Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

A METHOD FOR DIGITAL IMAGE REGISTRATION USING A MATHEMATICAL PROGRAMMING TECHNIQUE

Stanton S. Yao

Lockheed Electronics Company, Inc.
NASA - Johnson Space Center
Houston, Texas

I. ABSTRACT

A new algorithm based on a nonlinear programming technique to correct the geometrical distortions of one digital image with respect to another is discussed. This algorithm promises to be superior to existing ones in that it is capable of treating localized differential scaling, translational and rotational errors over the whole image plane. A series of piece-wise "rubber-sheet" approximations are used, constrained in such a manner that a smooth approximation over the entire image can be obtained. The theoretical derivation is included. The result of using the algorithm to register four channel S065 Apollo IX digitized photography over Imperial Valley, California, is discussed in detail.

II. INTRODUCTION

In processing and analyzing remotely-sensed data, it is often necessary to make point to point comparisons of data gathered from the same scene by different data-gathering platforms at different times of the year over different spectral bands and from different vantage points. When pattern recognition techniques are applied to the analysis of multispectral remote sensing imagery, one basic assumption is that corresponding data points from different channels of the images are spatially aligned. The science of how to efficiently bring image data into spatial alignment is called correlation/registration.

Correlation/registration of digital imagery is sometimes referred to as geometric corrections, however the process involves more than merely a reassignment of data points to different geometric locations. Inherent in correlation/registration are problems such as radiometric corrections, different resolution-cell sizes from different sensors, extrapolation from a small, overlayed image strip between two images, etc. Solutions to most of these problems are not yet well developed. This paper represents a preliminary attempt in trying to tackle some of these problems using well established mathematical techniques.

Techniques for correlating images with approximately equal resolution-cell sizes are available and numerous (1, 2). They will not be the subject of discussion in this paper. The number of image registration techniques in existence, on the other hand, is rather limited (3, 4). In general, all geometrical distortions are represented by a combination of the two-dimensional translation, rotation and scaling problems. When matching points, or "checkpoints", are located on two images; mathematical functions based on certain goodness criterion can be used to model the distortions and thereby bring the two images into registration according to the model established. The most common function used is the bi-variate polynomial, and the most tractable goodness criterion is the least-squares process. The different terms in the bi-variate polynomial may or may not have geometrical significance in relation to the distortions that exist, but they are not usually considered in detail.

For example, let (x,y) be the coordinate system of the reference image, and (u,v), the overlay image. Let it be assumed that a linear relationship exists between (x,y) and (u,v):

$$u = ax + by + c$$
$$v = dx + ey + f$$

Given three or more pairs of matching points $((x_i, y_i), (u_i, v_i))$, $i \geq 3$, six coefficients in the above equations can then be solved for by the standard least square process.

When the geometric distortions that exist between two sets of digital imagery are "smooth" or slow varying in nature, a low-order bi-variate polynomial based on the least-squares criterion usually yields a good approximation of the actual distortion. This assumes that the matching points are uniformly distributed throughout both images as is required by any approximation scheme. When distortions between two images have relatively large localized differential variations, higher-order bi-variate polynomials normally are used in order to achieve minimum error bounds. The coefficients associated with the high-order terms in the polynomial often change appreciably as different sets matching points are used or an extra point is added or deleted in the calculation. In other words, the approximation error often becomes a strong function of the location of the matching points when there are high-order terms in the bi-variate polynomial with significant values for the coefficients. This is clearly undesirable.

The algorithm proposed here is intended to alleviate this type of shortcoming. It does its job over one "patch" of the image at a time, and only lower order bivariate polynomials are used. The boundaries between the patches are constrained to be continuous. It is also desirable that the directional derivative across the boundaries of adjacent regions be continuous in order to have a smoothly registered image.

These boundary-continuity conditions turn out to be easily met due to certain properties of bi-variate polynomials. If constraints are placed on the coefficients of the polynomials calculated in different regions, an optimum solution based on the goodness criterion, can be obtained utilizing the mathematical programming technique. In this sense, the geometrical distortion modeling problem can be considered as a series of constrained approximations with the coefficients of the bi-variate polynomial in each region so computed that they not only approximate the distortion in that region, but also effect a smooth transition across the boundaries of adjacent regions. Mathematical formulation of this constrained approximation problem for a second-order bi-variate polynomial with six unknown coefficients over rectangular regions are discussed in detail.

The registration technique described in this paper assumes that, initially, the two images are fairly closely correlated. A set of matching points which may be arranged in the form of a grid structure over the image plane is computed using well-established correlation methods such as the Fast Fourier Transform Technique. Based on the correlation results, the registration algorithm is applied point-by-point to bring one image into spatial alignment with the other.

III. THEORETICAL DERIVATION

Consider a one-dimensional problem where the theory can most easily be demonstrated. Let $p(t')$ be the original geometrically distorted function and $p(t)$ be the "correct" or "registered" function after the distortions on the axis t' has been removed. Let $f(t)$ be the reference function. For simplicity, consider $f(t) = p(t)$ (see Figure 3). The distortion is characterized by performing a cross correlation of $p(t)$ with $f(t)$ at $t = 0, t_1, t_2, t_3, \ldots$ etc. The correlation procedure generates a table showing the deviation of the t' from t at $t = 0, t_1, t_2, t_3, \ldots$ etc. If a relation $t' = g(t) + t$ can be found based on the correlation results, where $g(t)$ indicates the error in registration, then $p(t) = p(t' - g(t))$ can be computed, and $p(t)$ and $f(t)$ are thus in registration.

The task is to find a suitable functional representation of the error $g(t)$ based on the correlation results at discrete points $t = 0, t_1, t_2, t_3, \cdots$ etc. It is reasonable to assume that the nature of the geometrical distortion is smooth. In other words, referring to Figure 4, the functional relationship between t and t' looks more like the smooth solid line than the broken piece-wise linear line. In order to approximate closely this function $g(t)$ between $t = 0$ and $t = 14$ based on the correlation results at $t = 0, t_1, t_2, t_3, \cdots$ etc. using a polynomial in t, a relatively high order polynomial would be required. If this polynomial approximation problem is two-dimensional in nature, and a close fit for all matching points are needed, going to higher and higher order two-dimensional polynomials will not be the answer. An alternate method is to approximate $g(t)$ (the error function) between $t = 0$ and $t = t_1$ by a low order polynomial $g_1(t)$, and between $t = t_1$ and $t = t_2$ by another low order polynomial $g_2(t)$, and to make sure that the two polynomials join smoothy at $t = t_+$, a point somewhere in the vicinity of t_1. That is, $g_1(t_+) = g_2(t_+)$, $g_1'(t_+) = g_2'(t_+)$ where the prime denotes the derivative of the two functions evaluated at $t = t_+$. The generalization of this approach to the two dimensional case will be discussed in detail later.

Since both $g(t)$ and $f(t)$ are digital images, the assumption that $g(t)$, the error function between t and t' is not zero at t_1, is a reasonable one. The correlation results at t_1 are always written in terms of an integral number of the sample-spacing Δ. The actual misregistration may be anywhere within $\pm\Delta/2$. The requirement that not only $g(t)$ itself, but that also its derivative be continuous at the "breakpoint" t_+, insures that a smooth transition can be brought about for the error function between one region and the next.

Next consider the problem of mathematically determining the "best" approximating second order polynomial in a region between the two correlation grid points t_1 and t_2 shown in Figure 5. Let the misregistrations at t_1, t_2, and t_3 be $z_1\Delta$, $z_2\Delta$ and $z_3\Delta$ respectively, where Δ is the sample spacing and $z_i = 0, \pm 1, \pm 2, \cdots, i = 1, 2, 3$. It is required that the coefficients of the approximating polynomial, $g_2(t)$ between t_1 and t_2 be determined in such a way that

$$J = W||g_2(t_2) - z_2\Delta||^2 + ||g_2(t_3) - z_3\Delta||^2$$

be minimized subject to the constraint that

$$g_2(t_1) = t_+$$

$$g_2'(t_1) = t_+'$$

where $|| \;||^2$ indicates square norm in a metric space, and
$W \geq 1$ is a scaler weighting factor
t_+ is the value of the previous approximating function $g_1(t)$ between t_0 and t_1 evaluated at t_1 (functional continuity)
t_+' is the derivative of the previous approximating function $g_1(t)$ between t_0 and t_1 evaluated at t_1 (derivative continuity).

Let $g_2(t) = at^2 + bt + c$, be a second order polynomial in t, and let λ_1, λ_2 be the Lagrange multipliers. It is required that the stationary points of the following equation be found.

$$H = W||g_2(t_2) - z_2\Delta||^2 + ||g_2(t_3) - z_3\Delta||^2 + \lambda_1(g_2(t_1) - t_+) + \lambda_2(g_2'(t_1) - t_+')$$

$$= W||at_2^2 + bt_2 + c - z_2\Delta||^2 + ||at_3^2 + bt_3 + c - z_3\Delta||^2 + \lambda_1(at_1^2 + bt_1 + c - t_+) + \lambda_2(2at_1 + b - t_+')$$

Taking partial derivatives of H with respect to a, b, c, λ_1 and λ_2, the following is obtained:

$$\frac{\partial H}{\partial a} = 0$$

or

$$2W(at_2^2 + bt_2 + c - z_2\Delta)t_2^2 + 2(at_3^2 + bt_3 + c - z_3\Delta)t_3^2 + \lambda t_1 + 2\lambda_2 t_1 = 0 \quad (1)$$

$$\frac{\partial H}{\partial b} = 0 \Rightarrow 2W(at_2^2 + bt_2 + c - z_2\Delta)t_2^2 + 2(bt_3 + at_3^2 + c - z_3\Delta)t_3 + \lambda_1 t_1 + \lambda_2 = 0 \quad (2)$$

$$\frac{\partial H}{\partial c} = 0 \Rightarrow 2W(at_2^2 + bt_2 + c - z_2\Delta) + 2(at_3^2 + bt_3 + c - z_3\Delta) + \lambda_1 = 0 \quad (3)$$

$$\frac{\partial H}{\partial \lambda_1} = 0 \Rightarrow t_+ = at_1^2 + bt_1 + c \quad (4)$$

$$\frac{\partial H}{\partial \lambda_2} = 0 = t'_+ = 2at_1 + b \quad (5)$$

In order to determine the coefficients a, b, and c , the above five independent linear equations with five unknowns (a, b, c, λ_1, λ_2) must be solved. However, a particular easy solution follows if t_1 is set to zero. There is no restriction why t_1 cannot be set to zero, since t_1 is just a reference "time mark" with respect to $g_2(t)$. When $g_3(t)$ between t_2 and t_3 are to be computed then t_2 could very well be set to zero to simplify the computation. Setting $t_1 = 0$, equations (4) and (5) yield $c = t_+$, $b = t'_+$ and from equation (1) it follows that

$$a(t_2^4 W + t_3^4) = z_2\Delta W t_2^2 + z_3\Delta t_3^2 - Wt_2^2(bt_2 + c) - t_3^2(bt_3 + c)$$

therefore,

$$a = \frac{Wt_2^2(z_2\Delta - t_+) + t_3^2(z_3\Delta - t_+) - t'_+(Wt_2^3 + t_3^3)}{Wt_2^4 + t_3^4}$$

A special case occurs when $t_3 - t_2 = t_2 - t_1 = t_2$, that is, equally spaced grid points with $t_3 = 2t_2$, then it follows that

$$a = \frac{W(z_2\Delta - t_+) + 4(z_3\Delta - t_+) - t'_+ t_2(W + 8)}{(W + 16)t_2^2}$$

After coefficients a, b, and c have been calculated, $g_2(t_2) = t_{+2}$ and $g'_2(t_2) = t'_{+2}$ could be evaluated so that the same procedure can be used to find the approximating polynomial $g_3(t)$ between grid points t_2 and t_3 , and so on.

A few comments about the method are in order before generalizing it into two-dimensions:

1. Almost exactly the same approach can be used to determine the coefficients of a third or even higher order approximating polynomial if so desired.
2. The reason that the correlation result at t_3 is used when determining the error approximating polynomial between t_1 and t_2 is to guarantee that a smooth transition takes place from one region of the approximation to another. The method can be easily extended to the case where t_4 or t_5 are also used in the compution of $g_2(t)$. The

weighting factor W is used to emphasize the importance of a good approximation at t_2 as compared to that at t_3.

3. When $g_2(t)$ is computed for the functional approximation between t_1 and t_2, $|g_2(t_2)-z_2\Delta|$ should be also evaluated to make sure that $|g_2(t_2) - z_2\Delta| < \Delta/2$. If the inequality does not hold, which means that the approximating function $g_2(t)$ is not satisfactory, either the weighting factor W should be increased or a higher than second order polynomial approximating function should be employed.

4. The amount of computational effort used in the determination of the coefficients of the error approximating functions in different regions is indeed rather small.

Before discussing how the error approximating functions are used for image registration, the two-dimensional version of the above derivation will be presented.

In order to solve the problem of the two-dimensional geometric correction using the technique discussed above, some modifications are necessary. The most obvious fact is that a two-dimensional region involves boundary continuity instead of point continuity. Nevertheless, the same philosophy and reasoning will govern the derivation. Let the following functions be defined:

p(u, v): The two dimensional geometrically distorted image to be corrected. u, v represent the geometrically distorted input coordinate system.

p(x,y): The output "registered" version of p(u,v). x,y are the geometrically corrected (or reference) coordinate system.

f(x,y): The reference image where p(u,v) is to be registered.

The functional relationships between the input and output coordinate systems are assumed to have the following form:

$$u = g_c(x,y) + x$$

where the subscript c stands for column, and

$$v = g_r(x,y) + y$$

where the subscript r stands for row. $g_r(x,y)$ and $g_c(x,y)$, the two-dimensional error functions, can both be bivariate polynomials in x,y. A second order bivariate polynomial has the following form:

$$g(x,y) = ax^2 + by^2 + cxy + dx + ey + f \quad .$$

Six coefficients are to be determined instead of only three coefficients as in the one-dimensional case. If a third order bivariate polynomial is used, 10 coefficients must be determined.

Consider the correlation grid structure obtained for errors in "column" registration shown in Figure 6. For simplicity, a regularly spaced grid with horizontal and vertical grid spacing of Θ and ∇ respectively is used.

Assume that a two-dimensional bivariate polynomial in (x,y) is to be used. Then $g_c(x,y)$ is the error function approximating the column geometrical distortion in region A surrounded by four grid points $A_{11}, A_{12}, A_{21}, A_{22}$ with registration errors $Z_{11}, Z_{12}, Z_{21}, Z_{22}$ respectively (see Figure 6). Assume further that the error functions approximating the distortion in regions C and B to the left and above A, both two-dimensional bivariate polynomials, have been computed. These two functions determine the upper and left boundaries of region A since boundary continuity is required. It will be shown that this continuity requirement determines five out of the six coefficients that uniquely define the function $g_c(x,y)$ in region A .

Again, a convenient origin for the coordinate system for region A is assigned as shown in Figure 7. A_{11} is at the origin, A_{12} is at $(\theta, 0)$, etc. if

$$g_c(x,y) = ax^2 + by^2 + cxy + dx + ey + f$$

then the equation of the upper boundary is given by

$$g_c(x,0) = ax^2 + dx + f$$

and the equation of the left boundary is given by

$$g_c(0,y) = by^2 + ey + f$$

Because of the requirement that these boundaries be continuous between B and A and C and A, the five coefficients a, b, d, e, and f can be determined from the previously computed coefficients for the functions approximating the distortions in regions B and C.

The remaining coefficient c in $g_c(x,y)$ can be determined by formulating and solving the following simple minimization problem:

$$J = W||g(\theta,\nabla) - z_{22}\Delta||^2 + ||g(2\theta,\nabla) - z_{23}\Delta||^2 + ||g(\theta,2\nabla) - z_{32}\Delta||^2$$

where $|| \quad ||^2$ denotes the square norm in a metric space.

$W \geq 1$ is a scalar weighting factor

$z_{22}\Delta$, $z_{23}\Delta$, and $z_{32}\Delta$ are the column misregistrations at A_{22}, A_{23}, A_{32} respectively, obtained from the correlation results.

The reason that grid points A_{23} and A_{32} are used in determining the coefficients in region A lies in the fact that once again, very smooth "surfaces" that extend from region A to the neighboring regions are desired.

A simple calculation yields the formula for computing c as

$$c = \frac{1}{(W + 8)\theta\Delta}\left[Wz_{22}\Delta + + 2(z_{23}\Delta + z_{32}\Delta) - (W + 10)(a\theta^2 + b\nabla^2) - (W + 6)(d\theta + e\nabla) - (W + 4)f\right].$$

When the correlation results which indicate the magnitude of the mis-registration are obtained not at the four corners of the region A, but rather, at K points scattered inside region A, and L points outside region A; a similar formulation yields:

$$J = W\sum_{k=1}^{K}||g(x_{ik}, Y_{ik}) - Z_k\Delta||^2 + \sum_{l=1}^{L}||g(x_{ml}, y_{nl}) - z_l\Delta||^2$$

where $o \leq i \leq \theta$ and $m > \theta$

$0 \leq j \leq \nabla$ and $n > \nabla$

and $$C = \frac{W\sum^{K} x_{ik}\, y_{jk}(z_k\Delta - ax_{ik}^2 - by_{jk}^2 - dx_{ik} - ey_{jk} - f_k) + \sum^{l} x_{ml}\, y_{nl}(z_l\Delta - ax_{ml}^2 - by_{nl}^2 - dx_{ml} - ey_{nl} - f_l)}{W\sum_{K} x_{ik}^2\, y_{jk}^2 + \sum_{l} x_{ml}^2\, y_{nl}^2}$$

After $g_c(x,y)$ in A has been determined, the following three quantities should be computed in order to determine not only how good the approximation has turned out, but also to provide information concerning the continuation of the procedure to other neighboring regions, in particular, to the right and below region A :

1. $|z_{22}\Delta - f(\theta,\nabla)| \leq \Delta/2$. $f(\theta,\nabla)$ is equivalent to t_+ in the one-dimensional case. If this inequality is satisfied, the approximating function $g_c(x,y)$ should be considered acceptable. Otherwise the weighting factor W would be increased.

2. $g_c(x,\nabla) = ax^2 + x(d + c\nabla) + (f + b\nabla^2 + e\nabla) = ax^2 + d'x + f'$

$g_c(x,\nabla)$ is the equation of the lower boundary of region A at $y = \nabla$. The coefficients a, d', and f', so calculated become the coefficients of the upper boundary equation of the region immediately below A.

3. $g_c(\theta,y) = by^2 + y(e + c\theta) + (f + a\theta^2 + d\theta) = by^2 + e'y + f''$.

$g_c(\theta,y)$ is the equation of the right boundary of region A at $x = \theta$. The coefficients b, e', and f" so calculated become the coefficients of the left boundary equation of the region to the right of region A.

To summarize, the following steps are needed in order to compute the coefficients of a second order bivariate polynomial approximating the distortion errors in row or in column in region A defined by four grid points A_{11}, A_{21}, A_{22} and A_{12} (This polynomial must be determined in order to be able to carry out the proposed registration procedure.):

1. From the previously calculated and stored coefficients of the equations for the upper and left boundaries of the region, obtain values a, b, d, e, and f of the bivariate polynomial.

2. Compute coefficient c based on a, b, d, e, and f, the weighting factor W and the correlation information at A_{22}, A_{23}, and A_{32}, or other points inside and outside region A.

3. Compute $|z_{22}\Delta - f(\theta,\nabla)|$ to check the validity of the polynomial approximation at (θ,∇), if z_{22} is available.

4. Compute and store coefficients a, d', f', for the upper boundary equation of the region immediate below A.

5. Compute and store coefficients b, e', f" for the left boundary equation of the region immediately to the right of A.

6. Repeat steps 1 to 5 either for regions immediately to the right of A or immediately below A.

Thus, within a certain region A the following relationships between the coordinate frameworks will be available after modeling the row and column distortion errors with two functions $g_c(x,y)$ and $g_r(x,y)$:

$$u = g_c(x,y) + x, \quad x \text{ from } 0 \text{ to } \theta$$

$$v = g_r(x,y) + y, \quad y \text{ from } 0 \text{ to } \nabla$$

where $g_c(x,y)$ and $g_r(x,y)$, which are valid only in region A, are both second order bivariate polynomials whose coefficients have been already calculated. At each position (x_0,y_0) on the "output" image plane, a corresponding pair of values (u_0,v_0), are computed based on the above equations. When the data value at $p(u_0,v_0)$ is placed on (x_0,y_0) it becomes $p(x_0,y_0)$, which is in registration with the reference image $f(x,y)$ at (x_0,y_0).

Before the raw data for the distorted and reference images are read in line-by-line from tape, all coefficients of the bivariate polynomials approximating the column and row distortion errors in different regions are stored in arrays. These arrays are ordered so that the coefficients correspond to regions over the image to which they apply. Then, with respect to one line of the reference data just read in, a cyclic buffer containing several lines of the distorted image is filled. The number of lines in the buffer depends on the maximum expected line misregistrations between the two images. Next, for each column sample on the reference line, say at (x_0,y_0), the corresponding line and column values (u_0,v_0) are determined using the above equations and the knowledge concerning the particular region where (x_0,y_0) is located. A data value at a location closest to (u_0,v_0) is then fetched from the buffer. This process is continued through all column samples on the reference line before the next reference line is read in and the buffer refilled.

IV. EXPERIMENTAL VERIFICATIONS

A set of four Apollo IX 70 mm S065 photography were separately digitized by a microdensitometer with 25 micrometer resolution cells and 25 micrometer distance between cells. The four sets of digital data were then combined to form four channels on a data tape. It was found by correlation that misregistration exists between channels. To illustrate this, let A, B, and C be three (out of the four) frames of the supposedly same scene. Let frame A be the reference frame, with respect to which, frame B and C are to be registered. Along line 1800, for example, it was found after performing a correlation procedure, that the error curves for column registration shown in Figures 1 and 2 exist.

It should be noted that there exists no registration error at columns 100 and 2100. However, there are localized distortions close to the center of line 1800. The misregistration between frames B and C can be as many as eight pixels. The average agriculture field is, however, only 4 pixel by 4 pixel.

The row-wise misregistration of the S065 digitized photography is found to follow almost the same pattern as column by column misregistration. Clearly, a more subtle nonlinear two-dimensional geometrical correction scheme is called for.

In order not to be over-burden by the amount of data involved, only a portion of the image is considered. Approximately the first 400 lines of the data are found to be useful in identifying agriculture features. Each line of data consists of 222 sample points, and each sample point is represented by one of 256 levels of gray. Using channel 1 as reference, channels 2, 3 and 4 are separately correlated with channel 1 to form the grid structures shown in Figures 8, 9 and 10.* An automatic correlation technique (1) is used so that the grids are equally spaced with 52-pixel column spacing and 70-pixel line spacing. A total of 20 pairs of "matching points" is thus established over the image. It is shown from the correlation grid structures that channel 2 is fairly well spatially aligned with channel 1; on the other hand channels 3 and 4 are not. In addition, considering row deviations, channels 3 and 4 register opposite errors with respect to channel 1 in the lower portion of the image, and the error can be as large as 4 lines either way. It is also noted that the registration error propagates rather smoothly over the grid structure. Also noted in Figures 9 and 10 is the fact that the correlation coefficients at grid points in the upper portions of the images, where the registration error is smaller, are higher in absolute values. These values tend to become smaller toward the lower end of the images. The correlation coefficients between channels 1 and 3 are generally small and negative, showing that a negative correlation peak indicated the best fit. The explanation is that channel 3 is an infrared channel while channel 1 covers mostly the visible region of the spectrum. It is well known that vegetation shows opposite contrast in infrared as compared to that in the visible. Channel 4, being a "red" channel, shows much better correlation with channel 1.

Due to the small registration errors between channels 1 and 2, only channels 3 and 4 are considered to be misregistered with respect to channel 1. The algorithm discussed in the previous section is employed to bring the columns and the rows in these two channels in spatial alignment with those of channel 1. A new data tape with all channels in registration is generated in LARSYS II format. The same correlation procedure as before is next used to correlate channels 3 and 4 with respect to channel 1. The row and column correlation results, with the same grid structures as those in Figures 9 and 10, are shown in Figures 11 and 12 with the exception that no data is available on line 330. Note that particular good registration is obtained for channel 4.

Also worth emphasis is the increase in correlation coefficients over the entire image for channel 4 data. This increase in correlation coefficients supplementing the zero registration error

*Numerals in parenthesis are row and column misregistration in units of sample spacings at the particular grid points shown. The number below the parenthesis shows the correlation coefficient at that location.

gives confidence to the conclusion that channels 4 and 1 are indeed spatially aligned. Channel 3 also shows improvement in registration, especially in the lower portion of the imagery, but the results are not as prominent as those in channel 4. The explanation lies partly in the difficulties encountered in trying to correlate essentially a negative image to a positive counterpart. The correlation values in the grid structure upon which the whole registration procedure depends are not reliable.

Finally, correlation procedures are used on different grid structures to serve as a double check for the validity of the registration, especially in areas far from the original "matching points." The overall small error, in particular for channel 4, where all but one place the row deviation registers a "1" instead of "0" indicates the success of this algorithm.

V. DISCUSSION

In the preceding two sections, the mathematical foundation of the newly proposed registration scheme was derived together with the discussion of the results of an experiment verification using the S065 digitized photography. The proposed registration scheme, which makes use of the mathematical programming technique, is not advantageous for all image registration problems. For example, when ERTS MSS imageries are to be overlaid on base maps made from rectified, scaled photography, a first order bi-variate polynomial correction applied over the entire image frame will suffice. The reason is that the ERTS MSS imageries are geometrically well-corrected. To look at the problem from a different point-of-view, the proposed image registration scheme necessitates the division of an image into several regions. Each individual region has its own characteristic distortion and is approximated by a bi-variate polynomial tailored to that distortion. The number of regions that must be subdivided cannot be determined until the whole image is first examined. If the image, after examination, is so "smooth" in nature that only one region-the entire image frame--needs to be considered, then the proposed registration scheme is no different from the ordinary low order bi-variate polynomial correction applied over the entire image. On the other hand, when different types of distortions resulting, for example, from pitch and yaw of an inherently unstable low-flying aircraft scanner platform were present in the image, then the proposed registration scheme should yield a much better error approximation than possible with a global bi-variate polynomial fit over the entire image. Some typical aircraft scanner data with these types of distortion present are currently under investigation.

Finally, some questions concerning the implementation of the proposed algorithm are answered in the following:

Question 1: How does one start up the coefficients computation in a region where no boundary information from adjacent regions is as yet available?

Answer: There are two possible methods one can use.

(a): The first method is to use the one-dimensional approach discussed in the first portion of the theory section to approximate the distortion error on the first grid line (upper-most) and the first grid column (left-most) of the grid structures. This provides enough information to get started on the upper-left region of the image. The process for computing the rest of the coefficients then proceeds in a left to right, up then down, region-by-region fashion. This approach is compatible with most of the existing data tape formats where the image is stored line-by-line on tape.

(b): One can also begin the one-dimensional approximation at some line and column where enough correlation information is available to give a good estimate of the distortion along that line and column. One then proceeds in all four directions from the intersection of that line and column. Note that for any given region, the distortion error can be modeled if the equations of any two adjacent boundaries are known. The boundaries do not have to be the upper and the leftmost ones as discussed in the Theory section of this report.

Question 2: Does the formula given in the Theory section of this report for calculating coefficient c of the second order bivariate polynomial no longer valid if one reaches the rightmost grid column or the lowest grid line?

Answer: Yes, but the formula can be modified in a straight-forward manner. For example, when reaching one of the right column grids, by lettering $Z_{23}\Delta$ equal zero, coefficient c can be computed by the following formula:

$$c = \frac{1}{\nabla\theta(W+4)} \left[W_{Z22}\Delta + 2z_{23} - (W+8)a\theta^2 - (W+4)d\theta - (W+2)\,(b\nabla^2 + e\nabla + f)\right]$$

Question 3: Can boundaries of a region be "updated" independently from their neighboring regions and still be continuous? This update may become necessary when two different types of distortions on the same image are encountered in or near that particular region. An example is the sudden drop in altitude of an aircraft which causes the image scale to undergo an abrupt change in that particular region.

Answer: Again the answer to "update" is yes, if one is careful. For example, region B has an updated upper boundary. Region P which is directly above region B now has three out of its four boundaries known. In region P, in order to provide continuity at the lower boundary, a higher order bivariate polynomial with more unknown coefficients must be used to model the distortion error. These extra coefficients are necessary so that all the known coefficients in the boundary equations can be satisfied. The appendix discusses such a case. It is to be noted, however, that some of the given coefficients in the boundary equations must be consistent among themselves to begin with. Higher order bivariate polynomials can also be used if in addition to boundary continuity, the directional derivatives across the boundaries are also required to be continuous. The formulas for calculating these extra coefficients can be derived in a straight forward manner.

The author concludes that the proposed digital image registration scheme using the mathematical programming technique is effective in treating certain geometrical distortion correction problems encountered in remote sensing applications. In particular, it is useful when the images from unstable sensor platforms such as an aircraft scanner are encountered.

XI. REFERENCES

1. Anuta, P. E.: Guide to Use of the Fast Fourier Transform Algorithm for Two-Dimensional Imagery Correlation. LARS Information Note 121069, Purdue University, January 1970.

2. Yao, S.: Notes on Image Correlation and Registration System Improvements. LARS Technical Memorandum, Purdue University, June 1972.

3. Anuta, P. E.: Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fourier Transform Techniques. LARS Information Note 052270, Purdue University, June 1970.

4. Barnea, D. I., Silverman, H. F.: A Class of Algorithm in Fast Digital Image Registration. IEEE Transaction on Computer, Vol. C-21, No. 2, February 1972.

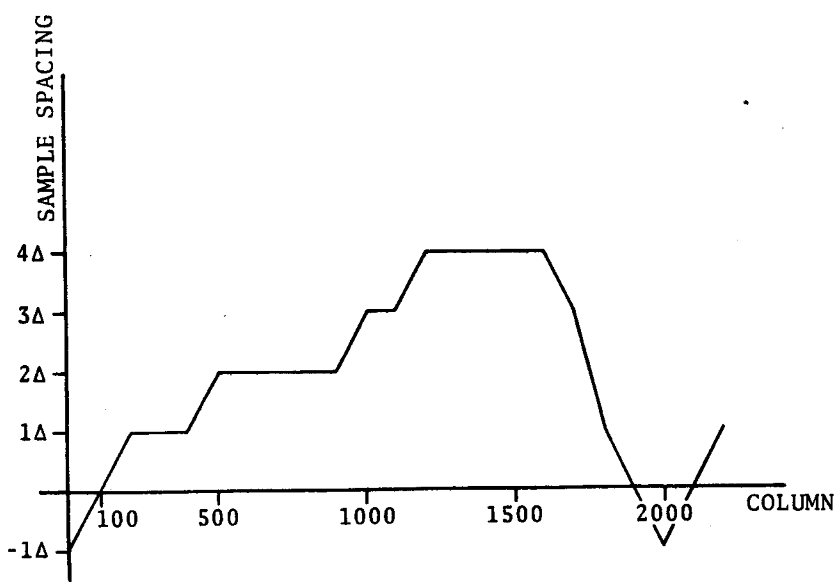


Figure 1. – Column registration error between frame A and B at line 1800.

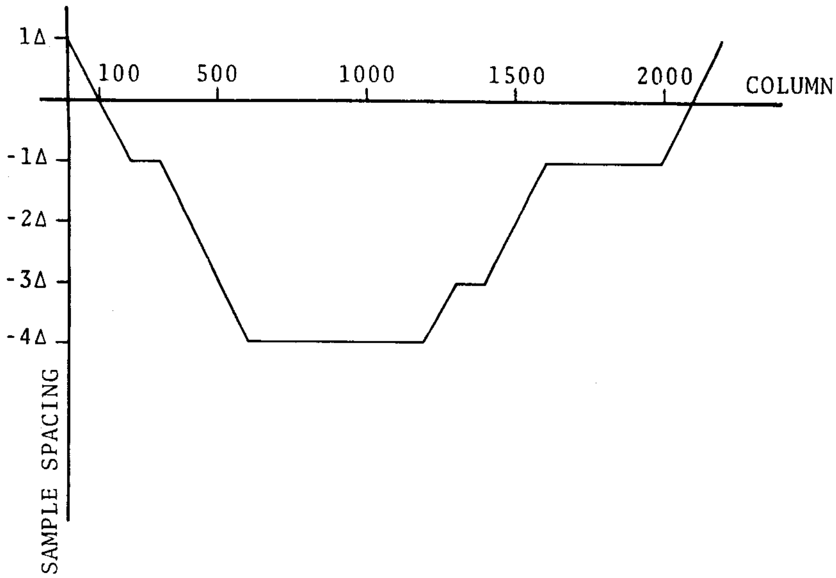


Figure 2. — Column registration error between frame A and C at line 1800.

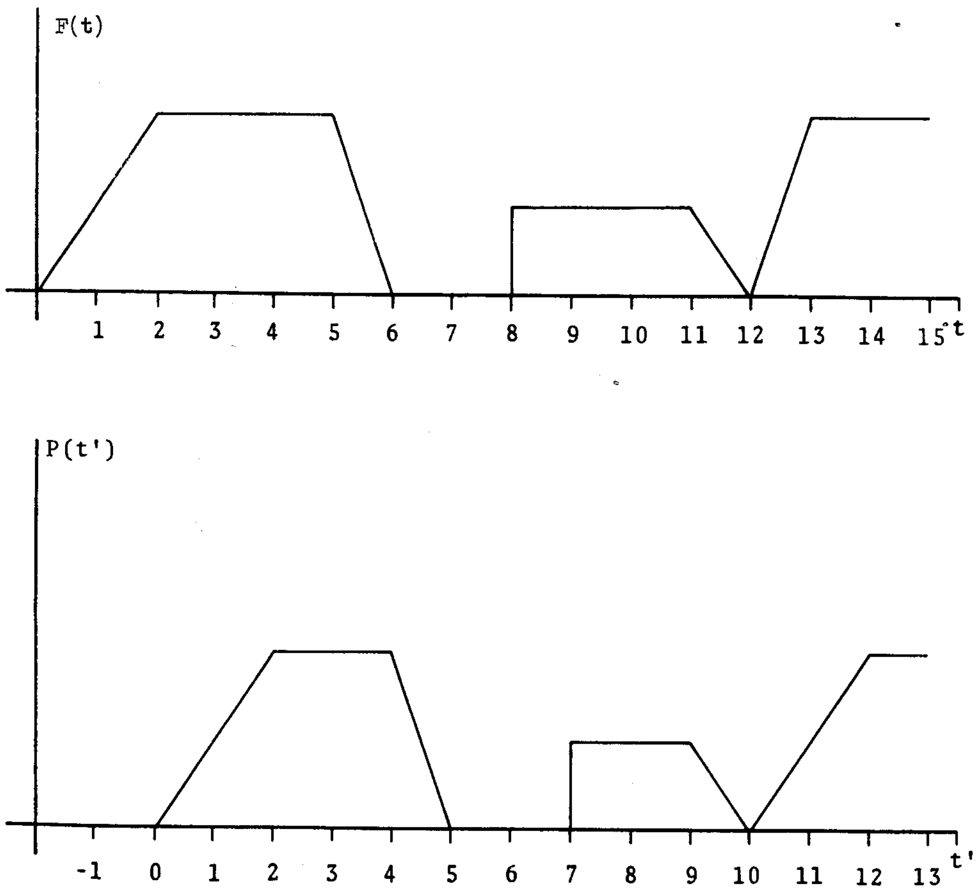


Figure 3. — Reference function f(t) and distorted function P(t').

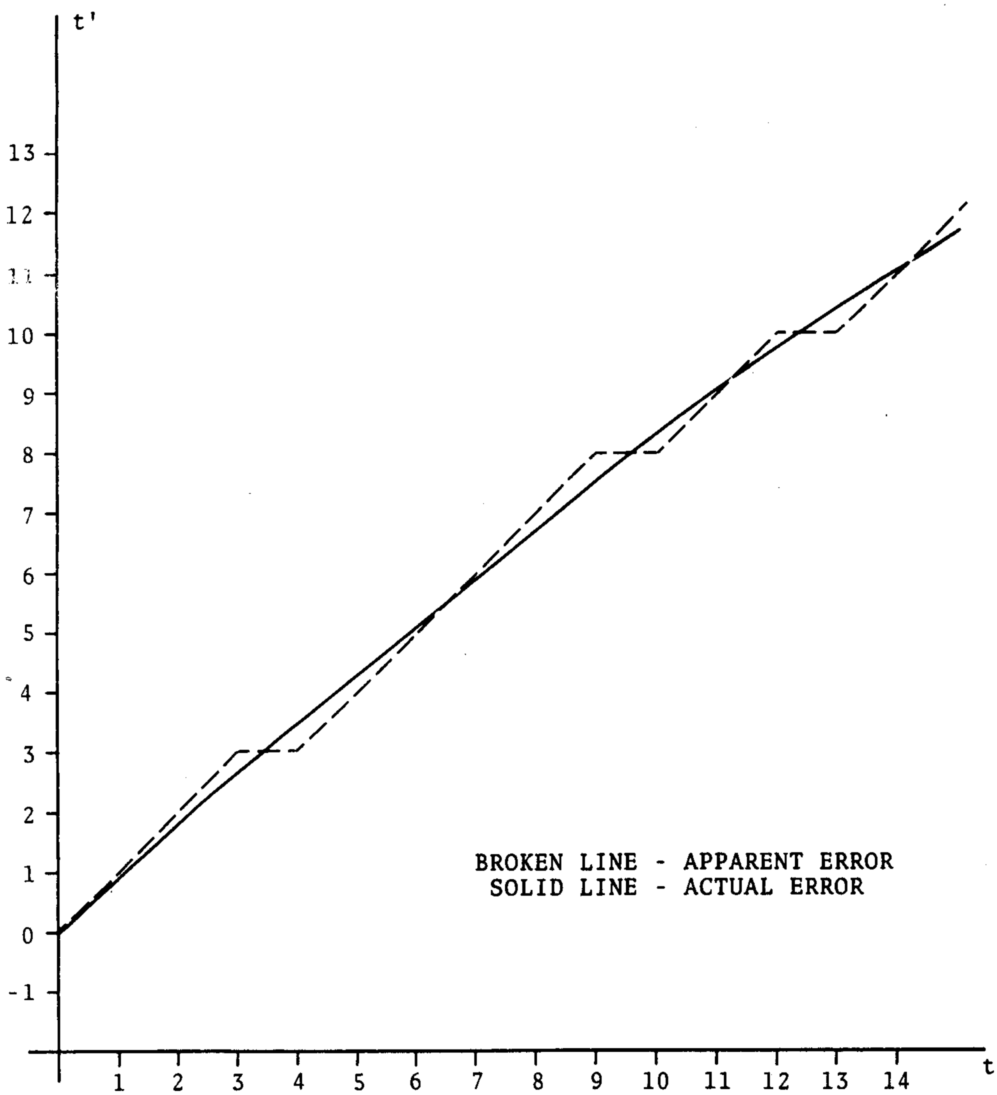


Figure 4. — Error between the coordinates t and t'.

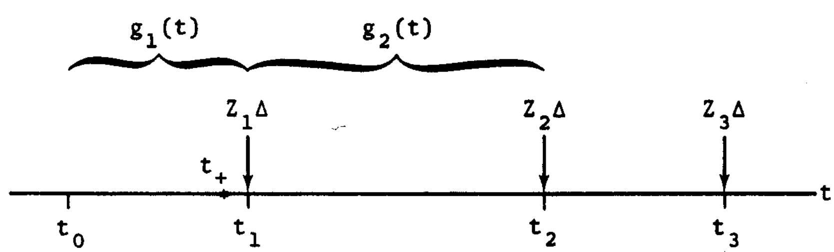


Figure 5. — The functions $g_1(t)$ and $g_2(t)$.

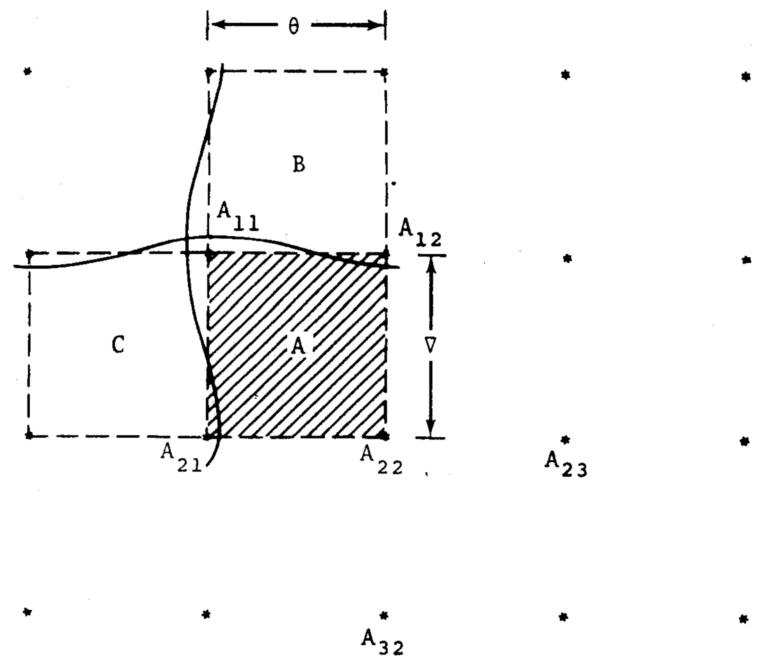


Figure 6. – Correlation grid structure showing regions A, B, and C.

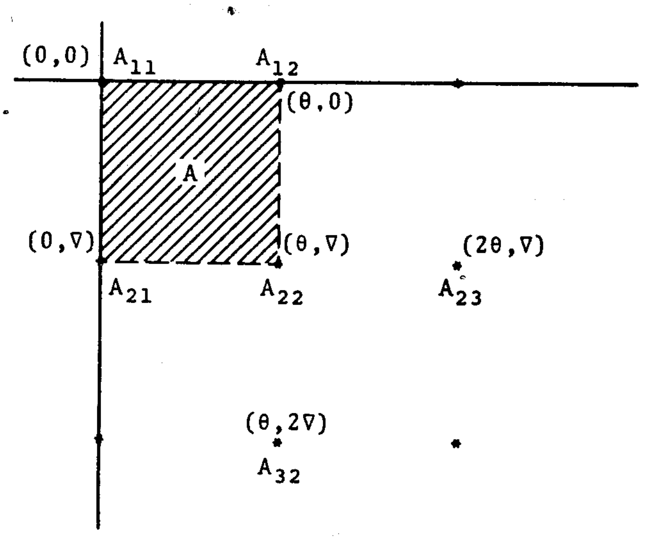
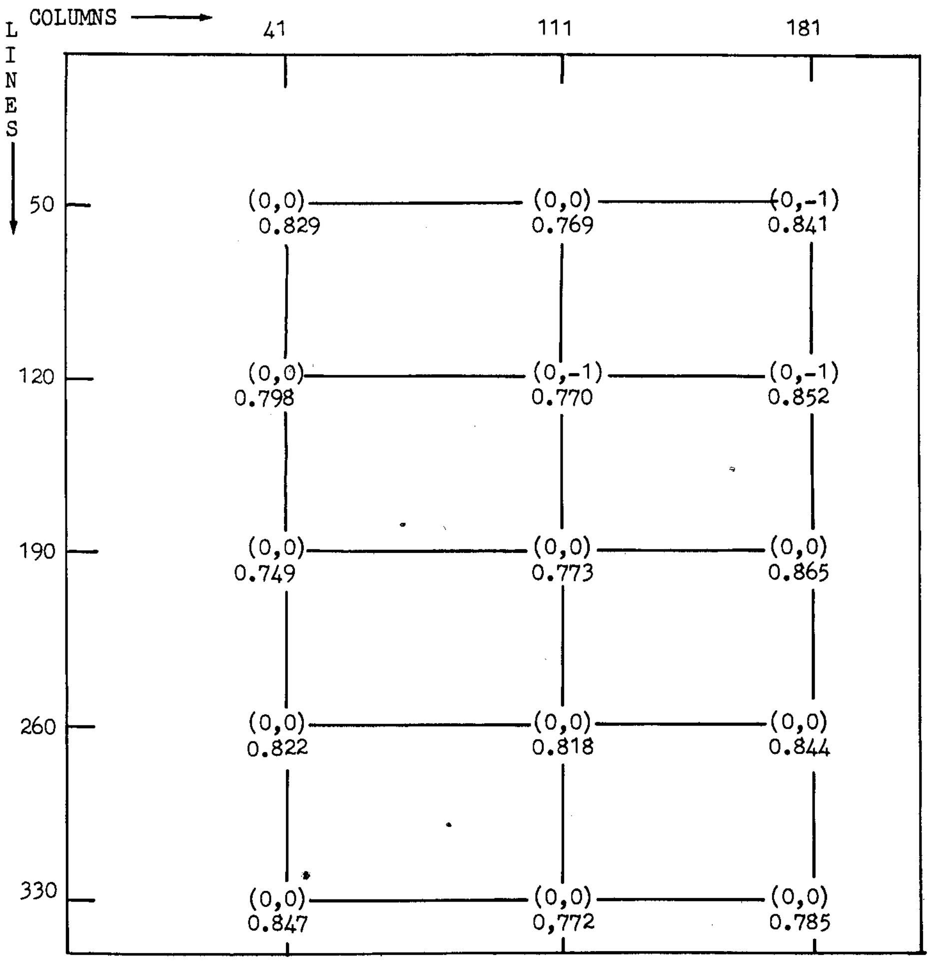


Figure 7. – Two-dimensional coordinate system for region A.



CORRELATION COEFFICIENTS FOLLOW ROW AND COLUMN MIS-REGISTRATION

Figure 8. - Correlation Grid for Channels 1 & 2 before registration

LINES \ COLUMNS →	33	85	137	189
50	(1,0) -0.334	(0,-1) -0.340	(0,-1) -0.406	(1,0) -0.204
120	(-1,-1) -0.181	(-1,0) -0.341	(-1,0) -0.344	(0,0) -0.370
190	(-1,-1) -0.241	(-1,0) -0.431	(-1,0) -0.448	(-1,0) -0.290
260	(-2,-1) -0.249	(-3,0) -0.349	(-3,0) -0.341	(-2,1) -0.247
330	(-3,1) 0.223	(-3,1) 0.277	(-3,2) 0.177	(-3,2) 0.130

Figure 9. - Correlation grid for Channels 1 & 3 before Registration

	33	85	137	189
	(-1,0) 0,891	(-1,0) 0.867	(0,0) 0.906	(1,0) 0.819
	(-1,1) 0.773	(0,1) 0.826	(0,0) 0.818	(1,0) 0.846
	(0,2) 0.652	(0,2) 0.671	(1,1) 0.687	(1,0) 0.652
	(0,3) 0.650	(1,3) 0.632	(2,2) 0.679	(3,2) 0.597
	(1,3) 0.711	(2,4) 0.342	(3,4) 0.225	(4,4) 0.366

Figure 10. - Correlation grid for Channels 1 & 4 before registration

LINES \ COLUMNS →	33	85	137	189
50	(1,0) -0.322	(0,0) -0.357	(0,0) -0.425	(0,-1) -0.213
120	(0,0) -0.139	(0,0) -0.354	(0,0) -0.355	(0,0) -0.371
190	(-1,0) -0.217	(0,0) -0.462	(0,0) -0.487	(0,1) -0.305
260	(0,1) -0.151	(0,0) -0.476	(-1,0) -0.376	(-1,1) -0.194

Figure 11. - Correlation grid for channels 1& 3 after registration

	33	85	137	189
	(0,0) 0.891	(0,0) 0.899	(0,0) 0.904	(0,0) 0.843
	(0,0) 0.863	(0,0) 0.862	(0,0) 0.832	(0,0) 0.901
	(0,0) 0.823	(0,0) 0.818	(0,0) 0.841	(0,0) 0.864
	(0,0) 0.881	(0,0) 0.848	(0,0) 0.876	(0,0) 0.816

Figure 12. - Correlation grid for channels 1 & 4 after registration

VII. APPENDIX

In this appendix it is shown that, given a rectangular region on an image plane where the geometric distortion can be approximated by a bivariate polynomial, all coefficients of the polynomial can be determined even if the equations of three out of the four boundaries surrounding the region are predetermined.

Let $g(x,y)$ be the bivariate polynomial. Let the origin of the coordinate system be located at the upper-leftmost corner of the region. In addition,

Let the upper boundary equation be $a_1x^2 + a^2x + c_1 = 0$

Let the left-hand boundary equation be $b_1y^2 + b_2y + c_1 = 0$

Assume the equation of the lower boundary at $y = \nabla$ is determined by some other means to be $d_1x^2 + d_2x + f_1 = 0$

Note all coefficients $a_1, a_2, b_1, b_2, c_1, d_1, d_2$ and f_1 are known constants.

Modeling the distortion in this region by a third order bivariate polynomial
$g(x,y) = ax^3 + by^3 + cx^3y + dxy^2 + ex^2 + fy^2 + hx + iy + j$

Then $g(x,0) = ax^3 + ex^2 + hx + j = a_1x^2 + a_2x + c_1 \Rightarrow a = 0,\ e = a_1,\ h = a_2,\ j = c_1$

$g(0,y) = by^3 + fy^2 + iy + j = b_1y^2 + b_2y + c_1 \Rightarrow b = 0,\ f = b_1,\ i = b_2.$

Further, $g(x,\nabla) = d_1x^2 + d_2x + f_1 = ax^3 + b\nabla^3 + cx^2\nabla + dx\nabla^2 + ex^2 + f\nabla^2 + i\nabla + j$

$= x^2(c\nabla + e) + x(d\nabla^2 + h) + (f\nabla^2 + i\nabla + j)$

Comparing coefficients

$$c + e = d_1 = c = \frac{d_1 - a_1}{\nabla}$$

$$d\nabla^2 + h = d_2 = d = \frac{d_2 - a_2}{\nabla^2}$$

and $f_1 - f\nabla^2 + i\nabla + j = b_1\nabla^2 + b_2\nabla + c_1$

Therefore, provided $f_1 = b_1\nabla^2 + b_2\nabla + c_1$ initially, all the coefficients of $g(x,0)$ can be expressed on the following as a function of the known constants.

$a = 0$, $b = 0$, $c = \dfrac{d_1 - a_1}{\nabla}$, $\alpha = \dfrac{d_2 - a_2}{\nabla^2}$, $e = a_1$, $f = b_1$, $h = a_2$, $i = b_2$, $j = c_1$

and the fourth boundary will have the equation

$$g(\theta,y) = \left(\theta\,\frac{(d_2 - a_2)}{\nabla^2} + b_1\right)y^2 + \left(\theta^2\,\frac{(d_1 - a_1)}{\nabla} + b_2\right)y + (a_1\theta^2 + a_2\theta + c_1)$$