RECTIFICATION OF SINGLE AND MULTIPLE FRAMES OF SATELLITE SCANNER IMAGERY USING POINTS AND EDGES AS CONTROL

by

Fidel C. Paderes, Jr.
Edward M. Mikhail
Wolfgang Förstner*

School of Civil Engineering
Purdue University
West Lafayette, Indiana 47907

NASA SYMPOSIUM ON
MATHEMATICAL PATTERN RECOGNITION AND IMAGE ANALYSIS

June 6-8, 1984

* Dr. Wolfgang Förstner is an Assistant Professor of Photogrammetry, Stuttgart University, West Germany; he was a Visiting Scientist at Purdue University at the time of writing this paper.

ABSTRACT

Rectification of single and overlapping multiple scanner frames is carried out using a newly developed comprehensive parametric model. Tests with both simulated and real image data have proven, that this model in general is superior to the widely used polynomial model; and that the simultaneous rectification of overlapping frames using least squares techniques yields a higher accuracy than single frame rectification due to the inclusion of tie points between the image frames. Used as control, edges or lines, which are much more likely to be found in images, can replace conventional control points and can easily be implemented into the least squares approach. An efficient algorithm for finding corresponding points in image pairs has been developed which can be used for determining tie points between image frames and thus increase the economy of the whole rectification procedure.

1. INTRODUCTION

1.1 General

Imaging, using scanners as sensors, yields the sensed data about the object in the form of pixels. Knowledge of the relative and/or absolute locations of these pixels in the object space is necessary for mapping, classification, and change detection or monitoring. Of primary interest is scanner imagery of the surface of the earth. The process of finding the location of pixels on the ground for this type of imagery is called rectification. If the reference is another image, the process is known as registration. This research covers rectification and registration of scanner imagery produced by satellite-borne scanners such as LANDSAT MSS imagery. An important element of this research concerns correspondence between two images or between an image and a representation of the terrain (i.e. a map).

If the position of the sensor platform (i.e. satellite) and the attitude of the sensor at the moment of sampling a given pixel is known, and if the interior geometry of the scanner at the same instant can be reconstructed, then the ground position of a pixel can be derived with some assumptions regarding the shape of the terrain. The satellite position can be derived from satellite tracking data. The sensor attitude can be supplied by attitude sensors on-board the satellite. The geometry of the sensor is recon-

structed using calibration data and the imagery. Unfortunately, the accuracy of the satellite position and sensor attitude measurements is not sufficient to produce sub-pixel rectification accuracy.

An alternative method for rectifying satellite scanner imagery is through the use of control information. Control can be in the form of points or edges with known ground and image locations. In this method, given a suitable mathematical model, the parameters needed for relating the image positions of pixels to their ground positions are first computed using control points and applying an appropriate adjustment procedure. Then the ground positions of pixels are computed using the same model and the derived parameters. The same method, with slight modifications, can also utilize edges as control instead of points.

The above method can be further subdivided into two approaches. The first is the interpolative or the surface fitting approach. This approach uses a mathematical series (e.g., polynomial, harmonic) to approximate the true mathematical model relating the image position of pixels to their corresponding ground position. This approach requires an excessive number of control points for uniform rectification accuracy.

The second approach is commonly called "parametric". In this approach, the mathematical model used is based on the geometry of the imaging process. Because of this, it is possible to develop highly accurate models. However, usually simplifying

assumptions are made to make the resulting model tractable, since the geometry of the satellite scanner image is very weak. In this approach, it is possible to exploit a-priori knowledge of the satellite position and sensor attitude, effectively combining the two main methods discussed above.

Both methods mentioned are normally used for rectifying single frames of satellite scanner imagery. This requires that some assumptions be made regarding the shape of the terrain covered by one image frame. Improvement in accuracy can be gained if overlapping frames of imagery are rectified simultaneously in a procedure commonly known as block adjustment.

1.2 Review of the Literature

The earliest approach to rectification utilized interpolative or surface fitting models such as polynomials. This model is easy to implement and gives results comparable to most early forms of the parametric model for satellite imagery (Forrest [13], Trinder [42], Bähr [1], Dowman [10]).

The parametric model based on the geometry of the scanner imaging process has many variations depending on the simplifying assumptions made. The simplest model, which is really designed for aircraft scanner data, assumes that the satellite orbit is a straight line and that the earth is projected onto a mapping plane (Kratky [23],

Konecny [21], Dowman [10]). The orbit of the satellite has been assumed to be a circle (Forrest [12], Levine [26], Synder [41]) or an ellipse (Bähr [2], Sawada [39]). The earth has been assumed to be a sphere (Caron and Simon [9], Bähr [2], Sawada [39]) or an ellipsoid of revolution (Puccinelli [36], Forrest [12], Levine [26], Synder [41]).

The satellite orbit and position can be defined simultaneously in terms of the satellite position and velocity vectors (Caron and Simon [9], Puccinelli [36]). The position of a satellite along an assumed orbit can be defined in terms of time varying orbital parameters (Bähr [2]). Alternatively, the orbital parameters can be assumed constant which results in an ideal orbit. Small deviations of the actual satellite position from the ideal are then modeled as arbitrary functions of time, usually a polynomial series (Levine [26], Mikhail and Paderes [32]).

The sensor, without the scanning action, is nominally pointed along the vertical. Small deviations of the sensor attitude with respect to the vertical are modeled as polynomial functions of time. Bridging of long strips with control at each end only is feasible through the effective use of a-priori attitude information (Friedmann [16]).

1.3 Scope of Investigation

In the early phase of this research, we derived a comprehensive model considering that the earth is an ellipsoid of revolution and the orbit of the satellite is an ellipse (Mikhail and Paderes [32]). All three components of the deviation of the satellite position from the ideal and the three components of the deviation of the sensor attitude from the nominal are incorporated into the model.

Using this model, we developed a system for simulating scanner image data both in the direct and inverse modes. In the direct mode, given the parameters defining the orbit, time, satellite position deviation, sensor attitude and internal sensor geometry, and given the ground coordinate of points of interest, the corresponding image row and column numbers are derived. In the inverse mode, the ground planimetric coordinates of points are computed given the corresponding image row and column numbers, the parameters mentioned above, and the shape of the terrain.

This model has been extensively tested using simulated data and reported on in last year's Symposium (Mikhail and Paderes [32]). Five different sets of experiments were performed to study the following factors: (1) the effect of error in parameter estimates on rectification accuracy; (2) the relative performance between our extensive model, three special cases with simplifying assumptions, and the polynomial model; (3) the effect of different control densities on rectification accuracy; (4) the effect of errors in

derived image position on rectification accuracy; and (5) the effect of errors in measured ground position of control points on rectification accuracy.

In Chapter 2 of this report, additional tests of this model using two frames of real data and the corresponding frames of simulated data employing the same characterizing parameters as the real data are included. Previous conclusions using purely synthetic data were generally confirmed.

With the comprehensive model fully developed and tested for the rectification of single images, effort was directed to the implementation of an extensive block adjustment program. It is based on the same mathematical model and is designed to accommodate data from overlapping satellite scanner imageries. Block adjustment reduces the required amount of control needed to meet a specified level of rectification accuracy. Synthetic data was used to verify the algorithm and the results are included in this report.

From the experience gained by analyzing both synthetic as well as real data, acceptable rectification results require from 20 to 30 control points. Securing this number of points is often difficult and costly because well identifiable "point" features are not abundant. Furthermore, high image and ground positional accuracy for control points is difficult to achieve. Therefore, the research effort was next directed toward an alternative type control. In Chapter 3 of this report, the novel concept of "edge point" is

developed and tested and found to be quite promising. The idea is rather simple in that a control point can be equivalently thought of as a pair of perpendicular edges. Therefore, one edge, which may be considerably easier to find and locate accurately, can be used as control. A point on an edge, which we shall term "edge point", will have 2x2 covariance matrix which is almost singular. This is because such a point provides precise information only in the direction normal to the edge.

Having generalized somewhat the approach to control by introducing the edge points, effort is then directed to the overall problem of correspondence. Chapter 4 reviews the general problem of correspondence and develops an algorithm for locating corresponding objects in image pairs. The algorithm is based on a robust estimation procedure for the parameters of an affine transformation. It has been tested on real image data with simulated distortions, and this early result is given.

2. IMAGE RECTIFICATION

2.1 Theory

The comprehensive model we derived (Mikhail and Paderes [32]) can be used for simulating data both in the direct and inverse modes and for rectification. This model has the following form:

(1)
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda M \begin{bmatrix} X-X_s \\ Y-Y_s \\ Z-Z_s \end{bmatrix}$$

where:

x, y, z — are the coordinates of a given point in the image space. These coordinates are functions of image row and column numbers and the internal sensor geometry;

X, Y, Z — are the corresponding ground coordinates of the given point;

X_s, Y_s, Z_s — are the ground coordinates of the satellite position when the pixel containing the given point is sampled. These coordinates are the sum of the ideal or predicted satellite position and the deviation of the actual satellite position from the predicted one. The ideal position is a function of orbital parameters and time (t) while the deviations are functions of time (t) only;

t — is time which is a function of pixel row and column numbers and the internal sensor geometry;

M is an orthogonal rotation matrix which brings the ground coordinate system into the sensor coordinate system. This is a function of time, sensor attitude, deviation of the satellite position from the ideal, orbital parameters and earth geometry;

λ is a proportional constant which varies from pixel to pixel (i.e. a scale factor).

In this model, small deviations of the satellite position from the ideal (3 components) and the small sensor attitude deviations from nominal (3 components) are modeled as third degree polynomial functions of time. Usually, λ, which is a nuisance parameter, is eliminated resulting in:

$$f_1 = \frac{x}{z} - \frac{m_{11}(X-X_s) + m_{12}(Y-Y_s) + m_{13}(Z-Z_s)}{m_{31}(X-X_s) + m_{32}(Y-Y_s) + m_{33}(Z-Z_s)} = 0$$

(2)

$$f_2 = \frac{y}{z} - \frac{m_{21}(X-X_s) + m_{22}(Y-Y_s) + m_{23}(Z-Z_s)}{m_{31}(X-X_s) + m_{32}(Y-Y_s) + m_{33}(Z-Z_s)} = 0$$

These two equations, which are now in a form suitable for rectification, are then linearized with respect to four groups of variables: (1) the row and column numbers of a given point; (2) the parameters defining time and satellite orbit; (3) the parameters defining the satellite position deviation from the ideal and the sensor attitude; and (4) the ground coordinate of the corresponding point. Other variables defining the internal geometry of the sensor and the geometry of the earth's shape are held constant. Variables in the first and fourth groups vary from point to point, while variables in the

second and third groups are constant throughout a whole frame. The linearized equation has the following general form:

$$A\,v + B_2\Delta_2 + B_3\Delta_3 + \ddot{B}\ddot{\Delta} = f \tag{3}$$

where:

v is a 2 element vector of residuals for the first group of variables (i.e. observed row and column numbers for a given point);

A is a 2x2 matrix of partial derivatives with respect to the first group of variables;

Δ_2 is an 8 element vector of corrections to the approximations for the second group of variables;

B_2 is a 2x8 matrix of partial derivatives with respect to the second group of variables;

Δ_3 is a 24 element vector of corrections to the approximations for the third group of variables;

B_3 is a 2x24 matrix of partial derivatives with respect to the third group of variables;

$\ddot{\Delta}$ is a 3 element vector of corrections to the approximations for the fourth group of variables (i.e., ground coordinates);

$\ddot{B}$ is a 2x3 matrix of partial derivatives with respect to the fourth group of variables;

f is a 2 element vector of constants resulting from the linearization.

The first and fourth group of variables in the linearization are known because they are supplied by ground control points. In rectification, the values of the unknown

parameters in the second and third group of variables are recovered in an adjustment procedure using control points and the linearized model shown in equation (3).

Because of weak satellite scanner image geometry, not all the unknown parameters can be solved for simultaneously. Instead, unknown parameters in the second group of variables are first recovered under the assumption that all parameters in the third group are zero. This is reasonable since the model is designed such that the parameters in the third group are as close to zero as possible. Then, using the same set of control points and the computed values of the parameters in the second group, estimates of the parameters in the third group are derived. Once estimates of all unknown parameters are available, the ground coordinates of any other image point can be solved for with some assumptions regarding the shape of the terrain.

2.2 Experiments With Real Single Frame Data

Two MSS frames taken by LANDSAT 2 are used in this experiment. The first frame covers Kansas State which is relatively hilly. It has 153 uniformly distributed control points. The second frame principally covers the state of Louisiana which is flat. About 1/3 of this frame on the south-east corner is over the sea. It has 192 well distributed control points, although not as uniformly as in the Kansas frame.

Ten cases were run for each frame corresponding to two types of model (collinearity and polynomial) and five control configurations. For each case, withheld control points were used as check points. Table 1 shows the results. The collinearity model is superior to the polynomial model when the control points are few especially in hilly terrain such as the Kansas frame. Also, increasing the number of control points beyond 25 has only a marginal effect on rectification accuracy. This confirms in general our previous results using simulated data (Mikhail and Paderes [32]). Two additional cases for each frame were also run where all the control points were exercised in the adjustment. The RMS of the residuals on control points for the Kansas frame were 58.8 and 57.8 m for the collinearity and polynomial models, respectively. The corresponding values for the Louisiana frame were 61.2 and 60.1 m. These values are the upper bounds of the quality of the data. They are used in the second experiment to determine the precision of the image measurements input into the simulation.

2.3 Experiments With Single Frame Synthetic Data

Using our extensive simulation program, the characteristics of the two real image frames were used to produce simulated images which reproduce as closely as possible the real images with respect to control configuration and accuracy. Simulation was done in the inverse mode, where perfect ground coordinates are calculated from the

Table 1 RMS Error on Check Points in Meters Using Real Data

Number of Control Points	Kansas		Louisiana	
	Collinearity	Polynomial	Collinearity	Polynomial
10*	68.8	117.1	90.4	96.6
15*	67.9	73.6	72.3	71.7
25	67.6	70.4	69.3	67.3
40	67.9	69.5	66.0	65.4
81/70**	63.8	65.5	68.4	68.4

* When the number of control points is low, the number of parameters in the model is reduced to avoid convergence problems.

** 81 control points for Kansas frame and 70 for Louisiana frame.

given image coordinates and derived rectification parameters. Then the calculated ground position of control points for both frames were perturbed using normal distribution with 15 m standard deviation in each of the three coordinates. The image positions were perturbed using a combination of normal and uniform distribution. The uniform distribution used for perturbing both frames has a range of -0.5 to +0.5 pixel, and is used to account for round off errors. The normal distribution used for perturbing the Kansas frame has standard deviations of 0.44 pixel in row and 0.40 pixel in column direction. These are the values which when used in the simulation program produced the RMS values given at the end of the preceding section for the full-control case. The corresponding standard deviations for the Louisiana frame were 0.40 pixel in row and 0.64 pixel in column direction. Several sets of simulated data with the described perturbations but with different "seeds" in the random number generator were produced and rectified. Table 2 shows the results of rectification using a representative simulated data set. Comparing Tables 1 and 2, it can be seen that the trends in Table 1 which resulted from rectification of real data are duplicated in Table 2.

Simulated data using the control configuration of the two real data frames but without perturbations were produced (i.e. perfect data sets). The rectification results using this perfect data set are shown in Table 3. From this table, two significant results can be seen. First, it is possible to recover the correct set of exterior orientation

Table 2 RMS Error on Check Points in Meters Using Simulated Data

Number of Control Points	Kansas		Louisiana	
	Collinearity	Polynomial	Collinearity	Polynomial
10*	84.0	134.4	80.9	89.9
15*	76.9	82.0	78.7	79.6
25	75.4	74.8	72.5	73.8
40	64.6	64.6	65.0	64.8
81/70**	61.9	62.9	60.5	61.0

* When the number of control points is low, the number of parameters in the model is reduced to avoid convergence problems.

** 81 control points for Kansas frame and 70 for Louisiana frame.

Table 3 RMS Error in Check Points in Meters Using Perfect Data

Number of Control Points	Kansas		Louisiana	
	Collinearity	Polynomial	Collinearity	Polynomial
10*	11.8	102.5	10.9	15.4
15*	0.6	13.2	0.3	11.2
25	0.5	10.8	0.3	9.6
40	0.5	10.4	0.3	9.6
81/70**	0.5	9.9	0.3	9.8

* When the number of control points is low, the number of parameters in the model is reduced to avoid convergence problems.

** 81 control points for Kansas frame and 70 for Louisiana frame.

elements using the collinearity model if the data is perfect. Second, and more importantly, it shows that the systematic error inherent in the polynomial model is about 10 meters.

2.4 Theory of Block Adjustment

Given overlapping strips of scanner imagery, instead of performing rectification frame by frame, all frames can be rectified simultaneously in one block adjustment. The main advantage of this approach is that conventional points and edge points common to many frames, even those with unknown ground coordinates, can be exploited to increase rectification accuracy. These points are known as tie points. Another advantage in using this method is that mosaicking of large areas is facilitated.

We implemented a block adjustment procedure for satellite scanner imagery utilizing the same mathematical model used for simple frame rectification. In block adjustment, each point appearing in any frame results in a pair of equations similar to equation (3). This is the linearized form of the mathematical model used for single frame rectification.

Using the method of least squares adjustment (Mikhail [34]), the resulting system of normal equation is of the form:

$$\begin{bmatrix} \dot{N} & \bar{N} \\ \bar{N}^T & \ddot{N} \end{bmatrix} \begin{bmatrix} \dot{\Delta} \\ \ddot{\Delta} \end{bmatrix} = \begin{bmatrix} \dot{t} \\ \ddot{t} \end{bmatrix} \tag{4}$$

where:

$\dot{N}$, $\ddot{N}$, $\bar{N}$ are submatrices of the normal equations coefficient matrix;

$\dot{\Delta}$ is a vector of corrections to the approximations for the unknown parameters in all frames (i.e., Δ_2 and Δ_3);

$\ddot{\Delta}$ is a vector of corrections to the approximations for the ground coordinates of all points;

$\dot{t}$ and $\ddot{t}$ are the resulting constant vectors.

As an example, consider the block of overlapping imagery shown in Figure 1. There are 5 image strips overlapping by approximately 60%. Every strip has 4 frames of imagery and every frame has 9 points in it. The frames are numbered consecutively in the vertical direction along the direction of the strips. The detailed form of the normal equations coefficient matrix is shown in Figure 2.

The contribution to the normal equations of the coordinates of ground points ($\ddot{\Delta}$) are usually eliminated first, resulting in a set of reduced normal equations, which has the form:

$$N\dot{\Delta} = t \tag{5}$$

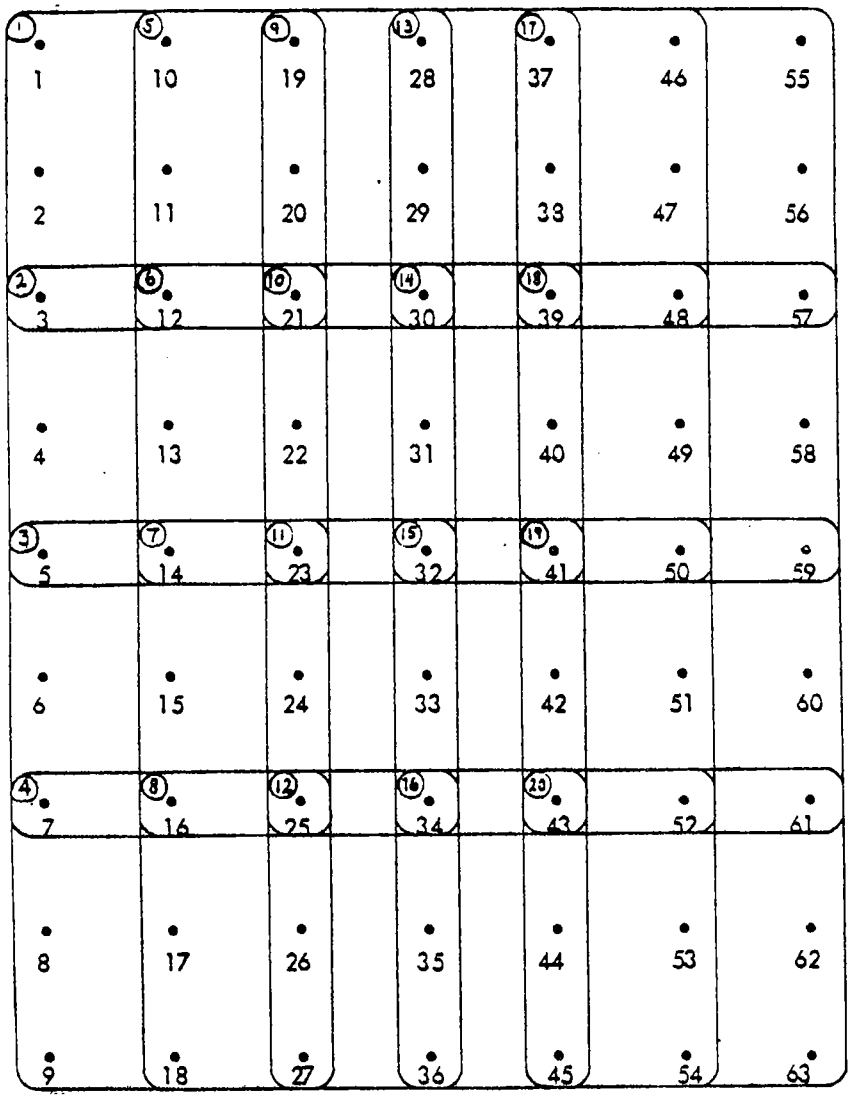

Figure 1 A Block of 5 Strips With 4 Frames Per Strip.

NOTE: (1) the number in circles are the frame numbers
(2) the dots represent common points and the numbers below them are the point numbers
(3) strips are in the vertical direction

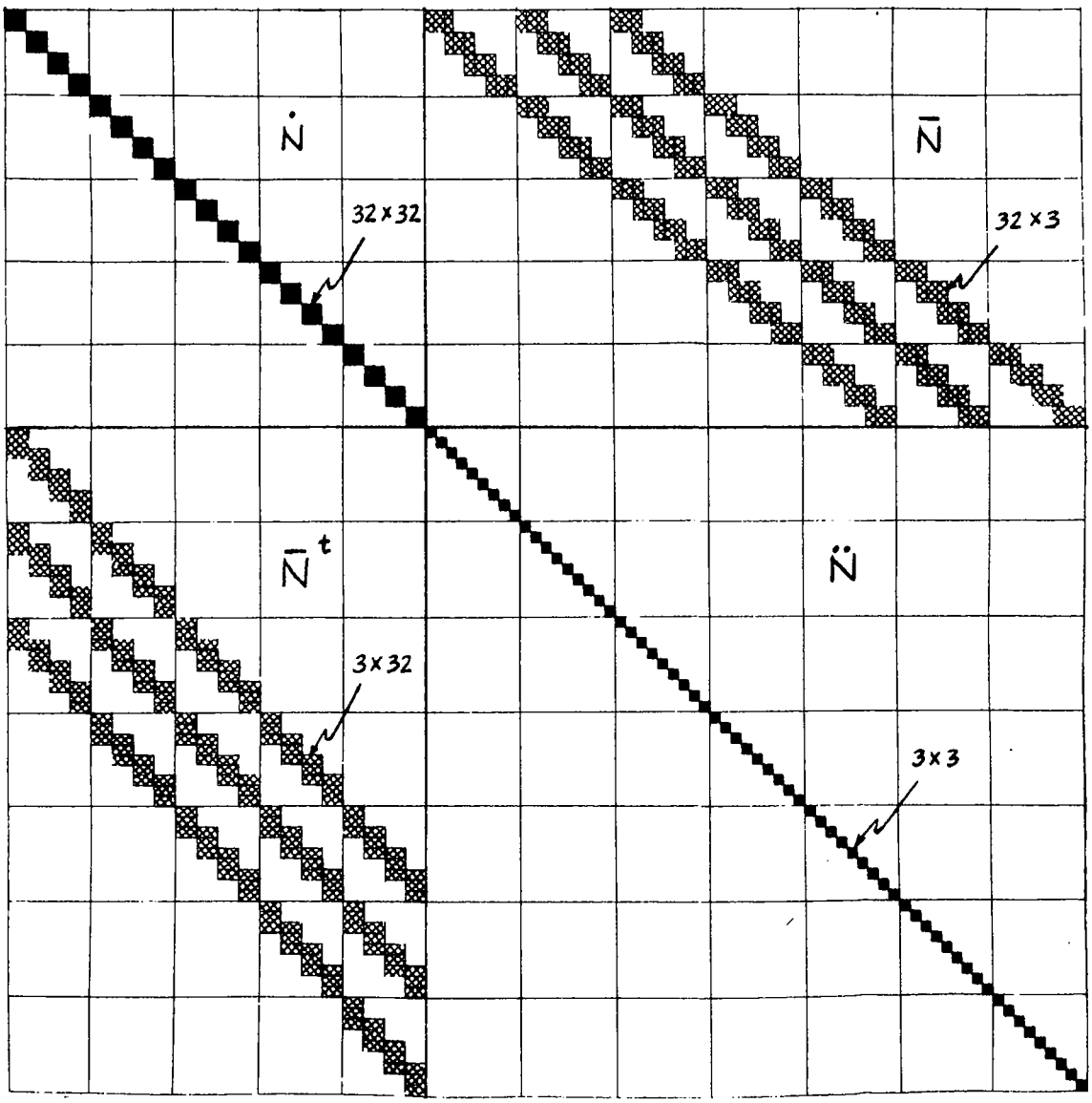

Figure 2 Detailed Form of the Normal Equations Coefficient Matrix.

where:

$$N = \dot{N} - \bar{N}\ddot{N}^{-1}\bar{N}^{T}$$

$$t = \dot{t} - \bar{N}\ddot{N}^{-1}\ddot{t}$$

The reduced normal equations can be formed directly without having to form the total normal equation. Proper numbering of frames results in a banded structure for the reduced normal equation coefficient matrix N. For the block shown in Figure 1, the detailed structure of the reduced normal coefficient matrix N is shown in Figure 3. Each off-diagonal sub-block in Figure 3 is due to points common to a given frame pair. The existence of these subblocks is the reason why block adjustment is more efficient than single frame rectification. As a matter of fact, block adjustment without tie points is equivalent to multiple single frame rectification. Efficient algorithms exist to solve for $\dot{\Delta}$ in equation (5).

2.5 Experiments With A Block of Overlapping Synthetic Image Data

A block of a total of 9 frames, composed of 3 adjacent strips and 3 frames per strip were simulated. The center of the block is approximately at 58.5°N latitude. The frames have about 60% sidelap between strips and 15% overlap along each strip. There are 454 control points at a grid interval of 20 km, and 453 check points also at a grid interval of 20 km. The check point grid is displaced by 10 km in both Easting and

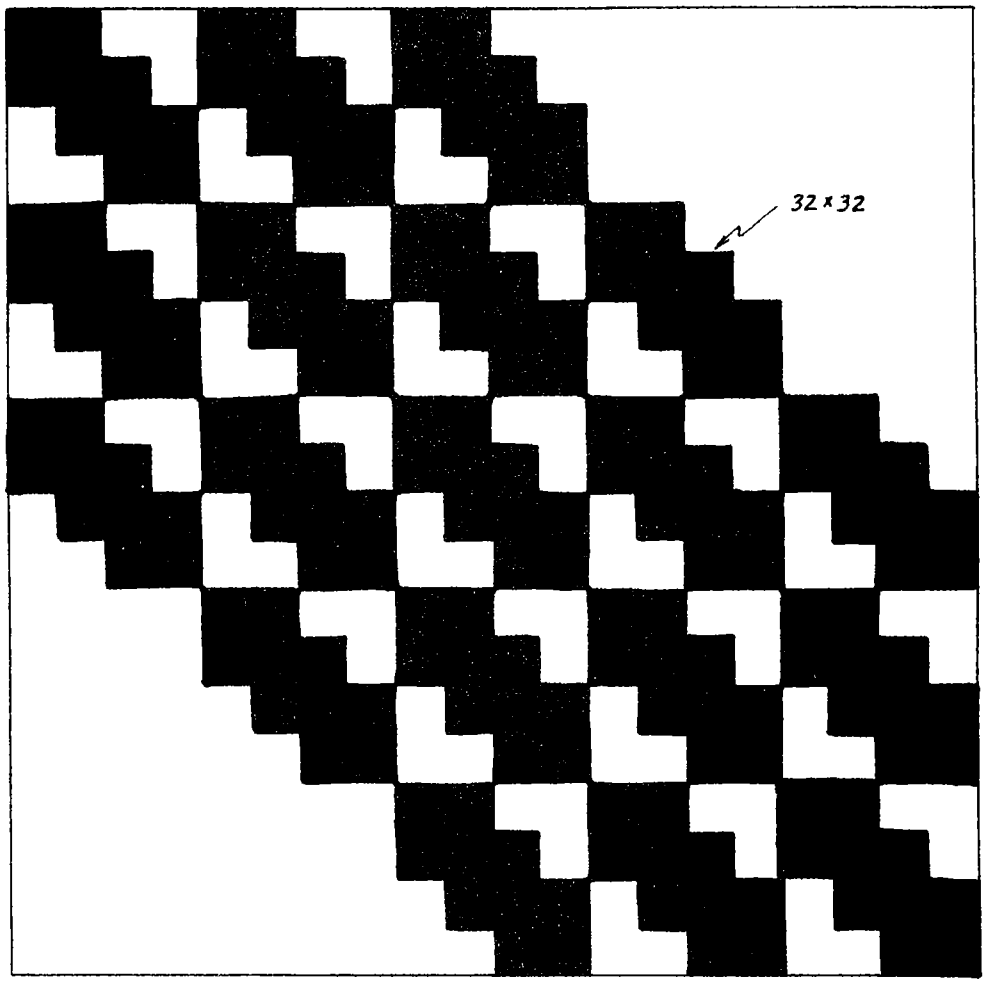

Figure 3 Detailed Form of the Reduced Normal Equations Coefficient Matrix.

Northing with the result that each control point is surrounded by 4 check points and vice versa. The ground position of both sets of points were perturbed by 15 m standard deviation in each of the three coordinate directions using the normal distribution. The image position of both sets were also perturbed using a combination of uniform and normal distribution. The uniform distribution has a range of -0.5 to +0.5 pixel. The normal distribution has a standard deviation of 0.5 pixel in both row and column direction. Five cases of block adjustment were run with different control configuration. Table 4 shows the number of control and check points for each frame and for the whole block for each of the 5 cases. It also shows the number of tie points in the block for all cases. A tie point is any point common to two or more image frames which has known image positions but unknown ground position and is included in the block adjustment. In this experiment, the ground elevation of tie points were constrained to its a-priori value. This is necessary because it was previously shown that elevations cannot be recovered with sufficient accuracy using block adjustment techniques for aircraft scanner data (McGlone and Mikhail [30]) and aircraft scanner imagery has a much stronger geometry compared to satellite scanner imagery.

Table 5 shows a relative comparison of RMS errors on check points on a frame by frame basis between block adjustment and single frame rectification for all five cases. The case where the parameters are perfectly known is included as a reference. It

Table 4 Number of Control, Tie, and Check Points Used in Block Adjustment Experiments

Cases / Frames	Number of Control Points					Number of Check Points
	1	2	3	4	5	
1	11	15	27	45	91	90
2	9	13	24	39	91	89
3	9	13	26	41	88	87
4	11	15	25	45	89	86
5	11	15	26	42	89	86
6	10	14	26	41	90	86
7	10	14	25	42	88	88
8	10	15	26	44	89	87
9	11	16	26	42	85	88
Block *	42/ 224	66/212	125/180	214/134	454/0	453

* Control points/tie points.

Table 5 A Comparison of Check Point RMS Error Between Block Adjustment (σ_{BA}) and Single Frame Rectification (σ_{SF}).

Cases / Frames	The Ratio σ_{BA}/σ_{SF} in Meters					Perfect Parameters
	1*	2*	3	4	5**	
1	93/-	79/92	66/76	67/70	66/66	65
2	77/-	76/-	68/79	74/80	69/69	62
3	117/-	100/-	73/81	80/79	79/79	68
4	87/-	77/98	65/73	67/66	65/65	63
5	76/-	74/142	67/73	70/72	68/68	64
6	79/-	74/142	63/69	69/70	63/63	62
7	113/-	70/85	65/72	65/68	65/65	59
8	92/-	97/-	64/81	69/76	68/68	60
9	83/-	72/82	65/78	67/69	68/68	62
Ave.	90.8/-	79.9/-	66.2/75.8	69.8/72.2	67.9/67.9	62.8

* Single frame rectification did not converge because of few control points (no model parameter reduction is exercised in this case).

** Block adjustment for case 5 is the same as single frame rectification because there are no tie points.

clearly shows that tie points, which are much more readily available (and less expensive) than control points, have a beneficial effect on rectification accuracy especially when control points are few. This improvement in accuracy is essentially due to tie points because block adjustment without tie points is equivalent to single frame rectification.

3. EDGES AS CONTROL

3.1 Edge Points

For a typical image frame, the necessary number of control points with the desired distribution and accuracy is difficult and sometimes impossible to secure because features that can serve as control points are few. By comparison, edges and lines occur more often and in combination with points, the necessary amount of control can more easily be satisfied if a method is devised that can utilize lines and edges as control.

A straight edge or line can be represented by a single point on that edge, preferably near the middle, and a direction. We call that point an *edge point.* Edge points on the ground, or maps representing the ground, can be identified and transferred into the corresponding image manually. The position of edge points on the image can then be measured in a direction perpendicular to the edge with an accuracy comparable to conventional points or even better. The covariance matrix for the position of the edge point in the (l,p) coordinate system is

$$\Sigma_{lp} = \begin{bmatrix} (\kappa\,\sigma_p)^2 & 0 \\ 0 & \sigma_p^2 \end{bmatrix} \tag{6}$$

where:

l is parallel to the line

p is perpendicular to the line

σ_p is the standard deviation of edge point position perpendicular to the edge

σ_l is the standard deviation of edge point position along the edge

κ is equal to σ_l/σ_p and is assigned a very large value

The direction of the edge, θ, can also be measured on the image. The covariance matrix of the edge point in the (r,c) coordinate system is

$$\Sigma_{rc} = R_\theta \, \Sigma_{lp} \, R_\theta^t \tag{7}$$

where:

r is the row direction in the image

c is the column direction in the image

R_θ is the rotation matrix with argument θ

Another method of finding the edge point on the image is through the use of digital correlation. First a window centered on the edge point on the map is digitized approximately in the row-column direction of the image. This window is then correlated with the image, with or without image pre-processing such as edge detection, resulting in image position of the edge point. The corresponding position covariance matrix can then be computed using (Forstner [14]):

$$(8) \qquad \Sigma_{rc} = \sigma_n^2 \begin{bmatrix} \Sigma\,(\partial g/\partial r)^2 & \Sigma\,(\partial g/\partial r)\,(\partial g/\partial c) \\ \text{sym.} & \Sigma\,(\partial g/\partial c)^2 \end{bmatrix}^{-1}$$

where:

σ_n	is the standard deviation of image noise
g	is the density of the image inside a window containing the edge point
r, c	are the row and column numbers
$\partial g/\partial r$, $\partial g/\partial c$	are the partial derivatives of g with respect to r and c.

Before the location of edge points are transferred into the image, their locations are first defined in the map or ground, hence edge points can be treated as ordinary points as far as their ground positions are concerned. Once their image postions are defined, edge points can be easily incorporated into existing rectification programs.

In theory a single edge point is enough to represent a straight edge segment, but in practice more than one point may be necessary, especially if the segment is not really straight.

3.2 Experiments With Edge Points as Control In Single Frame Rectification

In our experiments using edges as control for rectification, we ran ten cases with

different edge distributions. Figure 4 shows a schematic representation of all ten cases. In case (1), edge point pairs have the same coordinates and are positioned at a regular grid. The angle between the edges in an edge pair is fixed at 90°. Case (2) is the same as case (1), except that the acute angle between the edges in an edge pair varies randomly within the range 60° to 90°. Case (3) is the same as case (2), except that the range for this case is from 30° to 90°. Case (4) is the same as the previous cases except that the direction of edges in this case is totally arbitrary. Cases (5) to (8) are the same as in cases (1) to (4), respectively, except that the position of one edge point in an edge pair is randomly perturbed within the range -100 to +100 pixel. Case (9) is the same as in case (1) except that the position of each edge pair is now randomly distributed over the whole image frame. Case (10) is the most general case. In this case both the position of the edge points and the direction of edges are totally arbitrary.

The amount of contamination applied to all ten cases to simulate random errors was the same. In the image, the ideal coordinates of edge points were perturbed using a mixture of uniform and normal distribution along the edge direction and perpendicular to it. The uniform distribution has a range of -0.5 to +0.5 pixel in both directions representing the discretization errors. The normal distribution has a standard deviation of 0.5 pixel perpendicular to the edge and 25 pixels along the edge representing the identification errors. The ground position of edge points were perturbed using the nor-

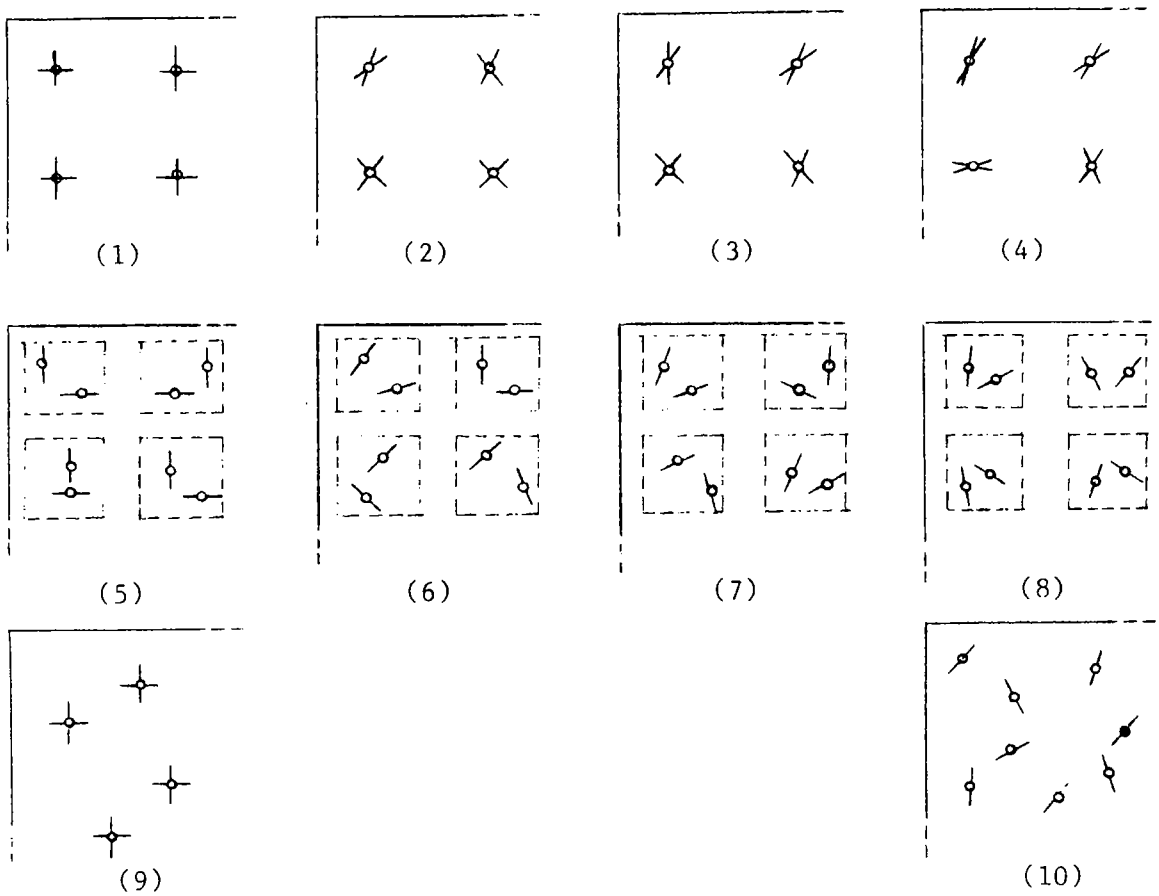

Figure 4 Distribution of Edges for Different Cases of Rectification With Edges as Control.

NOTE: (1) Pairs of edge points, having the same coordinates, are positioned at regular grid. The angle between the edges is 90°.
(2) Same as (1), except that the angle is at least 60°.
(3) Same as (1), except that the angle is at least 30°.
(4) Same as (1), except that the angle is arbitrary.
(5)-(8) Same as (1)-(4) respectively, except that the coordinates of edge points randomly deviate from regular grid up to 100 pixel.
(9) Same as (1), except that the position of a pair of edge points is totally random.
(10) Both the position of an edge point and the direction of the corresponding edge are arbitrary.

mal distribution with standard deviation of 15 meters in each of the three coordinate directions. The number of edge pairs for all cases varied from 25 to 145.

Check points were used to measure the accuracy of rectification. There were 144 check points situated on a uniform grid. For comparison purposes, the same check points were used for all cases. The image position of check points were perturbed in the same manner as edge points, except that the perturbations were applied in the row and column direction instead of along the edge and perpendicular to it and that the standard deviation for the normal component for the row and column direction were both 0.5 pixel. The ground position of check points were perturbed in exactly the same manner as those for edge points.

Each case in Figure 4 is replicated ten times using independent perturbations. A tabulation of the average rectification accuracy and the corresponding standard deviation are shown in Table 6. The average rectification accuracy for all cases are also shown in Figures 5 to 8. In these figures, the abscissa is the number of edge pairs and the ordinate is the average rectification accuracy of the ten replicates in meters. Each curve corresponds to the case number as annotated in the figures.

Figure 5 shows the results from cases (1) to (4). The only difference between these cases is the angle between the edges in an edge pair. It can be seen from the figure that decreasing the angle between edge pairs results in a corresponding decrease in

Table 6 Mean and Standard Deviation in Meters of the RMS Errors at Check Points Using Edge Points as Control

(Each Case Consists of Ten Replicates)

	Case 1		Case 2		Case 3		Case 4		Case 5	
No. of Line Pairs	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
25	73.17	1.13	79.85	2.22	77.92	1.28	105.66	8.86	69.92	0.74
41	69.24	0.67	75.24	2.38	74.51	1.21	85.76	2.82	67.94	0.56
81	65.81	0.74	69.26	1.22	70.42	0.76	75.22	1.46	65.63	0.52
145	64.59	0.71	-	-	-	-	69.29	1.08	-	-

	Case 6		Case 7		Case 8		Case 9		Case 10	
No. of Line Pairs	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
25	77.23	2.32	80.41	2.39	95.35	4.57	154.34	27.06	211.57	29.27
41	72.33	1.50	76.69	1.79	81.52	3.06	72.71	2.30	94.14	5.15
81	66.70	1.27	71.75	1.35	70.47	1.21	66.49	1.41	73.29	1.56
145	-	-	-	-	67.61	0.83	64.30	1.53	68.83	1.14

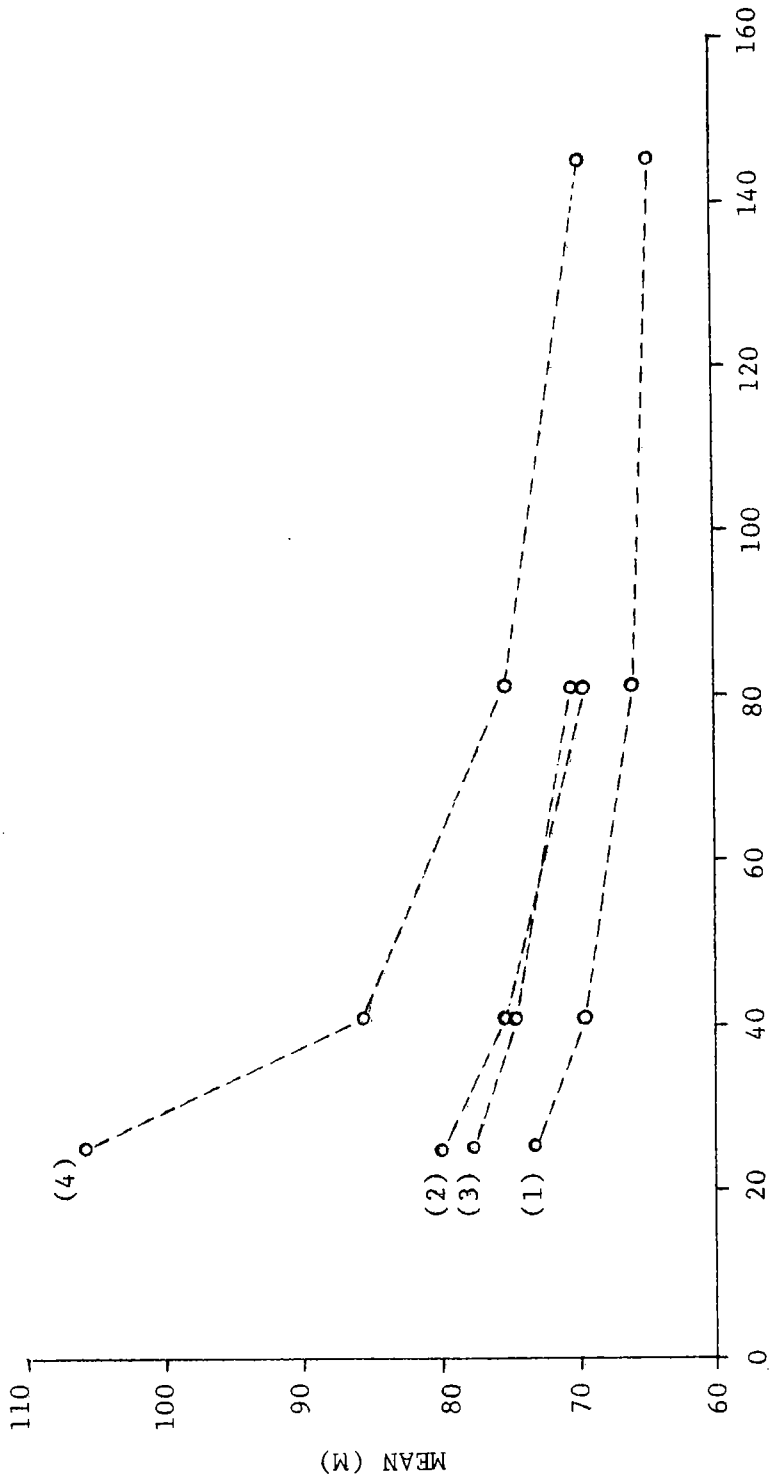

Figure 5 Plot of Rectification Results for Cases (1) to (4)

NOTE: Case (1): Pairs of edge points, having the same coordinates, are positioned at regular grid. The angle between the edges is 90°.

Case (2): Same as (1), except that the angle is at least 60°.

Case (3): Same as (1), except that the angle is at least 30°.

Case (4): Same as (1), except that the angle is arbitrary.

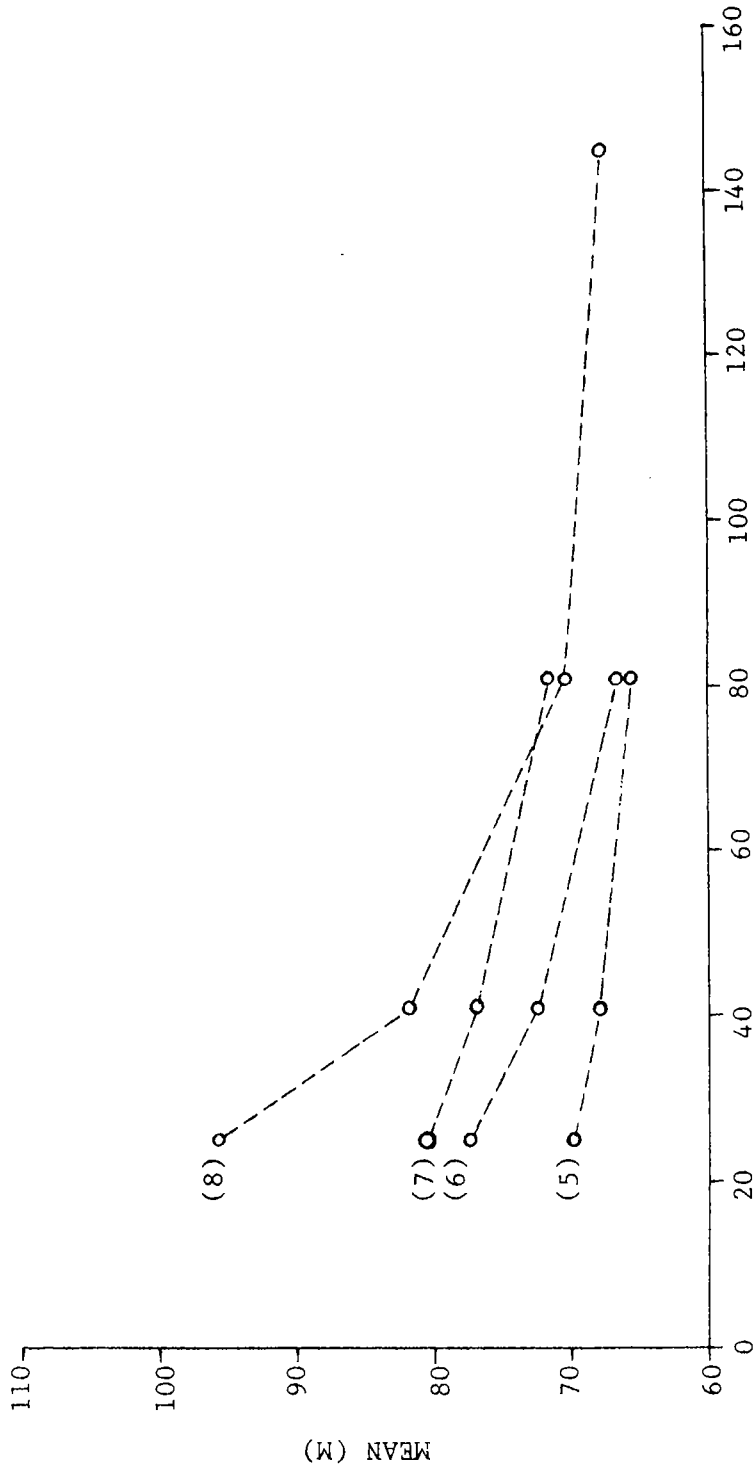

Figure 6 Plot of Rectification Results for Cases (5) to (8).

NOTE: Case (5): Edge points in an edge pair deviate randomly from regular grid up to 100 pixel. The angle between edges is 90°.
Case (6): Same as (5), except that the angle is at least 60°.
Case (7): Same as (5), except that the angle is at least 30°.
Case (8): Same as (5), except that the angle is arbitrary.

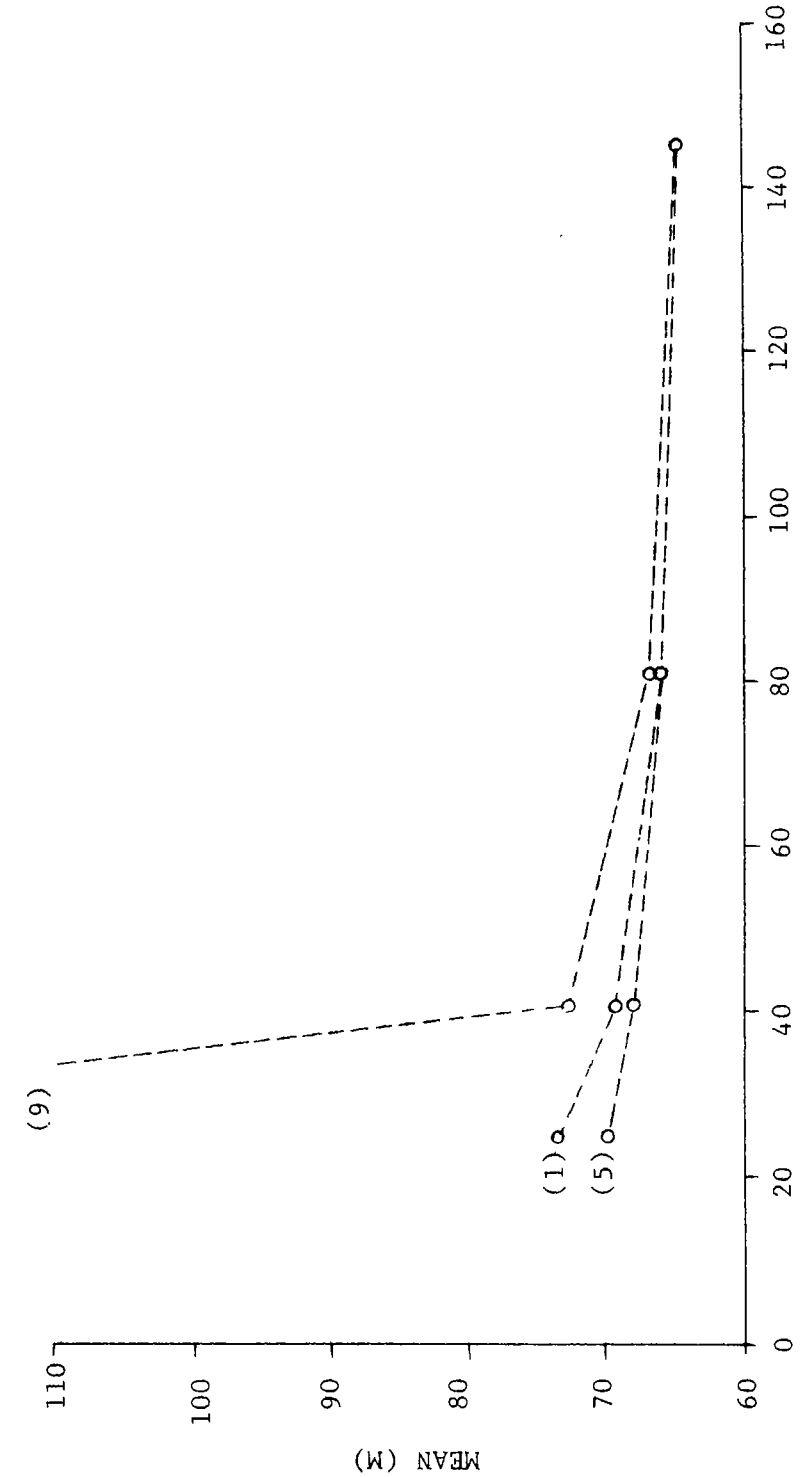

Figure 7 Plot of Rectification Results for Cases (1), (5), and (9).

NOTE: Case (1): Pairs of edge points, having the same coordinates, are positioned at regular grid. The angle between edges is 90°.

Case (5): Same as (1), except that the coordinates of edge points randomly deviate from regular grid up to 100 pixel.

Case (9): Same as (1), except that the position of a pair of edge points is totally random.

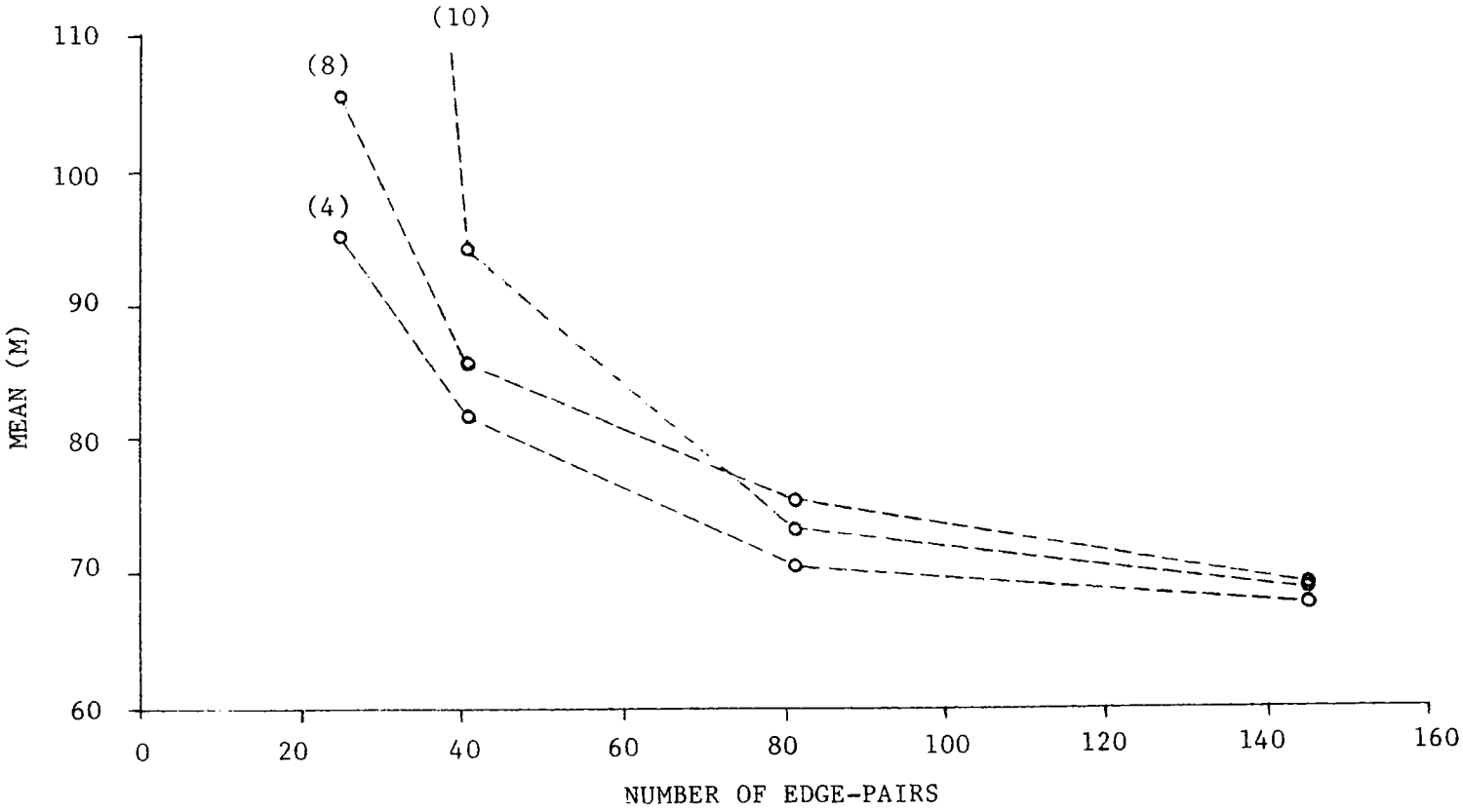

Figure 8 Plot of Rectification Results for Cases (4), (8), and (10).

NOTE: Case (4): Pairs of edge points have the same coordinates and are positioned at a regular grid. The angle between edges is arbitrary.

Case (8): Same as (4), except that the coordinates of edge points randomly deviate from regular grid up to 100 pixel.

Case (10): The position of edge points and the direction of edges are totally arbitrary.

rectification accuracy. This result is essentially repeated in Figure 6, because the only difference between these two figures is that the distance between edge points in an edge pair in all cases shown in Figure 5 is fixed at 0 while that for Figure 6 ranges up to 200 pixels. Comparing cases (1) and (4), or (5) and (8) in Figures 5 and 6 shows , that about 2 times more edge pairs are necessary to achieve the same accuracy as with conventional control points alone.

The effect of the distance between edge points in an edge pair is shown in Figures 7 and 8. Figure 7 shows cases (1), (5), and (9) where the angle between edges in an edge pair is fixed as 90°. Figure 8 shows cases (4), (8), and (10) where the angle is totally arbitrary. Separating the edges in edge pairs is beneficial up to a certain point. Total random distribution of edges over the whole image frame is inferior to other distribution when control edges are few.

Figure 9 is a comparison between cases (9) and (10). In case (9), where edge points in an edge pair have the same image coordinates and the pair of edges intersect at 90°, an edge pair is equivalent to a single control point. Case (10), where edges have totally arbitrary direction and distribution over the whole image frame, is the most extreme of all the ten cases studied. It can be seen from the figure that in order to achieve rectification accuracy when using edges comparable to that achieved when using conventional points, the number of edge pairs should be approximately 3 times the number

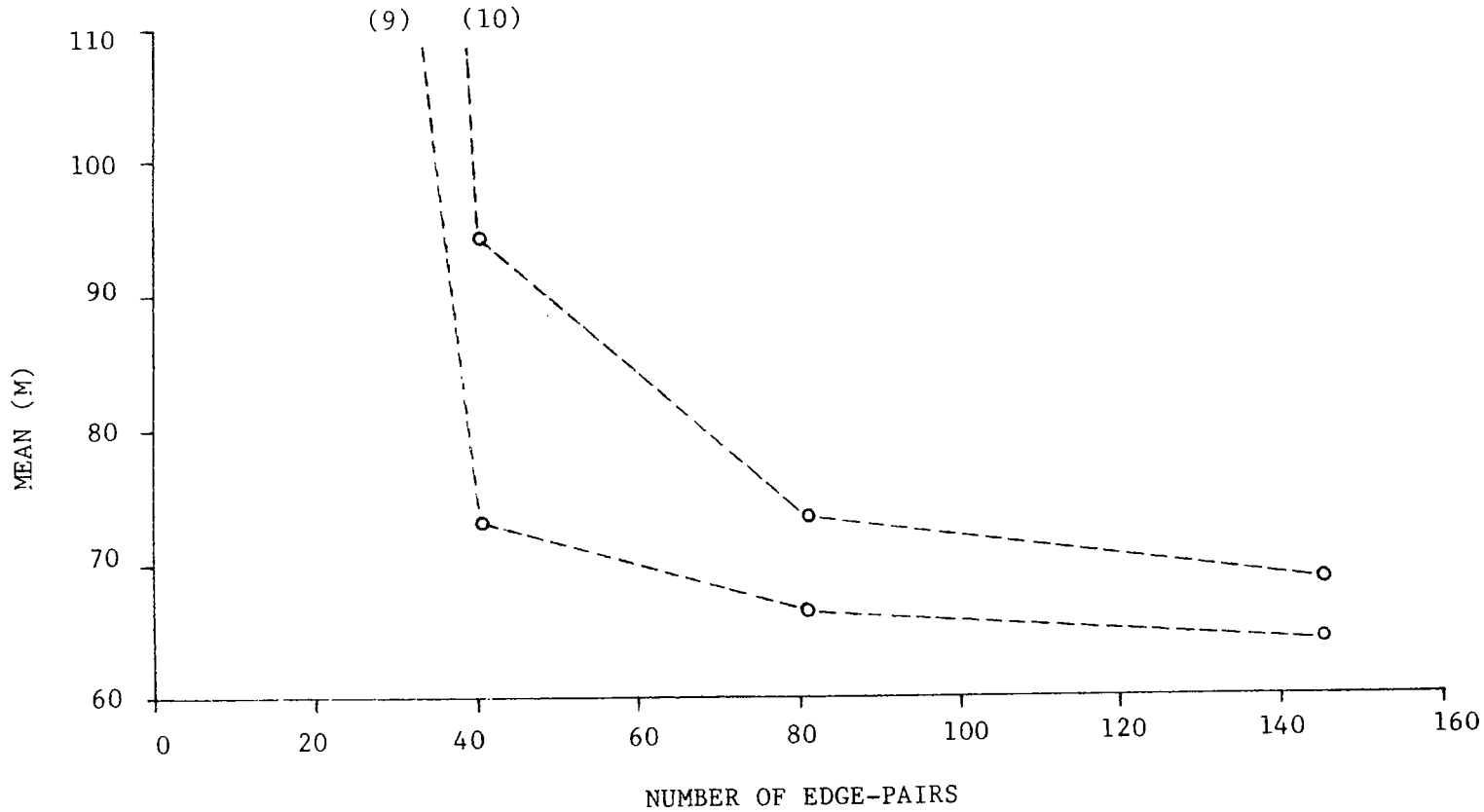

Figure 9 Plot of Rectification Results for Cases (9) and (10).

NOTE: Case (9): Pairs of edge points, having the same coordinates, are randomly distributed over the whole image frame. The angle between edges is 90°.

Case (10): The position of edge points and the direction of edges are totally arbitrary.

of control points. Thus it is worthwhile to attempt selecting well distributed control features.

Summarising the results of this approach, it has been shown that edge points can efficiently replace conventional control points. As they are much more likely to be found in an image and can be measured with at least the same precision as conventional points, one can expect that the overall rectification accuracy may even be improved.

For a practical implementation, especially to decrease the requirements on the skill of the operator, one should provide automatic algorithms for finding corresponding control features, for both conventional and edge points. This addresses the problem of scene matching. The next section is devoted to this problem and presents an algorithm, which is developed first for finding corresponding tie points in overlapping image frames.

4. ROBUST ESTIMATION FOR CORRESPONDENCE

4.1 A Hierarchical Approach to Correspondence

Scene matching is a basic requisite for different tasks which use the geometric properties of images, such as terrain classification, the derivation of digital height models or, map production. It is also the first step in applications where images are used for determining individual points in three dimensions as in photogrammetric triangulation. In all these cases either one image is related to another image (i.e. registration) or to a topographic map, (i.e. rectification).

Obviously there is no simple way to accomplish this task in one step. One rather has to pass several levels in a hierarchical way, where the results of one are the approximations for the next level. This is similar to the way the human visual system is believed to behave (Marr [28]). If one starts with a satellite image with a relative resolution of, say, 10^{-4}, i.e. 10^4 pixels per line, one could imagine a 4 step procedure, where each step increases the precision of rectification by about one order of magnitude:

1. A global image match which defines the position and the orientation to an accuracy of 2-10%, i.e. 200-1000 pixels, and 1–5°. This task is usually done by an operator but may use the very efficient algorithm by Lambird et.al. [25] (see also Stockmann, et.al. [40]).

2. In order to approximately compensate for unknown sensor position and attitude, and for relief displacements due to undulations of the terrain, one might continue with the matching of image patches. The number of these patches will depend on the roughness of the terrain in comparison to the flying height, and range from a few, say 5 or 10 to a hundred. The size of the image patches will be chosen in a way that the expected displacement will be less than half the linear patch size, thus between 4 and 20% of the side of the image. In order to keep the amount of data in a reasonable range one will use a reduced resolution, say between 1/2 and 1/8, leading to linear patch sizes of 50-200 pixels. The algorithm should be able to compensate for at least linear, i.e. affine distortions, and should lead to accuracies of 2-10 pixels, referring to the original image. Since high accuracy is not required, one might effectively use methods of structural pattern recognition to advantage by extracting scene features. One of the most promising algorithms for this step is the one by (Barnard and Thompson [4]).

3. Since fine correlation using differential methods requires approximate values which are within 1.5 pixels of the final match (Förstner [14]), an intermediate step is necessary. Here, all correlation-based methods can efficiently be used as the search area is very small. The window size will range beween 16 and 32 pixels (linear). The aim in this step is to choose a fast, robust, and reliable algorithm.

Possible candidates for this task are sequential algorithms (Barnea and Silverman [5]), the phase correlation technique (Kuglin and Hines [24]) or binary correlation. But of course the algorithm of step 2 could be used here too.

4. Fine correlation in the last step may yield subpixel accuracy, if the texture in the image allows and if it is required for the final product. Here, differential algorithms (Cafforio and Rocca [8], cf. Förstner [14]) are most efficient. The window size, depending on the texture, will range between 8 and 32 pixels. If the pixel size is adapted to the (spatial) spectrum of the images, accuracies of 0.2 pixels or better can be reached under production conditions (Bernstein [7], McGillem and Svedlow [29], Förstner [15]).

This sequence of steps has of course to take the special boundary conditions of the production into account, and may be varied accordingly. The main steps, however, will have to use similar algorithms. The concept is quite different from those used for the rectification of aerial images for orthohoto production, e.g. using the Gestalt Photo Mapper or the approach by Panton [35]. These systems do not have to cope with the weak geometry of satellite imagery, thus need only few control points. They can therefore use the internal geometry of the stereo pair for recursively updating the approximations for the fine correlation. Unlike these procedures, the above described hierarchical set up allows a great deal of parallelism in steps 2-4.

From the above mentioned algorithms the one by Barnard and Thompson needs further discussion. Its general line of thought can also be found in the approach by Lambird et.al. [25] and in the procedure by Marr, Poggio and Grimson (Marr [28], Grimson [17], cf. Kak [20]) .

With respect to its application in registration and rectification, the generality of Barnard and Thompson's geometric model turns out to be a disadvantage due to the resulting high numerical effort. Since the second step in the hierarchical procedure for registration and rectification is decisive for its reliability, this step has to be designed such that the actual data structure is taken into account, and it must also be flexible. Therefore a new algorithm has been developed, which can be used for registration and in particular for selecting tie points for rectifying overlapping image frames.

We will first formulate the problem of matching image patches of moderate sizes in subsection 4.2, discuss two of the algorithms and sketch the new one. Subsection 4.3 then describes the concept of the algorithm in detail. Subsection 4.4 is devoted to the actual implementation and subsection 4.5 contains an example, to demonstrate the performance.

4.2 The Problem of Correspondence

Let two images, or image patches, I' and I'' be given. Points i' and i'' in images I' and I'' may have the coordinates $z' = (x',y')^T$ and $z'' = (x'',y'')^T$, thus, z' and z'' are vectors, where T stands for transposed.

It is assumed that if i' and i'' are corresponding points, their coordinates can be related by

$$z'' = t(z';p) \tag{9}$$

where:

t is an arbitrary mapping function; it might reflect the knowledge about the geometric relation between the images I' and I''; and

p is a vector of unknown parameters $p_1, \ldots, p_u$

It may be viewed as a severe restriction, that the mapping function must have an analytical form. But one should keep in mind that also a stochastic and/or segment-wise continuous function can be brought into the form of eq. (9). Eq. (9) will cause no difficulties, particularly in small scale imagery.

For an arbitrary pair of points (i', i'') there are two states of interest:

A. i' and i'' are corresponding points

B. i' and i'' are not corresponding points

The problem of correspondence now simply consists of: 1) finding the corresponding points; and 2) determining the parameter vector, p, of the mapping function. Theoretically the solutions to 1) and 2) are equivalent, as 1) implies 2) if applied to all pixels, say in I'. But this is neither feasible nor necessary, as the mapping function can reasonably be assumed to be smooth, i.e. roughly speaking bandlimited, and only a small number of corresponding points is sufficient to describe the mapping function. Whereas these pairs of corresponding points might replace the parameters p, the mapping function is necessary, if interpolation is required.

The known approaches actually use only a very limited number of points and explicitly or implicitly a mapping function of the type in eq. (9). In order to reduce the numerical effort and at the same time increase the reliability, objects o' and o'' are used in both images with feature vectors f' and f'' in addition to the coordinates z' and z'' attached to it:

$$o' = o'(z';f') \quad \text{and} \quad o'' = o''(z'';f'') \tag{10}$$

The procedures typically consist of three steps:

a. selection of appropriate objects o' and o'';

b. determining the similarity between all objects o' in image I' and all objects o'' in image I'', yielding possible candidates for corresponding objects

c. using some context information to find the pairs (o', o'') of corresponding objects.

4.2.1 The LNK-Method (Lambird et.al. [25], Stockmann et.al. [40])

a. Using edge detection procedures, this method selects objects which are either points or point pairs. Points belong to 4 classes. Pairs of such points are called abstract edges, abstract because the connecting line need not be a real edge in the image. For simplicity, we restrict the discussion to the point objects. Thus, an object o', say, in image I' is represented by its coordinates z' and its class $f' = \omega'$: $o' = o'(z', \omega')$.

b. Among all possible points (o', o'') of objects, only those which belong to the same classes are selected as possible candidates. Thus, if $\omega' = \omega''$ the objects o' and o'' are said to be similar.

c. The aim of the procedure is to determine the unknown parameters of the geometric transformation, which in this case consist of the two shifts in x- and y-directions. Each pair of similar objects leads to an equation $t(z', p) = z' - p$ which can be solved for p. The estimate $\hat{p}$ for the true shift p is taken from the histogram of all $p = z' - z''$ by searching for the peak-value representing the most probable shift. At the same time one obtains a classification of the object pairs

into the two classes: ω_A of corresponding points and ω_B of non-corresponding points.

The approach is a direct solution, where no iterations are necessary. A further advantage is the sharp peak in the histogram, which guarantees a reliable solution even if the numbers of objects is large, i.e. even when the background noise is considerable. On the other hand, the method requires that eq. (9) in an extended form $(z'',f'') = t(z',f';p)$ is solvable for p. Thus, if more than two parameters have to be estimated, the object has to contain additional geometric features such as length and orientation, in f' and f'', thus, requiring more complex objects, such as lines, triangles, etc., to be extracted from the image. This might not only increase the number of combinations but also requires an additional dimension of the histogram for each additional unknown parameter. Nevertheless, a primary advantage is the absence of requirements for approximate values. Thus, with say 4 parameters, the images might have any relative orientation and scale. This method is therefore highly recommended for step 1 in the hierarchical scene matching procedure.

4.2.2 The Barnard-Thompson Algorithm [4]

a. This algorithm starts from objects which are represented by the gray level matrix g', say, centered at distinct points: $o'(z',g')$. The selection uses the interest operator by Moravec, namely the minimum variance of the gray level differences

in the four main directions. This guarantees that no points on edges are selected, which are not discernible from neighbouring points on the same edge.

b. The similarity measure uses the coordinates and the gray level differences under consideration, deriving an initial probability that two objects $o'(z',g')$ and $o''(z'',g'')$ correspond, i.e. (o',o'') belongs to the class ω_A of corresponding points: $P((o',o'') \in \omega_A) = f(z'-z'',g'-g'') \simeq 1/|g'-g''|^2$ if the shift $|\ z'-z''\ |$ is less than a threshold and $P((o',o'') \in \omega_A) = o$ elsewhere.

c. The model of the geometric transformation is a differential one. They assume that the scene is regionwise smooth: $z''-z' = t(z')$ with the derivative $\partial t/\partial z'$ being bounded, except for the borders of the regions. The bound for $\Delta t/\Delta z'$ (being 1 pixel for $\Delta z' < 15$ pixels) is used to update the initial probabilities using a relaxation scheme (Rosenfeld et.al. [38]).

The model is extremely flexible, due to the randomness of the derivative within the admissible bound. The method can further be generalized by using more complex objects, e.g. the abstract edges of the LNK-method and thus can be an excellent solution for step 2 in the hierarchy. The numerical effort and the quality of the result, however, are highly scene dependent. In particular, the number and distribution of the selected objects are critical for the reliability of the result. Also the complexity of the geometrical model might not be necessary for satellite or aerial imagery of moderate

scale (say, <1:20,000), thus questioning whether the numerical effort resulting from the relaxation process, cannot be reduced, if one takes the simpler geometry of "far-range" imagery into account.

Though both procedures follow the same general concept, their techniques are essentially different. The simple geometric model on which the LNK-Method is based, allows a fully consistent line of thought. This makes a statistical evaluation of the result feasible, e.g. using the broadness of the peak in the histogram. On the other hand, though the procedure of Barnard and Thompson is excellently motivated, it is heuristic. This prevents a thorough evaluation of its results.

4.2.3 Outline of the New Procedure

The new solution for the correspondence problem essentially aims at a maximum-likelihood estimation of the unknown parameters p of the geometrical transformation. It follows the same three steps of the procedures described above. An attempt has been made to derive the three steps on a common theoretical basis, and at the same time make it amenable to generalizations for rectification:

a. The same objects are used as in the Barnard-Thompson algorithms, namely points with their gray level matrix. The selection is guided by the theoretical precision expected from cross-correlation. It turns out, that this selection principle is closely related to Moravec's interest operator.

b. The similarity of pairs of objects is also based on the theoretical precision. In addition to the gray level difference between the two objects, the texture is taken into account, namely the variance of the gradient. Moreover, the formulation allows the introduction of correlation measures from any feature vectors, possibly including structural features. Thus, very general similarity measures can be used without losing the relation to the geometrical model.

c. The maximum likelihood estimation for the parameters p of the mapping function requires the knowledge of the probability density function of the observations. Observations in this case are the coordinate differences Δz from the modified form $\Delta z = z'' - z' = t(z'; p)$ of eq. (9). The coordinate differences of corresponding points can reasonably be assumed to be normally distributed, whereas the coordinate differences of non-corresponding points are approximately equally distributed between -d and +d where d is the dimension of the image patch. These observations therefore can be interpreted as outliers or blunders with respect to the model eq. (9). As the redundancy of the system is rather high, robust estimation procedures should work efficiently in this case. The high percentage of outliers, i.e. non-correspondence is compensated by the non-similarity of the objects, which lead to a low initial weight of these observations.

4.3 Mathematical Model

This section provides the mathematical model for the correspondence algorithm. We will start with the mapping functions and the robust estimation procedure for the determination of the unknown parameters. The similarity measure then leads us to the interest operator used for the point selection.

4.3.1 Mapping Functions

The relation between two image segments of a satellite or aerial image can be approximated by a low degree polynomial:

Shift only

(11 a,b) $$\underline{z}'' = a + \underline{z}' \quad \text{or} \quad \underline{\Delta z} = a$$

(Stochastical variables are underscored.)

Affine transformation

(12 a,b) $$\underline{z}'' = a + B\,\underline{z}' \quad \text{or} \quad \underline{\Delta z} = a + \bar{B}\,z'$$

Second order polynomial

(13 a,b) $$\underline{z}'' = a + B\,\underline{z}' + C\,\underline{z}' \otimes \underline{z}' \quad \text{or} \quad \underline{\Delta z} = a + \bar{B}z' + C\,z' \otimes z'$$

with

$$a = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix};\ B = \begin{bmatrix} p_3 & p_4 \\ p_5 & p_6 \end{bmatrix};\ C = \begin{bmatrix} p_7 & p_8 & p_8 & p_9 \\ p_{10} & p_{11} & p_{11} & p_{12} \end{bmatrix}$$

$$(z' \otimes z')^T = (x'x' \quad x'y' \quad y'x' \quad y'y')$$

and

$$\bar{B} = B - I.$$

By introducing conditions on the parameters p_i, one may restrict the mapping functions to conformal ones. For example, the conditions $p_3 = p_6$ and $p_4 = -p_5$ in eq. (12) lead to a similarity transformation with shifts, scale, and rotation only. The transformation parameters only occur linearly in the mapping function, thus could be solved in one step using the least squares technique.

4.3.2 Robust Estimation

The least squares technique starts from the linear (or linearized) model

$$E(\underline{l}) = A\,\tilde{x} = \sum_{i=1}^{n} a_i^T \tilde{x}\ ; \quad D(\underline{l}) = C_{ll} = \sigma_o^2\, Q_{ll} \tag{14}$$

where the nx1 vector $\underline{l}$ contains the observations, in our case the coordinate differences $\underline{\Delta z}$, with their covariance matrix C_{ll}. It is usually split into the unknown variance factor σ_o^2 and the known coefficient matrix Q_{ll}. The nxu design matrix A, having rows a_i, is supposed to be known. $\tilde{x}$ are the unknown parameters.

If the observations can be assumed to be uncorrelated, then one uses the weights w_i or the weight matrix $W = \text{diag}(w_i) = \text{diag}(1/q_{ii})$ to advantage, to solve the minimum problem

$$\Sigma (a_i' \hat{\underline{x}} - \underline{l}_i)^2 w_i = \Sigma \underline{v}_i^2 w_i \rightarrow \min. \tag{15}$$

It is known that the estimated parameters $\hat{\underline{x}}$ are sensitive to errors in the model eq. (9), especially gross errors or outliers, in the observations. This is due to the fact that the solution to eq. (15) is also the maximum likelihood estimator for $\tilde{x}$, if the observations are normally distributed. Observations with outliers, however, can be viewed to belong to longer tailed distributions. Examples are the Laplace-Distribution $f(x) = c\, e^{-|x|}$ and the Cauchy-Distribution $f(x) = c/(1+x^2)$.

In order to eliminate the effect of outliers on the result one can use maximum-likelihood type estimators. Then, instead of the sum of the squares of the residuals v_i the sum of a less increasing function $\rho(v_i)$ is minimized (Huber [19]):

$$\sum_i \rho(a_i^T \hat{\underline{x}} - \underline{l}_i) = \sum_i \rho(\underline{v}_i) \rightarrow \min \tag{16}$$

Discussion:

1. Choosing $\rho(v) = v^2/2$ gives the least squares estimator

2. Choosing $\rho(v) = \frac{1}{p} |v|^p$ yields the estimator minimizing the L_p-norm. A special case is obtained for $p = 1$: Minimizing $\rho(v) = |v|$ is the well known least sum method, being the ML-estimate for the Laplace-Distribution. It is the multiparameter version of the median. Barnea and Silverman [5] used it for cross correlation.

3. The choice of ρ can be guided by the evaluation of the "Influence-Curve" IC(v) (Hampel [18]) being proportional to the derivative $\psi(v) = \partial\rho/\partial v$ of the minimum function. IC(v) or $\psi(v)$ give an indication of how strong is the influence of an outlier on the estimate $\underline{\hat{x}}$.

4. The solution of eq. (16) can use existing programs for least squares solution, by either modifying the residuals, $v^* = \sqrt{\rho(v)}$ or by modifying the weights:

$$\Sigma\, \rho(v_i) = \Sigma \frac{\rho(v_i)}{v_i^2/2} \frac{v_i^2}{2} = \Sigma\, w(v_i) \frac{v_i^2}{2} \rightarrow \min. \tag{17}$$

using the weight function

$$w(v_i) = \frac{\rho(v_i)}{v_i^2/2 + c} \quad (c \ll 1) \tag{18}$$

In an iterative solution the weights of all observations are updated depending on their residuals from the previous iteration:

$$w_i^{(\nu+1)} = w_i^{(0)}\, w(v_i^{(\nu)}) \tag{19}$$

5. If the function $\rho(v)$ is convex, thus $\psi(v)$ non-decreasing, and the model is linear, then convergence is guaranteed under broad conditions.

Minimizing the L_1-norm thus seems to be optimal, as it is robust, and convergence is guaranteed. This method however has two disadvantages:

1. $\rho(v)$ has no derivative at 0, thus, the influence curve is not continuous, which does not guarantee a unique solution.

2. The influence curve $\psi(v) = \text{sign}(v)$ is not zero for large values, thus large outliers have still an influence onto the result, which is not desirable.

We therefore propose to use the following weight functions.

1. In order to ascertain convergence we slightly modify the minimum function of the L_1-norm (cf. Figure 10).

(20 a) $$\rho_1(v) = 2\,(\sqrt{1+v^2/2} - 1)$$

(20 b) $$w_1(v) = \frac{4\,(\sqrt{1+v^2/2} - 1)}{v^2}$$

(20 c) $$\psi_1(v) = \frac{v}{\sqrt{1+v^2/2}}$$

$\rho_1(v)$ is strictly convex with decreasing curvature for large v.

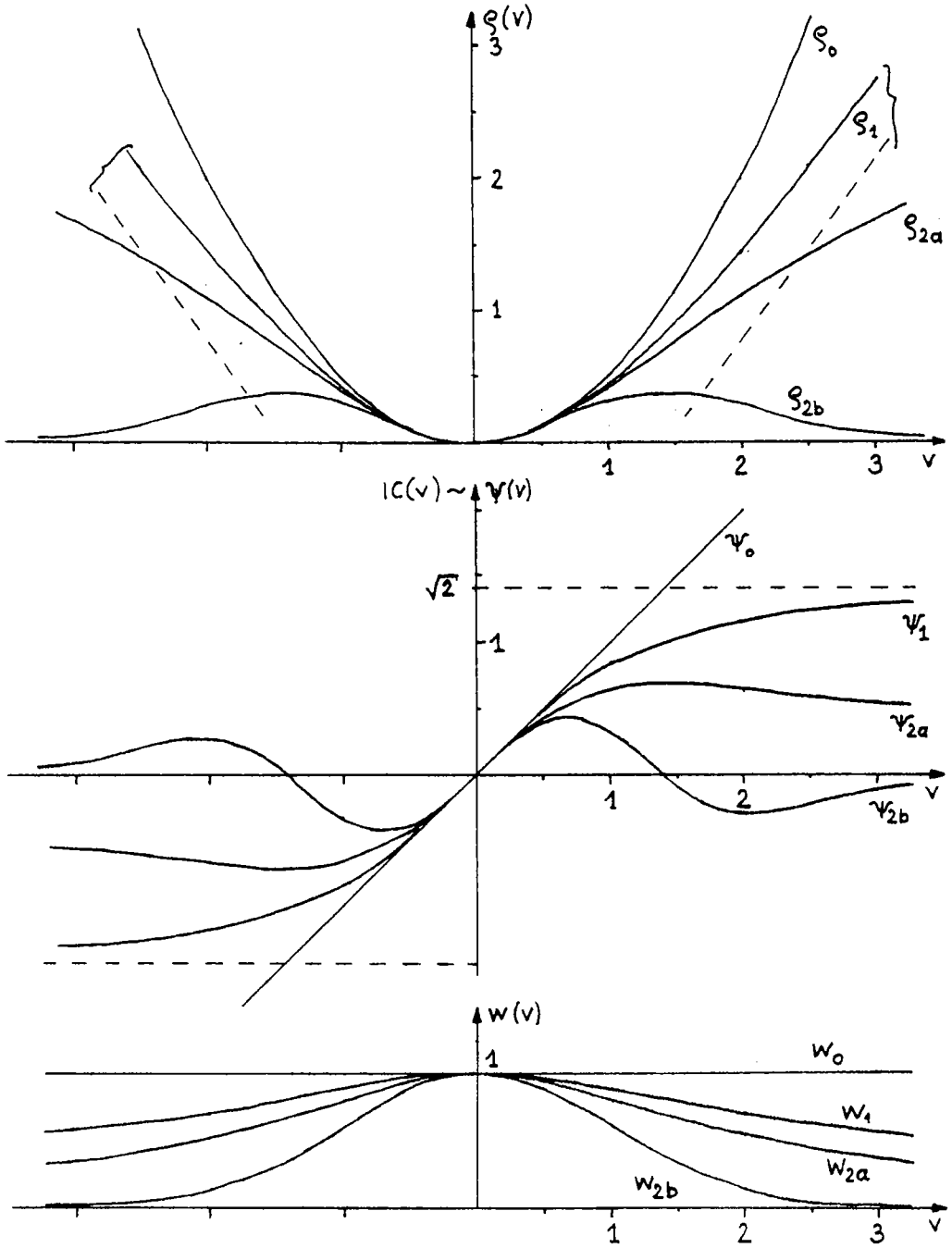

Figure 10 Minimum, Influence, and Weight Functions $\rho(v)$, $IC(v) \sim \psi(v)$, and $w(v)$.

NOTE: 0: least squares, non robust ($\psi(v)$ not bounded)

1: L_1-Norm, robust, convergence guaranteed

2: redescending IC,
   a: ML-estimator for Cauchy-Distribution
   b: exponential weight-function (Krarup [22])

2. After having reached convergence, one can assume to have good approximate values for the parameters. In order to eliminate the influence of large outliers one could take one of the following two minimum functions:

ρ_{2a} leads to maximum-likelihood estimators starting from a Cauchy-Distribution

(21 a) $$\rho_{2a}(v) = \ln(1+v^2/2)$$

(21 b) $$w_{2a}(v) = \frac{2\ln(1+v^2/2)}{v^2}$$

(21 c) $$\psi_{2a}(v) = \frac{v}{(1+v^2/2)}$$

No convergence is guaranteed in the *general* case. Also, as ψ is descending for large v, no unique solution is guaranteed if arbitrary approximate values are allowed. This is meaningful as the Cauchy-Distribution has neither mean nor variance.

The following minimum function is proposed by Krarup et.al. [22] which considerably reduces the weights of false observations due to its exponential form:

(22 a) $$\rho_{2b}(v) = \frac{v^2}{2} e^{-v^2/2}$$

(22 b) $$w_{2b}(v) = e^{-v^2/2}$$

(22 c) $$\psi_{2b}(v) = v \ (1 - \frac{v^2}{2}) \ e^{-v^2/2}$$

This weight function fulfills practically all requirements for a well behaved weight function. (Hampel [18], Werner [43]). The functions are shown in Figure 10 together with the minimum weight and influence function of the least squares, $\rho_o(v) = v^2/2$.

4.3.3 Similarity Measure

The estimation procedure requires initial weights for the observations which in our case are the coordinate differences Δz of object or point pairs, which need not correspond. Hence, the majority of the observations are outliers and assuming equal weight would prevent the solution from getting started.

Reasonable weights can be obtained from the covariance matrix of the estimated shifts $\underline{\Delta z}$, if we would apply cross correlation to all pairs of points. It is given by (Förstner [14])

$$(23) \qquad V(z''-z') = \mathrm{Cov}\begin{bmatrix} x''-x' \\ y''-y' \end{bmatrix} = \hat{\sigma}^2_{\Delta g}\begin{bmatrix} \Sigma g_x^2 & \Sigma g_x g_y \\ \Sigma g_y g_x & \Sigma g_y^2 \end{bmatrix}^{-1} = \hat{\sigma}^2_{\Delta g}\ Q$$

where:

g is the gray level function of the object, restored from g' and g''

$\hat{\sigma}^2_{\Delta g}$ the estimated variance of the gray level differences, and

g_x, g_y are the gradients of g in x- and y-directions respectively.

The covariance matrix fully describes the precision of the match between the gray level function g' and g'' of the two objects o' and o''. This precision depends on:

1. The number of pixels used.
2. The noise variance.
3. The texture of the object, namely the edge business. It can be shown that this measure is directly related to the bandwidth of the signal and the curvature of the cross correlation function (Förstner [14]).

The covariance matrix can be visualized by an error ellipse (cf. Mikhail [34]), giving the precision of the match for all directions. A good match therefore must fulfill the following two requirements:

C1: The error ellipse should be close to a circle, otherwise the match is not well determined in one direction, e. g. at an edge.

C2: The error ellipse should be small.

Both criteria will be used for the measure of similarity between two objects and the selection of interesting points.

If the ellipse is close to a circle the weight can be directly derived from the trace of the covariance matrix

$$w = \frac{1}{tr(V)} = \frac{1}{\hat{\sigma}^2_{\Delta g}\ tr\ (Q)} \tag{24}$$

Observe, that the trace is invariant to rotations. Taking the gray level differences directly to estimate $\hat{\sigma}^2_{\Delta g}$ has the disadvantage of being biased if the two images have different brightness and contrast. The correlation coefficient is known to be a better measure. Now, if one for simplicity assumes the images g' and g'' to be related to the true image g by $g' = a'\ (g+n') + b'$ and $g'' = a''\ (g+n'') + b''$, with $\sigma^2_{n'} = \sigma^2_{n''} = \sigma^2_n$ where a and b represent contrast and brightness, the signal to noise ratio $SNR^2 = \sigma^2_g/\sigma^2_n$ is functionally related to the correlation coefficient by:

(25 a) $$\rho = \frac{\sigma_{g'g''}}{\sigma_{g'}\sigma_{g''}} = \frac{\sigma_g^2}{\sigma_g^2+\sigma_n^2} = \frac{\mathrm{SNR}^2}{\mathrm{SNR}^2+1}$$

or

(25 b) $$\mathrm{SNR}^2 = \frac{\sigma_g^2}{\sigma_n^2} = \frac{\rho}{1-\rho}$$

By using the approximations

(26) $$\sigma_g^2 \simeq \sigma_{g'}\,\sigma_{g''}$$

(27) $$\mathrm{tr}\ Q = \sqrt{\mathrm{tr}\ Q'\ \ \mathrm{tr}\ Q''}$$

and

(28) $$\sigma_{\Delta g}^2 = 2\,\sigma_n^2$$

we obtain the following relation for the weight of the observation Δz:

(29) $$w(o',o'') \simeq \frac{1}{2}\ \frac{\rho}{1-\rho}\ \frac{1}{\sigma_{g'}\sigma_{g''}\cdot\sqrt{\mathrm{tr}\ Q'\ \ \mathrm{tr}\ Q''}}$$

Discussion:

1. The weight depends on two terms. The first term reflects the similarity between the two objects and needs to be calculated for all object pairs. The second term depends on values obtained separately from both images.

2. The traces tr Q' and tr Q'' measure the distinctness or the locatability of the objects and are critical for the selection of appropriate points. The reason is, the noise level $\hat{\sigma}^2_{\Delta g}$ can be realistically assumed to be constant in both images.

3. The weight is a generalization of the one used by Barnard and Thompson. It differs in two ways. First, it is independent of brightness and contrast, as we are only interested in the weight ratios. Second, it takes the texture of the object into account.

4. A simple and reasonable criterion to reject object pairs based on the correlation coefficient is $\rho \geq \frac{1}{2}$. This is equivalent to requiring the SNR to be larger than 1.

5. The main advantage of the separation of the different terms in eq. (29) lies in its ability to include other measures for similarity. The correlation coefficient need not be derived from the gray levels but may use other features f' and f'' of the objects, e.g.:

   a. One could use rotation and scale invarant features, as the moments proposed by Wong and Hall [44].

b. One could use a small set of features just to decrease the computation time, e.g. the low frequency terms of a cosine transform.

c. One could use structural information, the result of a classification or a linguistic description in combination with statistical measures. The only requirement for the measure is to have the properties of a correlation coefficient.

The separation of the correlation coefficient from the variance and the texture of the gray level function, allows one to generalize the weight determination without losing the information about the geometric distinctness of the object.

4.3.4 Interest Operator

We have assumed that the error ellipse representing the covariance matrix of the coordinate difference is close to a circle. Moreover, we require that the point can be well located. Measures of both requirements should, in a simple way, be derivable from the gray level function of the image patch, as they have to be determined for all pixels. They should also be invariant to rotation; a scale factor will not change too much the ranking of the different pixels.

As the eigenvalues of the covariance matrix are invariant to rotations, and the trace equals the sum of the eigenvalues, we will use them also for determining the closeness of the error ellipse to a circle. Moreover, the eigenvalues of the coefficient matrix, say Q', and those of its inverse $N' = (Q')^{-1}$ are related by $\lambda_i(Q') = 1/\lambda_i(N')$. Thus, let λ_1 and λ_2 be the eigenvalues of N', then the ratio

$$q = \frac{4 \det N'}{(\mathrm{tr}\, N')^2} = \frac{4\,\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2} = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 \tag{30}$$

is an adequate measure for the closeness of the error ellipse to a circle. If $q = 0$ (and not both λ_1 and λ_2 are zero), then $\det N'$ is zero and the matrix is singular. This means that g_x and g_y are linearly dependent thus the point may lie on an edge. The case $g = 1$ is reached, only if the eigenvalues are equal ($\lambda_1 - \lambda_2 = 0$) thus representing a circular error ellipse. The calculation of q need not use the eigenvalues, but rather the determinant and the trace of N':

$$\det N' = \Sigma(g_x')^2 \,.\, \Sigma(g_y')^2 - (\Sigma g_x' g_y')^2 \tag{31 a}$$

$$\mathrm{tr}\, N' = \Sigma(g_x')^2 + \Sigma(g_y')^2 \tag{31 b}$$

The sums can be readily derived from the squared and multiplied gradient images by convolution.

Similarly, one can derive an expression for tr Q':

$$\text{tr } Q' = \frac{\text{tr } N'}{\det N'} \tag{32}$$

Thus the selection of interesting points can be accomplished for both images separately in the following steps:

1. Determination of Σg_x^2, $\Sigma g_x g_y$, and Σg_y^2;

2. Determination of tr Q and q using eq. (30) - (32);

3. Determination of the interest value, being a preliminary weight,

$$\bar{w} = \begin{cases} \dfrac{1}{\text{tr } Q} = \dfrac{\det N}{\text{tr } N} & \text{for } q > \text{threshold} \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

   for each pixel;

4. Suppressing all non-maxima in the function $\bar{w}(i,j)$;

5. All values $\bar{w}(i,j)$ give rise to an object o.

4.4 Algorithmic Solution

4.4.1 The Selection of Objects of Interest

The interest oprator eq. (30) to (33) requires the variance and covariance of the gradient image at each pixel. The used window size should be adaptable to the texture of the image patch. If one uses a square (in general, a rectangle) window the number of operations per pixel needed for the interest operator can be made independent of the window size. This is due to the fact, that the array $I(\Sigma g_x^2)$, say, containing the sums Σg_x^2 can be derived from g_x^2 by convolution with a separable window of size n_{s1} x n_{s2}, $W(i,j) = 1$ with $w = e_1\, e_2^T$ and $e_i^T = (11...1)$ containing n_{si} elements 1. As the convolution with e, or e^T, needs only 2 additions, if done recursively, only 4 additions per pixel are necessary for the determination of the array $I(\Sigma g_x^2)$ independent of the window size. The gradients g_x and g_y are calculated with the Roberts operator.

Now two thresholds $q_{min.}$ and $\overline{w}_{min.}$ are necessary to check the form and the size of the ellipse:

C1: $q_i > q_{min.}$ (form)

C2: $\overline{w}_i > \overline{w}_{min.}$ (size)

If both conditions are fulfilled, the interest value of that pixel is set to the preliminary weight $\overline{w} = 1/\mathrm{tr}\ Q$, otherwise it is zero.

The threshold $q_{min.}$ is scale independent, a value of $q_{min.} = 0.25$ turned out to be reasonable. The condition C2 should also be independent of scale. Therefore, we used $w_{min.} = f \cdot \Sigma\bar{w}_i/n$, relating the preliminary weights $\bar{w}_i$ to their mean value. A value $f = 1.5$ was chosen for all tests performed.

From the resulting interest values, $\bar{w}$ or 0, the relative maximum within a certain window $n_m \times n_m$ are extracted. The window size n_m for this non-maximum suppression is independent from the one used for the sums. If the window size n_m is larger than 3 the non-maximum suppression is accomplished in two steps, the first using a 3x3 window and the second performing the comparisons in a spiral manner in the large window to keep the number of comparisons independent of the window size n_m. The selected objects are then stored in a list, containing the coordinates and the preliminary weight $\bar{w} = 1/\mathrm{tr}(Q)$. They are needed for the similarity measure.

4.4.2 The Selection of Object Pairs

The initial weight $\bar{w}$ from eq. (29) in addition to tr Q' and tr Q'', requires the standard deviations σ_g' and σ_g'' and the correlation coefficient $\rho = \sigma_{g'g''}/(\sigma_{g'}\sigma_{g''})$ where:

(34 a)
$$\sigma_g^2 = \frac{\Sigma(g')^2 - (\Sigma g')^2/n}{n-1}$$

(34 b)
$$\sigma_{g'}^2 = \frac{\Sigma(g'')^2 - (\Sigma g'')^2/n}{n-1}$$

and

$$\sigma_{g'g''} = \frac{\Sigma g' g'' - (\Sigma g')(\Sigma g'')/n}{n-1} \qquad (34\ c)$$

The sums $\Sigma g'$, $\Sigma g''$, $\Sigma (g')^2$, and $\Sigma (g'')^2$ are calculated for each object. The mixed sum $\Sigma g' g''$ is only calculated for pairs of objects with a distance $|z'-z''|$ less than a given threshold $d_{max.}$, which reflects the maximum expected distances between corresponding objects. All pairs of objects for which the correlaion coefficient ρ is greater than 0.5 are collected in a list, containing their coordinates z' and z'' and their weights.

4.4.3 The Selection of Corresponding Points

The selection of the corresponding points is based on the assumed geometrical relation between the images. In our context an affine transformation seems to be adequate, and therefore, has been employed. The robust adjustment is split into two steps. First, only the shift between the images is determined. This leads to better approximations, both for the shifts and the weights in the following 6 parameter transformation. Both adjustments have the same structure.

In each iteration, the parameters, the residuals, the precision of the shift, and the average weight are determined, and the weights are adapted for the next iteration. If a weight is smaller than a certain percentage (say, 10%) of the average weight, it is set to zero, eliminating that observation. The first 4 iterations are performed with the weight

function given by eq. (20b), after which th redescending function in eq. (22b) is applied. The agorithm stops if either the required precision of the shift is reached, not enough corresponding points are left, or a pre-set number of iterations is reached. The residuals of the last iteration are tested, and with all residuals passing this test one additional iteration with equal weights is performed to obtain the final transformation parameters.

The obtained list of corresponding points may then still be ambiguous, as the same point in one image might correspond to several points of a cluster in the other image. The list of pairs of points is then cleaned keeping those correspondences which have the smaller residuals.

4.5 Two Examples

The following two examples are presented to show the performance of the new algorithm. In both cases, the two images I' and I'' are derived from an original image by extracting two separate windows and distorting them by an affine transformation according to eq. (12b) with random numbers in $\bar{B}$ ranging up to 0.15. Thus, the average linear distortion is approximately 10% or 6°. The extracted windows are contaminated with white Gaussian noise with a standard deviation of $\sigma = 15$ gray levels. Both windows are then smoothed with a 3x3 Hanning filter $(1\ 2\ 1)^T * (1\ 2\ 1)$.

1. The first example is based on an artificial image (cf. Figure 11). It may represent a part of a rural scene with some light roads between fields of different brightness. The dark pixels are the points selected by the interest operator. Table 7 contains the preliminary weight $\overline{w}$ and the values of q in percent describing the closeness of the error ellipse to a circle.

   Observe, that some points, e.g. point 1 in the right image, lie on an edge, but due to the irregularity of the edges have been selected. Both values, $\overline{w}$ and q are small ($\overline{w}$ = 311, q = 36 % in this case). From the 15x16 = 240 possible point pairs 59 were selected as possible candidates for correspondence. Their weights vary considerably, namely due to the correlation coefficient (cf. Table 8). The robust shift adjustment yields the pairs listed in Table 9. It shows the ambiguity of the result, as for example point 3 in the left image is connected with points 7 and 9 in the right image. As the residuals of the pair (3,7) are smaller than those of (3,9) the pair (3,7) is kept. The cleaned list in Table 9b would be the result with the shift parameters only, showing that even with a wrong geometric model nearly all corresponding objects can be found (cf. Table 10b).

   The result of the robust affine transformation (cf. Table 10) shows a slightly different result. The final correspondences are shown in Figure 12. If one compares the final result with the list of the candidate pairs (Table 8), obviously the

Figure 11 Example 1: Artificial Image Pair With Selected Points (black pixels)

NOTE: Window size of interest operators: 7

Window size for non-maximum suppression: 3

Left Image					Right Image				
i	x	y	$\bar{w}$	q	i	x	y	$\bar{w}$	q
1	4	4	357	91	1	14	25	311	36
2	10	34	410	49	2	16	45	300	40
3	12	45	907	87	3	17	42	288	39
4	13	31	364	43	4	18	36	688	71
5	14	15	410	67	5	21	7	586	72
6	14	17	450	59	6	21	22	469	65
7	16	4	714	88	7	21	33	423	62
8	16	21	483	39	8	24	18	445	61
9	16	31	345	56	9	24	33	708	89
10	21	45	961	62	10	26	8	280	48
11	24	41	964	48	11	29	38	325	80
12	26	4	316	37	12	35	8	718	74
13	26	19	622	64	13	37	27	1374	65
14	26	40	963	51	14	37	45	368	46
15	30	23	448	53	15	38	36	508	50
16	31	38	1871	94	16	38	38	540	54
17	36	25	1518	82	17	39	11	314	37
18	40	22	867	50	18	42	6	289	86
19	41	45	311	85	19	42	11	396	42
20	43	40	640	92	20	43	4	301	81
21	45	14	324	50	21	45	5	299	77
22	45	19	327	57	22	45	10	446	59
23	45	39	686	78	23	45	24	1683	88
					24	45	29	344	45

Table 7 Example 1: List of Selected Points

NOTE: x,y coordinates

$\bar{w}$ interest value

q measure for isotropy of error ellipse (in percent)

	i	j	w	$\bar{w}_i$	$\bar{w}_j$	rho
1	201		4.757	410.	311.	0.813910
2	206		18.507	410.	469.	0.934145
3	307		6.047	907.	423.	0.743084
4	309		9.129	907.	708.	0.769720
5	406		3.555	364.	469.	0.733859
6	408		8.666	364.	445.	0.872919
7	505		27.938	410.	586.	0.932111
8	510		2.052	410.	280.	0.572882
9	605		16.436	450.	586.	0.883114
10	608		2.755	450.	445.	0.623202
11	708		16.076	714.	445.	0.895975
12	808		11.647	483.	445.	0.899352
13	810		1.838	483.	280.	0.589571
14	906		4.533	345.	469.	0.777062
15	908		11.360	345.	445.	0.899266
16	911		2.187	345.	325.	0.599452
17	1002		2.891	961.	300.	0.611310
18	1003		2.454	961.	288.	0.611800
19	1004		2.578	961.	688.	0.519524
20	1007		2.692	961.	423.	0.562607
21	1102		2.335	964.	300.	0.613341
22	1103		8.660	964.	288.	0.874139
23	1104		55.190	964.	688.	0.966564
24	1107		2.528	964.	423.	0.601318
25	1113		135.224	964.	1374.	0.983662
26	1114		3.310	964.	368.	0.681337
27	1115		13.244	964.	508.	0.891956
28	1116		7.161	964.	540.	0.812974
29	1212		4.827	316.	718.	0.776955
30	1312		22.732	622.	718.	0.919691
31	1313		7.178	622.	1374.	0.764347
32	1317		1.891	622.	314.	0.579353
33	1403		2.995	963.	288.	0.689592
34	1404		11.141	963.	688.	0.843669
35	1407		5.303	963.	423.	0.745271
36	1409		2.978	963.	708.	0.557126
37	1413		15.047	963.	1374.	0.861033
38	1415		4.759	963.	508.	0.732863
39	1416		3.142	963.	540.	0.638196
40	1512		7.280	448.	718.	0.800946
41	1515		2.030	448.	508.	0.584841
42	1516		2.125	448.	540.	0.589507
43	1517		12.494	448.	314.	0.908975
44	1519		24.556	448.	396.	0.943966
45	1522		11.948	448.	446.	0.874251
46	1609		9.858	1871.	708.	0.746461
47	1611		3.551	1871.	325.	0.556535
48	1623		129.159	1871.	1683.	0.970396
49	1706		16.139	1518.	469.	0.879777
50	1708		2.468	1518.	445.	0.533513
51	1717		6.066	1518.	314.	0.764804
52	1719		4.692	1518.	396.	0.683444
53	1810		1.567	867.	280.	0.504529
54	1824		5.183	867.	344.	0.779709
55	1914		3.144	311.	368.	0.693980
56	1915		7.749	311.	508.	0.843612
57	1916		6.314	311.	540.	0.810626
58	2013		5.123	640.	1374.	0.651036
59	2323		29.446	686.	1683.	0.911693

Table 8 Example 1: List of Selected Pairs

NOTE: ij point No. in left and right image (201 ≙ (2,1)), w initial weight, $\bar{w}$ preliminary weights, rho correlation coefficient.

estimated shift: 11.077 -12.615

a)

i	left	right	xl	yl	xr	yr	dx	dy
1	2	6	10.000	34.000	21.000	22.000	-0.077	0.615
2	3	7	12.000	45.000	21.000	33.000	-2.077	0.615
3	3	9	12.000	45.000	24.000	33.000	0.923	0.615
4	4	8	13.000	31.000	24.000	18.000	-0.077	-0.385
5	8	10	16.000	21.000	26.000	8.000	-1.077	-0.385
6	9	8	16.000	31.000	24.000	18.000	-3.077	-0.385
7	11	13	24.000	41.000	37.000	27.000	1.923	-1.385
8	13	12	26.000	19.000	35.000	8.000	-2.077	1.615
9	14	13	26.000	40.000	37.000	27.000	-0.077	-0.385
10	15	17	30.000	23.000	39.000	11.000	-2.077	0.615
11	15	19	30.000	23.000	42.000	11.000	0.923	0.615
12	15	22	30.000	23.000	45.000	10.000	3.923	-0.385
13	16	23	31.000	38.000	45.000	24.000	2.923	-1.385

clean list

b)

i	left	right	xl	yl	xr	yr	dx	dy
1	2	6	10.000	34.000	21.000	22.000	-0.077	0.615
2	3	9	12.000	45.000	24.000	33.000	0.923	0.615
3	4	8	13.000	31.000	24.000	18.000	-0.077	-0.385
4	8	10	16.000	21.000	26.000	8.000	-1.077	-0.385
5	13	12	26.000	19.000	35.000	8.000	-2.077	1.615
6	14	13	26.000	40.000	37.000	27.000	-0.077	-0.385
7	15	19	30.000	23.000	42.000	11.000	0.923	0.615
8	16	23	31.000	38.000	45.000	24.000	2.923	-1.385

Table 9 Example 1: Result of Robust Shift Adjustment

a) uncleaned list containing ambiguities
b) cleaned list

NOTE: 6 iterations

$$\begin{pmatrix} 1.19529 & 0.10307 & ; & 3.58 \\ -0.08967 & 0.88201 & ; & -6.45 \end{pmatrix} = (\hat{B} ; \hat{a})$$

a)

i	left	right	xl	yl	xr	yr	dx	dy
1	2	6	10.000	34.000	21.000	22.000	-1.961	0.637
2	3	7	12.000	45.000	21.000	33.000	1.563	-0.840
3	3	9	12.000	45.000	24.000	33.000	-1.437	-0.840
4	4	8	13.000	31.000	24.000	18.000	-1.684	1.722
5	8	10	16.000	21.000	26.000	8.000	-1.129	2.633
6	9	8	16.000	31.000	24.000	18.000	1.902	1.453
7	11	13	24.000	41.000	37.000	27.000	-0.505	0.556
8	13	12	26.000	19.000	35.000	8.000	1.618	-0.027
9	14	13	26.000	40.000	37.000	27.000	1.782	-0.505
10	15	17	30.000	23.000	39.000	11.000	2.811	0.142
11	15	19	30.000	23.000	42.000	11.000	-0.189	0.142
12	15	22	30.000	23.000	45.000	10.000	-3.189	1.142
13	16	23	31.000	38.000	45.000	24.000	-0.448	0.282
14	5	5	14.000	15.000	21.000	7.000	0.862	-1.480
15	13	17	26.000	19.000	39.000	11.000	-2.382	-3.027
16	6	5	14.000	17.000	21.000	7.000	1.068	0.284
17	4	6	13.000	31.000	21.000	22.000	1.316	-2.278

clean list

b)

i	left	right	xl	yl	xr	yr	dx	dy
1	2	6	10.000	34.000	21.000	22.000	-1.961	0.637
2	3	9	12.000	45.000	24.000	33.000	-1.437	-0.840
3	8	10	16.000	21.000	26.000	8.000	-1.129	2.633
4	9	8	16.000	31.000	24.000	18.000	1.902	1.453
5	11	13	24.000	41.000	37.000	27.000	-0.505	0.556
6	13	12	26.000	19.000	35.000	8.000	1.618	-0.027
7	15	19	30.000	23.000	42.000	11.000	-0.189	0.142
8	16	23	31.000	38.000	45.000	24.000	-0.448	0.282
9	6	5	14.000	17.000	21.000	7.000	1.068	0.284

Table 10 Example 2: Result of Robust Affine Transformation

a) uncleaned list, containing ambiguities
b) cleaned list, final result
(cf. Figure 12)

NOTE: 6 iterations

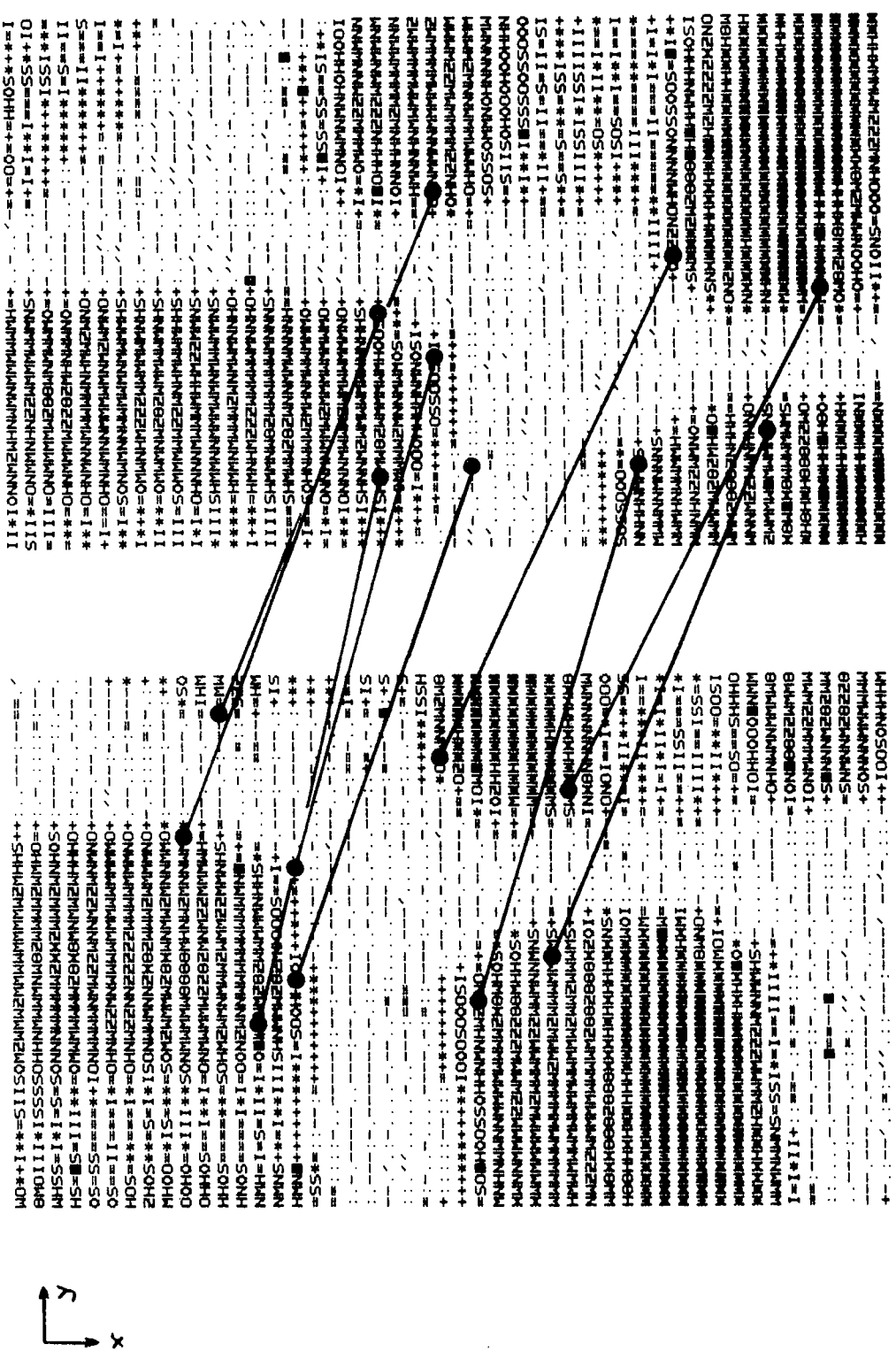

Figure 12 Example 1: Result of Correspondence Algorithm.

most similar objects are also correspondent ones. The object pair 3(8,10) with the largest residuals is found by chance, as both points are just above the level of distinctness. But observe that the objects in pair (5,5) are more similar than those in pair (6,5). The context, i. e. the common geometrical model, however, selects the pair (6,5) due to its better fit, which seem to be reasonable as can be seen from Figure 12. The final transformation parameters show scale differences up to 20% between the two images.

2. The second example is based on an image from the Arizona Test Area. The resolution of the original image has been reduced by a factor of two, yielding pixel sizes of 50 μm. The selected windows of 80x80 pixels with the interesting points are shown in Figure 13.

   39 and 50 points have been selected, almost all having error ellipses close to a circle (cf. Table 11). From the 1950 possible pairs 127 were retained as candidates (cf. Table 12). Observe that the weights in this case do not vary so much as in the first example, and are considerably smaller. The final result yields 18 object pairs and is shown in Figure 14 (cf. Tables 13 and 14). Also in this case the scale difference is approximately 20%, but in addition a rotation of approximately 10° in both axes becomes apparent from Figure 14. The shifts of $+4$ and $+18$ pixels correspond to an overlap of the two windows of approximately 70%.

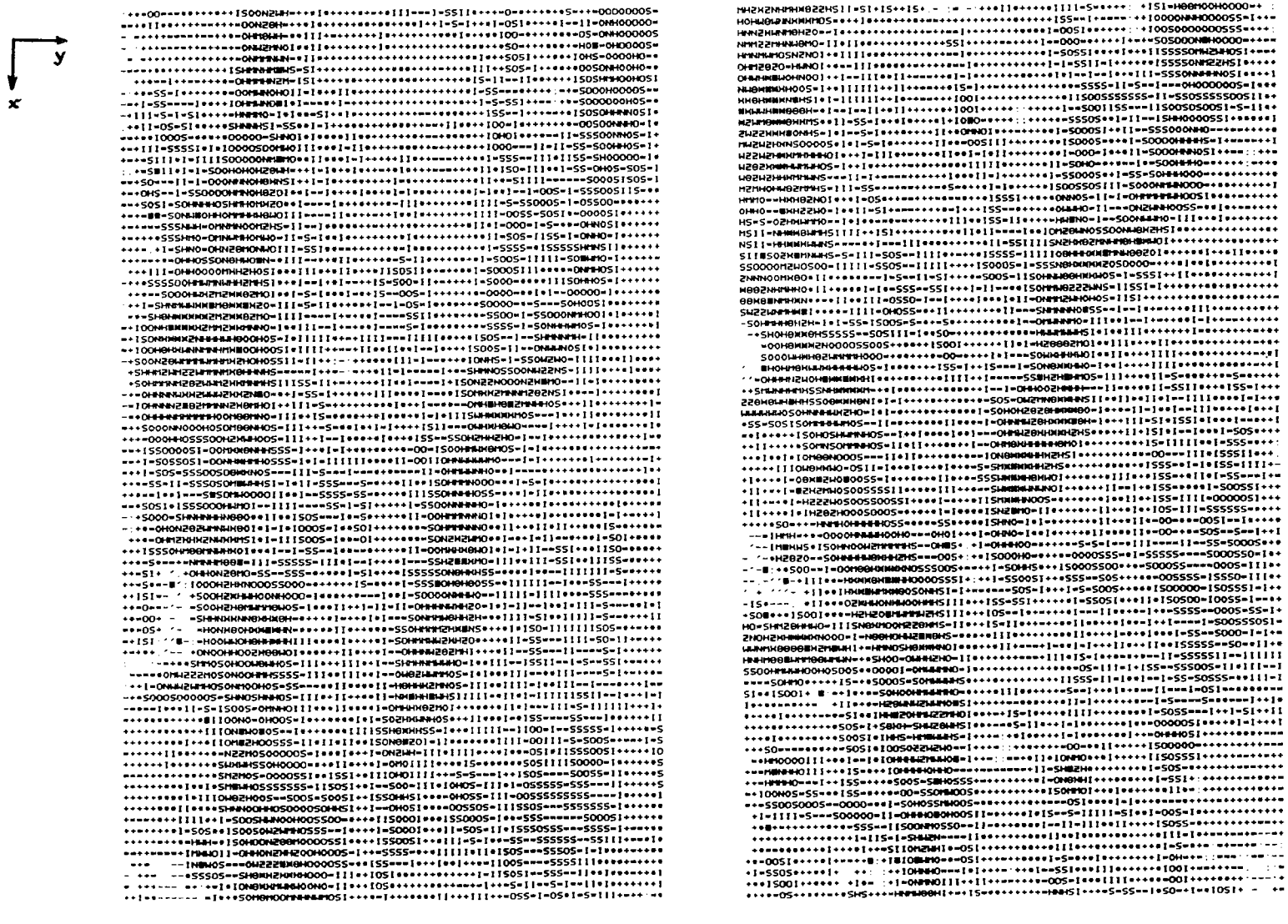

Figure 13 Example 1: Image Pair from Arizona Test Area With Selected Points (black pixels).

Left Image					Right Image				
i	x	y	$\bar{w}$	q	i	x	y	$\bar{w}$	q
1	4	71	280	79	1	4	70	206	96
2	6	24	439	83	2	7	6	474	72
3	9	24	392	82	3	9	9	673	89
4	14	24	399	61	4	11	35	231	75
5	15	6	184	66	5	12	8	660	91
6	19	5	271	81	6	19	7	589	91
7	19	6	271	84	7	20	50	355	87
8	19	12	327	89	8	22	60	521	75
9	23	21	319	57	9	23	4	496	94
10	23	69	196	68	10	23	9	622	87
11	27	14	520	85	11	23	55	475	81
12	27	19	385	64	12	27	5	489	88
13	29	8	802	85	13	28	10	363	39
14	31	18	320	55	14	28	52	275	95
15	34	63	287	53	15	33	4	950	94
16	35	21	180	57	16	34	13	597	96
17	36	10	208	93	17	34	16	778	96
18	36	55	642	91	18	34	45	348	87
19	36	58	530	86	19	34	49	335	55
20	43	18	563	69	20	36	8	328	98
21	44	14	355	92	21	38	50	612	82
22	50	18	552	70	22	43	12	563	86
23	50	50	397	67	23	43	16	261	81
24	50	51	397	57	24	44	9	494	89
25	52	8	964	87	25	46	41	669	86
26	52	47	307	87	26	49	7	959	74
27	56	22	832	94	27	49	31	197	78
28	56	51	365	49	28	51	4	751	91
29	57	9	1258	80	29	52	7	899	69
30	62	43	475	68	30	52	22	686	88
31	62	46	355	74	31	53	19	629	82
32	64	13	287	95	32	55	4	635	77
33	65	18	381	72	33	55	22	568	75
34	65	21	324	93	34	57	27	411	74
35	66	18	381	65	35	58	10	617	55
36	66	42	450	58	36	58	14	495	68
37	70	17	519	76	37	59	7	392	82
38	77	12	978	91	38	62	12	404	67
39	77	24	660	93	39	63	32	204	90
					40	64	23	745	98
					41	66	28	330	97
					42	68	32	216	84
					43	69	5	355	90
					44	69	49	539	92
					45	70	29	310	98
					46	73	29	279	88
					47	74	4	224	96
					48	77	19	367	96
					49	77	23	350	96
					50	77	26	339	92

Table 11 Example 2: List of Selected Points

NOTE: x,y coordinates

$\bar{w}$ interest value

q measure for isotropy of error ellipse (in percent)

no	i	j	w	no	i	j	w	no	i	j	w
1	118		5.643	51	2109		9.838	101	3238		5.483
2	314		4.394	52	2115		4.774	102	3245		2.673
3	321		2.968	53	2213		18.835	103	3343		4.388
4	417		5.736	54	2227		3.340	104	3423		3.961
5	419		2.998	55	2239		4.372	105	3522		4.566
6	624		2.901	56	2242		4.866	106	3527		2.576
7	626		7.761	57	2334		5.221	107	3540		4.418
8	726		6.448	58	2340		3.506	108	3543		5.201
9	903		5.922	59	2344		6.389	109	3625		18.249
10	921		2.360	60	2350		6.412	110	3631		3.709
11	923		5.297	61	2408		2.662	111	3722		6.343
12	927		1.790	62	2419		13.688	112	3725		4.658
13	1008		2.880	63	2427		4.470	113	3726		8.724
14	1112		3.905	64	2434		4.626	114	3731		4.081
15	1205		4.182	65	2440		3.616	115	3734		4.062
16	1213		2.141	66	2444		5.546	116	3737		2.934
17	1217		4.604	67	2528		3.528	117	3740		5.449
18	1219		2.110	68	2529		4.389	118	3741		5.045
19	1227		3.880	69	2538		10.427	119	3743		4.729
20	1304		2.721	70	2548		4.565	120	3826		18.680
21	1317		4.953	71	2618		22.025	121	3834		3.980
22	1322		5.687	72	2649		4.370	122	3837		3.512
23	1326		8.102	73	2717		39.595	123	3840		9.094
24	1334		4.166	74	2719		3.926	124	3843		4.102
25	1410		6.034	75	2722		7.434	125	3930		15.418
26	1413		2.392	76	2726		7.321	126	3935		3.754
27	1425		5.155	77	2734		9.565	127	3944		8.294
28	1427		2.200	78	2735		4.318				
29	1508		13.592	79	2740		5.840				
30	1511		2.105	80	2743		6.578				
31	1521		2.783	81	2744		15.460				
32	1525		2.925	82	2814		2.866				
33	1603		9.843	83	2821		23.014				
34	1605		2.750	84	2825		3.689				
35	1621		3.211	85	2827		3.425				
36	1623		4.138	86	2839		2.692				
37	1627		2.230	87	2915		5.572				
38	1639		9.894	88	2949		4.512				
39	1726		3.290	89	3017		4.004				
40	1737		4.264	90	3031		3.112				
41	1741		2.716	91	3040		4.683				
42	1804		8.803	92	3041		3.685				
43	1807		4.351	93	3044		8.097				
44	1819		3.058	94	3050		2.444				
45	1834		4.684	95	3121		12.112				
46	1911		39.671	96	3123		2.489				
47	1921		2.798	97	3127		2.834				
48	1939		2.405	98	3130		3.692				
49	2010		17.253	99	3139		3.171				
50	2025		14.639	100	3229		2.622				

Table 12 Example 2: List of Selected Pairs

NOTE: ij point No. in left and right image (201 ≅ (2,1))
w initial weight

estimated shift: -18.111 -3.333

i	left	right	xl	yl	xr	yr	dx	dy
1	18	7	36.000	55.000	20.000	50.000	2.111	-1.667
2	19	11	36.000	58.000	23.000	55.000	5.111	0.333
3	24	19	50.000	51.000	34.000	49.000	2.111	1.333
4	26	18	52.000	47.000	34.000	45.000	0.111	1.333
5	27	17	56.000	22.000	34.000	16.000	-3.889	-2.667
6	28	21	56.000	51.000	38.000	50.000	0.111	2.333
7	34	23	65.000	21.000	43.000	16.000	-3.889	-1.667
8	36	25	66.000	42.000	46.000	41.000	-1.889	2.333
9	38	37	77.000	12.000	59.000	7.000	0.111	-1.667

Table 13 Example 2: Result of Robust Shift Adjustment

NOTE: 7 iterations, list had not been cleaned

$$\begin{pmatrix} 0.81926 & 0.13883 & ; & -14.90 \\ 0.17803 & 1.17112 & ; & -19.92 \end{pmatrix} = (\hat{B} ; \hat{a})$$

i	left	right	xl	yl	xr	yr	dx	dy
1	18	7	36.000	55.000	20.000	50.000	2.225	0.899
2	19	11	36.000	58.000	23.000	55.000	-0.359	-0.588
3	24	19	50.000	51.000	34.000	49.000	-0.861	-0.293
4	26	18	52.000	47.000	34.000	45.000	0.222	-0.622
5	27	17	56.000	22.000	34.000	16.000	0.029	-0.188
6	28	21	56.000	51.000	38.000	50.000	0.055	-0.225
7	34	23	65.000	21.000	43.000	16.000	-1.737	0.244
8	36	25	66.000	42.000	46.000	41.000	-1.002	0.015
9	12	5	27.000	19.000	12.000	8.000	-2.147	-0.864
10	15	8	34.000	63.000	22.000	60.000	-0.303	-0.088
11	20	10	43.000	18.000	23.000	9.000	-0.177	-0.186
12	21	9	44.000	14.000	23.000	4.000	0.087	0.307
13	22	13	50.000	18.000	28.000	10.000	0.558	0.060
14	29	15	57.000	9.000	33.000	4.000	0.043	-3.234
15	35	22	66.000	18.000	43.000	12.000	-1.334	0.908
16	38	26	77.000	12.000	49.000	7.000	0.845	0.840
17	9	3	23.000	21.000	9.000	9.000	-2.146	-0.234
18	39	30	77.000	24.000	52.000	22.000	-0.489	-0.107

Table 14 Example 2: Result of Robust Affine Transformation

NOTE: 6 iterations, cleaned list (final result), only 2 correspondencies were ambiguous.

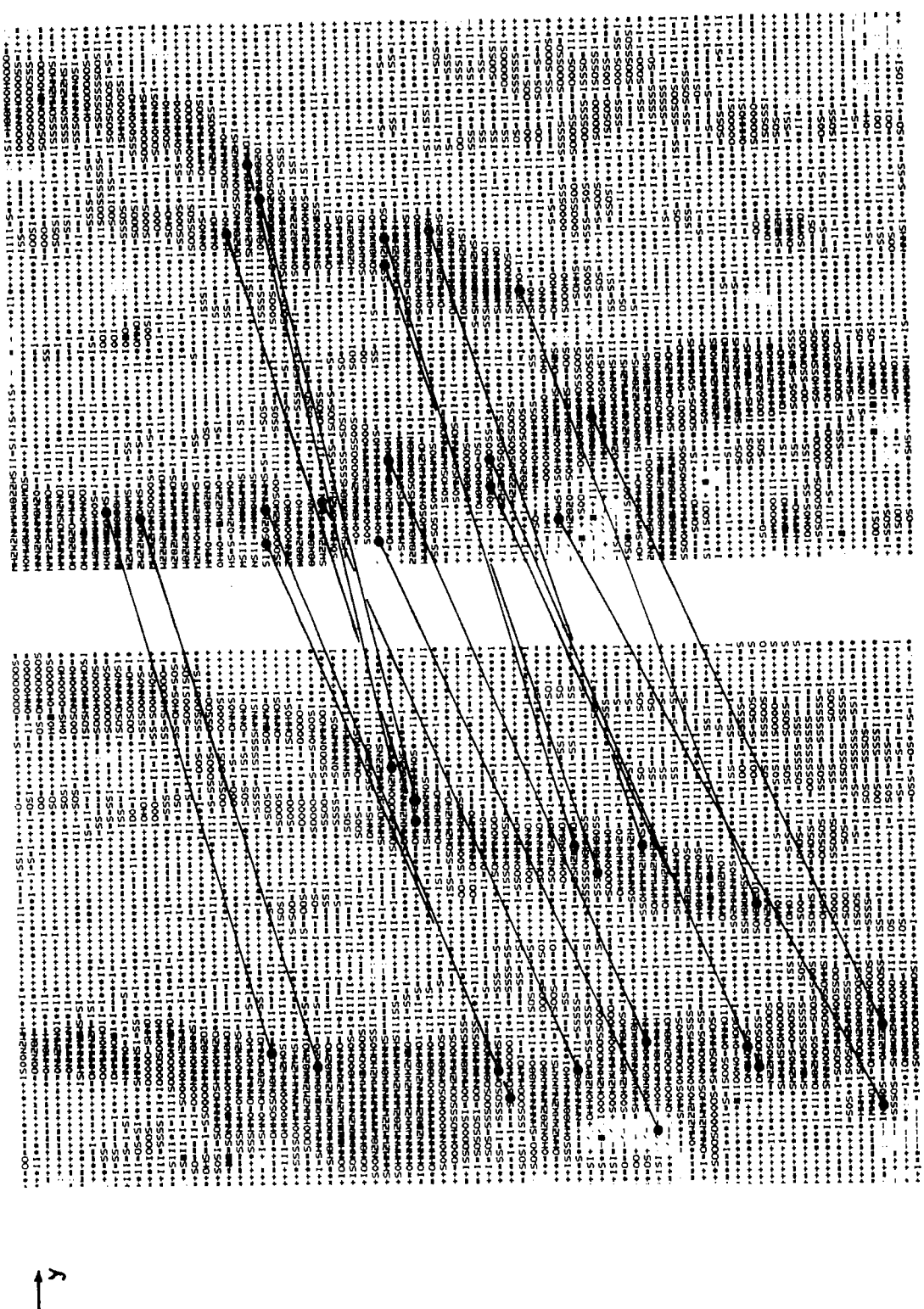

Figure 14 Example 2: Result of Correspondence Algorithm

4.6 Preliminary Conclusions

From other experiments with simulated and real data the following preliminary conclusions may be drawn:

a. The algorithm in its present form works well if the relative distortions of the images are not larger than 20-30% (corresponding to a rotation of up to 20°) and the overlapping area contains enough distinct points. These conditions can always be met if an operator provides the approximate values, or the images are oriented with an automatic procedure as the LNK-Method.

b. The results are accurate up to 1-2 pixels, if the deviation of the geometrical model from the real distortion is not too large.

c. The algorithm is fast enough to replace the first iterations in a correlation-based algorithm for high precision registration or rectification. The total computing time for a pair of images with 128x128 pixels is approximately 2-3 seconds on a VAX 11/780 and is nearly proportional to the number of pixels.

d. The limitation of the algorithm in its present form results from the similarity measure, namely the correlation coefficient, which is not scale or rotation independent. As already pointed out, other measures, as for example invariant moments might solve this problem.

e. Further research should be directed towards a link with the features of the LNK-method. There are two ways to do that which are complementary: One could use robust estimation procedures to refine the estimation of the LNK-method and one could use the abstract features, especially the abstract lines, as input for the correspondence algorithm. In this case each abstract line would give rise to four observation equations, derived from the coordinates of one end point of the line and the coordinate differences to the other end point. If consideration is restricted to rotation and scale differences only, the angular difference and the logarithm of the scale ratio of the pairs of abstract edges would lead to a robust estimation of the means of the shift, the rotation and the logarithm of the scale difference of the two images. The inclusion of line features into the algorithm would allow its application for rectification of satellite images.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

From the research performed so far, the following are the conclusions to be drawn:

1. The collinearity (or parametric) model is superior to the polynomial (or interpolative) model particularly when the number of control points is small.

2. Through simulations, it is shown that the parametric model adequately describes the real data.

3. Rectification of single image scanner data is more sensitive to image position errors than ground position errors.

4. Uncertainty in attitude estimate is the main source of error in system-corrected images.

5. In general, when more than about 25 well distributed control points are used, the effect in rectification accuracy is marginal.

6. The distribution of control features is critical to the rectification accuracy; to obtain the same accuracy about three times the number of well distributed control features are needed when such features are randomly distributed.

7. The block adjustment procedure based in the parametric rectification model was successful. Tie points between overlapping images improved rectification accuracy, particularly when few control points are used.

8. Edges proved to be an effective type of control for single image rectification. In general, about three edge pairs are needed for each conventional control point.

9. An efficient new algorithm for finding corresponding points in image pairs has been developed. The unknown parameters of the geometric transformation between the two images are derived using robust estimation techniques.

10. Tests with simulated and real data show that the present correspondence algorithm can accomodate geometric distortions up to 20 to 30 %, which corresponds to an average distortion of 3 to 7 pixels in an image of size 128x128 pixels.

11. The correspondence algorithm incorporates a new operator for finding distinct objects in an image based on the expected precision of locating such object by cross correlation.

5.2 Recommendations

1. Continue to investigate other non-conventional control such as geometric constraints and relative control (e. g. distances, angles etc.).

2. Extend the block adjustment program to accomodate edge control and perform tests.

3. Continue to develop the correspondence algorithm and apply it to remote sensing data both for registration and rectification.

4. Study the rectification/registration sequence.

5. Investigate rectification accuracy assessment.

6. Analyse blunder detection and identification procedures.

7. Research the problem of merging remote sensing data and digital terrain models.

6. ACKNOWLEDGEMENTS

This research was supported by NASA Contract No. 9-16664 through Subcontract No. L200074 with Texas A&M Research Foundation. The authors wish to thank Dr. R. P. Heydorn of NASA Johnson Space Center and Professor L. F. Guseman, Jr., of Texas A&M University for their cooperation and counsel. They would also like to express their appreciation to Dr. D. W. Mooneyhan and Dr. David Dow of NASA National Space Technology Laboratories for providing the two real data sets.

References

1. Bähr, H. P., "Geometrical Analysis and Rectification of LANDSAT MSS Imagery: Comparison of Different Methods", Nachr. Kart. u. Vermess. wes., No. 36, Institut fur Angewandte Geodasie, Frankfurt, 1978, pp. 25-46.

2. Bähr, H. P., "Geometrical Models for Satellite Scanner Imagery", ISP Commission III Congress, Helsinki, July, 1976.

3. Bähr, H. P., "Interpolation and Filtering of ERTS Imagery", ISP Commission III Symposium, Stuttgart, September, 1974.

4. Barnard, S. T. and Thompson, W. B., "Disparity Analysis of Images", IEEE, Vol. PAMI-2, 1981, pp. 333-340.

5. Barnea, D. T. and Silverman, H. F., "A Class of Algorithmus for Fast Digital Image Registration, IEEE, Vol. C-21, 1972, pp. 179-186.

6. Bernstein, R., "Digital Image Processing of Earth Observation Sensor Data", IBM J. Research Development, January, 1976, pp. 40-57.

7. Bernstein, R., "Scene Correction (Precision Processing) of ERTS Sensor Data Using Digital Image Processing Techniques", 3rd ERTS Symposium, Vol. 1-A, NASA SP-351, 1973.

8. Cafforio, C. and Rocca F., "Tracking Moving Objects in Television Images", Signal Processing 1, 1979.

9. Caron, R. H. and Simon, K. W., "Attitude Time-Series Estimator for Rectification of Spaceborne Imagery", J. Spacecraft, Volume 12, Number 1, January, 1975, pp. 27-32.

10. Dowman, I. J. and Mohamed, M. A., "Photogrammetric Applications of LANDSAT MSS Imagery", International J. Remote Sensing, Volume 2, Number 2, 1981, pp. 105-113.

11. Ewing, C. E. and Mitchell, M. M., *Introduction to Geodesy*, New York: American Elsevier, 1969.

12. Forrest, R. B., "Simulation of Orbital Image-Sensor Geometry", Photogrammetric Engineering and Remote Sensing, Volume 47, Number 8, August, 1981, pp. 1187-1193.

13. Forrest, R. B., "Geometric Correction of ERTS-1 MSS Images",ISP Commission III Symposium, Stuttgart, September, 1974.

14. Förstner, W., "Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation Techniques", ISP Commission III Congress, International Archives of Photogrammetry, Volume 25-III, Rio de Janeiro, July, 1984.

15. Förstner, W., "On the Geometric Precision of Digital Correlation", International Archives for Photogrammetry, Vol. 24-III, Helsinki, 1982, pp. 176-189.

16. Friedmann, D. E. , Friedel, J. P., Magnussen, K. L., Kwok, R., and Richardson, S., "Multiple Scene Precision Rectification of Spaceborne Imagery With Very Few Ground Control Points", Photogrammetric Engineering and Remote Sensing, Volume 48, Number 12, December, 1983, pp. 1657-1667.

17. Grimson, W. E. L., *From Images to Surfaces: A Computational Study of the Human Early Visual System*, M.I.T. Press, 1981.

18. Hampel, F. R., "Robust Estimation: A Condensed Partial Survey", Zeitschrift fur Wahrscheinlichkeits Theorie, 1973.

19. Huber, P. J., *Robust Statistics*, New York, 1981.

20. Kak, A. C., *Depth Perception for Robots*, School of Electrical Engineering, Purdue University, TR-EE 83-44, 1983.

21. Konecny, G., "Mathematical Models and Procedures for the Geometric Restitution of Remote Sensing Imagery", ISP Commission III Congress, Helsinki, July, 1976.

22. Krarup, T., Juhl, J., and Kubik, K., "Gotterdammerung Over Least Squares", International Archives for Photogrammetry, Vol. 23-B3, Humburg, 1980, pp. 369-378.

23. Kratky, V., "Photogrammetric Solution for Precision Processing of ERTS Images", ISP Commission II Congress, Ottawa, July, 1972.

24. Kuglin, C. D. and Hines, D. C., "The Phase Correlation Image Alignment Method", Proceedings of the IEEE International Conference on Cybernetics and Society, 1975, pp. 163-165.

25. Lambird, B. A., Lavine, D., Stockmann, G. C., Hayes, K. C., and Kanal, L. N., Study of Digital Matching of Dissimilar Images, ETL Report, No. ETL-0248, 1980.

26. Levine, I., "Computational Aspects of Geometric Correction Data Generation in the LANDSAT-D Imagery Processing", Sixth Annual Flight Mechanics/Estimation Theory Symposium, NASA/GSFC, October, 1981.

27. Levine, I., "The MSS Control Point Location Error Filter for LANDSAT-D", Sixth Annual Flight Mechanics/Estimation Theory Symposium, NASA/GSFC, October, 1981.

28. Marr, D., *Vision: A computational Investigation of the Human Representation and Processing of Visual Information*, W. H. Freemann, San Francisco, 1981.

29. McGillem, C. D. and Svedlow, M., "Image Registration Error Variance as a Measure of Overlay Quality", IEEE, Vol. GE-14, 1976, pp. 44-49.

30. McGlone, J. C. and Mikhail, E. M., *Photogrammetric Analysis of Aircraft Multispectral Scanner Data*, Technical Report Number CE-PH-81-3, School of Civil Engineering, Purdue University, West Lafayette, Indiana, October, 1981.

31. Mikhail, E. M., Akey, M. L., and Mitchell, O. R., "Detection and Sub-Pixel Location of Photogrammetric Targets in Digital Images", Presented Paper, Specialist Workshop on Pattern Recognition in Photogrammetry, Graz, 1983.

32. Mikhail, E. M. and Paderes, F. C., "Simulation Aspects in the study of Rectification of Satellite Scanner Data", Proceedings of the NASA Symposium on Mathematical Pattern Recognition and Image Analysis, Johnson Space Center, Houston, Texas, June, 1983, pp. 413-483.

33. Mikhail, E. M. and McGlone, J. C., "Current Status of Metric Reduction of (Passive) Scanner Data", ISP Commission III Congress, Hamburg, July, 1980, pp. 504-514.

34. Mikhail, E. M., *Observations and Least Squares*, New York: Harper and Row, 1976.

35. Panton, D. J., "A Flexible Approach to Digital Stereo Mapping", Proceedings of the DTM Symposium, ASP, St. Louis, 1978, pp. 32-60.

36. Puccinelli, E. F., "Ground Location of Satellite Scanner Data", Photogrammetric Engineering and Remote Sensing, Volume 42, Number 4, April, 1976, pp. 537-543.

37. Rifman, S. S., Monuki, A. T., and Shortwell, C. P., "Multi-Sensor LANDSAT MSS Registration", 13th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, April, 1979.

38. Rosenfeld, A., Hummel, R. A., and Zucker, S. W., "Scene Labeling by Relaxation Operations", IEEE, Vol. SMC-6, 1976, pp. 420-433.

39. Sawada, N., Kidode, M., Shinoda, H., Asada, H., Iwanaga, M., Watanabe, S.,Mori, K., and Akiyama, M., "An Analytic Correction Method for Satellite MSS Geometric Distortions", Photogrammetric Engineering and Remote Sensing, Volume 47, Number 8 ,August, 1981, pp. 1195-1203.

40. Stockmann, G. C., Kopstein, S., and Benett, S., "Matching Images to Models for Registration and Object Location via Clustering", IEEE, Vol. PAMI-4, No. 3, 1982, pp. 229-241.

41. Synder, J. P., "Geometry of a Mapping Satellite", Photogrammetric Engineering and Remote Sensing, Volume 48, Number 10, October, 1982, pp. 1593-1602.

42. Trinder, J. C. and Nasca, S. U., "Tests on the Mapping Application of LANDSAT Imagery", ISP Commission III Congress, Helsinki, July, 1976.

43. Werner, H., "Automatic Gross Error Detection by Robust Estimators", International Archives for Photogrammetry, Vol. 25-III, Rio de Janeiro, 1984.

44. Wong, R. and Hall, E., "Scene Matching with Invariant Moments", CGIP, Vol. 8, 1978, p. 16.