COMPARATIVE ANALYSIS OF AREA-BASED IMAGE MATCHING TECHNIQUES FOR HIGH RESOLUTION SATELLITE IMAGERY

T. Fuse [a], C. S. Fraser [b], P. Dare [c]

[a] Dept. of Civil Engineering, University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo 113-8656, Japan, - fuse@civil.t.u-tokyo.ac.jp
[b] Dept. of Geomatics, University of Melbourne, Victoria 3010, Australia, - c.fraser@unimelb.edu.au
[c] Airborne Research Australia, Flinders University, Salisbury South, South Australia 5106, Australia, - Paul.Dare@AirborneResearch.com.au

Commission III, WG III/8

KEY WORDS: Image Matching, Image Processing, High resolution Satellite Imagery (HRSI), IKONOS, Accuracy Comparison

ABSTRACT:

Recently, it has become much easier to obtain commercial high-resolution satellite imagery (HRSI), and application of IKONOS satellite imagery to 3D positioning has recently been investigated, with sub-meter accuracy being achieved. Image matching is one of the most fundamental processes in geopositioning from multi-image HRSI. Least squares matching and cross correlation matching are the most basic and popular techniques applied by photogrammetrists. These techniques, however, have most often been used independently. This paper presents the initial results of a comparative analysis of these two image matching techniques. Initially, the relationship between the two techniques is described. Maximising the cross correlation coefficient is equal to minimising the sum of squared distances between observed points and a regression line in feature space. This interpretation leads to principal component regression. By using principal components, the cross correlation coefficient can be maximised in the same way as in least-squares estimation. Accordingly, we develop a cross correlation matching which includes not only translation but also deformation (affine and projective transformation). The least squares matching and cross correlation matching are applied to an IKONOS stereo pair, and the characteristics and results of applying the matching techniques are discussed.

1. INTRODUCTION

Recently, it has become much easier to obtain commercial high-resolution satellite imagery (HRSI), and application of IKONOS satellite imagery to 3D positioning has recently been investigated, with sub-meter accuracy being achieved (e.g. Fraser et al, 2002 and Fraser & Hanley, 2003).

Image matching is one of the most fundamental processes in geopositioning from multi-image HRSI. So far, various techniques have been developed in order to automatically match multiple images (Brown, 1992). These techniques can also be applied to camera stabilization, object detection, the tracking of moving objects, camera motion determination, image compression, and so on. Least-squares matching and cross correlation matching, as area-based approaches, are the most basic and popular techniques applied by photogrammetrists. These techniques, however, have most often been used independently.

Least-squares matching is a mature development and it has proven to be a most powerful matching technique (Gruen, 1985, Gruen and Baltsavias, 1988). The method can deal with scene deformation. On the other hand, cross correlation matching includes only translation. Accordingly, cross correlation matching has an obvious limitation.

This paper presents the initial results of a comparative analysis of these two image matching techniques. Initially, the relationship between the two is described. Based on this relationship, the cross correlation coefficient can be maximised in the same way as in least-squares estimation. Accordingly, we develop a cross correlation matching which includes not only translation but also deformation. It is important to point out here that our final goal is a contribution to the matching of multi-temporal and multi-resolution images, be they IKONOS, QuickBird or aerial images.

2. LEAST-SQUARES MATCHING

2.1 Least-Squares Approach

Assumed are two image windows given as discrete functions $f(x, y)$, $g(x', y')$, where f is a master image, g is a slave image and x, y, x', y' are the image coordinate with window size of M x N pixels. Objective function of the least squares matching is as follows:

$$E_1 = \sum_{i=1}^{M}\sum_{j=1}^{N} e^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(f\left(x_i, y_j\right) - g\left(x'_i, y'_j\right)\right)^2 \rightarrow \min. \quad (1)$$

Affine transformation is applied with respect to coordinates (x, y) and (x', y') (Gruen, 1985, Gruen and Baltsavias, 1988):

$$x' = ax + by + c, \; y' = dx + ey + f. \quad (2)$$

Projective transformation is also applied (Szeliski, 1996):

$$x' = \frac{ax + by + c}{px + qy + 1}, \quad y' = \frac{dx + ey + f}{px + qy + 1} \quad (3)$$

2.2 Optimization Method

The Gauss-Newton method has been widely used to perform the minimization of Equation 1, since the equation represents an unconstrained nonlinear optimization problem. The method, however, does not display good convergence properties when the functions have inflection points. The Levenberg-Marquardt method works very well in practice and has been becoming a standard for nonlinear least-squares routines (Press et al, 1988). The method requires computation of the partial derivatives of e_i with respect to the unknown parameters p_i (affine: i = 1 – 6, projective: i = 1 – 8),

$$\frac{\partial e_i}{\partial p_k} = \frac{\partial g}{\partial x'}\frac{\partial x'}{\partial p_k} + \frac{\partial g}{\partial y'}\frac{\partial y'}{\partial p_k}. \tag{4}$$

When the affine transformation is applied, the partial derivatives are derived as follows:

$$\begin{aligned} &\frac{\partial e_i}{\partial a} = \frac{\partial g}{\partial x'}x_i, \quad \frac{\partial e_i}{\partial b} = \frac{\partial g}{\partial y'}y_i, \quad \frac{\partial e_i}{\partial c} = \frac{\partial g}{\partial x'}, \\ &\frac{\partial e_i}{\partial d} = \frac{\partial g}{\partial y'}x_i, \quad \frac{\partial e_i}{\partial e} = \frac{\partial g}{\partial x'}y_i, \quad \frac{\partial e_i}{\partial f} = \frac{\partial g}{\partial y'}. \end{aligned} \tag{5}$$

In the case of projective transformation they become:

$$\begin{aligned} &\frac{\partial e_i}{\partial a} = \frac{\partial g}{\partial x'}\frac{x_i}{D_i}, \quad \frac{\partial e_i}{\partial b} = \frac{\partial g}{\partial x'}\frac{y_i}{D_i}, \quad \frac{\partial e_i}{\partial c} = \frac{\partial g}{\partial x'}\frac{1}{D_i}, \\ &\frac{\partial e_i}{\partial d} = \frac{\partial g}{\partial y'}\frac{x_i}{D_i}, \quad \frac{\partial e_i}{\partial e} = \frac{\partial g}{\partial y'}\frac{y_i}{D_i}, \quad \frac{\partial e_i}{\partial f} = \frac{\partial g}{\partial y'}\frac{1}{D_i}, \\ &\frac{\partial e_i}{\partial p} = -\frac{x_i}{D_i}\left(x'\frac{\partial g}{\partial x'} + y'\frac{\partial g}{\partial y'}\right), \quad \frac{\partial e_i}{\partial q} = -\frac{y_i}{D_i}\left(x'\frac{\partial g}{\partial x'} + y'\frac{\partial g}{\partial y'}\right), \end{aligned} \tag{6}$$

where $D_i = px_i + qy_i + 1$.

To compute partial derivatives $\frac{\partial g}{\partial x'}$ and $\frac{\partial g}{\partial y'}$, a consistent gradient operator (Ando, 2000) is applied. When the window size is set as 3 x 3, the gradient operator with respect to x and y is represented as follows:

$$\begin{bmatrix} -0.112737 & 0 & 0.112737 \\ -0.274526 & 0 & 0.274526 \\ -0.112737 & 0 & 0.112737 \end{bmatrix},$$

$$\begin{bmatrix} -0.112737 & -0.274526 & -0.112737 \\ 0 & 0 & 0 \\ 0.112737 & 0.274526 & 0.112737 \end{bmatrix}. \tag{7}$$

From these partial derivatives, the Levenberg-Marquardt algorithm computes an approximate Hessian matrix **A** and the weighted gradient vector **b** with components

$$a_{kl} = \sum_i \frac{\partial e_i}{\partial p_k}\frac{\partial e_i}{\partial p_l}, \quad b_k = -\sum_i e_i \frac{\partial e_i}{\partial p_k} \tag{8}$$

and the following linear equation is obtained:

$$\left(\mathbf{A} + \lambda \mathbf{I}\right)\delta\mathbf{p} = \mathbf{b}, \tag{9}$$

where λ is a time-varying stabilization parameter, and **I** is the identity matrix. As λ approaches 0, Equation 4 goes over to the Gauss-Newton method. On the other hand, when λ is very large, it corresponds to the steepest descent method. Normally, an initial value of λ is given as around 0.001 and the parameter estimate **p** is updated by an amount

$$\delta\mathbf{p} = \left(\left(\mathbf{A} + \lambda\mathbf{I}\right)^t\left(\mathbf{A} + \lambda\mathbf{I}\right)\right)^{-1}\left(\mathbf{A} + \lambda\mathbf{I}\right)^t\mathbf{b}. \tag{10}$$

3. CROSS CORRELATION MATCHING

3.1 Relationship between Least-Squares Matching and Cross Correlation Matching

The function of Equation 1 can be expanded as follows,

$$\sum_{i=1}^{M}\sum_{j=1}^{N}\left(f^2\left(x_i, y_j\right) - 2f\left(x_i, y_j\right)g\left(x'_i, y'_j\right) + g^2\left(x'_i, y'_j\right)\right). \tag{11}$$

Since the first term is constant, it can be ignored for the moment. The third term depends upon (x', y'). If it is constructed such that the value depending upon it is not to change, the normalized cross correlation can be obtained.

$$\gamma = \frac{\sigma_{MS}}{\sigma_M \sigma_S} \tag{12}$$

where

$$\begin{aligned} \sigma_M &= \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(f\left(x_i, y_j\right) - \overline{f}\right)^2}{MN - 1}}, \\ \sigma_S &= \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(g\left(x'_i, y'_j\right) - \overline{g}\right)^2}{MN - 1}}, \\ \sigma_{MS} &= \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(f\left(x_i, y_j\right) - f\right)\left(g\left(x'_i, y'_j\right) - \overline{g}\right)}{MN - 1}, \end{aligned} \tag{13}$$

and $\overline{f}$, $\overline{g}$ are the means of the intensity of the master and slave image, respectively. When the evaluation value for optimization is set as per Equation 9, we can apply cross correlation matching.

3.2 Signal-to-Noise Ratio

Consider the cross correlation between f and g, with white noise n:

$$\sum_{i=1}^{M}\sum_{j=1}^{N}\left(f\left(x_i, y_j\right)g\left(x'_i, y'_j\right) + f\left(x_i, y_j\right)\cdot n_i\right) = \mathbf{f}^t\mathbf{g} + \mathbf{f}^t\mathbf{n}. \tag{14}$$

The signal-to-noise ratio (SN ratio) R is computed as follows:

$$R = \frac{\left(\mathbf{f}^t\mathbf{g}\right)^2}{E\left\{\left(\mathbf{f}^t\mathbf{n}\right)\left(\mathbf{f}^t\mathbf{n}\right)^T\right\}} = \frac{\left(\mathbf{f}^t\mathbf{g}\right)^2}{\mathbf{f}^t\boldsymbol{\Sigma}_n\mathbf{f}}. \quad (15)$$

The maximum SN ratio means that optimal matching is accomplished. From $\frac{\partial R}{\partial \mathbf{f}} = 0$, optimal matching is achieved when

$$\mathbf{f} = \left(\frac{\mathbf{f}^t\boldsymbol{\Sigma}_n\mathbf{f}}{\mathbf{f}^t\mathbf{g}}\right)\boldsymbol{\Sigma}_n^{-1}\mathbf{f} = \left(\frac{\mathbf{f}^t\mathbf{f}}{\mathbf{f}^t\mathbf{g}}\right)\mathbf{f}. \quad (16)$$

If the slave image **g** is equivalent to the master image **f**, the SN ratio becomes a maximum. The cross correlation is then optimal in the sense of the SN ratio, and is robust against noise.

3.3 Expansion of Regression Model

So far, since it has been assumed that the errors exist only in the slave image (*g*-axis), a minimization of the sum of squared differences from a regression line in the *g*-axis has been performed (Figure 1).

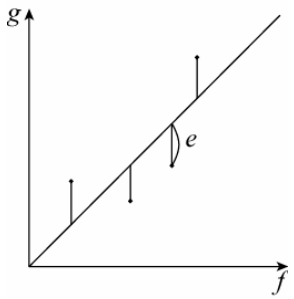

Figure 1. Errors only in slave image

However, it is plausible that the errors exist in both the master and slave image. In such a case, we have to minimize the sum of squared distances from observed points to the regression line (Figure 2).

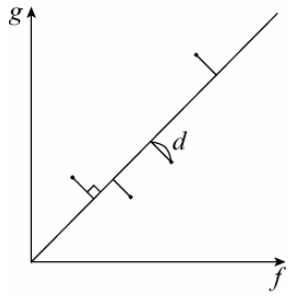

Figure 2. Errors in both master and slave images

This kind of regression model is known as principal component regression (Greene, 2000). The principal component regression is applied to the intensity values of both master and slave images:

$$\mathbf{X} = \begin{bmatrix} f(x_1, y_1) & f(x_2, y_2) & \cdots & f(x_{MN}, y_{MN}) \\ g(x'_1, y'_1) & g(x'_2, y'_2) & \cdots & g(x'_{MN}, y'_{MN}) \end{bmatrix} \quad (17)$$

$$\mathbf{z} = \mathbf{PX} \quad (18)$$

where, **z** is the principal component vector, $\mathbf{P} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$ is an orthogonal matrix consisting of the eigen vectors of $\mathbf{XX}^t$. Since only the z_2-axis is related to minimization of $\sum d^2$, the objective function of the principal component regression is formulated as follows:

$$E_2 = \sum_{i=1}^{M}\sum_{j=1}^{N} d^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(w_{21} f(x_i, y_j) + w_{22} g(x'_i, y'_j)\right)^2 \rightarrow \min. \quad (19)$$

To perform the minimization, the Levenberg-Marquardt method is again applied:

$$\left(\mathbf{A} + \lambda\mathbf{I}\right)\delta\mathbf{p} = \mathbf{b}, \quad (20)$$

$$\mathbf{A} = \left[a_{kl}\right] = \left[\sum_i \frac{\partial d_i}{\partial p_k}\frac{\partial d_i}{\partial p_l}\right], \quad \mathbf{b} = \left[b_k\right] = \left[-\sum_i d_i \frac{\partial d_i}{\partial p_k}\right], \quad (21)$$

where,

$$\frac{\partial d_i}{\partial p_k} = w_{22}\left(\frac{\partial g}{\partial x'}\frac{\partial x'}{\partial p_k} + \frac{\partial g}{\partial y'}\frac{\partial y'}{\partial p_k}\right). \quad (22)$$

The optimization can be carried out in the same way as described in Section 2.2.

3.4 Matching Process

The results obtained using both methods, least-squares matching and cross correlation matching, will very much depend on the initial values of parameters. In order to assign initial values, a matching process is followed:

1. roughly specify conjugate points at the four corners of the images by manual means;
2. transform the master image to slave image by using affine transformation;
3. set the image patch at the feature points in the master image;
4. apply cross correlation matching, including only translation to search for the initial location;
5. apply least squares matching or cross correlation matching (including deformation); and
6. continue the iterative procedure of the Levenberg-Marquardt method until the solution converges or a given number of iterations is reached.

4. EXPERIMENT

4.1 Comparison of least-squares and cross correlation matching

Least-squares matching and cross correlation matching have been evaluated using an IKONOS Geo panchromatic stereo image pair covering an area of 7 x 7 km over central Melbourne (Figure 3). Figure 4 shows an enlarged portion of the stereo image with feature points, these having been selected to as clearly image identifiable points.

We applied the following methods to the IKONOS stereo image:

a) Least-squares matching including affine transformation,
b) Least-squares matching with projective transformation,
c) Cross correlation matching with affine transformation,
d) Cross correlation matching with projective transformation.

The window size and search area to set the initial position were determined through trial and error:

- Window size: 15 x 15 pixels,
- Search area: 30 x 30 pixels.

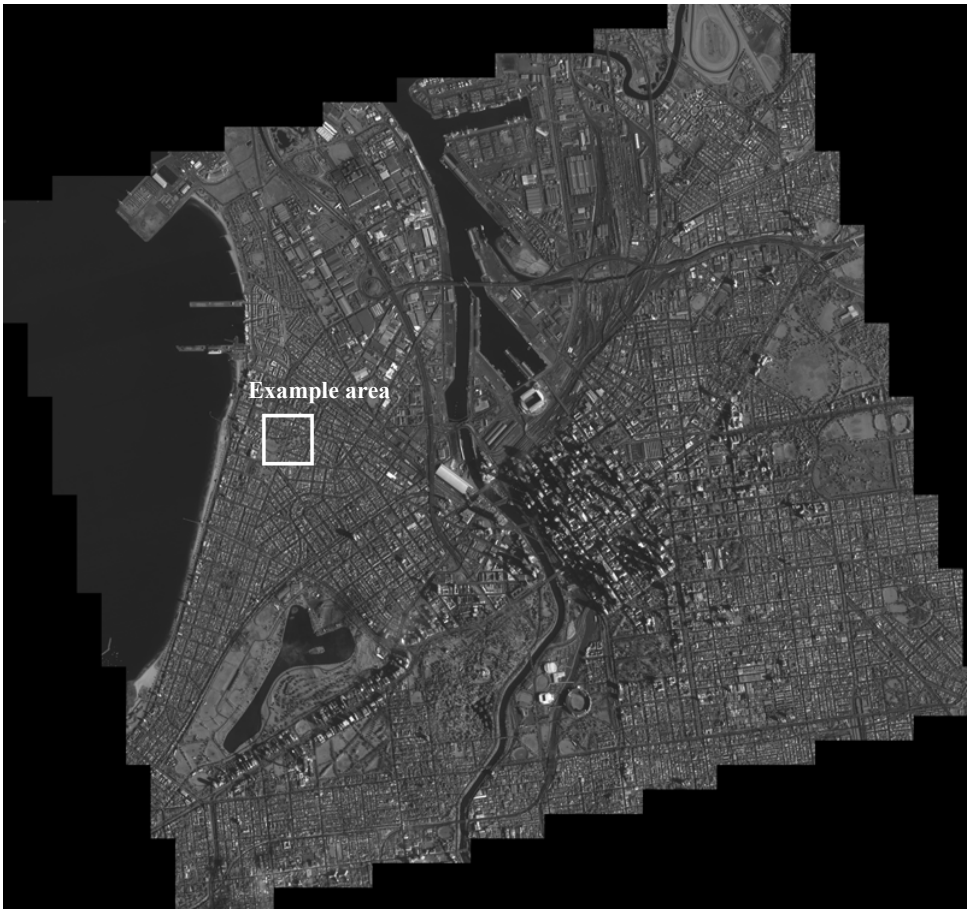

Figure 3. IKONOS Geo panchromatic nadir image of Melbourne

(a) master image (b) slave image

Figure 4. Area within the IKONOS stereo image pair.

Table 1. Results using least-squares matching

	Coordinates (master)	Coordinates (slave)	Affine ransformation	Discrepancies (pixels)	Projective transformation	Discrepancies (pixels)
1	(381, 220)	(379, 238)	(379.6, 238.4)	0.69	(379.2, 238.0)	0.24
2	(205, 350)	(205, 369)	(206.7, 367.7)	2.10	(208.3, 367.6)	3.62
3	(91, 285)	(91, 304)	(94.9, 303.8)	3.88	(93.7, 303.8)	2.73
4	(112, 107)	(111, 125)	(110.2, 123.1)	2.07	(109.7, 123.4)	2.04
5	(246, 39)	(237, 55)	(237.1, 55.6)	0.65	(237.1, 56.6)	1.61
Average				1.88		2.05

Table 2. Results using cross correlation matching

	Coordinates (master)	Coordinates (slave)	Affine ransformation	Discrepancies (pixels)	Projective transformation	Discrepancies (pixels)
1	(381, 220)	(379, 238)	(379.1, 238.3)	0.33	(379.0, 238.2)	0.22
2	(205, 350)	(205, 369)	(205.3, 368.7)	0.48	(205.0, 369.0)	0.00
3	(91, 285)	(91, 304)	(91.0, 304.0)	0.00	(91.0, 303.8)	0.24
4	(112, 107)	(111, 125)	(110.5, 125.1)	0.54	(110.4, 125.2)	0.64
5	(246, 39)	(237, 55)	(237.5, 54.9)	0.51	(237.2, 53.9)	1.10
Average				0.38		0.44

The experimental results are summarized in Tables 1 and 2 for the example shown in Figure 4.

The results obtained using the cross correlation matching, all of which accomplished sub pixel accuracy, are better than those achieved using least-squares matching. It was initially assumed that projective function would result in the best improvement, however affine transformation was more effective in some cases.

5. CONCLUSIONS

The conclusions of this paper are as follows:

- An improvement to the standard method of least-squares matching is possible, in regard to optimization.
- The relationship between least-squares matching and cross correlation matching has been confirmed.
- Cross correlation matching can be formulated to include image deformations.
- An experimental comparison between least-squares matching and cross correlation matching by using IKONOS stereo imagery has been demonstrated.

Geometrically constrained matching could be expected to further improve accuracy, though this paper has not dealt with geometric constraints. Nevertheless, the cross correlation matching yielded positive results. It indicated that the method has potential to be a powerful matching strategy.

Verification of the stability of the method needs to be further investigated by applying various conditions (illumination and location, including central city with high buildings, suburbs, mountain area, etc.). So far, this application has been relatively restricted.

As mentioned above, our final goal is to make a contribution to the matching of multi-temporal and multi-resolution images, including IKONOS, QuickBird and aerial imagery. It will require a combination with other techniques, especially those related to change detection and resampling.

6. REFERENCES

Ando, S., 2000. Consistent gradient operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3): 252-265.

Brown, L. G., 1992. A survey of image registration techniques. *ACM Computing Surveys*, 24(4): 325-376.

Fraser, C.S., Hanley, H.B. and Yamakawa, T., 2002. 3D geopositioning accuracy of IKONOS imagery. *Photogramm. Record*, 17(99): 465-479.

Fraser, C.S. and Hanley, H.B., 2003. Bias compensation in rational functions for IKONOS satellite imagery. *Photogramm. Engineering and Remote Sensing*, 69(1): 53-57.

Greene, W. H., 2000. *Econometric Analysis*. Prentice Hall, New Jersey, pp. 255-259.

Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogramm., Remote Sensing & Cartography*, 14(3): 175-187.

Gruen, A. and Baltsavias, E.P., 1988. Geometrically constrained multiphoto matching. *Photogrammetric Eng. and Remote Sensing*, 54(5): 633-641.

Dare, P., Pendlbury, N., and Fraser, C.S., 2002. Digital orthomosaics as a source of control for geometrically correcting high resolution satellite imagery. *Proceedings of 23rd Asian Conference on Remote Sensing*, Kathmandu, Nepal, CD-ROM.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., 1988. *Numerical Recipes in C*. Cambridge University Press, Cambridge, pp. 681-688.

Szeliski, R., 1996. Video mosaics for virtual environments. *IEEE Compter Graphics and Applications*, 16(2): 22-30.