. The relative accuracy of this system was within approximately 10 cm, while the absolute accuracy was between one to three meters. The GPSVanTM successfully showed how land-based multi-sensor systems improve the efficiency of GIS and mapping data collection. The absolute accuracy of the object space points, however, was too poor for many applications, especially when compared with competing technologies, such as total station. Furthermore, the dead-reckoning sensors in the GPSVan were not suitable for bridging GPS signal outages.

The VISAT system was subsequently developed in order to obtain more accurate mapping results. This system was developed by the Department of Geomatics Engineering at the University of Calgary. The VISAT system consists of eight digital cameras, a dual frequency carrier phase differential GPS, and a navigation grade IMU, used for improving the accuracy of the mapping solution during GPS signal outages. The relative accuracy of this system was within 0.1m, with an absolute accuracy of 0.3 meters (El-Sheimy, 1996; Schwarz and El-Sheimy, 1996).

Smartphone motion sensors and consumer-grade cameras, in principle, are the essential elements of MMS. Al-Hamad (2014a) introduced the "Mobile Mapping Using Smartphones" method, where smartphone sensors (i.e., camera, GPS, IMU, and Magnetometer) are used as MMS. The initial position and orientation of the camera at the time of exposure was determined

using low-cost GPS and motion sensors. These position and orientation measurements were used within a bundle adjustment as initial values for EOPs, and were then corrected using imaging techniques (i.e., epipolar geometry). The Speeded Up Robust Features (SUFT) algorithm was used to determine automatic matched points between each two consecutive images. The epipolar line, along with these matched points, was then used as a constraint to enhance the relative position and orientation of each captured image with respect to the first captured image. Although the relative position and orientation accuracy of each image increased, because the first image had not been corrected, and because it relied on the poor accuracy of the GPS, the mapping result were shifted by the error in the EOP of the first image. Consequently, some control points were used to rescale the mapping solution (Al-Hamad et al., 2014b). Al-Hamad et al. (2014c) suggested possible methods for solving the shifted solution problem, such as integrating the absolute and relative navigation systems (GPS/IMU), or developing a system using non-linear least square estimator to correct initial IOP and EOP values. Although this work was implemented in a small testing situation, it successfully illustrated that “Mobile Mapping Using Smartphones” is a promising low cost solution with the potential to expand the range of MMS technologies by creating new opportunities for low-cost and easy area coverage, into research fields.

2.2 Image-based 3D point cloud Generation

Generally, dense matching techniques can be divided into three types: Local, Global, and Semi-Global Matching (SGM). The difference between these techniques centres on the method used to locate corresponding pixels in image stereo-pairs. The local disparity optimization technique is based on finding the best matched pixel using a Winner-takes-all (WTA) strategy, and does not operate using images with uniform areas (He et al., 2015). Several global methods have been deployed to overcome the inherent disadvantages of the local technique such as Graph-Cut based (Boykov et al., 2001), Belief Propagation (Sun et al., 2003), and Dynamic Programming (Forstmann et al., 2004), but these methods are fairly inefficient in terms of computational time. The SGM technique provides a trade-off between the two previous techniques and is based on minimizing the matching cost along several 1D directions in the image. This technique generally follows four steps: 1) matching cost computation; 2) cost aggregation; 3) disparity map optimization, and; 4) refinement (Hirschmüller, 2005, 2008). The SGM technique is used in this paper for generating the 3D point cloud.

3. METHODOLOGY

3.1 Direct geo-referencing using a smartphone:

The direct geo-referencing procedure is conducted using smartphone motion sensors and a low-cost camera, as shown in Equation (1), and illustrated in Figure 1:

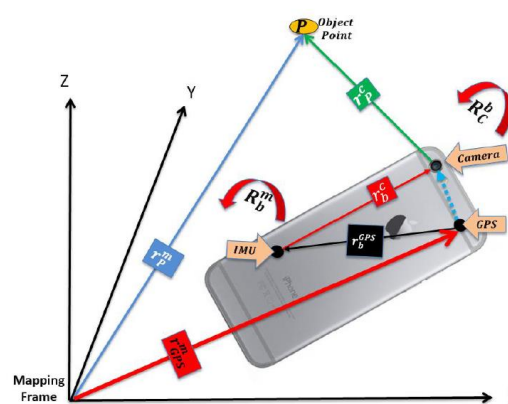


Figure 1. Position of the object point with respect to the mapping system

$$\mathbf{r}_p^m = \mathbf{r}_{GPS}^m(t) + \mathbf{R}_b^m(t) * (\mathbf{S}^p * \mathbf{R}_c^b * \mathbf{r}_p^c + \mathbf{r}_b^c - \mathbf{r}_b^{GPS}) \quad (1)$$

Where \mathbf{r}_p^m and \mathbf{r}_{GPS}^m are the object point and GPS or (INS) positions in the mapping coordinate frame, respectively, (t) is time, $\mathbf{R}_b^m(t)$ is the rotation matrix between the IMU (body) frame and the mapping frame, \mathbf{r}_b^c is the lever arm, a position vector between the IMU-body unit and the camera, and \mathbf{r}_b^{GPS} a position vector from the IMU-body unit to the GPS receiver. \mathbf{S}^p is the scale factor between the camera and the mapping coordinate systems for each point, \mathbf{R}_c^b is the boresight angle, the rotation matrix between the camera and IMU coordinate systems, and \mathbf{r}_p^c is the coordinate vector of the object point image coordinates in the camera frame.

3.2 Data collection app

Apple app software has been developed to capture and synchronize images with motion sensors at the time of exposure, as shown in Figure 2:

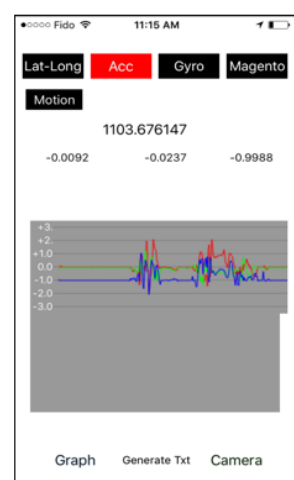


Figure 2. Developed Apple App

3.3 Workflow

As shown in Figure 3, the workflow of the proposed methodologies begins with the calibration of the smartphone's camera in order to determine its IOPs. Concurrent to this initial calibration, the collection and synchronization of captured images using their position and orientation measurements at time

of exposure, using the low-cost motion sensors from the smartphone. Consequently, the initial EOPs of each image are determined directly

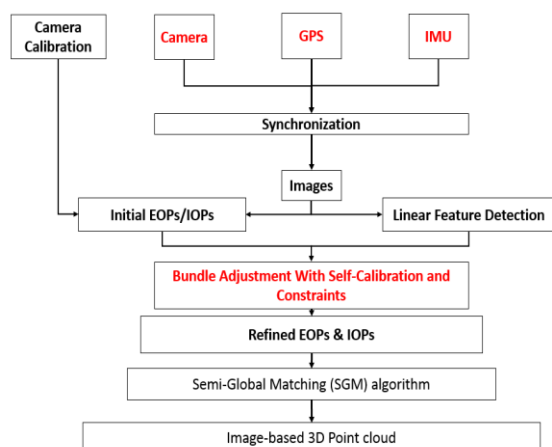


Figure 3. Methodologies Chart

3.4 Camera Calibration

Most smartphone uses cameras with unstable IOPs, consequently, the stability of this camera must be tested frequently and compared frequently with a reference value. Therefore, robust calibration procedure is required before using the phone in mapping, in order to obtain ideal IOP values. The camera in the iPhone 6 is calibrated in the laboratory using test fields (coded targets) with known ground control and tie points, as shown in Figure 4 (Lari et al., 2014) through the collinearity conditions (Equations (2-3)) resulting in the IOPs listed in Table 2. The IOP values, especially the focal length, are very close to the values provided in the *iPhone-6* manual.



Figure 4. Coded Targets

IOPs Parameters	(mm)
x_p Principle point	0.0723
y_p Principle point	0.0077
c focal length	4.2029
k_1 radial lens distortion	0.00533
k_2 radial lens distortion	-0.000406
P_1 decentric lens distortion	0.001376
P_2 decentric lens distortion	0.00092
Pixel size	0.0015

Table 1. IOPs Result

3.5 3D Object Reconstruction (Collinearity Equations)

Collinearity equations (conditions) represent the functional (mathematical) mode, exploiting the general relationship between the image and the ground coordinate systems. Therefore, the image point, the object point, and the perspective center of the camera, must lie on a line in the 3D space in order to fulfil this condition of collinearity (Mikhail et al., 2001; Wolf and Dewitt, 2000). The collinearity concept is illustrated in Figure 5 and represented by equation.

$$x_a = x_p - c \frac{r_{11}(X_A - X_0) + r_{12}(Y_A - Y_0) + r_{13}(Z_A - Z_0)}{r_{31}(X_A - X_0) + r_{32}(Y_A - Y_0) + r_{33}(Z_A - Z_0)} + \Delta x \quad (2)$$

$$y_a = y_p - c \frac{r_{21}(X_A - X_0) + r_{22}(Y_A - Y_0) + r_{23}(Z_A - Z_0)}{r_{31}(X_A - X_0) + r_{32}(Y_A - Y_0) + r_{33}(Z_A - Z_0)} + \Delta y \quad (3)$$

Where c is the perspective distance of the camera. The image coordinates of the principal point are x_p, y_p ; the image coordinates of the object point are x_a, y_a ; the perspective centre ground coordinates are X_0, Y_0, Z_0 , and the object point ground coordinates are X_A, Y_A, Z_A .

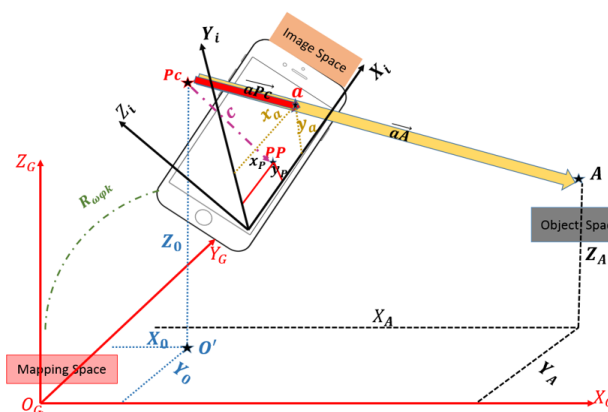


Figure 5. Collinearity Equation Concept

3.6 Mathematical Model (Bundle Adjustment)

The Bundle adjustment technique is based on the non-linear least square method, used to estimate and recover object point coordinates, EOPs, and occasionally IOPs. Equations (4, 5, and 6) show the Gauss Markov observation model of the non-linear least squares method, demonstrating the general relationship between a set of observation quantities, and the unknown parameters.

$$\delta z = A \delta x + v \quad (4)$$

Where z is the observation vector, A is the design matrix describing the mathematical relation between the observations and the unknown's vector, x is the unknown vector, and v is the measurements error vector.

Given that the bundle adjustment is not a linear problem, the final least square equation can be expressed using Equation (5)

$$\delta \hat{x} = (A^T R^{-1} A)^{-1} A^T R^{-1} \delta w \quad (5)$$

Where $\delta \hat{x}$ is a correction for the expected unknown's vector, (A) is the design matrix, derived by the partial derivative of the measurements with respect to the unknowns, and δw is the misclosure vector, derived by the difference between the

measured and the expected observations using the expected unknown vector. Equation (5) can be written in a more general structure as shown in Equation (6).

$$\delta \hat{\mathbf{x}} = \mathbf{N}^{-1} * \mathbf{u} \quad (6)$$

Where \mathbf{N} is the Normal matrix, and (\mathbf{u}) is known as the Normal vector.

The unknown's parameters can be split into two homogeneous groups, namely: EOPs denoted with subscript (EOPs) and object points (OP) parameters, denoted with subscript (OP), as shown in Equation (7).

$$\begin{pmatrix} \delta \hat{\mathbf{x}}_{\text{EOPs}} \\ \delta \hat{\mathbf{x}}_{\text{OP}} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{EOPs}} & \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{OP}} \\ \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{EOPs}} & \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{OP}} \end{pmatrix}^{-1} * \begin{pmatrix} \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{w} \\ \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{w} \end{pmatrix} \quad (7)$$

3.6.1 Vertical and Horizontal Linear Features Constraints in Bundle Adjustment: Geometric information in the images can be used to enhance coordinate determination and to achieve a higher quality, more reliable, solution. A common example of geometric information are straight lines, either vertically or horizontally. The constraints for two points of line can be expressed using Equation (8) (El-Sheimy, 1996).

$$\begin{aligned} \mathbf{X}_i - \mathbf{X}_j &= \mathbf{Y}_i - \mathbf{Y}_j = \mathbf{0} \\ \mathbf{Z}_i - \mathbf{Z}_j &= \mathbf{0} \end{aligned} \quad (8)$$

Where i, j are any two points on straight line.

As illustrated in Figure 6, the only change between any two points along the blue vertical line is in height, while (x, y) dimensions for each point are similar. Similarly, any two points along the red horizontal line have the same value for height, while the (x, y) dimensions for each point are different.



Figure 6. Straight Linear Feature Constraints

These linear features are independently determined observations, capable of being added to the system equations as constraints in the Normal matrix. These linear features conditions are denoted with subscript c, and are added to the object space group as they are measured in the object space domain. Therefore, Equation (7) can be rearranged as follows:

$$\begin{pmatrix} \delta \hat{\mathbf{x}}_{\text{EOPs}} \\ \delta \hat{\mathbf{x}}_{\text{OP}} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{EOPs}} & \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{OP}} \\ \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{EOPs}} & \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{A}_{\text{OP}} + \mathbf{A}_c^T \mathbf{R}^{-1} \mathbf{A}_c \end{pmatrix}^{-1} * \begin{pmatrix} \mathbf{A}_{\text{EOPs}}^T \mathbf{R}^{-1} \mathbf{w} \\ \mathbf{A}_{\text{OP}}^T \mathbf{R}^{-1} \mathbf{w} + \mathbf{A}_c^T \mathbf{R}^{-1} \mathbf{w}_c \end{pmatrix} \quad (9)$$

3.6.2 Free Network Adjustment: Since the smartphone uses low cost GPS and INS sensors, EOPs obtained by this system need to be corrected inside the bundle adjustment. Hence, if the bundle adjustment is performed without GCPs, then the datum seven parameters need to be defined using alternative method. Therefore, a free bundle-adjustment procedure is used to overcome the problem of datum deficiency, where the inner constraint matrix is used to remove the rank defect of the Normal matrix (Granshaw, 1980) using the estimated initial ground coordinate of the tie point's values as shown in Equation (10). Figure 7 demonstrate the general layout of the final result of using linear features constraints with free network adjustment method.

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & -Z_A & Y_A & X_A \\ 0 & 1 & 0 & Z_A & 0 & -X_A & Y_A \\ 0 & 0 & 1 & -Y_A & X_A & 0 & Z_A \end{pmatrix} \quad (10)$$

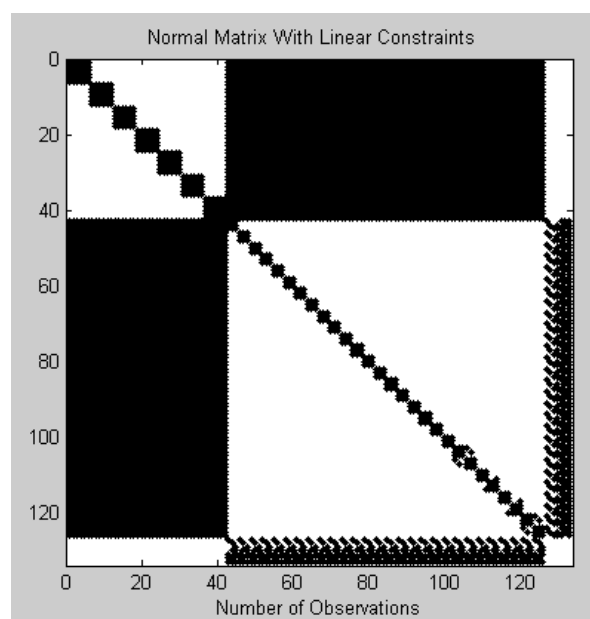


Figure 7. Final Form of Normal Matrix

3.6.3 Semi-Global Dense Matching (SGM): The enhanced EOPs of each image are obtained from the bundle adjustment procedure using initial values acquired directly through the smartphone sensors and geometric constraints technique. These enhanced EOPs, along with IOPs obtained through the camera calibration procedure, are used to generate rectified images. Then, the epipolar geometry between each rectified stereo-pair of images is reconstructed. By using epipolar geometry, the search for the corresponding pixel between each stereo-pair is minimized. Then, the matching cost is computed using the Normalized Cross Correlation (NCC) algorithm and cost aggregation, eventually resulting in the production of a coarse disparity map. Using this map, a dense 3D point cloud is finally generated using a linear spatial intersection of light rays from all matched pixels.

4. EXPERIMENT RESULT AND DISCUSSION

Seven images, along with their EOPs, are captured using the developed Apple app over an area with well distributed ground control points (GCPs), shown in Figure 8. The bundle adjustment procedure was implemented once with GCPs, and again without GCPs. In total, five different GCPs were used to recover the EOPs of the captured imagery, the IOPs of the iPhone 6's camera, and the ground coordinates of object points. Then, using the free network adjustment approach, along with the proposed linear feature constraints, resulted in enhanced accuracy of Euler angles (ω , ϕ , and κ).



Figure 8. Captured overlapped images using *iPhone6's* camera

Table 2 shows the difference in position for each camera station at the time of exposure with and without the use of GCPs. These positions are calculated using a single point positioning technique.

Position error	East (m)	North (m)	Height (m)
Image1	3.71	2.89	2.45
Image2	2.011	3.09	2.177
Image3	1.101	4.61	1.73
Image4	0.42	3.12	2.18
Image5	0.60	4.01	2.23
Image6	7.13	2.61	2.57
Image7	0.43	2.41	2.58

Table 2. Camera Position Error

The 3D reconstruction result for several object points, using a free network adjustment, is compared with ground truth data, and the root mean square error is then calculated, as illustrated by Table 3 and Figure 9.

Position error	East (m)	North (m)	Height (m)
RMS	3.82	4.27	3.85

Table 3. Final Solution Error

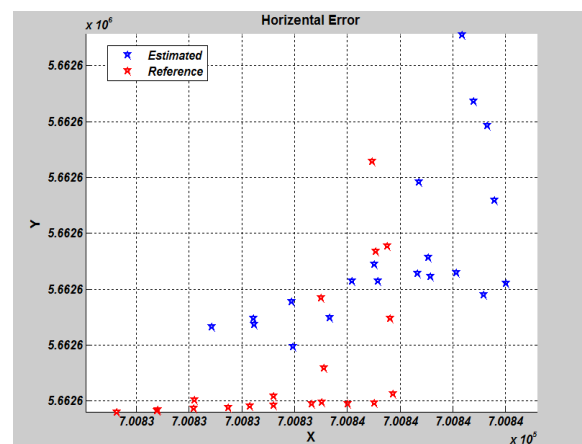


Figure 9. Horizontal error

4.1 Generated 3D point cloud

Figure 10 illustrates the smartphone-generated 3D point cloud, which clearly shows the potential of smartphone sensors. Although, seven images were used, someone can identify and recognize different features in 3D. Furthermore, higher quality 3D point cloud can be produced if dense matching technique and more images are used.



Figure10. Image-based 3D point cloud result

4.2 Measuring Lengths Application

Several feature lengths, shown Table 4 and Figure 11, are measured using the generated point cloud and these are compared to the truth data in order to ascertain the effectiveness of using the point cloud for mapping applications. The results indicate a relatively high level of accuracy for the final mapping solution.

Measured Feature number	True length (m)	Measured length (m)
1	0.90	0.98
2	1.02	1.05
3	2.56	2.61
4	2.24	2.33
5	1.09	1.012

Table 4. Measured length

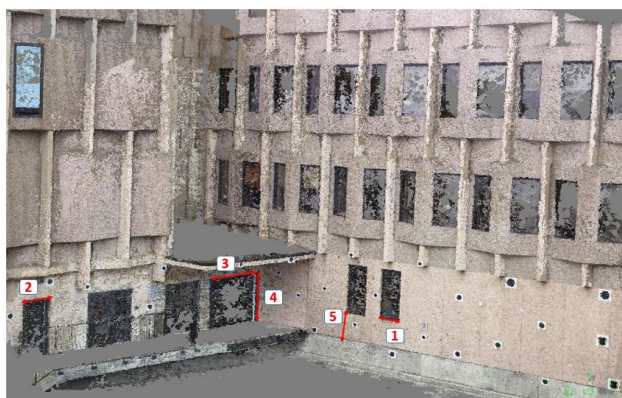


Figure 11. Image-based 3D point cloud result

5. CONCLUSION

In this paper, a new technique for increasing the direct geo-referencing accuracy of image-based 3D point cloud using a smartphone is proposed. Vertical and horizontal linear feature constraints are used in the bundle adjustment procedure to correct initial EOPs acquired through smartphone motion sensors. The results demonstrate the ability of smartphone sensors for generating 3D point clouds with a relatively acceptable level of accuracy.

ACKNOWLEDGEMENTS

This project is funded by research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors are grateful to Saudi Government for financing the first author's scholarship at the Department of Geomatics Engineering in University of Calgary, Canada.

REFERENCES

- Al-Hamad, A., 2014c. Mobile Mapping Using Smartphones, Thesis for degree of master of science. The University of Calgary.
- Al-Hamad, A., El-Sheimy, N., 2014a. SMARTPHONES BASED MOBILE MAPPING SYSTEMS. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 29-34.
- Al-Hamad, A., Moussa, A., El-Sheimy, N., 2014b. Mobile Mapping Using Smartphones. In: *The Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, Tampa, Florida.
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1222-1239.
- El-Sheimy, N., 1996. The development of VISAT-A mobile survey system for gis applications. University of Calgary.
- El-Sheimy, N., 2008. Land mobile mapping systems, *Advances in photogrammetry, remote sensing and spatial information sciences*. 2008 ISPRS congress book, pp. 85-100.
- Ellum, C., 2001. The development of a Backpack mobile mapping system, 2001. Thesis for degree of master of science The University of Calgary.
- Forstmann, S., Kanou, Y., Ohya, J., Thuering, S., Schmitt, A., 2004. Real-time stereo by using dynamic programming. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2004. CVPRW'04. IEEE, pp. 29-29.
- Granshaw, S., 1980. Bundle adjustment methods in engineering photogrammetry. *The Photogrammetric Record* 10, 181-207.
- He, F., Habib, A., Al-Rawabdeh, A., 2015. Planar Constraints for an Improved UAV-image-based Dense Point Cloud Generation. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, PP. 269-274.
- Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. CVPR 2005, pp. 807-814.
- Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, PP.328-341.
- Lari, Z., Habib, A., Mazaheri, M., Al-Durgham, K., 2014. Multi-camera System Calibration with Built-in Relative Orientation Constraints (Part 2) Automation, Implementation, and Experimental Results. *Journal of the Korean Society of Surveying Geodesy Photogrammetry and Cartography* 32, pp. 205-216.
- Mikhail, E., Bethel, J., McGlone, J., 2001. Introduction to modern photogrammetry. John Wiley & Sons Inc.
- Schwarz, K., El-Sheimy, N., 1996. KINEMATIC MULTI-SENSOR SYSTEMS FOR CLOSE RANGE DIGITAL HWIAGING. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vienna, pp. 774-785.
- Sun, J., Zheng, N., Shum, H., 2003. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, pp. 787-800.
- Wolf, P., Dewitt, B., 2000. Elements of Photogrammetry: with applications in GIS. McGraw-Hill New York.