INTEGRATING DEPTH MAP AND IMU DATA FOR 3D RECONSTRUCTION FORM A SINGLE IMAGE

Tzu-Fei Chen [a], Huan Chang [b] and Fuan Tsai [c]

[a,b] Graduate student, Department of Civil Engineering, National Central University,

[c]Associate Professor, Center for Space and Remote Sensing Research, National Central University,

300,Jhongda Rd., Jhongli, Taoyuan 32001, Taiwan;

Tel: +886-3-4227151#57619

E-mail: vivy7385@hotmail.com, 1984chang@gmail.com, ftsai@csrsr.ncu.edu.tw

KEY WORDS: Inertial measurement unit (IMU), Single image reconstruction, Depth map, accuracy assessment

Abstract: This research developed a method for reconstructing a 3D model from a single photograph with an estimated depth map and inertial measurement unit (IMU). In photogrammertry, exterior orientation parameters (EOP) can be calculated using multiple images based on relative geometry. However, if only one image is available, this approach will not work. The proposed method uses IMU to provide the relative orientations of EOP and assumes the camera position as the origin. The estimated depth map provides initial relative distances from feature points to the camera center. Therefore, collinearity condition equations can be used for determining 3D positions of feature points from a single image. In this study, there are four main steps for 3D building reconstruction from a single image: (1) camera and boresight calibration; (2) initial depth map generation; (3) coordinates of feature points calculation and iteration; (4) 3D model reconstruction. The proposed method begins with angles calibration between camera boresight and IMU system. Boresight angles can be predicted using the result of image back projection and simultaneous recorded orientation data from IMU. Using the relative distance obtained from the initial depth map, coordinates of the extracted feature points can be determined from collinearity condition equations. The depth information is further refined by co-planarity process of coplanar feature points. Refined depth map is generated from the calculated feature points coordinates and used to compute the new coordinates. The iteration stops when the system converges or the number of the iteration reaches the pre-defined maximum. The final coordinates of feature points are then used to reconstruct the 3D model.

1. INTRODUTION

Photogrammetry has many advantages as a technique for the acquisition of three-dimensional models for virtual reality. However, the traditional photogrammetric process to extract 3D geometry from multiple images is often labour-intensive and time consuming. On the other hand, single view reconstruction (SVR) techniques have a potential to extract 3D geometry from an image and can provide a more economic and efficient way for building model reconstruction. Many researchers have dealt with the problem of 3D reconstruction from a single image. The majority of the researchers have managed to solve this problem using a calibrated camera (Delage et al., 2006) or a partially calibrated camera (a camera that is set to a known height) (Guan & Pollefeys, 2009). However, the internal and external parameters of the camera play the vital role in the reconstruction. In this study exterior orientation parameters are obtained with extra equipment IMU data. According to El-Sheimy (1996), the transformations between camera and IMU have to be taken into account before catch the image.There are several important photogrammetry parameters, including ground coordinate, image coordinate and exterior orientation. The objective of this study is to obtain exterior orientation with extra equipment IMU data and reconstruct the model from a single perspective image with an estimated depth map. According to Wolf & Dewitt (2000), the collinearity condition is the condition using known exterior orientation parameters that any object point, and its photo image all lie along a straight line in three- dimensional space. In order to solve the coordinate of the object, in the field of the computer vision, depth map from the single image could provide an extra boundary condition for the collinearity equation. Saxena et al. (2008) proposed a method to reconstruct 3-D depth from a single image by applying supervised learning to predict the value of the depth map as a function of the image. However, this research further provides a novel method to reconstruct the scene from a single photograph with depth map and inertial measurement unit (IMU).

2. METHODOLOGY

The proposed method consists of four major parts for 3D scene reconstruction from a single image. The first part is to calibrate the IOP of a camera and to calibration angles between the camera boresight and IMU system. The second one is to generate the initial depth map from a single image. The third is to calculate coordinates of the extracted feature points using collinearity condition equations and to employ Least squares method to determine the coplanar surface. The depth information is also further refined in this step. The final step is to reconstruct the 3D building model. Fig. 1 illustrates the procedure of the proposed method.

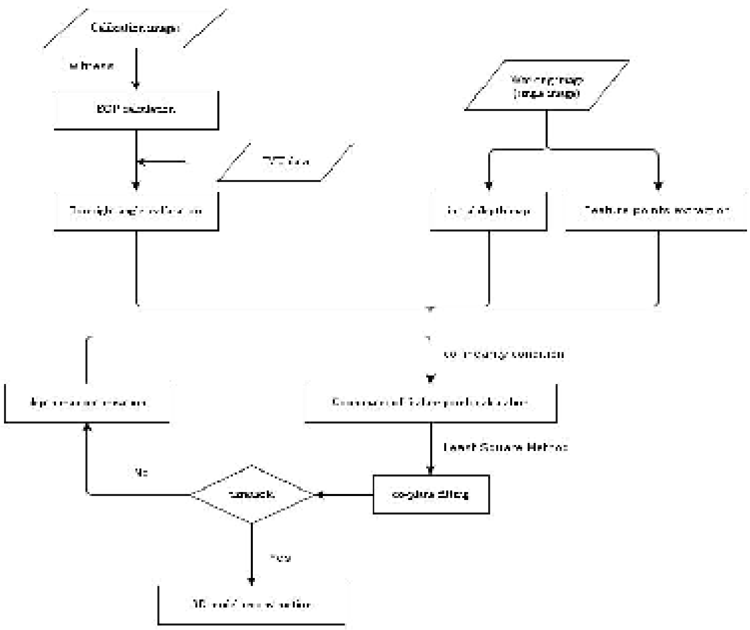


Fig.1. Procedure of the proposed method

2.1 Camera and boresight calibration

The interior orientation is required for the reconstruction of the bundle of rays. As the result, the first step is camera calibration. The focal length (**f**) of the camera in this case is determined by calibration before catching the photograph, and this study ignored the lens distortion.

After camera calibrating, an extra equipment, IMU, is used to attain three elements of exterior orientation parameters (**ω, φ, κ**) of a photograph. Boresight angle calibration between different sensor frames has to be taken into account, too. Broesight angles can be predicted using the result of image back projection and simultaneously recorded orientation data from the IMU. In addition to transformations between sensors, rotations between different sensor frames are also considered. The INS b-frame (gyro frame) cannot be aligned with the c-frame. According to El-Sheimy (1996), the constant rotation $\mathbf{R}_c^b$ between the two frames is again obtained by calibration. In this case, $\mathbf{R}_c^m(t)$ can be written as:

$$\mathbf{R}_c^m(t)=\mathbf{R}_b^m(t)\cdot\mathbf{R}_c^b \qquad \textbf{[2]}$$

Where $\mathbf{R}_c^b$ is the rotation between the camera c-frame and the INS b-frame as determined from a calibration process.

2.2 Depth map generation

Depth map provides the relative distance from feature points to the camera center and can be used as extra boundary conditions when solving the collinearity equations. To generate a reliable depth map from a single image usually requires a rough depth classified (Saxena et al., 2009). An approximated depth map can be generated by applying supervised learning to predict the value of the depth. The distance from object to camera center are defined as depth **(d)**, and the formulas is described as equation [3].

$$d=\sqrt{(X-X^c)^2+(Y-Y^c)^2+(Z-Z^c)^2} \qquad [3]$$

2.3 3D model reconstruction

After feature extracting, the coordinates of extracted feature points are determined from collinearity condition equations and with depth information from the depth map. The collinearity condition uses known exterior orientation parameters provided from IMU data that any object point, and its photo image all lie along a straight line in three- dimensional space. The equation can be represented as [4]

$$x=-f\frac{m_{11}(X-X^c)+m_{12}(Y-Y^c)+m_{13}(Z-Z^c)}{m_{31}(X-X^c)+m_{32}(Y-Y^c)+m_{33}(Z-Z^c)}$$

$$y=-f\,\frac{m_{21}(X-X^c)+m_{22}(Y-Y^c)+m_{23}(Z-Z^c)}{m_{31}(X-X^c)+m_{32}(Y-Y^c)+m_{33}(Z-Z^c)} \qquad \ldots\ldots\ldots\ldots\ldots[4]$$

Where

$$m_{11} = \cos\varphi\cos\kappa$$

$$m_{12} = \sin\omega\sin\varphi\cos\kappa + \cos\omega\sin\kappa$$

$$m_{13} = \cos\omega\sin\varphi\cos\kappa + \sin\omega\sin\kappa$$

$$m_{21} = -\cos\varphi\sin\kappa$$

$$m_{22} = -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa$$

$$m_{23} = \cos\omega\sin\varphi\sin\kappa + \sin\omega\cos\kappa$$

$$m_{31} = \sin\varphi$$

$$m_{32} = -\sin\omega\cos\varphi$$

$$m_{33} = \cos\omega\cos\varphi$$

2.4 Accuracy evaluation

In order to evaluate the quality of the reconstructed model from a single image, this research compares the geometrical performances between the model from a single image and a simulated real model. The results are used to evaluate the performance of the proposed method and the quality of the reconstructed model from a single image.

3. EXPERIMENT RESULTS

The first test data is a 510 by 765 image as shown in Fig. 2, and its focal length is 30mm. Depth from camera center to the target building is between 100~120m. An estimated depth map (as shown in Fig. 3) is also generated (using Saxbna method) for subsequent process of 3D reconstruction. Because the test area is simulated from a created 3D model, a few corner points are measured from the model to be used as check points to evaluate the accuracy of the reconstruction result. The experiments compare the geometrical performances between the model form a single image and the real model. The experimental results are to evaluate the quality of reconstructed model from single image.

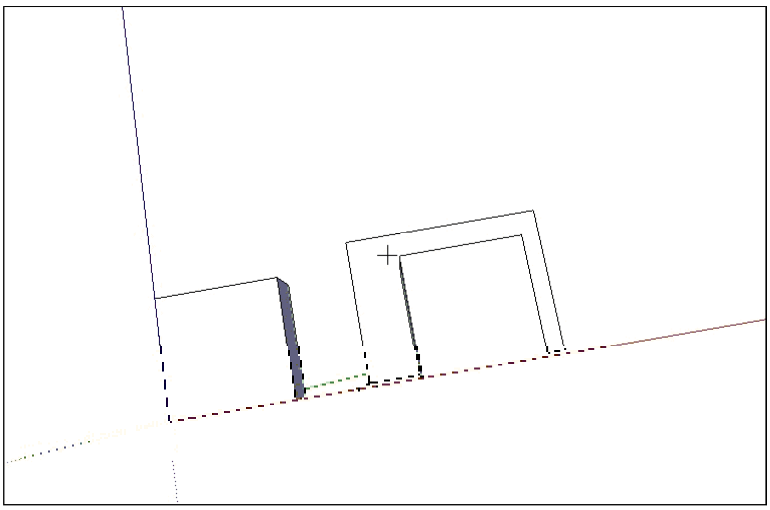


Fig. 2 Test image (working image) for 3D reconstruction from a single image.

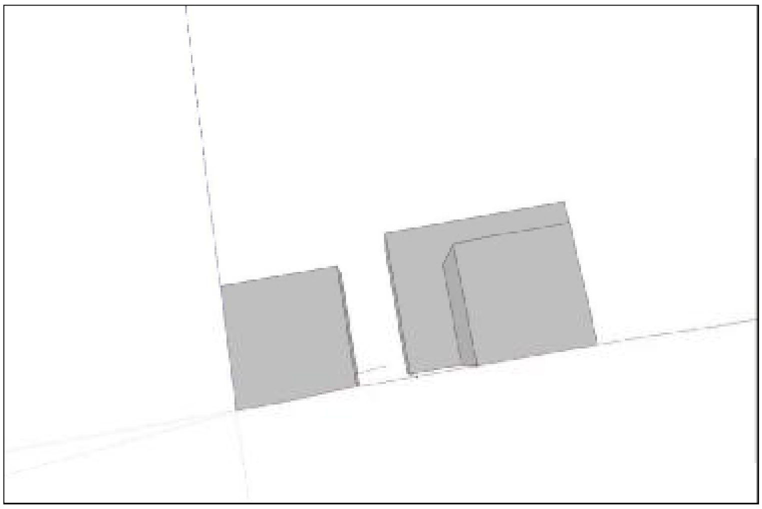


Fig. 3 Estimated (initial) depth map of the test image.

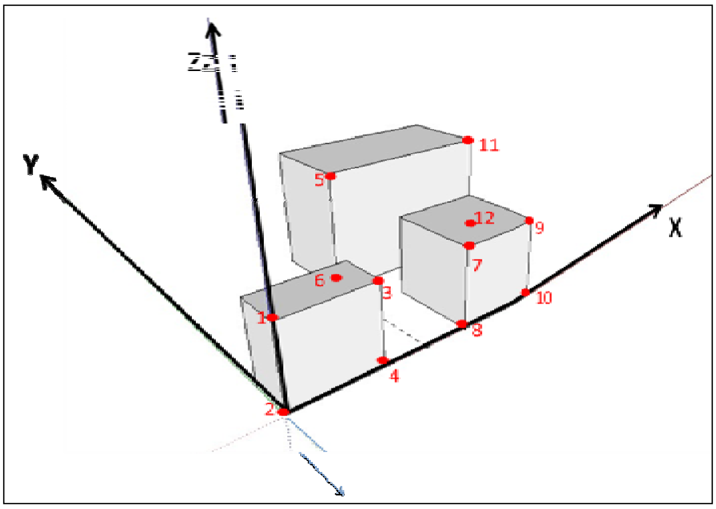


Fig. 4 Distribution of control points from real model of test image.

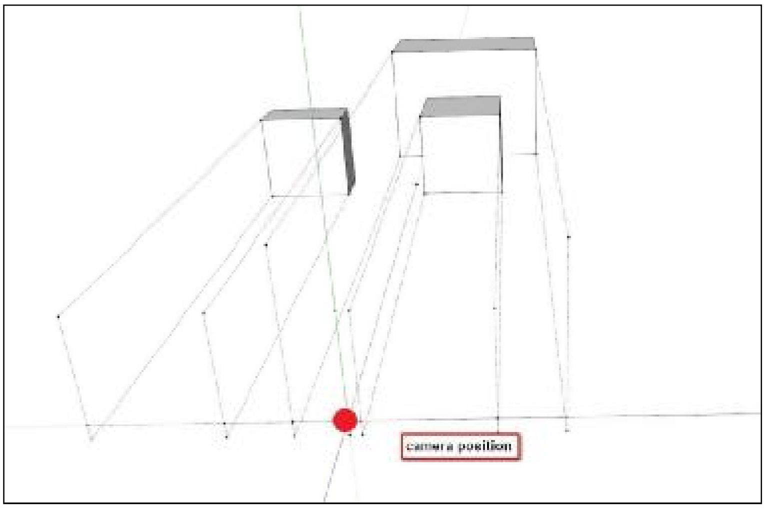


Fig. 5 Relationship from camera center and real model.

3.1 IMU data quality analysis

This study uses an extra equipment, IMU, to provide three elements relative orientations of exterior orientation parameters (**ω, φ, κ**) of a photograph. In order to evaluate the influence caused by IMU data quality, random errors are added into the real IMU measurement to simulate the effects caused by machine uncertainty. To evaluate the effect in different angle rotation (**ω, φ, κ**) in X, Y, and Z axes, an error of 2° is introduced and the root-mean-square-errors (RMSE) in the three axes are compared with the coordinates of the feature points determined by the simulated model and with the true model. The results are listed in Table 1.

Table 1: RMSE of the reconstructed model with IMU error.

	ω (with 2°)	**φ** (with 2°)	**κ** (with 2°)
RMSE_X (m)	0.023	0.041	0.034
RMSE_Y (m)	0.035	0.029	0.047
RMSE _Z (m)	0.654	0.240	0.241

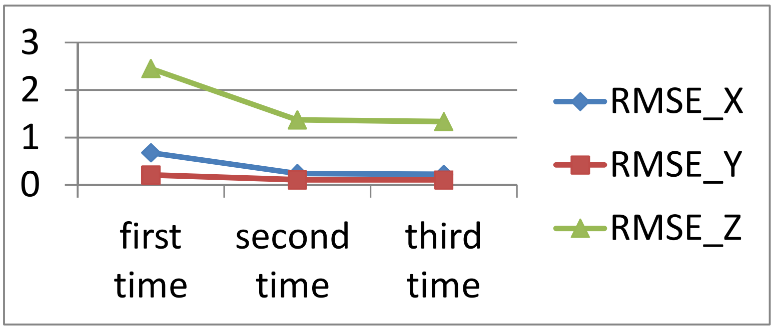
3.2 compare accuracy in using real depth and using depth map Real depth and depth map comparison

By incorporating the IMU data and depth map as boundary conditions of the collinearity equations, there is no degree of freedom. As the result, it is important that the depth map should be of high accuracy or to develop a method to refine the depth information. Otherwise, it may produce significant errors if the quality (accuracy) of the depth map is not adequate. An algorithm based on Least Squares Method (LSM) is developed in order to refine the depth map. The idea is to used the estimated depth map as an initial condition to reconstruct the model and a new depth map is generated from the reconstructed model. The process is repeated until the change of the depth map is insignificant or the number of iteration has reached the pre-defined threshold. Table 2 and Fig. 4 shows the relation between the accuracy and the number of iterations. As listed in the table, the accuracy has indeed improved as the number of iteration increased.

Table 2: Accuracy improvement of depth map iteration.

RMSE(m)	First iteration	Second iteration	Third iteration
X	0.6818	0.2420	0.2230
Y	0.2091	0.1091	0.1068
Z	2.4548	1.3696	1.3344

Fig. 6 Accuracy improvement of depth map iteration.



4. CONCULSION AND FUTURE WORKS

This study developed a method to construct 3D models from a single image. The method combines computer vision and close range photogrammertry and requires only a single image with IMU data and an estimated depth map to reconstruct 3D building models. Experimental results using an image of computer-generated model demonstrate that although the accuracy in the reconstruction of 3D building models using the initial depth map may not be satisfactory, the accuracy improves significantly as the depth map is refined using the proposed algorithms and procedure. Future improvement of the developed system will focus on applying the developed method to images of natural scenes and with complicated scence compositions.

REFERENCES:

Delage, E., Lee, H. and Ng, A.,2006. A dynamic bayesian network model for autonomous 3d reconstruction from a single indoor image. Computer vision and pattern recognition(CVPR).

Guan, L., Pollefeys, M.,2009. A unified approach to calibrate a network of camcorders and toF cameras. IEEE workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2), in conjunction with ECCV.

El-Sheimy, N., 1996. The development of VISAT - a mobile survey system for GIS applications. Ph.D. Thesis, UCGE Report No.20101, Department of Geomatics Engineering, The University of Calgary, Canada.

Wolf, P. R. and Dewitt, B. A., 2000. Elements of photogrammetry with Applications in GIS. McGraw

Hill

Saxena, A., Chung, S., and Ng, A., 2008. 3-D depth reconstruction from a single still image, IJCV, 76 (1), pp.53–69.

Saxena, A., 2009. Make3D: Learning 3-D Sence Structure from a single Still Image. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 31(5), pp. 824-840.

Wolf, P. R. and Ghilani, C. D., 2006. Adjustment computation spatial data analysis. John Wiley & Sons, New York.