Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

TECHNIQUES FOR IMAGE REGISTRATION

Dr. W. F. Webber

McDonnell Douglas Astronautics Company
5301 Bolsa Avenue
Huntington Beach, California 92647

I. ABSTRACT

Techniques are developed for determining spatial or geometric distortions between two images of the same scene. The first procedure is iterative linearized least squares estimation (LLSE) for determining small geometric distortions between images. Error variances for these estimators are derived which are interpreted as noise-to-signal ratios for translational and rotational registration. The natural measure of the signal strength of an image for translational registration obtained from these variances is used to establish threshold settings in a new algorithm for fast translational registration. This algorithm belongs to the class of sequential similarity detection algorithms (SSDA's) recently developed for translational registration. Finally, an implementation of an image registration system incorporating all these techniques is described.

II. INTRODUCTION

Image registration is a procedure to determine the spatial best fit between two images that overlap the same scene. Registration is basic to image processing systems since two images of the same scene cannot be meaningfully compared (to determine temporal changes, for example) without having the images in registration. Automatic analysis of remotely sensed imagery will necessitate accurate registration.

Several digital techniques have been used for registering imagery. Principal among these are cross-correlation, normalized cross-correlation and minimum distance criteria. Fast algorithms for determining translational differences between images have been developed in recent years for all these techniques. Efficient algorithms for determining other spatial or geometric distortions such as horizontal scale or rotational differences have not been extensively developed. The primary mechanisms and error types associated with such distortions have been investigated(Bernstein and Silverman, 1971). In some applications determining these distortions will be critical to the success of further analyses on the images.

In this paper a technique, iterative linearized least squares estimation, is derived for efficiently estimating all distortions between two images of the same scene. Also techniques are suggested for improving fast translational registration procedures. A particular implementation of an image registration system is discussed which incorporates all these procedures.

III. ITERATIVE LINEARIZED LEAST SQUARES ESTIMATORS

Two images W and Z of the same scene are to be registered. The image W, referred to as the reference image, covers only a portion of the total scene described by the search image Z. W is assumed to be of good quality, i.e., no clouds are present, contrast is good and geometrical distortions are negligible. This reference image is just one of a large set of such small image available for registering new images as they become available. The search image, on the other hand, is of relatively unknown quality. Some cloud cover may be present along with geometrical distortions. Techniques are developed in this section for estimating the remaining geometrical distortions present in Z after an initial translational registration has been completed.

Denote the reference image by $W = \{w(j,k)\}$ where $j = 1,2,\ldots,M_w$, $k=1,2,\ldots, N_w$ and $w(j,k)$ is the gray level of the pixel (picture element) at the array location (j,k). Similarly the search image to be registered is $Z = \{z(j,k)\}$ where $j = 1,2,\ldots,M_z$ and $k=1,2,\ldots,N_z$. Since the scene covered by W is assumed to be contained in that of Z, necessarily $M_z > M_w$ and $N_z > N_w$. The best match sub-image of Z compared with W obtained by an initial translational registration is $Z_o = \{z(j_o+j,k_o+k)\}$ where $j = 0,1,\ldots,M_w-1$ and $k = 0,1,\ldots,N_w-1$.

It is assumed that the relationship between points in W and Z_o is of the form

$$\alpha z(x,y) = w(x',y') + n(x,y). \tag{1}$$

The coordinate systems for both Z_o and W centered at the origins of these arrays with axes parallel to the array rows and columns are denoted (x,y). The coordinate system (x',y') referred to in (1) is assumed to be related to the (x,y) system of W by the linear transformation

$$\begin{aligned} x' &= (1+c_{11})x+c_{12}y+c_{13} \\ y' &= c_{21}x+(1+c_{22})y+c_{23} \end{aligned} \tag{2}$$

where the c's are all small unknown constants. Such a transformation corresponds to small two dimensional translational, horizontal scaling and rotational errors. Anuta (Anuta, 1971) assumed this transformation to explain the total distortion over an image. At the least, it should explain local distortion in an image,which is its purpose in this paper. The unknown scale factor in (1) approximately describes the differences in brightness levels between W and Z_o. The additive noise surface $n(x,y)$ in (1) describes the remaining differences between W and Z_o.

The registration problem in this formulation is to estimate the unknown c's. Linearized least squares estimators (LLSE's) of these terms will be discussed here. To obtain such estimates write the Taylor series expansion of $w(x',y')$ in terms of $w(x,y)$ as

$$w(x',y') = w(x,y) + (x'-x)w_x(x,y) + (y'-y)w_y(x,y) + 0\ (\Delta^2) \tag{3}$$

where $w_x = \partial w/\partial x$, $w_y = \partial w/\partial y$ and $0\ (\Delta^2)$ refers to the remaining terms which contain factors of the form $(x'-x)^p(y'-y)^q$ such that $p+q \geq 2 (p,q \geq o)$. Since the c's are all small, these terms will be assumed negligible. From (1) and (3) the difference

$$\alpha z(x,y) - w(x,y) = (x'-x)w_x(x,y) + (y'-y)w_y(x,y) + n(x,y) + 0\ (\Delta^2) \tag{4}$$

is approximately linear in the unknowns, i.e., using (2) and (4)

$$\alpha z(x,y) - w(x,y) = (c_{11}x+c_{12}y+c_{13})w_x(x,y) + (c_{21}x+c_{22}y+c_{23})w_y(x,y) + n(x,y) + 0\ (\Delta^2) \tag{5}$$

This suggests obtaining LLSE's for α and the c's by minimizing the quadratic

$$\phi(\alpha,c) = \sum_x \sum_y (\alpha z-w-(c_{11}x+c_{12}y+c_{13})w_x -(c_{21}x+c_{22}y+c_{23})w_y)^2. \tag{6}$$

The double summation is over all the points (x,y) of the sub-images W and Z_o. Also the arguments of all quantities such as $w(x,y)$ have been suppressed. These conventions will be used in the remainder of this section. A generalized least squares procedure taking into account the covariance structure of the noise surface will not be considered here because it is believed to be of limited importance in applications.

The minimization of (6) is readily accomplished by setting all partial derivatives of ϕ in respect to the unknowns to zero and solving these equations for the unknowns.

The estimation of only translational and rotational errors will be examined in detail in the remainder for simplicity of presentation. In this case, to the same order of approximation as above

$$\begin{aligned} x' &= (x+x_o)\cos\theta_o + (y+y_o)\sin\theta_o = x+x_o + y\theta_o + 0\ (\Delta^2), \\ y' &= (y+y_o)\cos\theta_o - (x+x_o)\sin\theta_o = y+y_o - x\theta_o + 0\ (\Delta^2) \end{aligned} \tag{7}$$

so that the quadratic form to be minimized is

$$\phi(x_o,y_o,\theta_o) = \sum_x \sum_y (u-x_o w_x - y_o w_y - \theta_o(y w_x - x w_y))^2, \tag{8}$$

where $u(x,y)= z(x,y)-w(x,y)$. The normal equations for the LLSE's are found to be

$$\begin{aligned}
\sum_x \sum_y w_x u &= \hat{x}_o \sum_x \sum_y w_x^2 + \hat{y}_o \sum_x \sum_y w_x w_y + \hat{\theta}_o \sum_x \sum_y (y w_x - x w_y) w_x \\
\sum_x \sum_y w_y u &= \hat{x}_o \sum_x \sum_y w_x w_y + \hat{y}_o \sum_x \sum_y w_y^2 + \hat{\theta}_o \sum_x \sum_y (y w_x - x w_y) w_y \\
\sum_x \sum_y (y w_x - x w_y) u &= \hat{x}_o \sum_x \sum_y (y w_x - x w_y) w_x + \hat{y}_o \sum_x \sum_y (y w_x - x w_y) w_y + \hat{\theta}_o \sum_x \sum_y (y w_x - x w_y)^2
\end{aligned} \tag{9}$$

If it is further assumed that only translational errors are present, the estimators for x_o and y_o are given by

$$\begin{aligned}
\hat{x}_o &= \Big(\sum_x \sum_y w_y^2 \sum_x \sum_y w_x u - \sum_x \sum_y w_x w_y \sum_x \sum_y w_y u\Big)/D \\
\hat{y}_o &= \Big(\sum_x \sum_y w_x^2 \sum_x \sum_y w_y u - \sum_x \sum_y w_x w_y \sum_x \sum_y w_x u\Big)/D
\end{aligned} \tag{10}$$

where $D = \sum_x \sum_y w_x^2 \sum_x \sum_y w_y^2 - (\sum_x \sum_y w_x w_y)^2$. Similarly if only θ_o is to be estimated the LLSE is given by

$$\hat{\theta}_o = \sum_x \sum_y (y w_x - x w_y) u / \sum_x \sum_y (y w_x - x w_y)^2 \tag{11}$$

A. ERROR ANALYSIS

The variances of the last three estimators can be evaluated in the important case that $\hat{x}_o=\hat{y}_o=\hat{\theta}_o=0$ and that the noise surface $\{n(x,y)\}$ is white with zero mean, i.e., $En(x,y)=0$ and $En(x,y)n(x',y')=\sigma_N^2 \delta_{x-x',y-y'}$, where $\delta_{x,y}$ is the Kronecker delta. In this situation $u(x,y)=n(x,y)$ so that $E\hat{x}_o=E\hat{y}_o=E\hat{\theta}_o=0$ and

$$\begin{aligned}
Var(\hat{x}_o) &= \sigma_N^2 \sum_x \sum_y w_y^2/D \\
Var(\hat{y}_o) &= \sigma_N^2 \sum_x \sum_y w_x^2/D \\
Var(\hat{\theta}_o) &= \sigma_N^2 \sum_x \sum_y (y w_x - x w_y)^2
\end{aligned} \tag{12}$$

where D was defined above. The mean square radial error corresponding to the translational error estimates is given by

$$E(\hat{x}_o^2+\hat{y}_o^2) = \sigma_n^2/(D/\sum_x \sum_y (w_x^2+w_y^2)). \tag{13}$$

This noise-to-signal type ratio is useful in summarizing the translational accuracy of a spatial registration system. The positive quantity $D/\sum_x \sum_y (w_x^2+w_y^2)$ can be interpreted as a measure of the two dimensional signal strength of the reference image W for purpose of translational registration. This quantity will be of primary importance in the next section for establishing threshold settings in sequential similarity detection algorithms used for translational registration. The quantity $\sum_x \sum_y (y w_x - x w_y)^2$, similarly, is to be interpreted as a measure of the two dimensional signal strength of W for rotational registration.

B. NON-ZERO MEAN NOISE

In applications there is no assurance that the noise is zero mean. The usual approach to correct for non-zero mean noise would be to remove the sample means from both W and Z_o before evaluating the error estimates. It is recommended, however, that least squares planes be removed from both of these arrays to protect against two dimensional trends as well as non-zero means in the noise surface. Details for this procedure are given in the next section.

C. PARTIAL DERIVATIVE ESTIMATES

The LLSE's for the spatial errors have been derived assuming that the partial derivatives $w_x(x,y), w_y(x,y)$ are known at all lattice points for which $w(x,y)$ is defined.

In general these derivatives will not be available requiring that estimates be used in their place. The simplest such estimates are of the form

$$\hat{w}_x(x+1/2,y+1/2)=(w(x+1,y+1\quad w(x,y))/\sqrt{2}$$

$$\hat{w}_y(x+1/2,y+1/2)=(w(x,y+1)-w(x+1,y))/\sqrt{2}$$

assuming the spatial distance between lattice points to be unity. These estimates actually correspond to a reference system translated by 1/2 grids units in x and y and rotated counter-clockwise through $\pi/4$ radians from the original reference system for W. The difference terms u(x,y) should also be adjusted to the new reference system, i.e., use

$$\hat{u}(x+1/2,y+1/2)=(u(x,y)+u(x+1,y)+u(x,y+1)+u(x+1,y+1))/4.$$

The estimates $\hat{x}_o,\hat{y}_o,\hat{\theta}_o$ obtained with these approximations can readily be transformed to the original system.

Computational requirements could be reduced if pre-computed estimates of the partial derivatives of W of the form

$$\hat{w}_x(x,y)=\sum_{k=1}^{m} a_k(w(x+k,y)-w(x-k,y))$$

$$\hat{w}_y(x,y)=\sum_{k=1}^{m} a_k(w(x,y+k)-w(x,y-k))$$

were to be used since no correction of the u(x,y) need be made in this case. Only the first estimates for w_x and w_y have been used to evaluate the LLSE technique.

D. ITERATION OF THE ESTIMATES

The many approximations and assumptions required to obtain the LLSE's make the actual worth of these estimators, at best, uncertain.

If the estimation procedure is useful,then iteration on the solution will generally improve the estimates by reducing the error due to the linearization assumption of (5).

The first step in the iteration procedure is to obtain the estimates $\hat{x}_o,\hat{y}_o,\hat{\theta}_o$ using the reference image W and the best match sub-image Z_o obtained in the initial translational registration. The best match sub-image, Z_1, corresponding to these error estimates is interpolated (using nearest neighbor four point linear interpolation for example) from the search image Z. The LLSE's, $\hat{x}_1,\hat{y}_1$ and $\hat{\theta}_1$, of the remaining spatial errors are then calculated using W and Z_1. Similarly the sub-image Z_n is interpolated from Z corresponding to the previous best estimates of the spatial errors $\hat{x}_{n-1}$, $\hat{y}_{n-1}$ and $\hat{\theta}_{n-1}$. The next LLSE's $\hat{x}_n,\hat{y}_n,\hat{\theta}_n$ are obtained from W and Z_n. This procedure can be continued until the estimates converge satisfactorily.

This iterative procedure has been evaluated using pseudo-random signal and noise surfaces. The reference image W (a 16x16 array) was obtained by linear interpolation from the search image to simulate offset and rotational errors. The rotation used to form W was 10 degrees and the offset between the sampling lattices of W and Z was 1/2 the sampling interval in both x and y. White noise, noise with the same spectrum as the search image and multiplicative noise to simulate white cloud covering were used to evaluate the sensitivity of the procedure. The multiplicative noise was simulated by first generating a correlated Gaussian noise surface and then setting the surface to zero if the original surface were below a threshold value and to one if it were above. The threshold value was selected to give a specified percent of cloud cover over the scene. In all cases very similar results were obtained. It was found that whenever the amount of noise was sufficiently low to allow accurate translational registration that the initial LLSE's reduced the spatial errors. However, these estimates were always smaller than the remaining errors. The iteration procedure converged to values that were roughly consistent with the variation predicted by the error analysis previously given.

The number of iterations required to reduce the translational error to within half a sampling interval in both x and y and the rotational error to within a degree of their final values was greater than fifteen in some cases. Several ad hoc methods were evaluated to increase the rate of convergence. One of these procedures resulted in convergence (as defined above) in only three or four iterations. This particular procedure is to use twice the horizontal error estimates and four times the rotational error estimates until the rotational error estimate changes sign. The actual error estimates are used in subsequent iteration steps.

These results indicate that the iterative LLSE approach to precision image registration would be useful for production processing of remotely sensed imagery. Techniques for increasing the rate of convergence of the interactive procedure should be investigated further to assure a technique is employed that is insensitive to the types of error sources expected in a given application.

IV. SEQUENTIAL SIMILARITY DETECTION ALGORITHMS

A very efficient class of algorithms for translational registration has recently been suggested (Barnea and Silverman, 1972). These sequential similarity detection algorithms (SSDA's) are reported to be one to two orders of magnitude faster than previously discussed algorithms such as fast cross-correlation (Anuta, 1971).

As an example of algorithms of this type consider that a reference image W and a search image Z, as described in the last section, are to be registered. One method useful for this purpose is the minimum p-distance criterion, which is to find the rectangular subset Z_o of Z which minimizes

$$d_p^p(j',k')=\sum_{j=1}^{M_w}\sum_{k=1}^{N_w}|Z(j'+j,k'+k)-w(j,k)|^p \tag{14}$$

where p>1. (This criterion with p = 1 or 2 is commonly used. The p = 2 case is referred to as the Euclidean distance criterion. Computational considerations usually determine the choice of p .) Now consider a random sequence related to the distance measure. Let (j_s,k_s) for s = 1,2..., M_wN_w be a sequence of random samples without replacement from the lattice of points on which W is defined. Define the random sum

$$S_p(r;j',k')=\sum_{s=1}^{r}|Z(j'+j_s,k'+k_s)-w(j_s,k_s)|^2 \tag{15}$$

for r = 1,2,...,M_wN_w which increases monotonically to $d^p(j',k')$ as r increases to M_wN_w. The essential feature of all SSDA's is that the rate of increase of $S_p(r\ ;j',k')$ is used in determining the similarity between the sub-image of Z and the reference image. If this quantity increases rapidly then the two images are dissimilar. Barnea and Silverman suggested using a threshold to measure an average rate of increase. In the simplest case a constant threshold is used and the number of steps required for each sub-image to reach this threshold is the measure of the rate of increase of S_p. The importance of this procedure is that the calculation of S_p terminates whenever the threshold is crossed, resulting in the substantial reduction in computational time quoted above. The primary difficulty with the technique is establishing a useful threshold level. Too high a threshold value reduces the technique to the original p-distance criterion, whereas too low a level can give rise to gross registration errors. Barnea and Silverman suggested adaptive threshold setting procedures to reduce the difficulties associated with a constant threshold.

Another approach is to base the threshold setting on the signal strength of W and the allowable noise level associated with W for accurate registration, i.e., a signal-to-noise criterion could be used to set the threshold. In the last section it was shown that a natural signal-to-noise ratio for translational registration assuming additive white noise with variance σ_N^2 is (using (13))

$$\sigma_S^2/\sigma_N^2=((\sum_x\sum_y w_x^2\sum_x\sum_y w_y^2-(\sum_x\sum_y w_xw_y)^2)/\sum_x\sum_y(w_x^2+w_y^2))\ \sigma_N^2 \tag{16}$$

defining the signal strength measure σ_s^2. Suppose that z(x,y)=w(x,y)+n(x,y) for some sub-image Z_o of the search image Z, i.e., assume that there is no geometric distortion. In this case the Euclidean distance criterion $d_2^2(j',k')$ will reduce to

$$d_2^2(j_o,k_o)=\sum_j\sum_k n^2(j,k)$$

for (j_o,k_o) corresponding to Z_o. The mean of this quantity is

$$Ed_2^2(j_o,k_o) = M_wN_w\sigma_N^2$$

so that at low to moderate noise levels the minimum value of the discrete function $d_2^2(j',k')$ is roughly proportional to σ_N^2. In order to assure a given level of registration accuracy the threshold should be set proportional to σ_S^2 as indicated by (16). The threshold must also be set sufficiently low that values of $d_2^2(j',k')$ away from the match point are always greater than the threshold even in the case of no noise.

A. TECHNIQUES FOR STABILIZING IMAGE SIMILARITY CALCULATIONS

The primary difficulty in using σ_s^2 to establish the threshold setting is that the minimum and shape of $d_2^2(j,k)$ are sensitive to low wave-number error sources. In general, signal energy content at low wave-numbers does not contribute to the signal strength of an image for registration (the quantity σ_s^2 is a function only of the partial derivatives of W) so that perhaps the best procedure to alleviate this sensitivity is to apply a digital high-pass filter to the reference image and the search image before the SSDA calculation. However, the computational advantages of the SSDA algorithm would be essentially lost if this approach were taken. Removing the mean from the reference image and each subset of the search image before the similarity calculations helps to reduce this sensitivity, but a better approach requiring only a moderate computational increase over mean removal is to remove the least squares plane from each sub-image.

Consider a sub-image $Z_o = \{z(x,y)\}$ where $x=j_o,j_o+1,\ldots,j_o+M_w-1$ and $y=k_o,k_o+1,\ldots,k_o+N_w-1$. The spatial center of this data set is $(x_o y_o)=(j_o+(M_w-1)/2,k_o+(N_w-1)/2)$. A plane defined on this lattice can be written as

$$P(x,y)=a(x-x_o)+b(y-y_o)+c$$

so that $P(x_o,y_o)=c$. The least squares best fit plane of this form to the sub-image Z_o minimizes

$$\phi(a,b,c) = \sum_x \sum_y (z-a(x-x_o)-b(y-y_o)-c)^2.$$

The solution is

$$\begin{aligned}
\hat{a} &= \frac{12}{M_w N_w (M_w^2-1)} \sum_x \sum_y (x-x_o)z, \\
\hat{b} &= \frac{12}{M_w N_w (N_w^2-1)} \sum_x \sum_y (y-y_o)z, \\
\hat{c} &= \frac{1}{M_w N_w} \sum_x \sum_y z.
\end{aligned} \tag{17}$$

B. SIMULATION RESULTS

Simulation studies have been used to evaluate the technique for setting the SSDA threshold level using the signal strength measure σ_s^2 and the least squares plane removal procedure. It was found that setting the threshold at three to four times σ_s^2 resulted in the true minimum being found in all cases at low to moderate noise levels. At high noise levels the original Barnea and Silverman constant threshold approach can be used. It is not necessary to know the noise level a priori in any case. An added advantage to using the method suggested here for setting thresholds is that whenever a sum is found that does not reach the threshold the search procedure can be stopped since the region of the registration position has been found. This early termination results in considerable time savings in many cases.

V. REGISTRATION SYSTEM IMPLEMENTATION

The implementation of a production image registration system has to be based on many considerations not covered in this paper. In particular the expected data quality and registration accuracy requirements will strongly affect design parameters. One implementation that has been extensively studied will be described here.

The initial translation registration is accomplished in two stages. The first stage is a coarse translation registration. The SSDA threshold is initially established using the signal strength measure σ_s^2 as discussed in the last section. A spiral search procedure starting with the a priori most probable sub-image of the search area is employed to locate the initial translation registration estimate. Let (x_o,y_o) be the center position of the sub-image at the center of the search image Z. The coordinates of the center points of subsequent sub-images in the spiral search procedure are $(x_o+r,y_o),(x_o+r,y_o+r),(x_o,y_o+r),(x_o-r,y_o+r),(x_o-r,y_o)$, $(x_o-r,y_o-r),(x_o,y_o-r),(x_o+r,y_o-r),(x_o+2r,y_o-r)$, etc. Taking $r = 2$ in this search pattern will significantly reduce the required number of calculations without compromising system performance in many applications. The parameters of the least squares planes of the sub-images can be efficiently updated for this search pattern and the resultant arithmetic error accumulation is minimized.

After the coarse registration is completed the residue at the best match point is used as the new threshold. A fine search procedure is then started about this best match point taking every sub-image (r=1). This procedure is continued until a local minimum is obtained.

At the completion of the fine search procedure the iterative least squares estimates of the remaining registration errors are computed. The actual parameters estimated in this procedure will depend on the types of geometrical distortions expected in the data. The number of iterations can be fixed or, if high precision is required, determined by the rate of convergence observed with the actual data set.

It should be noted that the user generally can select the set of reference images employed for registration. These should be chosen to have large signal strength, σ_s^2. The final registration estimates may not be accurate. This can roughly be determined by computing the sample variance of the difference between W and Z_n obtained in the final iteration of the LLSE's. This quantity can then be compared with σ_s^2 to assess the registration accuracy.

VI. ACKNOWLEDGEMENT

The research reported here was partially supported by the Air Force Office of Scientific Research (Air Force Systems Command) under Contract F44620-71-C-0052.

REFERENCES

P. E. Anuta, Spatial Registration of Multispectral and Multidimensional Digital Imagery Using Fast Fourier Transform Techniques, IEEE Transactions Geosci, Electron, GE-8, October 1970.

R. Bernstein and H. Silverman, Digital Techniques for Earth Resource Image Data Processing, American Institute of Aeronautics and Astronautics (AIAA) Paper No. 71-978, October 1971.

D. I. Barnea and H. F. Silverman, A Class of Algorithms for Fast Digital Image Registration, IEEE Transactions on Computers, C-21, pp. 179-186, February 1972.