AEJ
Alexandria Engineering Journal

Automatic extraction and matching of interest points on stereo aerial imagery

Mohamed I. Zahran
Surveying and Photogrammetry, Faculty of Eng. at Shoubra, Benha University, Benha, Egypt

This research presents a proposed strategy of automatic extraction and matching of interest points on stereo aerial imagery. Given a stereo image pair, the strategy starts with applying an interest point operator on both images, followed by suppression of local non-maxima. Each interest point in the master image is then matched against candidate interest points in the slave image. The matching entities are the normalized central moments of orders two and three of interest points. A similarity measure, between the master interest point and each slave candidate point, is estimated based on computed moments. Correct matches are indicated by minimum value of the measure. A stereo pair of digital aerial images covering an urban area is used for the research experimentation. They are standard aerial photographs that are scanned with a 100 dpi scanning resolution, leading to nearly 254-μm pixel size. The experimentation is carried out on the overlap area exhibited in two 450 pixels by 700 pixels patches of the image pair. A number of 1325 final matches are resulted at applying the presented matching approach. By employing the image coordinates of matched points in a relative orientation process, a global matching precision of 0.6 pixel is yielded.

يستعرض البحث اسلوب مقترح للاستخلاص والتوافق الأتوماتيكي للنقاط المميزة علي الصور الجوية المتداخلة. يبدأ هذا الأسلوب بتطبيق مستخلص للنقاط المميزة على كلا من الصورتين المتداخلتين متبوعا بعملية استبعاد للنقاط الأقل تمييزا داخل مساحة محدودة علي الصورة. بعد ذلك يتم عمل توافق لكل نقطة مستخلصة في الصورة الأساسية مع النقاط المرشحة في الصورة التابعة. تستخدم قيم العزوم المركزية ذات الدرجة الثانية والثالثة المحسوبة للنقاط المستخلصة في حساب معيار للتماثل بين كل نقطتين نبحث توافقهما. تم اجراء الجزء التجريبي في البحث بالاستعانة بزوج من الصور الجوية الرقمية المتداخلة تغطي منطقة حضرية وممسوحة رقميا بقدرة مسح تبلغ مائة نقطة في البوصة المربعة. تمت التجارب على جزء من منطقة التداخل تبلغ مساحته ٤٥٠ نقطة عرضا و٧٠٠ نقطة طولا وانتهت الى توافق ١٣٢٥ من النقاط المستخلصة في هذا الجزء. لقياس جودة التوافق وظفت احداثيات النقاط المتوافقة في اجراء توجيه نسبي حيث تم التوصل الى دقة توافق شاملة تبلغ ٠,٦ نقطة.

Keywords: Automated photogrammetry, Image matching, Object recognition, Moment invariants, Image orientation

1. Introduction

Both correlation matching and least-squares matching techniques cannot be reliably used in featureless areas of nearly homogenous brightness or in discontinuous areas with severe scale changes. As a result, matching a grid of points that is selected blindly leads to a waste of time and many false results as well. Therefore, a significant preprocessing phase before image matching is to extract eligible matching features, based on the image texture and contrast. This is necessary for reliable matching results. Such features can be points, lines and shapes. They are matched relying on the similarity of their attributes.

Point features are usually referred to as interest points. Interest points are well-defined points or corners, found by using an interest point operator. Many point operators have been developed in the domains of computer vision and pattern recognition [4, 9]. Point operators don't rely only on spatial information but also apply statistics to determine how well-defined the position of a particular point is. Detected interest points in stereo scenes are often matched in order to create the shape of the imaged surface. Automated detection and matching of interest points play a major role in automating most photogrammetric processes, especially triangulation and generation of digital terrain model [1, 2, 6].

Alexandria Engineering Journal, Vol. 46 (2007), No. 5, 759-766

For the purpose of matching, detected interest points require an object description technique providing the characteristics that uniquely represent their form. Object moments of different order are combined and adopted in this research to characterize extracted point features. Moment invariants are those quantities computed from object moments and are independent to object translation, scale and orientation. Translation invariance is achieved by computing moments that are normalized with respect to the object centre of gravity. Size invariant moments are derived from algebraic invariants. A set of moment invariants that are independent of rotation can be computed from the second and third order values of the normalized central moments.

2. Interest point operators

This section presents three interest point operators. The first two operators are the well-known Moravec's operator [6] and Forstner's operator [7]. The third one, which has some advantages over those two operators, is named as the ground operator [8]. Moravec's operator is the first interest point operator developed. It searches for points that have high variance between adjacent pixels. Here, the image is divided into 4 pixels by 4 pixels or 8 pixels by 8 pixels windows. The minimum of the variances in the horizontal, vertical and both diagonal directions is calculated for each window. Variances are determined by summing up the differences of gray levels of adjacent pixels along each of the four directions, respectively. Windows with scores exceeding a set threshold are considered interest points. The threshold is selected empirically to control the number of resulted interest points. Finally, suppression of local non-maxima is carried out to avoid clusters of points in highly-textured area.

Forstner's operator implements four steps to extract distinct points and corners. It first calculates the Robert gradients and matrix N (see eq. 1) for each window within the image. Recommended window sizes are 5 pixels by 5 pixels or 7 pixels by 7 pixels. The terms g_u and g_v in the equation are the gradient components in the diagonal directions.

$$N = \begin{bmatrix} \sum g_u^2 & \sum g_u g_v \\ \sum g_u g_v & \sum g_v^2 \end{bmatrix}. \quad (1)$$

Second, for each window the interest values q and w (weight) are calculated as follows:

$$q = \frac{4\,Det\,N}{(tr\,N)^2}. \quad (2)$$

$$w = \frac{Det\,N}{tr\,N}. \quad (3)$$

Third, windows having values of q and w exceeding certain thresholds indicate interest points. The threshold for q ranges from 0.5 to 0.75, whereas the range of the threshold for w is between $0.5\,w_{mean}$ and $1.5\,w_{mean}$. The magnitude of w_{mean} is the average of w values of all window positions in the image. At last, suppression of local non-maxima is carried out in a spiral approach in order to obtain fairly distributed interest points.

In applying both Moravec's and Forstner's operators, the process of search and computation is done point by point allover the entire image, which takes considerable time. In addition, reaching proper thresholds for Forstner's operator is somewhat difficult and time-consuming.

The ground operator, which has two versions, has rather simple form. In its first version, the gradient to each of the four neighboring pixels is computed at each point. If at least two of the absolutes of the gradients are greater than a set threshold, the point is marked as an interest point candidate. An Interest value for each candidate is calculated as the sum of the absolute differences of gray levels between the central pixel and the eight neighboring pixels. Again, suppression of local non-maxima is executed such that if an interest value within the suppression window exceeds the value of the window center, the center's value is set to zero. The size of the suppression window is selected according to the required density of interest points.

The second version of the ground operator resembles Forstner's operators with

improvements in speed and accuracy [8]. Here, interest point candidates are found as in the first version. However, the interest value q for each candidate is calculated as in Forstner's operators using eq. (1 and 2), but using only 8 adjacent points. The threshold for w is no longer necessary and a threshold ranging from 0.32 to 0.5 is used for q. Windows having q-values exceeding the set threshold indicate interest points. Lastly, suppression of local non-maxima is done similar to the first version.

3. Moments for feature matching

The coefficients determined by approximation of a figure in terms of some basis functions may be used as analytical shape descriptors. Those coefficients may be combined to attain invariance to position, scale and rotation. One of the popular methods in this regard is the moment theory-based method that uses region-based moments to characterize the shape of an object with a set of parameters that are invariant to geometric changes.

The ij^{th} order moment of a given figure S is defined as:

$$m_{ij} = \sum_{x \in S} \sum_{y \in S} x^i y^j . \tag{4}$$

In eq. 4, the coordinates (x,y) are related to a point inside or on the boundary of the figure. The zero order moment m_{00} gives the number of points (or the area) of the figure. Also, m_{10} and m_{01} yield the position of its centroid.

The moments can be made invariant to position of the figure by translating the origin to the centroid leading to the following new coefficients (central moments):

$$\mu_{ij} = \sum_{x \in S} \sum_{y \in S} (x - \bar{x})^i (y - \bar{y})^j , \tag{5}$$

where $\bar{x} = m_{10} / m_{00}$ and $\bar{y} = m_{01} / m_{00}$. Invariance to scale is given by using the following formulae:

$$\mu'_{ij} = \frac{\mu_{ij}}{\mu_{00}^{[(i+j)/2]+1}} . \tag{6}$$

Invariance to rotation is obtained by rotating the coordinate axes by an angle θ that is given as [9]:

$$\tan 2\theta = \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} . \tag{7}$$

In the case where the figure is a patch on a gray-scale image, the coefficients of eq. (5) turn into $\mu_{ij} = \sum_{x \in S} \sum_{y \in S} (x - \bar{x})^i (y - \bar{y})^j g(x,y)$, where $g(x,y)$ is the normalized image brightness [10].

A set of functions of second and third order moments that are invariant to rotation and reflection can be given as [9]:

$$I_0 = \mu_{20} + \mu_{02} . \tag{8}$$

$$I_1 = (\mu_{20} - \mu_{02})^2 + 4\mu^2_{11} . \tag{9}$$

$$I_2 = (\mu_{30} + 3\mu_{12})^2 + (3\mu_{21} + \mu_{03})^2 . \tag{10}$$

$$I_3 = (\mu_{30} + \mu_{12})^2 + (3\mu_{21} + \mu_{03})^2 . \tag{11}$$

$$\begin{aligned} I_4 = {} & (\mu_{30} - 3\mu_{12})^2 (\mu_{30} + \mu_{12})^2 [(\mu_{30} + \mu_{12})^2 \\ & - 3(\mu_{21} + \mu_{03})^2] \\ & + (\mu_{21} - 3\mu_{03})^2 (\mu_{21} + \mu_{03})^2 [3(\mu_{30} + \mu_{12})^2 \\ & - (\mu_{21} + \mu_{03})^2] . \end{aligned} \tag{12}$$

$$\begin{aligned} I_5 = {} & (\mu_{20} - 3\mu_{02})^2 [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ & + 4\mu_{11} (\mu_{30} + \mu_{12})^2 (\mu_{21} + \mu_{03})^2 . \end{aligned} \tag{13}$$

$$\begin{aligned} I_6 = {} & (3\mu_{21} - \mu_{03})^2 (\mu_{30} + \mu_{12})^2 [(\mu_{30} + \mu_{12})^2 \\ & - 3(\mu_{21} + \mu_{03})^2] \\ & + (3\mu_{12} - \mu_{30})^2 (\mu_{21} + \mu_{03})^2 [3(\mu_{30} + \mu_{12})^2 \\ & - (\mu_{21} + \mu_{03})^2] . \end{aligned} \tag{14}$$

Moment invariants can be utilized to form a multi-dimensional parameter space on which the correspondence between features of stereo imagery is based. Features u on one image and v on the other image of a stereo pair might be compared by the similarity measure DI_{uv} as follows [10]:

$$DI_{uv} = \sum_{k=0}^{6} \frac{\left| I_k^L(u) - I_k^R(v) \right|}{\max(I_k^L(u) , I_k^R(v))}, \quad (15)$$

where $I_k^L(u)$ = value of k^{th} invariant of the feature u on the left image, and

$I_k^R(v)$ = value of k^{th} invariant of the feature v on the right image.

Similarly, the similarity measure DM_{uv}, in the case of using n central moments instead, can be expressed as:

$$DM_{uv} = \sum_{k=1}^{n} \frac{\left| \mu_k^L(u) - \mu_k^R(v) \right|}{\max(\mu_k^L(u) , \mu_k^R(v))}, \quad (16)$$

where $\mu_k^L(u)$ = value of k^{th} central moment of the feature u, and $\mu_k^R(v)$ = value of k^{th} central moment of the feature v.

Given a feature m in the left image and its matching candidates n_1, n_2, ..., n_k in the right image, the conjugate feature will have the minimum of similarity measure values computed between m and each of n_1, n_2, ..., n_k.

4. Methodology

This section presents a proposed strategy of automatic extraction and matching of interest points on stereo aerial imagery. Given a stereo image pair, the strategy starts with applying the second version of ground interest operator on both images utilizing appropriate threshold. The amount of resulted interest points is reduced through suppression of local non-maxima. Each interest point in the master image is then matched against candidate interest points in the slave image. Here, all interest points located within a specified search window in the slave image are considered to be matching candidates. The size of the search window can be specified using available information regarding the image formation process.

The matching process begins with the computation of normalized central moments of orders two and three for a square window with suitable size around the master interest point and around each matching candidate point in the slave image. Having significant rotation between the images, moment invariants are rather calculated for the surrounding windows. The similarity measure is computed between the master interest point and each slave candidate point. The minimum value of the measure, if it passes a set threshold, indicates the correct match. The matching process is continued for each interest point extracted in the master image. The entire procedure is implemented by developing prototype software in the MATLAB environment.

In working with the proposed strategy, some difficulties have occurred and solutions are proposed. One example when more than one point in the master image match the same point in the slave image. Here, the couple with the minimum dissimilarity measure can be chosen as the correct match. Also, in the case of having a matched pair with significant y-parallax compared with the parallaxes of neighboring matched pairs, it has to be discarded. Moreover, some matches might result with almost similar value of dissimilarity measure. At this point, instead of choosing the match with the minimum measure value, it is suggested to reject those matches. Such concerns are actually considered during the implementation stage.

The global quality of matching process is assessed through an analytical relative orientation process utilizing matched pairs of interest points. The resulted precision of estimated parameters is taken as a measure for the matching quality. In the process of relative orientation, the relative angular attitude and positional displacement between the images are determined at the time of exposure. Here, the rotation angles ω_1, φ_1, κ_1 and the perspective center coordinates X_{L1} and Y_{L1} regarding the left image are set equal to zero values. Also, the flying height Z_{L1} of the left image and the perspective center X_{L2} of the

right image are set equal to the focal length (f), and the photo base (b), respectively. Consequently, the remaining five parameters of the right image are the unknowns to be resolved.

5. Experimentation and results

A test pair of digital aerial images covering an urban area is used for the research experimentation (see figs. 1 and 2). They are standard aerial photographs that are scanned with a 100 dot per inch (dpi) scanning resolution, leading to nearly 254-μm-pixel size [11]. The camera used in acquiring the images has a focal length of 153.380 mm. The flying height is nearly 830 m above datum. The experimentation is carried out on the overlap area exhibited in the two 450 pixels by 700 pixels patches shown in fig. 3. The two patches are cut out from the original test images such that there are shifts of few pixels in the coordinates of corresponding feature points.

The proposed strategy that is presented in section 4 is applied on the two test patches. First, the ground interest operator, described in section 2, is applied on both image patches such that a threshold of 40 is specified for marking a point as an interest point candidate. Accordingly, numbers of 46695 and 43230 point candidates are obtained in the left and right patches, respectively. The resulted patches including those candidates are displayed in fig. 4. After suppression of local non-maxima using 3 pixels by 3 pixels window, the numbers of detected interest points turn into 10761 in the left patch and 9681 in the right patch (see fig. 5).

In order to find matches for an interest point in the left patch, all interest points located within a 25-pixels by 25-pixels window in the right patch are considered to be matching candidates. The size of the window is chosen depending on the elevation range of the photographed area. Normalized central moments of orders two and three are estimated for 11-pixels by 11-pixels windows around the master interest point and each of its matching candidate points in the slave image. Consequently, the moment similarity measure, given by eq. (16), is computed between the master point and each candidate point. If the minimum of computed measure values passes a threshold of 3, it points to a match. Matching is executed for each interest point detected in the left patch. A number of 1325 matches are resulted by the end of the matching process. Table 1 list the number of detected interest points in the two test patches resulted by applying the three phases of the proposed strategy.

Fig. 1. The left image of the test pair.

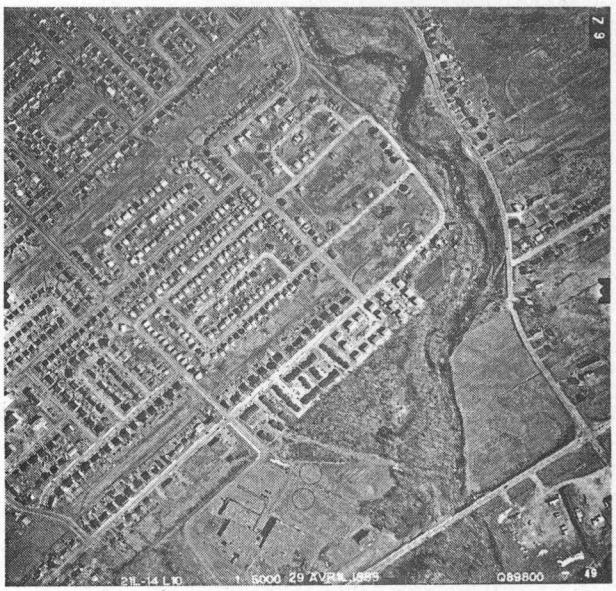

Fig. 2. The right image of the test pair.

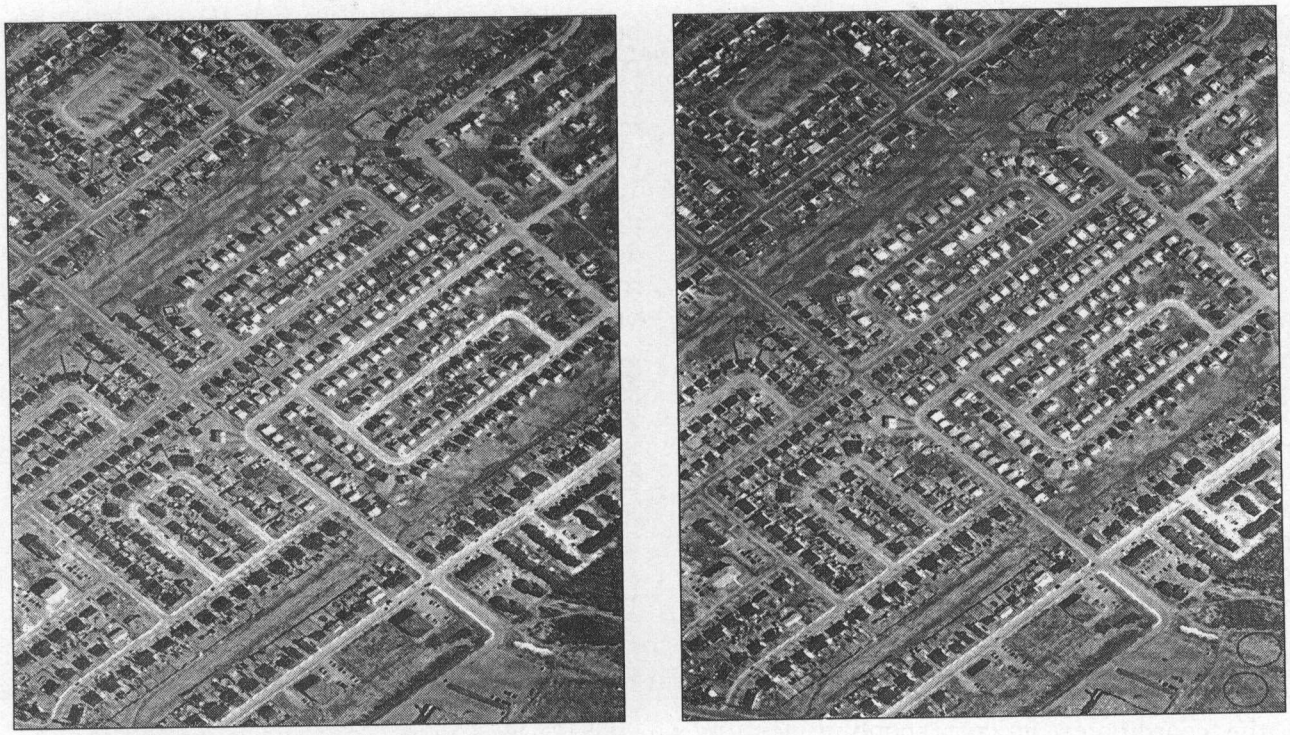

Fig. 3. The left and right image patches of the test pair.

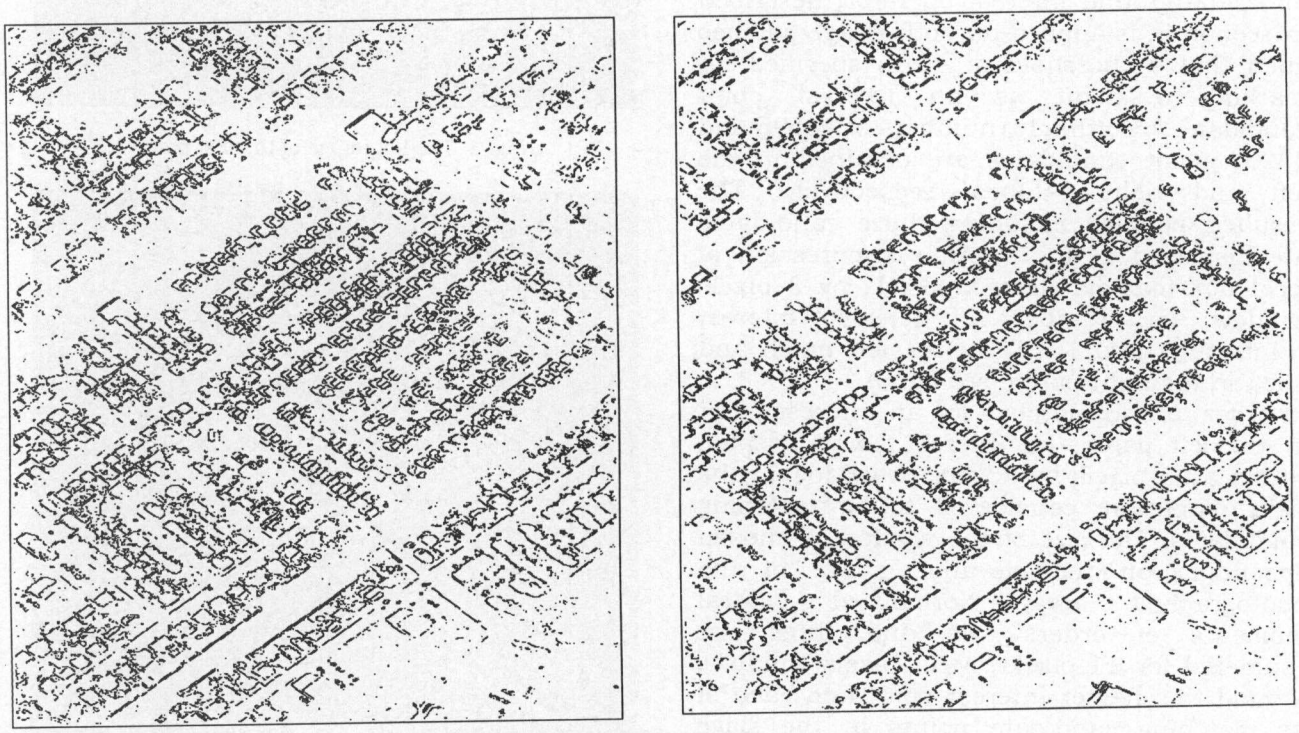

Fig. 4. All interest points detected in the two image patches.

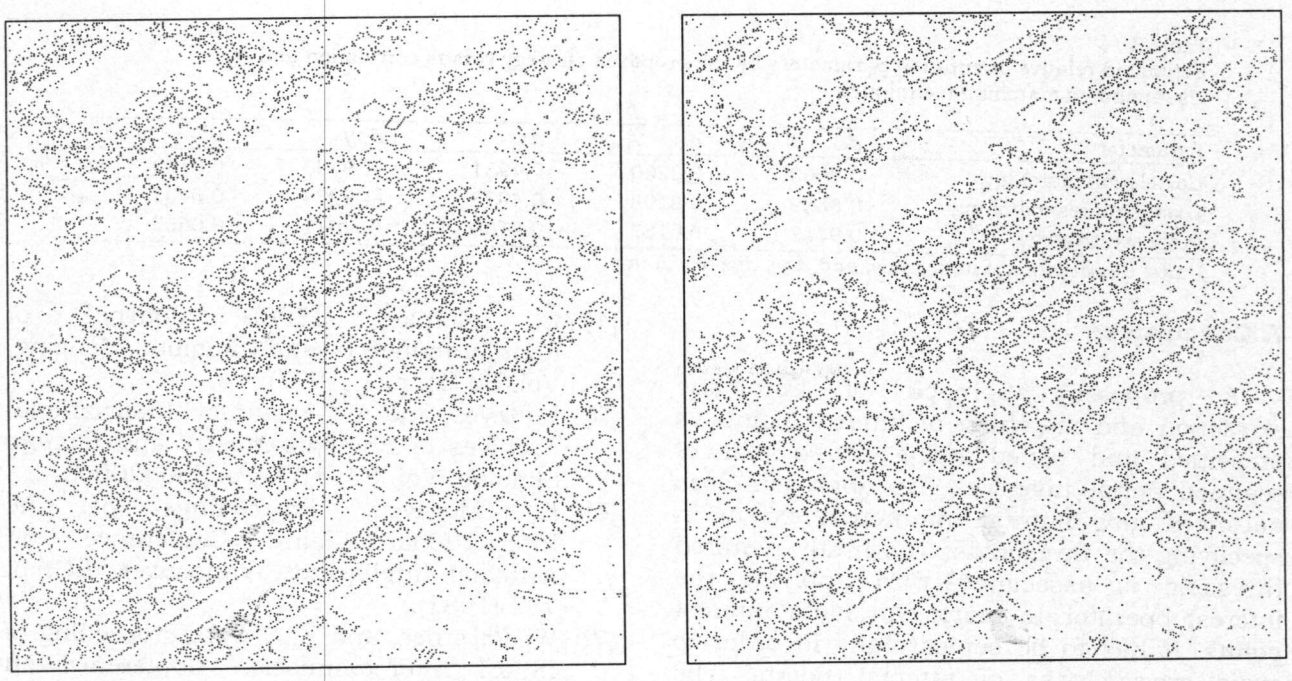

Fig. 5. Interest points detected in the two patches after suppression.

For each of the two image patches, an affine transformation is performed using the calibrated coordinates of fiducial marks, as well as their measured pixel coordinates, through a least-squares procedure. The resulted transformation parameters are employed to convert from the pixel coordinate system to the related image coordinate system, centered at the image principal point. Accordingly, pixel coordinates of matched points in both the left and the right patches, found by applying the proposed matching technique, are converted into equivalent image coordinates.

The converted coordinates of matched points in the two patches are utilized to solve for relative orientation parameters of the test image pair, as explained in section 4. The resulted standard errors of estimated orientation parameters are shown in table 2. The related estimate of standard error of unit weight precision is 0.150 mm, which is equivalent to a fraction of a pixel (0.6 pixel). This indicates a global matching precision that is better than the resolution of used imagery. table 3 lists the estimated relative orientation parameters by the proposed strategy as well as corresponding estimates resulted by correlation and least-squares matching techniques. The estimates resulted by those two techniques are obtained from an earlier work [12]. It was based on matching a set of fifteen distinct image points employing a finer resolution (600 dpi) version of the test images.

Table 1
Number of detected points in the two test patches in each phase

Strategy phase	Left patch	Right patch
Interest point detection	46695	43230
Non-maxima suppresion	10761	9681
Interest point matching	1325	1325

Table 2
Standard errors of estimated relative orientation parameters

Parameter	Dω	$d\varphi$	$d\kappa$	dby	dbz
Std. Error	0.0477	0.0688	0.0202	0.1373	0.0746

Units: $d\omega$, $d\varphi$, $d\kappa$ are in degrees; and dby, dbz are in mm.

Table 3
Estimated relative orientation parameters by the proposed strategy versus correlation and least-squares matching techniques

Parameter	$d\omega$	$d\varphi$	$d\kappa$	dby	dbz
Correlation matching	0.8903	0.0260	0.4951	-1.9333	0.0100
Least-squares matching	0.8889	0.0108	0.4975	-1.9274	0.0403
Proposed strategy	0.9137	0.0187	0.3708	-2.2811	0.0605

Units: $d\omega$, $d\varphi$, $d\kappa$ are in degrees; and dby, dbz are in mm.

7. Conclusions

A proposed approach of automatic extraction and matching of interest points is presented and tested using stereo pair of scanned aerial imagery. The approach does not need any manual interaction except in specifying the thresholds. The firstly required threshold is associated with applying the interest operator to mark candidate interest points. It has to be set carefully in order to avoid crowdedness of interest points. The secondly needed threshold is to set a minimum value of similarity measure to indicate a match. Multiple matches as well as obviously erroneous matches are handled instantaneously by the proposed approach. The approach has yielded a good global matching precision that reaches a fraction of a pixel.

References

[1] E.M. Mikail, J.S. Bethel and J.C. McGlone, Introduction to Modern Photogrammetry, John Wiley and Sons, Inc., New York, p. 186 (2001).

[2] R.M. Haralic, and L.G. Shapiro, Computer and Robot Vision, Addison-Wesley, Inc., Reading, MA (1992).

[3] C. Heipke and K. Eder, OEEPE Official Publication, No. 35 (1998).

[4] C. Heipke, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 52 (1), p. 1 (1997).

[5] L. Keyes, and A. Winstanley, International Archives of Photogrammetry and Remote Sensing, Vol. 33, B3/1, p. 580 (2000).

[6] H. Moravec, Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver, BC, p. 785 (1981).

[7] W. Forstner, and E. Gulch, Proceeding of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, p. 281 (1987).

[8] L. Yan, International Archives of Photogrammetry and Remote Sensing, Vol. 27, B3 (1988).

[9] R. Nevatia, Machine Vision, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, p. 69 (1982).

[10] Y.B. Blokhinov and D.A. Gribove, International Archives of Photogrammetry and Remote Sensing, Vol. 35, B3 (2004).

[11] DVP Geomatic Systems, DVP Software, DVP Geomatic Systems, Inc. (1997).

[12] M.I. Zahran, Scientific Bulletin of the Faculty of Engineering, Ain Shams University, Vol. 41 (2), p. 207 (2006).

Received December 28, 2006
Accepted September 3, 2007