SIFT-BASED MATCHING OF MULTIPLE AT IMAGES

Ming-Shiuan Lee[1], Jaan-Rong Tsay[*2], and Po-Tsung Su[3]
[1]Graduate Student, Department of Geomatics, National Cheng Kung University, 1 University Road,
Tainan, 70101 Taiwan; Tel: +886-6-2757575 ext. 63834,
Email: evenif0901@msn.com

[2]Associate Professor Dr.-Ing., Department of Geomatics, National Cheng Kung University, 1 University Road,
Tainan, 70101 Taiwan; Tel: +886-6-2370876 ext. 838,
E-mail: tsayjr@mail.ncku.edu.tw

[3]Graduate Student, Department of Geomatics, National Cheng Kung University, 1 University Road,
Tainan, 70101 Taiwan; Tel: +886-6-2757575 ext. 63834,
Email: patrick.n80250@msa.hinet.net

KEY WORDS: SIFT, Multi-Image Matching, Quality Filtering (QF)

ABSTRACT: To further increase the degree of automation of modern aerial triangulation and geomatic data acquisition, this paper proposes a new method for multi-image matching based on the Scale Invariant Feature Transform (SIFT) for aerial triangulation (AT), especially without the need on the input data such as block and strip data for providing image overlap information. To simultaneously deal with a large number of aerial images with large image format in a block area, we proposed both schemes of Quality Filtering (QF) and Affine Transformation Prediction (AFTP) for automatic tie point extraction and measurement with a better and satisfactory efficiency. Some tests are done by using aerial images taken with the RMK DX camera. Also, high precision ground check points are used to evaluate the quality of the results. They demonstrate that the automatic tie point selection and measurement is done efficiently even under the circumstance that no priori-knowledge on image overlap is available. Also, ground check points show that the accuracy of photo coordinates is 0.21 pixels, namely it reaches a subpixel level.

1. INTRODUCTIONS

With advances in sensors and storage devices, spatial information technologies and applications utilize digital images increasingly. And image orientation with high efficiency and high automation is one of the key research issues. Inevitably, tie point measurements are necessary in aerial triangulation, where image matching brings the feasibility of automation. Generally, there must be some known data as input for providing initial values on the practical application of the multi-image matching, such as block parameters and strip parameters for informing image overlap.

Nowadays, the research and application of multi-image matching is more and more important as camera sensors are designed with better radiometric and geometric quality. Compared to LiDAR techniques, aerial photogrammetry provides redundant observations, reliable and applicable image features, better horizontal accuracy and meaningful tie point measurements. More importantly, unknowns of image orientations and object coordinates can be solved by bundle block adjustment, which is a primary process of geomatic data acquisition (Heipke, 1997). To further increase the degree of automation of modern aerial triangulation and geomatic data acquisition, this paper proposed a new method for multi-image matching based on SIFT for a large number of automatic tie point measurements without the need on any image overlap information.

2. MULTI-IMAGE MATCHING STRATEGY

2.1 Main Scheme

Multi-image matching is a milestone of automatic digital image processing (Schenk, 2004). To focus on a more compatible architecture, we choose a scale and rotation invariant method for tie point measurements, namely the scale invariant feature transform (SIFT) technique. The SIFT method belongs to the class of feature-based matching, whose processing operations include two main parts – keypoint extraction and keypoint matching (Lowe, 2004). Keypoint extraction is achieved by the procedures of Gaussian Filtering and computation of DoG (Difference of Gaussian) at different scales in image pyramid to detect the extreme values. The locations corresponding to these extreme values are keypoints, described by means of a descriptor defined by a 128 dimensional vector. Then keypoint matching is simply to calculate the Euclidean distances from one keypoint descriptor on the left image to another keypoint descriptors on the right image, i.e. a pair of images at one time. If the distance ratio (the shortest Euclidean distance divided by the second short one) is smaller than the threshold, then the keypoint is matched. Thus, the one on the left image is matched to another one on the right image. Otherwise, the matching for this keypoint on the left image fails. The matching and searching operations will be done repeatedly until all keypoints are processed.

The proposed strategy of multi-image matching is illustrated in Figure 1. Without the need on priori knowledge on image overlap information, the first step is to process SIFT keypoint extraction to obtain the location (abbreviated as Loc.) and descriptor (abbreviated as Des.) of every kepoint on P input images (P ≧ 2). Step 2 will be the Keypoint Matching for C_2^P pairs of images where the loop number equals to the combination of selecting 2 images from *P* images, one image matching pair at each time. Then, the result table of each image matching pair stores locations (of matched points) and numbers of the left and right image for every image matching pair. And Step 3 will be Matched Point Connection via comparing the locations of matched points, rearranging and coding all the matched points into numbered result, eventually.

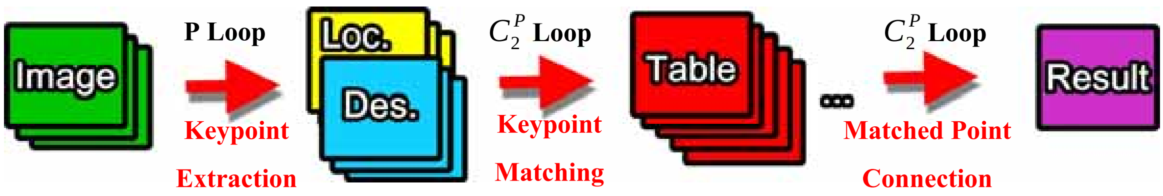

Figure 1. Multi-Image Matching Strategy

The format of temporary tables of Matched Point Connection is illustrated in Figure 2. Every table of single image matching pair's result contains locations (row, column) of matched points on left and right image, denoted by (r_L, c_L) and (r_R, c_R). And the table of multi-image matching result stores location, point number and index value ((r, c) , PN, Index) for every tie point in each image, which is done by means of location matching using the result table of image matching pair. The index value is used for descriptor inquiry, namely to inform that the descriptor belongs to the i-th keypoint on the j-th image. Therefore, the numbered tie points are connected, if their $|\Delta r|<10^{-6}$pixels and $|\Delta c|<10^{-6}$ pixels, and the repeated measurements are eliminated at this step.

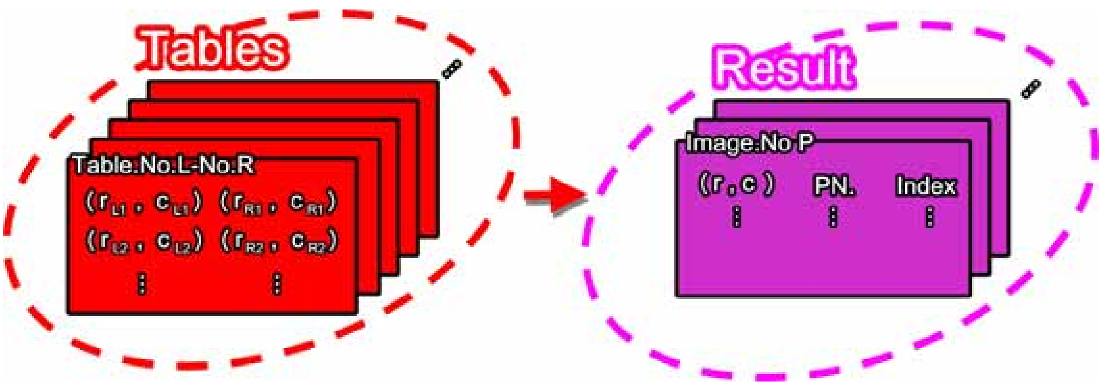

Figure 2. Strategy for Connection of Matched Points

2.2 Quality Filtering (QF)

In a process for a large number of aerial images with large image format, the calculation efficiency is impacted due to the extremely huge number of keypoints. In order to conquer the impact and to reduce the time consuming, Quality Filtering (QF) is attempting to reserve the best keypoints according to an image quality assessment procedure. According to Eq. (1), the Standard Deviation of Gray Levels, denoted by G_{std}, of every keypoint can be calculated for every image. The loop number equals to P, namely the number of input images. The Standard Deviation of Gray Levels is computed in a local image window of 15 x 15 pixels centered at the keypoint of interest. Generally, G_{std} stands for the contrast of the keypoint image. In case of less noise, it also indicates the amount of texture information (or so-called quality) on the keypoint.

$$\overline{G} = \frac{1}{15^2}\sum_{c=1}^{15}\sum_{r=1}^{15} G_{rc}, \quad G_{std} = \sqrt{\frac{1}{(15^2-1)}\sum_{c=1}^{15}\sum_{r=1}^{15}\left(G_{rc}-\overline{G}\right)^2} \qquad (1)$$

Assuming that the indicator values G_{std} of all keypoints in one image is normally distributed, the threshold for the selection of those best keypoints will be set to their average plus standard deviation of overall indicator values in one image. Thus, only about 16% keypoints are reserved for later multi-image matching. Since the indicator value of QF is exchangeable and the threshold is adjustable, the goodness and availability of the setting will be verified by the tests.

2.3 Affine Transformation Prediction (AFTP)

AFTP plays two important roles in the whole scheme. The first one is Searching Window Prediction, and the second is Overlap Estimation, as shown in Figure 3. Instead of using original resolution images, AFTP uses higher layer images with less number of keypoint in image pyramid to perform a fast pre-matching to maintain the efficiency and determine the necessity of follow-up process simultaneously. If the affine parameters $a{\sim}f$ of an image matching pair denoted in Eq. (2) with the system of observation equations $AX=L+V$ can be calculated by Least-Squares Adjustment (LSA), then these two images are overlapped. The locations of their corresponding image points are approximately described by the affine parameters, which can be utilized for prediction of searching window. Otherwise, this image matching pair has no overlap or rare overlap, and it will be skipped in the follow-up matching. In terms of whole matched point connection procedure, as long as better overlapped image matching pairs are processed, then the tie points can be connected correctly.

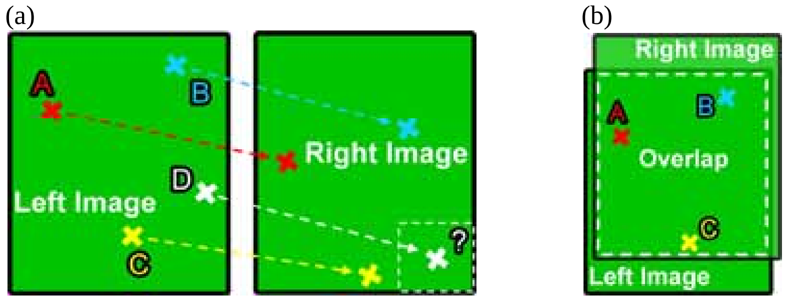

Figure 3. Concept of AFTP: (a) Searching Window Prediction, (b) Overlap Estimation

$$A=\begin{bmatrix} r_{L1} & c_{L1} & 1 \\ r_{L2} & c_{L2} & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}, X=\begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}, L=\begin{bmatrix} r_{R1} & c_{R1} \\ r_{R2} & c_{R2} \\ \vdots & \vdots \end{bmatrix}, V=\begin{bmatrix} v_{r1} & v_{c1} \\ v_{r2} & v_{c2} \\ \vdots & \vdots \end{bmatrix} \quad (2)$$

Based on the afore-mentioned strategy, the proposed multi-image matching procedure can be done by the only input of aerial images. After the transformation from image coordinates (r, c) to photo coordinates (x, y), a bundle block adjustment with data snooping (Baarda, 1968) procedure is used for preliminary error detection (Kruck, 1984) and validation is done by means of ground check points.

3. EXPERIMENTAL RESULT

3.1 Efficiency Analysis

In this study, test data includes in total 108 images captured by RMK DX camera in Taiwan, Chang-Hua, including 6 strips with two cross strips. First, a selected block of 11 images is used for efficiency and benefit analysis of AFTP and QF. There are three cases: Case I with only QF, Case II with only AFTP and Case III with both QF and AFTP. Table 1 shows the pre-process information and calculation time of the multi-image matching procedure. The "distance ratio of AFTP" means the distance ratio of SIFT set to a stricter smaller threshold in order to select less number of best keypoints as input points to the AFTP.

Several cases are processed with different distance ratio and their results are listed in Table 2 and illustrated in Figure 4, where the number of skipped points denotes the number of image points eliminated in the free network adjustment of AT. Thus, the ratio of the number of skipped points divided by the number of matched points, denoted as the skip rate, describes the goodness of matching in each case. As shown in Figure 4, when distance ratio is set to the threshold less than 0.3, Case III maintains better matching efficiency and best matching with lowest rate of skipped points. In general, the larger the distance ratio is, more points are matched per second, but the more the skipped points become. Moreover, Table 2 shows apparently that the calculation time is almost the same for the same case.

Table 1. Process Information on Keypoint Extraction, AFTP and QF

Process	Object	Calculation Time	Process Information
Keypoint Extraction	11 images	6575 Seconds	Calculation time is related to image size exclusively.
AFTP	55 image pairs	116 Seconds	When distance ratio of AFTP is 0.2, there are 28 image matching pairs available, i.e. loop number is reduced to 28.
QF	11 images	567 Seconds	Calculation time is related to the number of keypoints.

Table 2. Efficiency Analysis of QF and AFTP

Case	Distance Ratio	Distance Ratio of AFTP	QF	Matching Time (Unit: Seconds)	Number of Matched Points	Average Matching Efficiency (Unit: Points/ Seconds)	Number of Skipped Points	Skip Rate (Unit: %)
Ⅰ.1	0.20	N	Y	5915	2716	0.2080	19	0.70
Ⅱ.1	0.20	0.20	N	11621	37505	2.0481	512	1.37
Ⅲ.1	0.20	0.20	Y	307	7924	1.0474	45	0.57
Ⅰ.2	0.25	N	Y	5920	5767	0.4415	60	1.04
Ⅱ.2	0.25	0.20	N	11623	77426	4.2277	1198	1.55
Ⅲ.2	0.25	0.20	Y	303	15085	1.9951	145	0.96
Ⅰ.3	0.30	N	Y	5933	10026	0.7668	94	0.94
Ⅱ.3	0.30	0.20	N	11626	129161	7.0514	2293	1.78
Ⅲ.3	0.30	0.20	Y	303	23650	3.1279	218	0.92
Ⅰ.4	0.35	N	Y	5968	15214	1.1605	156	1.03
Ⅱ.4	0.35	0.20	N	11632	187295	10.2218	3867	2.06
Ⅲ.4	0.35	0.20	Y	304	32787	4.3358	422	1.29

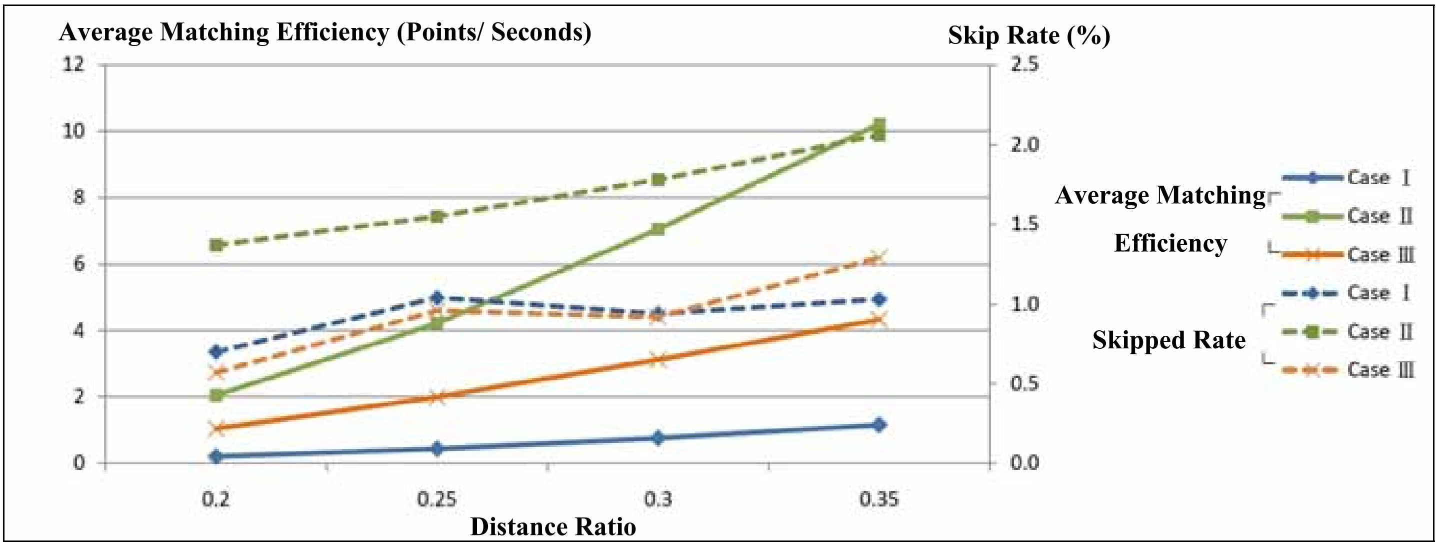

Figure 4. Efficiency Analysis of QF and AFTP

3.2 Bundle Block Adjustment

Now, all 108 RMK DX images and 71 known high precision ground points are used for quality validation. The efficiency of the new method is also compared to the commercial software LPS/ERDAS Imagine 2010. The latter provides automatic tie point generation process based on least-squared image matching and a default density of 25 points per image. These tie points are also first checked by means of free network adjustment, and the results are shown in Table 3, where the test value is the ratio of the a priori accuracy divided by posteriori accuracy. Then the bundle block adjustment with control data is performed, where 6 full control points and 65 independent check points shown in Figure 5 are used. The statistic figures of both bundle block adjustments computed by the new method proposed in this paper and by the commercial software LPS 2010 are listed in Table 4. They show that the new method is available to aerial triangulation with high precision and good efficiency. Especially, the new method don't need any information on image overlap. In other word, the new method don't need the well-known input data of block parameters and strip parameters which are often adopted by general commercial AT softwares.

Table 3. Results of free network adjustments done by the new method and LPS2010

Tie Points	Number of Points	Number of Skipped Points	Skip Rate (Unit: %)	Priori-accuracy	Test Value	Total Redundancy
New method	235002	2548	1.08	0.0015mm	1.05	193965
LPS 2010	9736	174	1.79	0.0021mm	1.00	14371

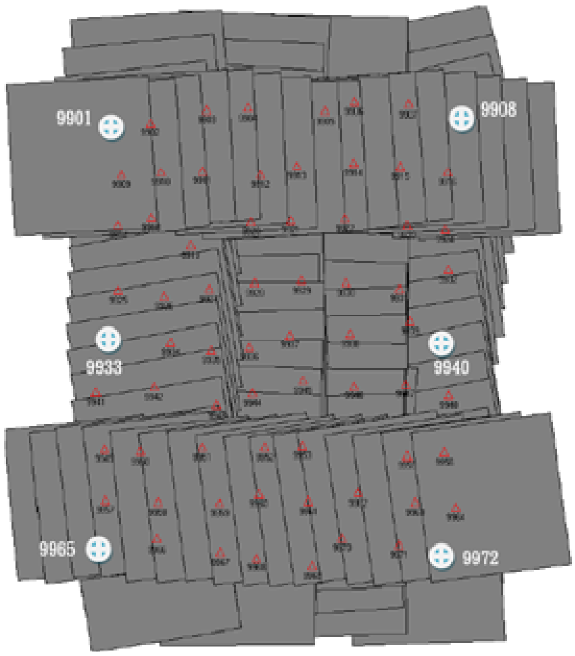

Figure 5. Coverage of RMK DX images, locations of control points and check points

4. CONCLUSIONS

The accuracy of photo coordinates of image points extracted and matched by the proposed new method of SIFT-based multi-image matching reaches a subpixel level and the accuracy of aerial triangulation is 0.21 pixels (the pixel size of large format aerial images taken with RMK DX camera is 7.2μm). The automatic extraction, selection and measurement of corresponding tie points is done efficiently, especially *without the need on priori knowledge on image overlap* information such as block parameters and strip parameters. The result also shows that a larger threshold of distance ratio adopted by SIFT causes more incorrect measurements. But due to the robustness of bundle block adjustment with data snooping operation, the errors can be detected and eliminated. Also, the efficiency and benefit of AFTP and QF processing have been verified. These two pre-process procedures still can be improved on efficiency and feasibility, especially the quality indicator for QF.

Table 4. Results of Bundle Block Adjustments done by the new method and LPS2010

Tie Points \ Residuals of Measurements	RMS(1/1000mm)		MAX(1/1000mm)	
	x	y	x	y
New method	0.8	1.2	5.8	5.9
LPS 2010	1.8	1.7	7.2	7.5

Tie Points \ Control Point Residuals	RMS(mm)			MAX(mm)		
	X	Y	Z	X	Y	Z
New method	11	8	112	17	11	142
LPS 2010	21	2	188	32	3	142

Tie Points \ Precision Values of Orientations	RMS(X,Y,Z-mm；ψ,ω,κ-1/1000 gon)					
	X	Y	Z	ψ	ω	κ
New method	70	67	103	4.1	3.9	0.4
LPS 2010	123	121	166	7.3	7.2	0.9

Tie Points \ Check Point differences	RMS(mm)		
	X	Y	Z
New method	28	19	119
LPS 2010	31	20	214

5. REFERENCES

Baarda, W., 1968. A Testing Procedure for Use in Geodetic Networks. Netherlands Geodetic Commission, Vol. 2, No. 5.

Heipke, C., 1997. Special issue: Automatic image orientation - Preface. ISPRS Journal of Photogrammetry and Remote Sensing, 52(3), 101-102.

Kruck, E., 1984. Ordering and Solution of Large Normal Equation Systems for Simultaneous Geodetic and Photogrammetric Adjustment. 15th ISPRS Congress, 17-29 June, 1984, Rio de Janeiro, BRASIL.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110.

Schenk, T., 2004. From point-based to feature-based aerial triangulation. ISPRS Journal of Photogrammetry and Remote Sensing, 58(5-6), 315-329.