AUTOMATED MOSAICKING OF UAV IMAGES BASED ON SFM METHOD

Hui Wang[1,2], Jonathan Li[1], Senior Member, IEEE, Liyong Wang[2], Haiyan Guan[1], Zexun Geng[2]

[1] Faculty of Environment, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
[2] Zhengzhou Institute of Surveying & Mapping, Zhengzhou, Henan, China 450052
Email: hollywang2014@gmail.com, junli@uwaterloo.ca)

ABSTRACT

Optical sensors onboard an unmanned aerial vehicle (UAV) can collect high resolution images with small dimensions. Image mosaicking is necessary to cover a larger geographic area. This paper presents a novel approach to mosaicking UAV images automatically. The "Orthophoto Map" is based on Structure From Motion (SFM). This method can fully automatic mosaic generate a wide range, and with well visual effects, and no evident deformation. This method can not only get a panoramic image of wide range of areas, and can get the corresponding three-dimensional terrain model.

Index Terms—Image mosaic, Structure From Motion, Digital Terrestrial Model, ortho-rectified

1. INTRODUCTION

In recent years, unmanned aerial vehicle (UAV) remote sensing has been widely used in agriculture, forestry, land resources survey, military reconnaissance, environmental protection, and other fields. Although we can obtain high resolution digital images by the way of UAV, it covers small area for each image because of the limit of aircraft load and altitude. Therefore, image mosaicking is required to stitch these small images to obtain a large coverage of areas of interest. The stitched image should be close to the original images farthest, with little image distortion and without evident seam. Image Mosaicking is applied in many fields, such as remote sensing image processing, geographic information system, video sequence mage mosaicking, medical image analysis, and computer vision.

In this paper, we propose a method, which automated mosaics digital images based on Structure From Motion (SFM). The principle is detailed as follows:

1) The preprocessing of UAV images and feature extracting and matching of image points.

2) The calculation of camera parameters and 3D (three-dimensional) coordinates of feature points for each image based on SFM in a local coordinate system.

3) The image ortho-rectification.

4) Image fusing according to the coordinates of each ortho-image in the local unified coordinate system and then generating a wide range image of areas.

This method can obtain not only a panoramic image covering a wide range of areas but also a corresponding 3D terrain model.

2. MAIN FLOW OF IMAGE MOSAICKING

The main flow of image mosaicking includes image preprocess, feature extraction and matching, transformation model construction, transformation coordinate unification, image fusion [1]. The forth step, transformation coordinate unification is the key of the whole flow. Transformation model construction is to set up the projection transformation relationship between each central projection image and the large mosaicking image to be processed. The mosaicking results will be different if different transformation models are used.

In the unified coordinate system, each image is ortho-rectified using exterior orientation parameters and digital terrestrial model of the image covering. The ortho-rectified images are mosaicked according to the overlapping areas.

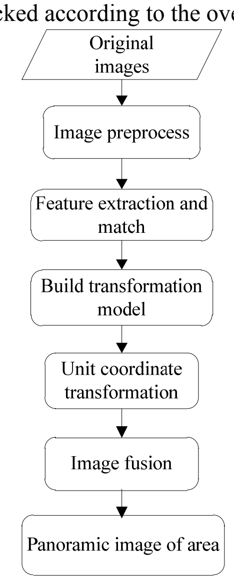

Fig.1 Flow diagram of image mosaic

3. PARAMETER MODEL OF DIGITAL CAMERA

Camera parameters, including interior and exterior parameters, describe the relationship between image points and ground points. Supposing axis of lens just go through the center of image plane and the pixel of camera is foursquare. Therefore, the camera parameters can be described with 7 parameters, namely, rotation parameters $\vec{\Theta} = (\theta_1, \theta_2, \theta_3)$, translation parameters $\vec{t} = (t_1, t_2, t_3)$, and the lens focus f.

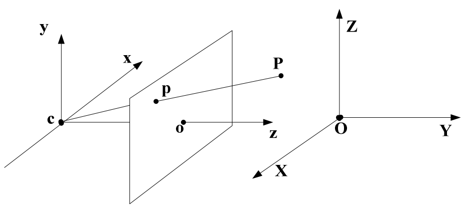

Fig.2 Schematic diagram of a digital camera

As shown in Fig.2, point p in image plane is the projection of the ground point P, the coordinate of ground point P in object space coordinate system is ($\overline{X}=(X,Y,Z)$), the coordinate of image point p in image space coordinate system is ($\vec{x}=(x,y,z)$). So, the relationship can be described:

$$\vec{x} = R\overline{X} + \vec{t} \tag{1}$$

Where R indicates the rotate matrix [2].

$$R = e^{[\theta]}, [\theta] = \begin{bmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{bmatrix}$$

The origin of the coordinate measured in digital image is at the top left corner, and the unit of coordinate is pixel in the pixel coordinate system. Given the pixel is foursquare, the size of it is α and principal point of photograph is at the same place of image center, the coordinate of principal point of photograph in pixel coordinate system is (u_0, v_0). The transformation from the pixel coordinate of one point to its image plane coordinate (x, y) can be indicated as formula (2):

$$\begin{cases} x = \alpha(u - u_0) \\ y = \alpha(v - v_0) \end{cases} \tag{2}$$

Generally, UAV images have large distortion because they are collected by a non-metric digital camera. The main part of image distortion is radial distortion [3], which is in proportion to the distance square of image point to the principal point of photograph. Supposing the distortions in directions of u, v are identical, the radial distortion model is regarded as follows:

$$\begin{cases} u' - u_0 = \dfrac{xk(r)}{\alpha} \\ v' - v_0 = \dfrac{yk(r)}{\alpha} \end{cases} \tag{3}$$

$$k(r) = 1 + k_1 r^2 + k_2 r^4 \tag{4}$$

Where, (u', v') indicates the pixel coordinates with distortion difference, $r = \sqrt{x^2 + y^2}$ is the distance square of image point to the principal point of photograph, $k(r)$ is the ratio factor from non-distortion coordinates to distortion coordinates, k_1 and k_2 are distortion parameters. Therefore, there are 9 parameters in the parameter model of camera.

4. STRUCTURE FROM MOTION

Structure from Motion (SFM) [4] is used to calculate cameral parameters and 3D coordinates of feature points in Computer Vision. It is a process of rebuilding 3D structure of target and calculating camera parameters based on images with the same scene but different shooting angles. The common calculating method takes advantage of matched points and sparse bundle block adjustment after the extraction of feature points and matching. The concrete process is as follows [2]:

1) Two images, which have enough matching points near to the center of the whole area, are selected as the initial stereo pair. The left image space coordinate system is regarded as the local coordinate system. The initial values of camera parameters and coordinates of homologous feature points are obtained through relative orientation and space intersection, and then optimized by sparse bundle block adjustment.

2) A new image is added, and its camera parameters are calculated (initial value of radial parameter is set as 0). Because of the 3D coordinates of the corresponding feature points in new image have been calculated in above step, the new camera parameters can be obtained using the space resection with the 3D points and their matched feature points of the corresponding images. Then, the adjustment is employed to optimize the result.

3) Repeat steps 1) and 2) until all the images are processed. The mage is removed if its number of matched points is quite small.

In this paper, the adjustment process is performed by Sparse Bundle Adjustment Software Package [5]. Fig.3 shows the concrete operation flow.

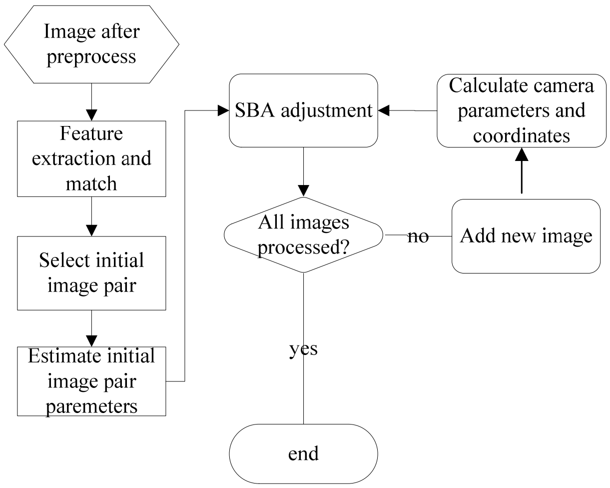

Fig.3 The camera parameters and feature point coordinate calculation flow chart

In the local coordinate system, the camera parameter $\vec{\Theta}_i$ ($0 \le i \le M$) of each image and 3D coordinates of all feature points $\vec{X}_j$ ($0 \le j \le N$) are obtained after all images are processed with adjustment.

Where, M is the number of images, N is the number of points.

5. ORTHO-RECTIFICATION

The process of ortho-images generation is defined ortho-reflection. Particularly, a process of ortho-rectification on digital image is also named digital differential rectification using the corresponding imaging formula or a certain mathematical model based on the correlative parameters and DTM [6]. It has two calculation methods: positive solution and inverse solution. The inverse solution is usually used because of the "blank points" and "repeated points" existing when using positive solution.

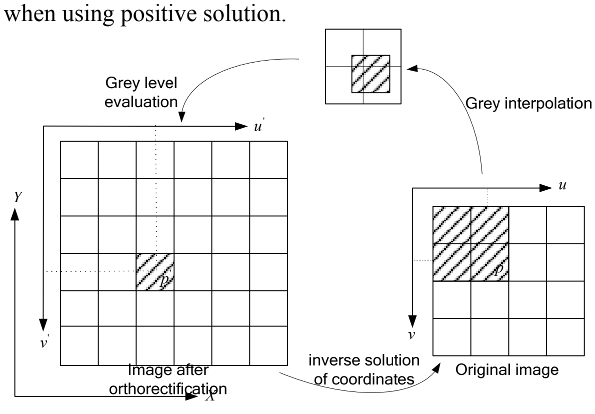

Fig.4 Inverse solution of digital differential rectification

6. TESTS AND ANALYSIS

With Structure Form Motion, the camera parameters and 3D coordinates of feature points are calculated for all images in a local unified coordinate system. Then, the interior and exterior parameters of each image are then calculated. A large number of the discrete 3D points generate an irregular point cloud group. The irregular DTM is then regularized after eliminating outliers from the generated 3D points. Finally, each image is rectified based on its camera parameters, interior and exterior parameters and the regularized DTM.

169 images shot collected by a UAV for one district are used. Fig.5 shows the feature points in the unified coordinate system. Fig.6 shows an ortho-rectified image. Fig.7 shows a 3D terrain model.

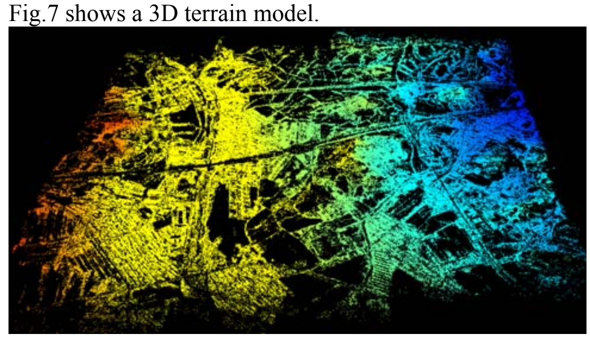

Fig. 5 3D coordinates of feature points hypsometric map

Fig. 6 Orthophoto Map after mosaicking

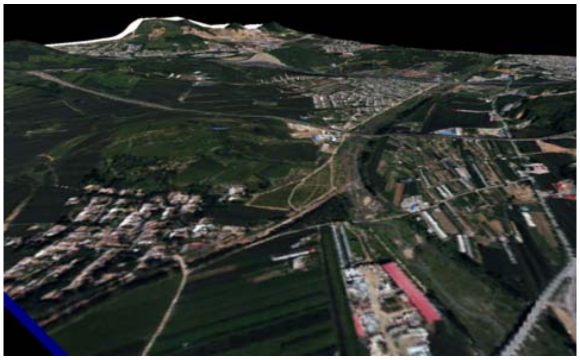

Fig.7 3D terrain model

It can be seen from Fig.5 and Fig.6 that the distribution of the extracted feature points are not even. The number of feature points in cropland is less than that of the residential area. Fig.7 shows that the terrain is up-and-down and complex. The image mosaicking result is shown well without evident errors because of ortho-rectification using the DTM generated by feature matching points.

7. CONCLUSIONS

An automated image mosaicking method based on SFM is proposed in this paper. It calculates the camera parameters and 3D coordinates of feature points using SFM on basis of a set of pre-processing, feature extraction, and image matching without the support of ground control points. And then, a regular DTM in a unified coordinate system is developed according to the feature points. Finally, the original images are ortho-rectified with the camera parameters and DTM. The ortho-rectified images present good visual effects (no evident distortion in a stitched large image). The proposed image mosaicking method is automated without ground control points.

8. REFERENCES

[1] Inampudi, R. B., *Image Mosaicking*, Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International, vol. 5, pp. 2363 - 2365, July 1998.

[2] Brown, M., Lowe, D.G., *Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets*, Fifth international conference on 3-D digital imaging and modeling, pp.56–63, June 2005.

[3] David P. Casasent, Ernest L. Hall; Juha Röning, *Intelligent Robots and Computer Vision XXV: Algorithms, Techniques, and Active Vision*, Proceedings of SPIE--the International Society for Optical Engineering ; vol. 6764, September, 2007.

[4] Dellaert, F. , Seitz, S.M. , Thorpe, C.E. , Thrun, S., *Structure from motion without correspondence*, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol.2, pp.557-564, June 2000.

[5] Lourakis, Manolis, I. A. , Antonis, A., *SBA: A software package for generic sparse bundle adjustment*, ACM Transactions on Mathematical Software (TOMS), vol.36(1), pp.1-30, March 2009.

[6] Kasser, Michel., *Digital photogrammetry*, London New York : Taylor & Francis, 2002.