

AUTOMATIC RECTIFICATION OF LANDSAT IMAGES
USING FEATURES DERIVED FROM
DIGITAL TERRAIN MODELS

by

JAMES J. LITTLE

Technical Report 80-10
December, 1980

Department of Computer Science
University of British Columbia
Vancouver, B.C., Canada V6T 1W5

ABSTRACT

Before two images of the same object can be compared, they must be brought into correspondence with some reference datum. This process is termed registration. The reference can be one of the images, a synthetic image, a map or other symbolic representation of the area imaged. A novel method is presented for automatically determining the transformation to align a Landsat image to a digital terrain model, a structure which represents the topography of an area. Parameters of an affine transformation are computed from the correspondence between features of terrain derived from the digital terrain model, and brightness discontinuities extracted from the Landsat image.

Based upon a thesis submitted to the Department of Computer Science, University of British Columbia, in partial fulfillment of the requirements of the degree of Master of Science.

© James Joseph Little, 1980

Table of Contents

1.	Introduction	1
2.	Previous Work	9
	2.1 Registration and Rectification	9
	2.2 Digital Terrain Models	12
	2.3 Synthetic Images	14
	2.3.1 Reflectance Functions	14
	2.3.2 Sun Position	18
	2.4 Feature Selection	18
	2.5 Matching	20
	2.5.1 Symbolic Matching	21
	2.5.2 Geometrically Constrained Matching	23
3.	The Registration Method	28
	3.1 The First Stage: Feature Extraction	28
	3.1.1 Extracting Features from Digital Terrain Models	28
	3.1.2 Extracting Features from the Landsat Image	29
	3.1.3 Selecting Feature Points	31
	3.1.4 Line Growing	35
	3.1.5 Approximations to Lines	36
	3.2 The Second Stage: Matching	44
	3.3 The Affine Transformation	46
	3.4 Construction of a Pairing	49
	3.5 Support for a Pairing	53
	3.6 Consistency of the Transform	56
	3.7 Verification	56
4.	Implementation and Testing	59
	4.1 The Input	59
	4.2 Programming Languages	60
	4.3 Data Structures	60
	4.4 Estimating the Affine Transform	62
	4.5 Examples of Registration and Results	63
5.	Discussion and Conclusions	74
	5.1 Discussion	74
	5.2 Further Work	76
	5.3 Conclusions	76
	Bibliography	77

List of Figures

Figure 1	Subsection of a Landsat image	3
Figure 2	Contour plot (100 meter interval) of the digital terrain model ..	4
Figure 3	Landsat features	5
Figure 4	DTM features	6
Figure 5	Registered Landsat image	8
Figure 6	Triangulated Irregular Network	13
Figure 7	The geometry of light reflection	16
Figure 8	Synthetic image	17
Figure 9	Sun position in terms of azimuth and elevation	19
Figure 10	Self-shadowed facets	30
Figure 11	Subsection of a Landsat image	32
Figure 12	Output of the edge detection step	33
Figure 13	Edge detector output after echo suppression	34
Figure 14	Lines before merging	37
Figure 15	Compatibility of segment directions in line merging	38
Figure 16	Lines after merging	39
Figure 17	Line generalization	40
Figure 18	Varying the detail level in generalization	42
Figure 19	The band of a curve	43
Figure 20	Landsat features	45
Figure 21	Structured and simple ridges	47
Figure 22	Close fit in a pairing	51
Figure 23	Loose fit in a pairing	52
Figure 24	Mismatches and support	54
Figure 25	Self-consistency of a transform	57
Figure 26	Synthetic image for September 14 1976	64
Figure 27	DTM features with matched points	65
Figure 28	Landsat features with matched points	67
Figure 29	Subsection of Landsat image	68
Figure 30	Synthetic image for January 8 1979	70
Figure 31	DTM features with matched points	71
Figure 32	Landsat features with matched points	72
Figure 33	Registered Landsat image for January 8 1979	73

Acknowledgement

I sincerely thank my supervisor Robert J. Woodham for his help. His perceptive criticism of my work in progress has not only improved this thesis but also has been very beneficial to my education. I also wish to thank Alan K. Mackworth for encouraging me originally to pursue this thesis topic.

1. Introduction

There are many image analysis tasks where the objects in the scene are known beforehand. Often industrial inspection and manipulation tasks involve determining the position and orientation of a known part within a given image (Hsieh and Fu, 1979; Agin, 1980; Myers, 1980). Similarly, biomedical applications, such as chest X-ray interpretation (Ballard et al., 1979), often deal with images whose general content is known. The interpretation of images acquired via satellite or aerial photography is facilitated by knowledge of the scene given in form of maps and other geographic data. Once the position and orientation of objects in a scene is determined, image analysis simplifies. Thus, registering the image to the scene model is an important first step in automatic image interpretation. In remote sensing, the spatial relations between the objects in the scene are precisely known, and the geometric relation between image and scene model can be characterized by a fixed mathematical transformation of known form but unknown parameters. In contrast, the number of ribs in a chest x-ray, for example, is given, as well as their general spatial relationships, but their precise size and position are not precisely known. The importance of registration has been demonstrated in the domain of Landsat images. When a new image has been brought into correspondence with surface data, the interpretation of ground cover is improved. For example, the effect of shading due to variations in surface topography and shadows can be estimated (Woodham, 1980). Thus far, registration has eluded complete automation. The object of this thesis is to present a method for automated registration of Landsat images.

A Landsat MSS image measures scene radiance in each of four spectral

bands, at a nominal ground resolution of 79 x 79 meters. The position and attitude of the Landsat MSS platform is known with limited precision. After systematic corrections based on the estimated platform position and attitude, the ground location of an image point may differ from its true position by as much as 10 kilometers. Since each picture element (pixel) of a Landsat MSS image has a nominal ground spacing of 56 meters in the across track direction and 79 meters in the along track direction, this represents an error of up to 179 pixels. Further processing is thus required to relate the image coordinate system to other coordinate systems.

A digital terrain model (DTM) represents surface elevation as a function of ground coordinates. A DTM can be accurately located in a geographic coordinate system. A Landsat image registered to a digital terrain model can be directly compared with other sources of geographic information, and other images.

An automatic method for registering Landsat images to digital terrain models is developed. As an example, figure 1 shows a 100 x 100 section of a Landsat image. Figure 2 shows a contour plot at 100 meter intervals of the digital terrain model. In the method presented here, a set of curvilinear features is determined from both the Landsat image and the DTM. Features from the Landsat image are shown in figure 3 while those selected from the DTM are depicted in figure 4. A correspondence between the elements of the two sets of curvilinear features is established which satisfies both geometric (shape) constraints and topological (adjacency) constraints. The matching between elements determines point pairs input to a least-squares estimator for the parameters of an affine transform. The image registration problem is transformed into the problem of matching sets of curves in the plane. The points on the features used for calculating the transform are



Figure 1

100 X 100 pixel subsection of Landsat image (band 7) from September 14, 1976, frame ID 11514-17153. Photographed from the screen of the COMTAL Vision 1.

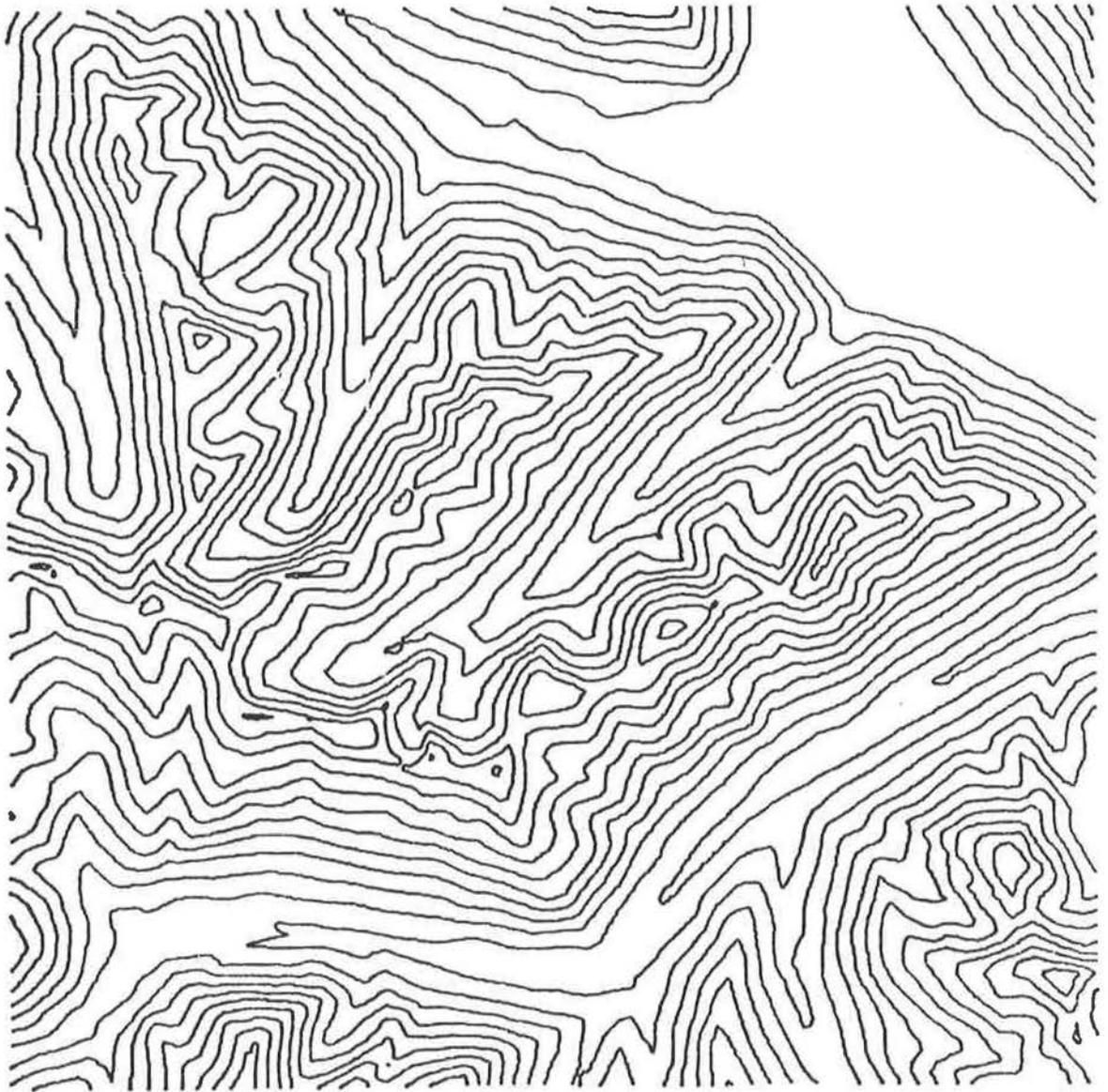


Figure 2

Contour plot (100 meter interval) of the digital terrain model sub-section used in the test case.



Figure 3

Features derived from Landsat image subsection
acquired on September 14, 1976.



Figure 4

Features derived from DTM using the sun's position at the time of image acquisition on September 14, 1976.

labelled 1-6 in figures 3 and 4. The registered Landsat image is shown in figure 5. The derived affine transform is:

$$\begin{aligned}x' &= a x + b y + c \\y' &= d x + e y + f\end{aligned}$$

where (x,y) are Landsat image coordinates, (x',y') are DTM coordinates, and where

$$\begin{aligned}a &= 0.555292 \\b &= 0.131612 \\c &= 1.944259 \\d &= -0.143495 \\e &= 0.773612 \\f &= 2.197464\end{aligned}$$

Chapter 2 reviews previous work in registration, feature detection, digital terrain models and matching methods.

Chapter 3 develops the method for registering images to digital terrain models. Knowledge of sun position is used to select features for registration. Geometric constraints are used to guide the registration process.

Chapter 4 presents the particular implementation used to realize the method.

Chapter 5 discusses the results and their relevance to other image understanding tasks.

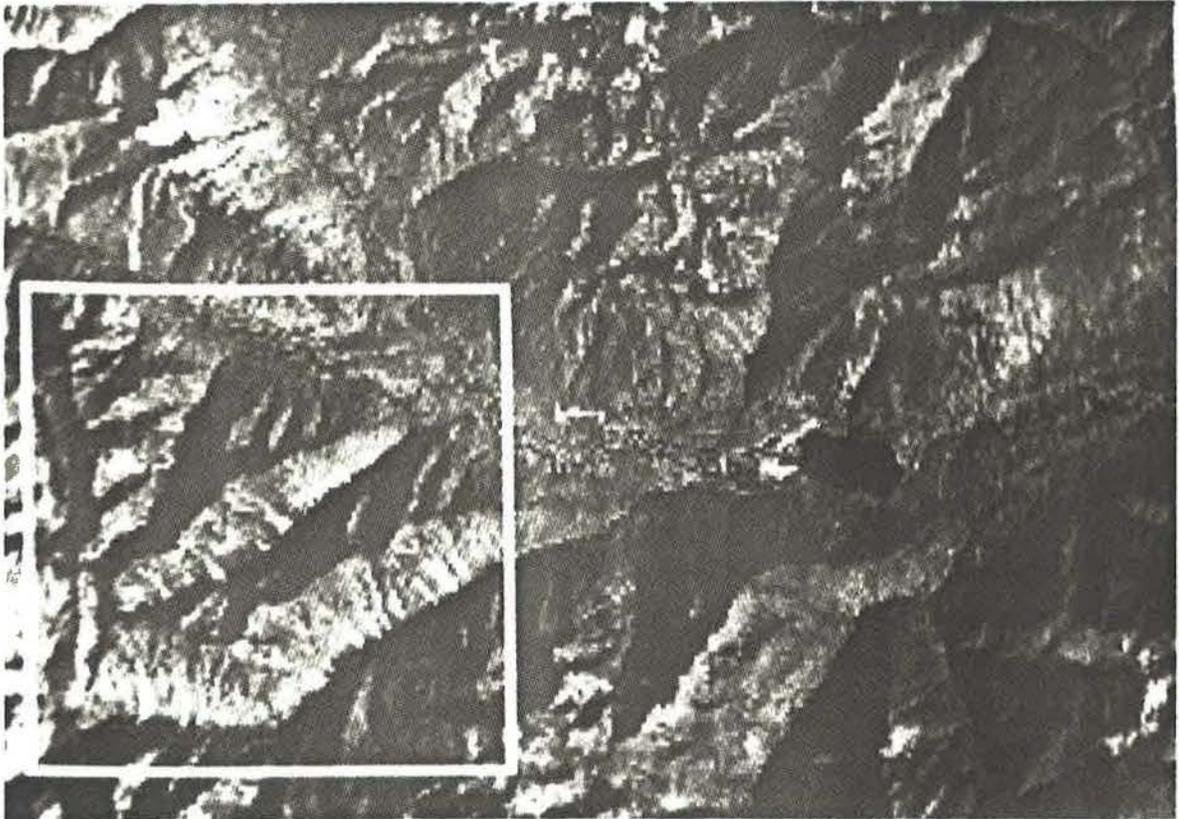


Figure 5

Registered Landsat image for September 14, 1976. The white square outlines the 10 km² area covered by the DIM. Photographed from the screen of the COMTAL Vision 1.

2. Previous Work

2.1 Registration and Rectification

Image registration is the process of determining the correspondence between elements of two or more images and applying a transformation to one image to align it with the other. Two satellite images would first be registered in order to proceed with change detection. However, it is often necessary to register images not just to each other but also to absolute ground coordinates. Registering an image to absolute ground coordinates is called image rectification. Often the term registration is used for both image-to-image registration and image-to-ground registration (i.e., rectification).

Commonly, two images are registered by manually selecting ground control points (GCP's) from each image (Bernstein, 1976). A ground control point is a distinctive ground feature detectable in an image. Typical GCP's are airports, land-water boundaries, field patterns and highway intersections. Parameters of an appropriate transformation are calculated from a subset of the selected GCP's. For each GCP in one of the images, the corresponding GCP must be located in the second. Manual selection of GCP's is time-consuming. Several techniques have been developed to automate partially the selection of ground control points.

As an initial improvement to the manual method, correlation techniques can be used to improve the estimates of the GCP locations. For each GCP in the reference image, a small $m \times n$ subsection of image surrounding the GCP is used as a template. The best matched position of the template determines

the location of the corresponding GCP. The best position is that at which the correlation of the template with the image is maximized. The correlation between an $m \times n$ template $S1$ and a subsection of an image $S2$ at (x,y) is:

$$\sum_{i=1}^m \sum_{j=1}^n S1(i,j) \cdot S2(x+i,y+j)$$

High output of this operation may result if one of the subsections has a high average gray level. For this and other reasons, it is convenient to normalize the correlation, resulting in the following formulation:

$$\frac{\sum_{i=1}^m \sum_{j=1}^n (S1(i,j) - \bar{S1}) \cdot (S2(x+i,y+j) - \bar{S2})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (S1(i,j) - \bar{S1})^2 \cdot \sum_{i=1}^m \sum_{j=1}^n (S2(x+i,y+j) - \bar{S2})^2}}$$

where

$$\bar{S1} = 1/(m \cdot n) \cdot \sum_{i=1}^m \sum_{j=1}^n S1(i,j)$$

and

$$\bar{S2} = 1/(m \cdot n) \cdot \sum_{i=1}^m \sum_{j=1}^n S2(x+i,y+j)$$

Then a perfect correlation corresponds to a value of 1. Sequential similarity detection algorithms (SSDA's) can be used to speed up template matching (Barnea and Silverman, 1972). This correlation process is repeated for each GCP. The refined GCP locations are used to determine the parameters of the registration transformation. This refinement technique can be embodied in a fully automatic registration procedure if a library of GCP templates is maintained for use in the registration of subsequent images of the same area.

The GCP method can be further automated by introducing automatic selection of the GCP's. As a first step toward this, the reference image can be regularly subdivided into overlapping subimages, each of which is

used as a template in the correlation technique. This automatic method has difficulties. There is no guarantee, for example, that the GCP templates can be located by correlation in other images. The GCP's produced by regular subdivision are random with respect to the content of the image. Davis and Kenue (1977) describe a method for automatically selecting ground control points in a reference image. Ground control points are selected where there is a strong connected set of brightness discontinuities. The algorithm thresholds the gradient of the image and finds connected sets of pixels in the thresholded gradient image. GCP's are selected from the resulting set so as to be as evenly scattered about the image as possible. This method improves upon regular subdivision of the reference image into templates, but it also suffers two major shortcomings:

- 1) Since the GCP's are chosen on the basis of image features, the GCP's have no necessary relation to ground features whose appearance can be expected to remain constant in other images.
- 2) In particular, no attempt is made to take account of possible changes in illumination between the images, which will systematically affect the appearance of the templates.

In the absence of a scene model, not much more can be done. However, digital terrain models, when available, can be used to select and verify GCP's for registration.

Horn and Bachman (1978) use synthetic images generated from digital terrain models to register Landsat images. The synthetic image represents the appearance of the terrain under the illumination conditions corresponding to the sun position at the time of image acquisition. Their published work assumes that the transformation between the synthetic image and the Landsat image can be described in terms of rotation, translation and scale change. A correlation of the real and synthetic images is used as measure of goodness of fit to guide the adjustment of rotation, translation and scaling. The correlation of the entire image is ultimately used. This

is computationally expensive. The authors avoid some of this expense by first using low resolution images to produce rough estimates of the transformation parameters. The full resolution of the data is used to compute the final refinements to the transformation parameters.

The method presented in this thesis follows the spirit of the work of Horn and Bachman. The known position of the sun is used to predict the terrain features which will appear as distinct image features. The symbolic features themselves are used to determine the transformation parameters.

2.2 Digital Terrain Models

A digital terrain model (DTM) represents the surface of the earth in a particular region. This is usually taken to mean that the DTM can be used to determine the elevation of the surface at any point in the region. Besides providing height information, a digital terrain model also represents the surface orientation since slope and aspect can be derived. Slope information is crucial for accurate calculation of the synthetic image. Because the DTM is defined in a ground coordinate system, an image registered to a DTM can be directly compared to other sources of geographic information.

A common representation of terrain is as a discrete grid of terrain elevations. Slope is determined in a grid representation by local differencing. Alternatively, terrain can be modelled as a set of contiguous non-overlapping triangular facets (figure 6), in a Triangulated Irregular Network (TIN) (Peucker et al., 1978). In the TIN, slope information is directly computed from the surface facets. Efficient procedures exist for converting the grid to a TIN and vice-versa (Peucker et al., 1978; Fowler

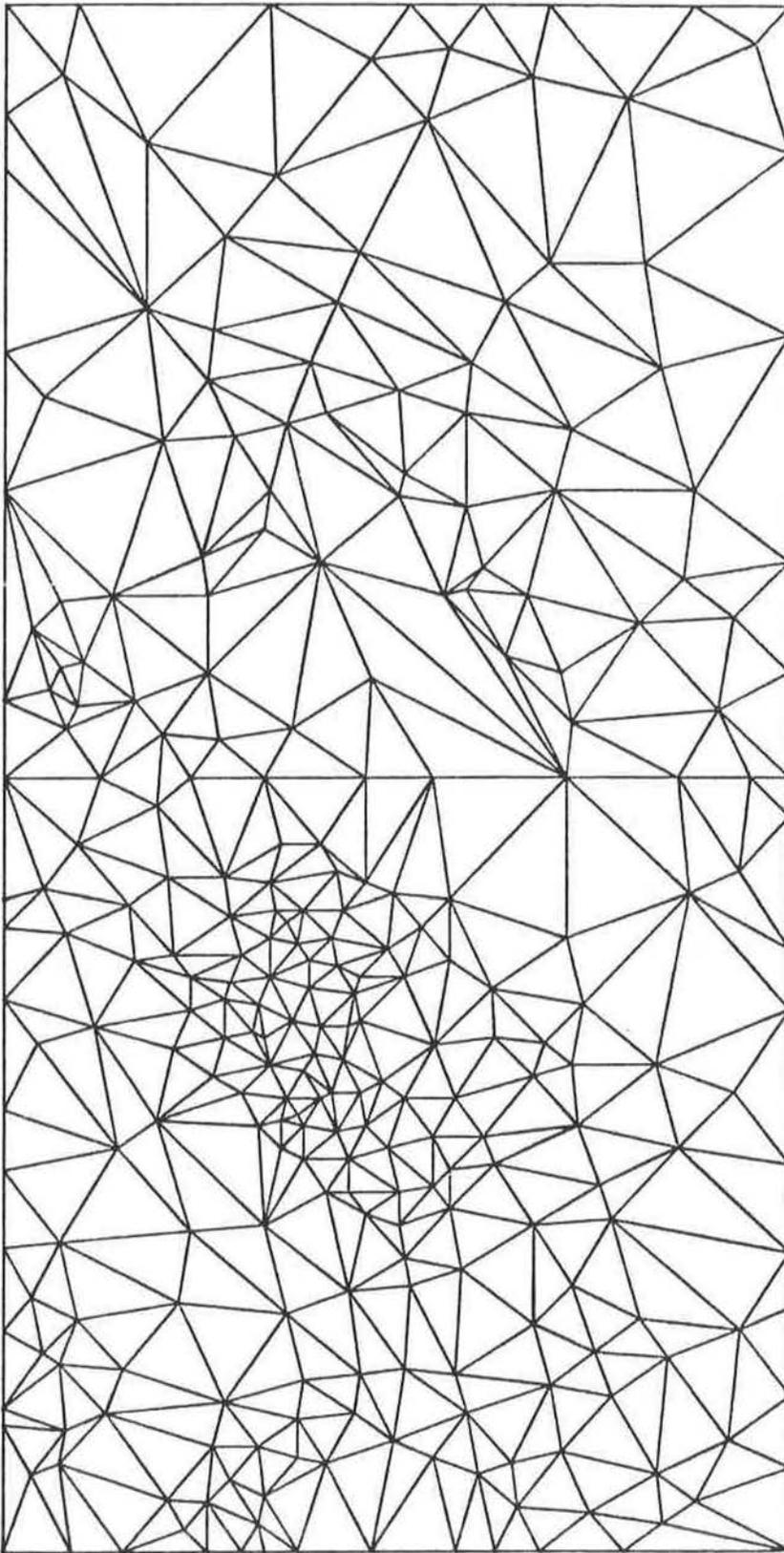


Figure 6

A Triangulated Irregular Network (TIN)

and Little, 1979). The digital terrain model used for the work described herein was constructed in the TIN format.

The structure of terrain can be modelled by the network of ridges and channels (divides and streams). The ridges are convex linear surface features which, theoretically, connect passes (saddle points) to peaks (relative maxima). In practice the set of ridges on a surface also includes convex linear features which connect to the main ridges that do join passes to peaks. Channels are concave linear features connecting passes to pits (relative minima). In addition, the surface behavior of the terrain between ridges and channels is modelled. This surface behavior includes lines along which the surface changes slope. These are termed breaks of slope. Actual production of a DTM, whether a grid or a TIN, often involves recording the terrain structure of ridges and channels.

Several methods exist for deriving the the location of ridges and channels from the grid representation (Peucker and Douglas, 1975; Toriwaki and Fukumaru, 1978). The TIN model is advantageous for feature selection since ridges, channels and breaks of slope are explicitly represented as the boundaries of triangular facets.

2.3 Synthetic Images

2.3.1 Reflectance Functions

Image irradiance at a given point depends on the object material imaged at that point and its orientation in space with respect to the light source(s) and viewer. Following Horn and Bachman, one model of image

formation uses a surface reflectance function:

$$\text{PHI}(I, E, G) = P * \text{COS}(I)$$

where I, E, and G are the incident angle(I), emittance angle(E), and the phase angle(G) (figure 7). In the above equation, P is an albedo factor depending on the surface composition, which, without additional a priori knowledge, is assumed to be constant. This reflectance function models a lambertian surface which, as a perfect diffuser, appears equally bright from all viewing directions. The incident angle(I) is the angle between the surface normal and the illumination direction. In the case of Landsat imagery, the principal light source is the distant sun, so that the illumination direction is effectively constant for all surface points. Diffuse illumination from the atmosphere and possibly from other scene elements is ignored in the synthetic image.

The DTM provides accurate estimates of the surface orientation for the test area. A synthetic image is produced under the assumption of an orthographic projection and a single light source at the known sun position. The brightness in the synthetic image at each picture element (pixel) is a function of the surface orientation at the appropriate point in the DTM. Figure 8 shows a synthetic image produced from a digital terrain model, using the simple reflectance function described above, with the sun from the northwest at 45 degrees elevation, as in standard cartographic convention.

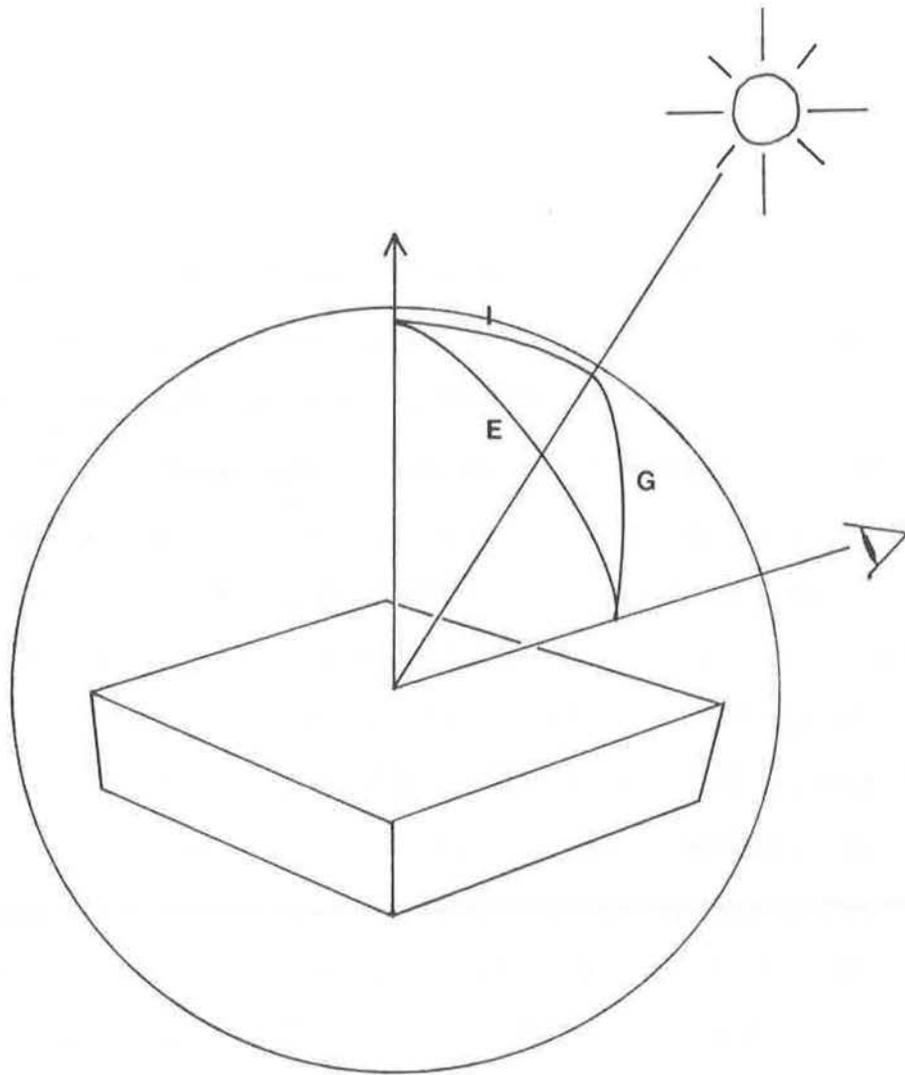


Figure 7

The geometry of light reflection from a surface element is governed by the incident angle, I , the emittance angle, E , and the phase angle, G . (after Horn and Bachman, 1978)

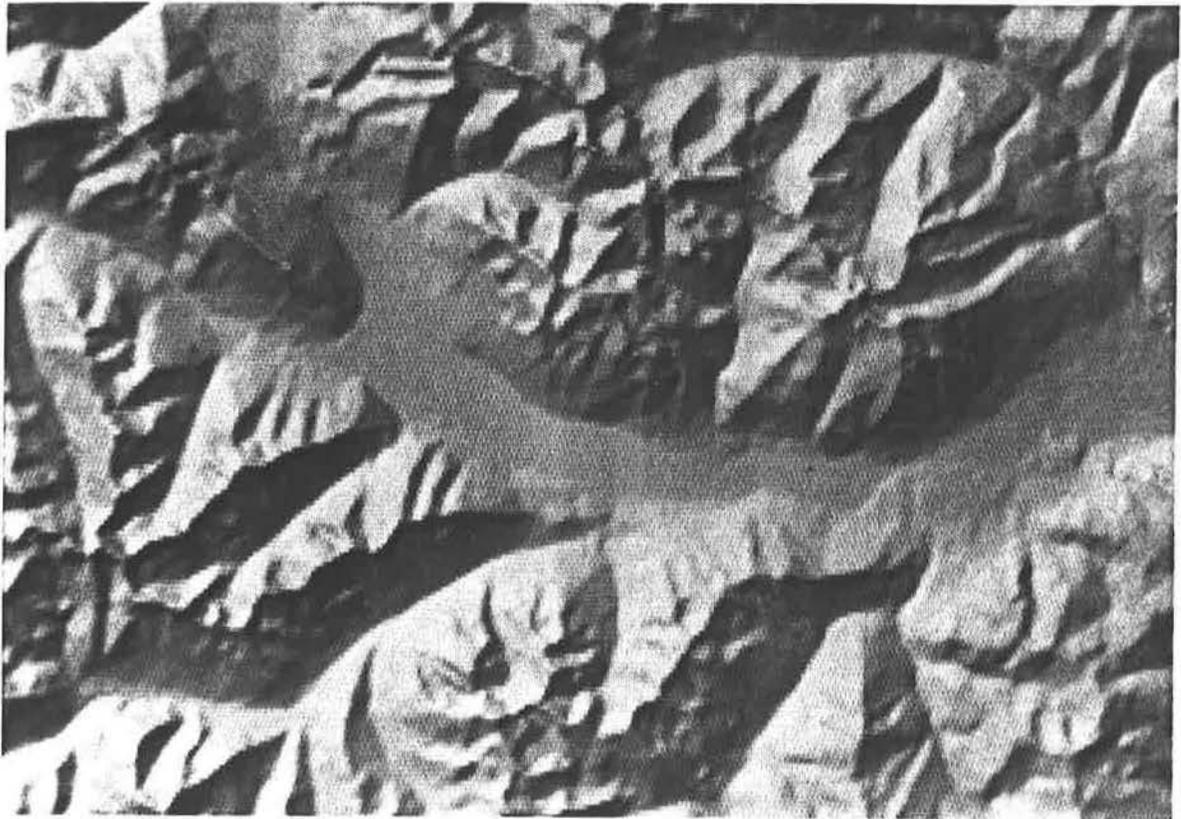


Figure 8

Synthetic image of the DTM. The light source is positioned in the northwest at 45 degree elevation as in cartographic convention. Photographed from the screen of the COMTAL Vision 1.

2.3.2 Sun Position

In order to produce a synthetic image corresponding to an actual imaging situation, it is necessary to determine the sun's position at the precise time of image acquisition. Fortunately, the time of acquisition of each Landsat scan line is accurately determined and recorded as part of the image annotation data. Standard tables or formulae can be employed to determine the position of the sun at a given date, time, latitude and longitude. Sun position is described in terms of azimuth, the angle of rotation about the vertical axis, in degrees clockwise from north, and elevation, the angle of rotation above the horizontal (figure 9) . In this description, standard cartographic convention situates the light source at azimuth 315 degrees, elevation 45 degrees.

2.4 Feature Selection

The literature in image analysis abounds with techniques for determining the position and orientation of brightness discontinuities in images (Davis, 1975). An operator which performs this task is called an "edge detector". Generally, edge detectors perform well in locating sharp boundaries where the relative brightness difference is large and the boundary is locally linear. For the purpose of the method presented in this thesis, the features to be found in Landsat images are restricted to discontinuities between relatively bright and dark regions. The intensity differences between regions are known to be large and the transitions between the regions sharp. The focus of the research is not on the design

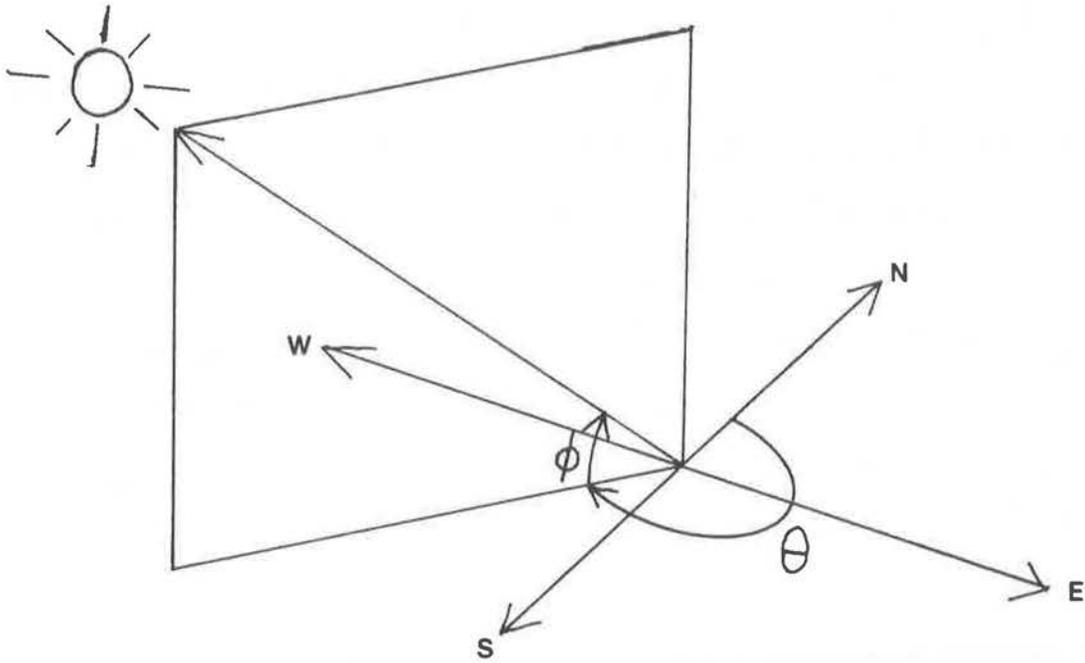


Figure 9

Definition of the position of the sun in terms of azimuth θ , and elevation ϕ .

of an optimal filter for detecting such features. Rather it is assumed that most edge-detecting operators will be robust enough to detect the required boundaries.

Similarly, the choice of method for joining pixels which display edge activity into lines is not critical. Portions of the image features sought will be relatively straight, and connecting these into curves is straightforward. Any rule for connection which prefers extending existing lines along the general line tendency is acceptable. It is presumed when the features are used that there will be gaps in the curves, so the degree to which a line-growing method is able to bridge these gaps is not critical. In sum, feature selection techniques were chosen from existing methods in the literature.

Feature selection in the domain of the digital terrain model derives from an investigation of the production of synthetic images (discussed in section 2.3). The method for DTM feature selection will be elaborated in section 3.1.1.

2.5 Matching

In the automatic registration method presented in the thesis, a one-to-one correspondence is derived between symbolic features of an image and a digital terrain model. This correspondence between features can be modelled as a matching of the features, depending upon the similarity of their descriptions. If the matching between features based on their symbolic descriptions is reliable, then matching methods can be used for automatic registration. Other researchers have considered the problem of matching an image to a model of the scene. Research on matching

descriptions of an image to models of the scene can be divided into those which handle symbolic descriptions only, and those which also manipulate geometric relations. The topics of interest in this work are both representations of object and image relations and the control structures used in matching.

2.5.1 Symbolic Matching

Barrow and Popplestone (1971) describe the adjacency relation of picture regions by a region adjacency graph (RAG). This graph is necessarily planar. A description of a model scene is also represented in terms of a RAG. Recognition of elements in the picture is performed by matching a region with a model component. Both region adjacency graphs are augmented by edges indicating relations between regions such as relative size, position (above-below, left-right), shape and convexity. The match is incrementally increased, one region at a time. The search space can be represented as a tree; nodes represent matchings and descendants of a node represent developments of the matching at the node by adding another region. The search space is probed for a solution using best-first search.

Barrow and Burstall (1976) use maximal matching of graphs for matching relational structures, such as image and scene descriptions represented as graphs. A graph is defined as a set N of nodes, and a relation R , a subset of $N \times N$. A matching from G_1 to G_2 is a subset S of $N_1 \times N_2$ which preserves the relations in each graph; for all pairs (a,A) and (b,B) in S , a is connected to b in G_1 if and only if A is connected to B in G_2 . A matching is maximal when no other matching has higher cardinality. Under this definition, a node in G_1 may be matched to more than one node in G_2 and vice versa. Two

pairs are 'compatible' if the pairs are in S . A graph is derived as follows: the elements of the graph (the set X) are elements of S and the relation (H) between elements of the graph is that of compatibility between pairs. Given a graph so constructed, a maximal matching of the original graphs can be obtained by finding a maximal clique (complete subgraph) of (X, H) . It is clear that in such a clique all pairs are compatible, by definition, and that its order is maximal. If the restriction that $a = b$ iff $A = B$ is added to the definition of compatibility, the correspondence generated by the maximal matching is one-to-one. To our dismay, however, this merely reduces a difficult problem to an NP-complete problem. However, the best clique-finding algorithms seem to perform efficiently in most cases.

Tanimoto (1976) offers an excellent discussion of the motivation for using graph matching and develops an algorithm for enumerating all maximal matchings of two graphs. A matching assigns labels to regions in the segmentation of an image. The labels form one set of nodes in a bipartite graph, and the regions the other set. Edges represent the compatibility of descriptions of a region with a label and hence restricted to a yes-no decision. A maximal matching of the bipartite graph so formed is the maximal set of edges from one set of nodes to the other, where a node occurs at most in one edge. Tanimoto notes that methods for constructing such bipartite graphs are 'neither usually obvious nor necessarily possible'. One approach is to allow edges which satisfy many constraints such as degree restrictions. Usually these are determined by 'local constraints, that is, those which only require examining the neighbors of a node. A maximal matching can be generated in $O(e * \sqrt{n})$ time, where e is the number of edges in the graph, and n the number of nodes. An algorithm is presented by

Tanimoto which can list the set of maximal matchings of the graph. A note of caution: the number of maximal matchings is potentially exponential in n . A recent paper by Itai, Rodeh and Tanimoto (1978) also characterizes the cases when matching problems with restrictions are NP-complete (Aho et al., 1974), and provides a discussion of the applicability of graph-matching vis-a-vis constraint propagation.

Maximal matching techniques are appealing because once the compatibility relation is constructed, generation of matchings is efficient. However, the effectiveness of maximal matching methods depends upon the extent to which the compatibility relation can constrain matchings to appropriate ones. If the compatibility relation is too general, many matchings will be generated which are incorrect. Computing compatibility becomes expensive when the interrelation of features extends more than just to local features. In the registration problem, in particular, the solution must satisfy global constraints, while compatibility testing must be rather local to allow maximal matching to be a successful alternative method. In addition, it is not clear that the relations among features can easily be characterized by descriptions such as 'left-right' or 'near-far'. Rather, metric relations such as angle and distance are appropriate in this domain, especially as they are precisely known for the DTM. Using metric relations becomes more important when matchings between intrinsic aspects of features, such as shape, length or position, are less reliable.

2.5.2 Geometrically Constrained Matching

Fischler and Eschinger (1973) detail a method for matching a reference image in raster format to a reference image which is described by a graph

composed of components (coherent pieces of the model). For each of these components, a local evaluation array (LEA) is computed. The LEA measures the goodness of fit of each component of the model at each point in the image, rather like the goodness of fit of a template at all points in the image. The model components are assumed to be joined together by springs. The cost of a matching is the amount of tension in the springs joining the components. This method is compatible with a 'rubber-sheeting' transformation of the image, in which direction is not globally preserved and scaling can vary across the image.

Dynamic programming is used to solve the matching problem. Using the LEA, the cost of orienting the constituent component subassemblies can be computed, recorded in tabular form, and used to find a global minimum cost matching. The cost of this method increases exponentially as the degree of interconnection of the components rises. As an alternative, the authors suggested an incremental method, very similar to Barrow and Popplestone's technique. The work is of note in two respects: first, it constructs a full transformation from one image to the other, and second, it uses geometric constraints as well as semantic constraints in the matching. The registration method presented in this thesis uses an incremental method.

Bajcsy and Chance (1975) studied the problem of establishing the correspondence between images of brain slices before and after chemical or physical operations in which there is appreciable shape distortion of the brain. The images are processed to extract the veins in the images. The nodes (vein junctions) are ordered by degree. A matching is generated by comparing the degree of junctions from the two images. Because it is likely that an edge in one graph will show up in another, this seems an appropriate strategy. The graphs are not totally matched in this process; rather, the

initial matching is used as a 'seed' to a registration procedure which perturbs the initial mapping slightly in order to reach an optimal match. The authors state that without such an initial match, the hill-climbing method of the registration is not sufficiently constrained and might 'walk off the edge of the image'.

Work at SRI (Bolles et al., 1979) models the transformation from the test image to the reference as a function of camera parameters, such as focal length, position, yaw, pitch and roll. The reference is a 'database' of highways and features of highways. An essential part of the SRI method is that there is a good 'a priori' estimate of the camera parameters and of the errors in these parameters. These estimates are used to predict the location and extent of the region in the image which is to be searched for an element from the reference image. The predicted search region for an element is termed its 'uncertainty region'. Once an element is located within its search region, the search regions for other elements are constrained in location and size. The pairing of reference element and image element provides new information which is used to improve the camera parameters and reduce the errors. Both linear and point features are hand-selected from the reference image for registration. Because highway structures, such as the boundaries of lanes, are locally very similar, it is possible to mistake a feature for one offset from the proper match. To prevent this situation, the system identifies features nearby which can be used to verify a match, and searches for them in the image. For example, highways are composed of several parallel lanes; in detection of a highway the system searches for locally offset lanes to confirm the matching of others. This notion is termed 'local support' for a feature match. The matching of elements provides information for the correspondence refinement

process which solves the nonlinear camera parameter estimation problem.

Bolles (1979) describes a further use of the maximal clique method for matching features in an image with a model. Nodes in the feature graph represent labellings of nodes. Arcs represent compatibility relations between the labellings of features. These relations are based on distance and orientation measures computed in the image and compared with model descriptions. A maximal clique in this graph represents a maximal match of the features of the image with the labels in the model. As with all maximal matching methods, there are difficulties with the combinatorial behavior of the problem and the inordinate size of the graphs; generating graphs for reasonable problems in itself is time-consuming. Bolles (p. 144) suggests several ways of improving the method:

- 1) Restrict the model to key features
- 2) Use geometric limits with respect to some feature to exclude unnecessary features.
- 3) Iteratively apply the maximal clique method to refine the assignment.

With respect to the last point, Bolles further states "the benefit of this approach is derived from the fact that the structural constraints are applied sequentially instead of all at once" (p.145).

In general, methods for maximal matching suffer from the difficulties encountered in Bolles's method. Explicit construction of the relations which may hold between elements of the model and the features of the image is itself an expensive task. The registration method presented here follows the spirit of Bolles's work. The method depends upon an analysis of the

model to find promising features of the model. These are found in the image. Nearby features of the model are used to confirm the initial matching. Then additional features are selected for matching, from the restricted set constrained by the previous matches. Local support for a feature acts as a breadth-first look-ahead to select promising matches, following which a depth-first search is conducted for further matches. In addition, once an estimate for the matching has been constructed, it is locally adjusted to improve the registration.

These facets of the method anticipate the suggestions of Bolles. The illumination conditions at the time of image acquisition are used to determine the features. Key features are selected based on the structural complexity of their components. The geometric constraints of the transform derivation guide its development. Lastly, the inspection and rejection of choices at early stages of the search deliver the benefits of sequential exploration of the possibilities.

3. The Registration Method

The registration method proceeds in several stages. The data consist of the raw Landsat image and a digital terrain model (DTM) for the area imaged. The time at which the image was acquired is known. In the first stage, features of the digital terrain model and the Landsat image are derived. The second stage considers matchings of three features from both the DTM and the image. Each match determines the parameters of an affine transformation. When one of the derived transforms predicts other ridge-to-Landsat feature pairings, with a sufficiently small total residual error, the transformation is accepted. Otherwise, registration fails.

3.1 The First Stage: Feature Extraction

3.1.1 Extracting Features from Digital Terrain Models

Since the sun's position corresponding to the Landsat image is known, it is possible to determine the location of convex breaks of slope which will appear in the image with strong brightness discontinuities, as follows: The slope of each surface facet is derived and the brightness of the facet determined using the reflectance function. For every location in the TIN where surface slope changes (represented by the junction of facets), the brightness of the surface facets adjoining is computed. The difference between these values indicates the relative magnitude of the brightness discontinuity to be expected at that position. In testing the method, only those edges are selected which are bounded, on one side, by a self-shadowed

facet (one which receives no direct illumination), and, on the other side, by a facet whose predicted brightness is relatively high (figure 10) . This restricts the features to a small subset of all edges which will generate brightness discontinuities. Many ridges will not be self-shadowed on one side, but will nevertheless appear as sharp discontinuities in the image. However, areas in self-shadow, if present, will be very dark in the image, and their appearance will be less sensitive to ground cover variation. Self-shadowed ridges tend to lie perpendicular to the azimuthal direction of the sun. This leads to strong constraint along the direction parallel to the azimuthal direction of the sun, but little constraint in the orthogonal direction. Pairings of features at junctions or endpoints of features provide the needed constraint in the orthogonal direction. Edges satisfying this criterion are merged into curves when their endpoints are adjacent and merging them does not cause the resulting curve to loop back upon itself. Only those curves are output which represent a strong brightness discontinuity. These curves will be termed "ridges" in the following discussion of the registration method, while noting that the curves can be generated both by terrain ridges as well as other convex breaks of slope. The features extracted from the DTM are shown in figure 4.

3.1.2 Extracting Features from the Landsat Image

In the Landsat image, some surface slope breaks appear as boundaries at transitions between bright and dark regions. Desirable boundaries are those formed by convex breaks of slope oriented perpendicular to the azimuthal direction of the sun's illumination. Shadow boundaries also appear as transitions between bright and dark regions, but, since the direction of the

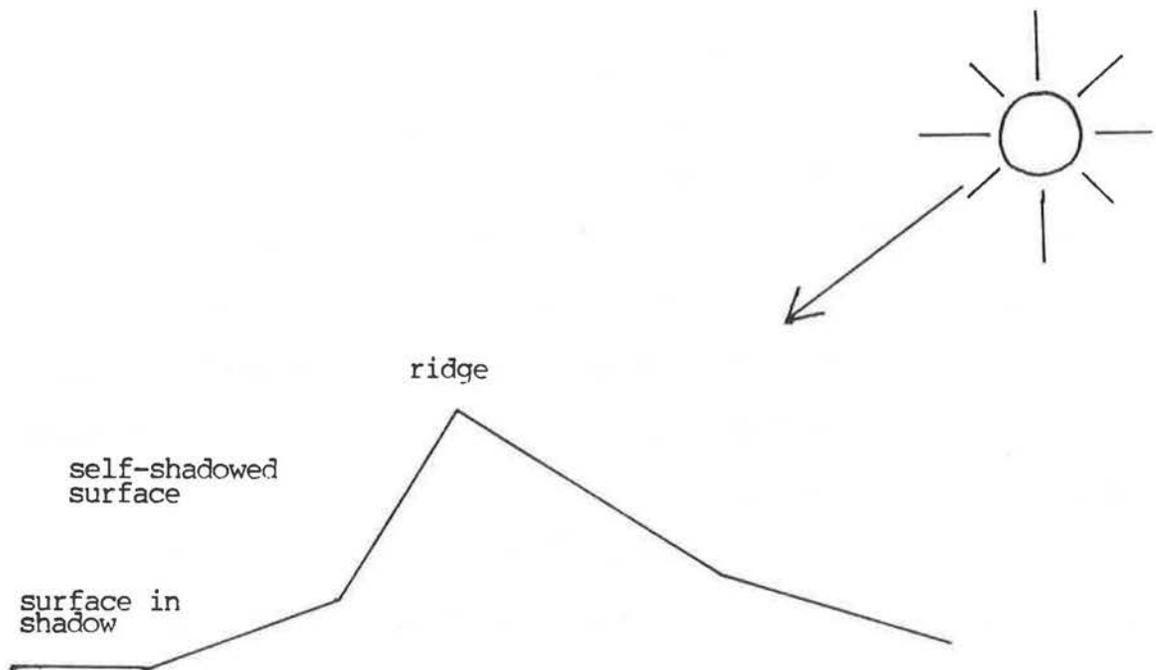


Figure 10
Self-shadowed surfaces.

incident illumination is known, they can be distinguished from the transition features formed by ridges. Shadows are dark on the side of the edge nearer the light source. Figure 11 shows the Landsat image subsection used for the registration tests.

3.1.3 Selecting Feature Points

The Landsat image is correlated with an edge detector composed of two orthogonal components. The 5x5 Sobel operator (Iannino and Shapiro, 1979) was used because it had been reported to yield acceptable results, in the literature. The ratio of the outputs of the components of the operator provides an estimate of the direction of the boundary element passing through the pixels tested.

The edge detector gives high values not only at discontinuities, but also at pixels offset from the discontinuities. This produces secondary lines, called echos, lying parallel to the original (figure 12). In order to eliminate these as early as possible a scheme of Nevatia and Babu (1979) is used. An edge element is judged to exist at a pixel if :

- a) the magnitude of the filter output is above a threshold
- b) its magnitude is higher than that of its two neighbors in the direction normal to the estimated edge direction, and
- c) the edge directions of these neighboring pixels are within 45 degrees of the direction at the central pixel.

If any of these conditions do not hold then no edge element is judged to exist. The effect of this process is to suppress the echo elements at an early stage, eliminating the need for later curve thinning procedures (figure 13).



Figure 11

100 X 100 pixel subsection of Landsat image (band 7) from September 14, 1976, frame ID 11514-17153. Photographed from the screen of the COMTAL Vision 1.



Figure 12

Result of applying the edge detector. *'s represent edge elements whose orientation is consistent with the sun's position.

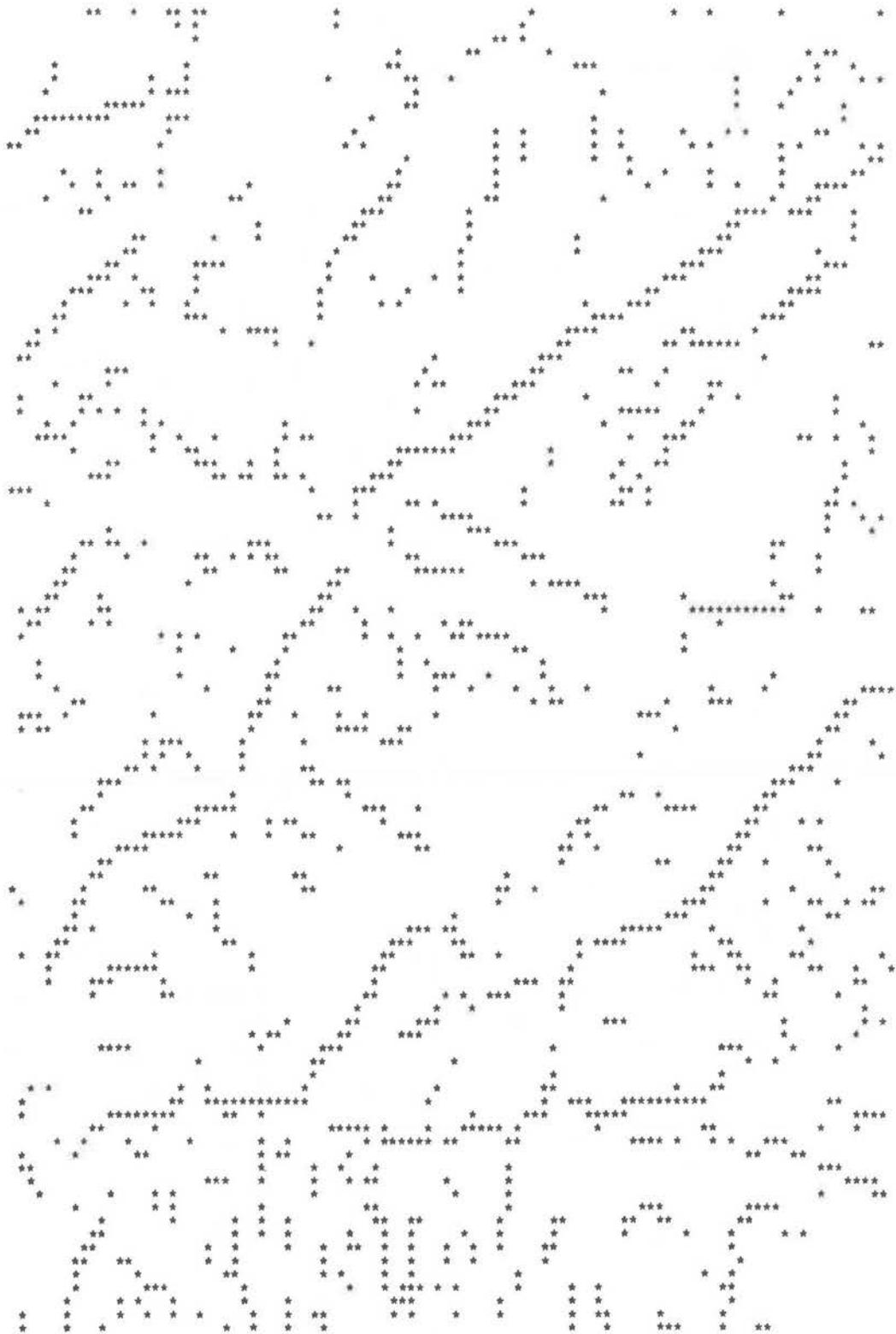


Figure 13

Edge detector output after echo suppression.

3.1.4 Line Growing

The output of the Sobel operator is used to construct the linear features, following the method of (Bajcsy and Tavakoli, 1976). The process is divided into two steps. First, pixels are connected into curves represented as sets of pixels. Next, a piecewise linear approximation is derived for these curves, and curves are merged when possible. In the discussion which follows, the terms "line" and "curve" will be used interchangeably to refer to a string of points connected by straight line segments.

In the first stage, a histogram of the values of the filter output is derived. This density histogram is used to direct the process so that lines are 'grown' from those points which had the highest output from the filtering step. A cumulative distribution function is derived from the density histogram. At each step in the line growing process, the threshold is relaxed so that five percent more pixels are above it. Initially the threshold is set at the 95 percent level.

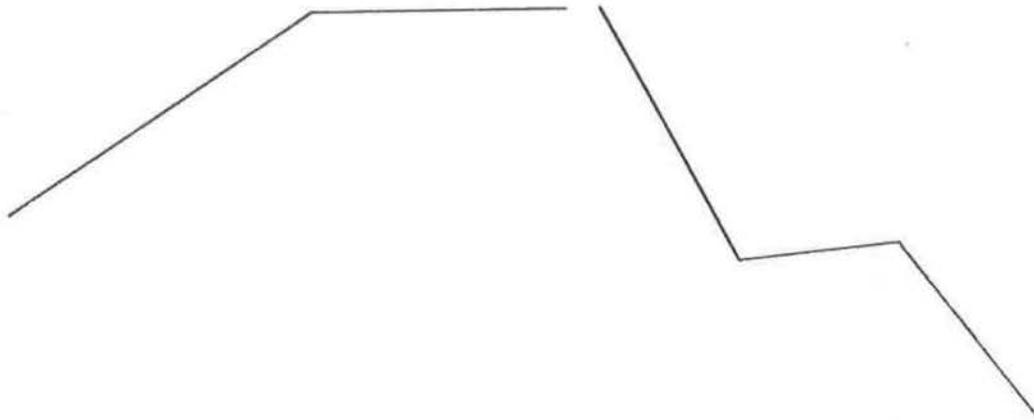
At each stage in the line construction process, the threshold is set at the proper level and all points in the image above the threshold and not already in a line are processed. The threshold is then lowered a level, and the process repeated, until the minimum level is reached. Lines are constructed incrementally in this first stage; at first a line consists of a single point. When an adjacent point lies above the current threshold, and cannot be joined to any existing line, it is joined to the single point and forms a two-point line. To ensure that the lines found have less than a certain maximum curvature, points are added to an existing line only if they are adjacent to the endpoints of the line and the segment connecting the new

point to the endpoint lies within 45 degrees of the direction of the nearest segment in the line.

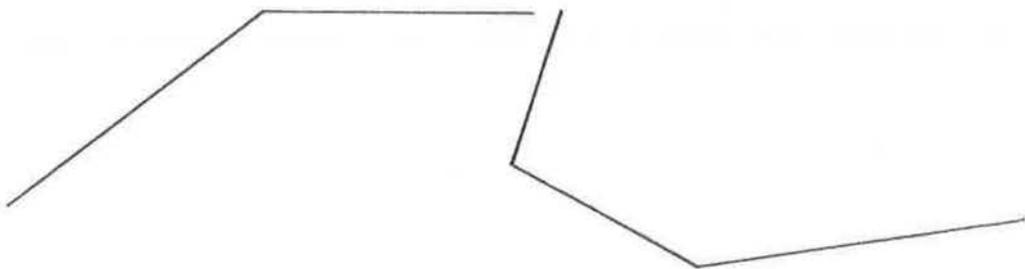
The result of the first stage is a set of lines each consisting of a set of connected pixels. Figure 14 shows these lines; at each pixel the last digit of the curve to which that pixel belongs is printed. In the next stage, these lines are merged into larger connected lines when two conditions hold: first, the lines are adjacent at their endpoints, and, second, the segment directions are compatible, that is, joining the two lines at their endpoints does not cause the resulting curve to loop back upon itself. The segments are compatible in direction if the dot product of the segments, considered as vectors originating at the common endpoint, is less than or equal to zero (figure 15). Figure 16 shows the curves after merging. To aid in curve merging, a piecewise linear approximation is derived for each of the curves.

3.1.5 Approximations to Lines

A piecewise linear approximation to a digital line (Ramer, 1972; Pavlidis, 1977) approximates a line to a given precision by a set of linear segments connecting points on the line. In its construction, the first and last points in the line are connected by a straight line segment. The extreme points lying farthest in perpendicular distance from the line segment are determined, B and D in the example shown in figure 17. The extreme points are included in the approximation if their distances from the segment are above the specified threshold, which will be termed the "detail level" of the line. The line is then subdivided into the three sets of points to the left, between and right of the selected points. The three



compatible segments



incompatible segments

Figure 15

Two lines can be merged if the dot product of the segments at the join point is negative.

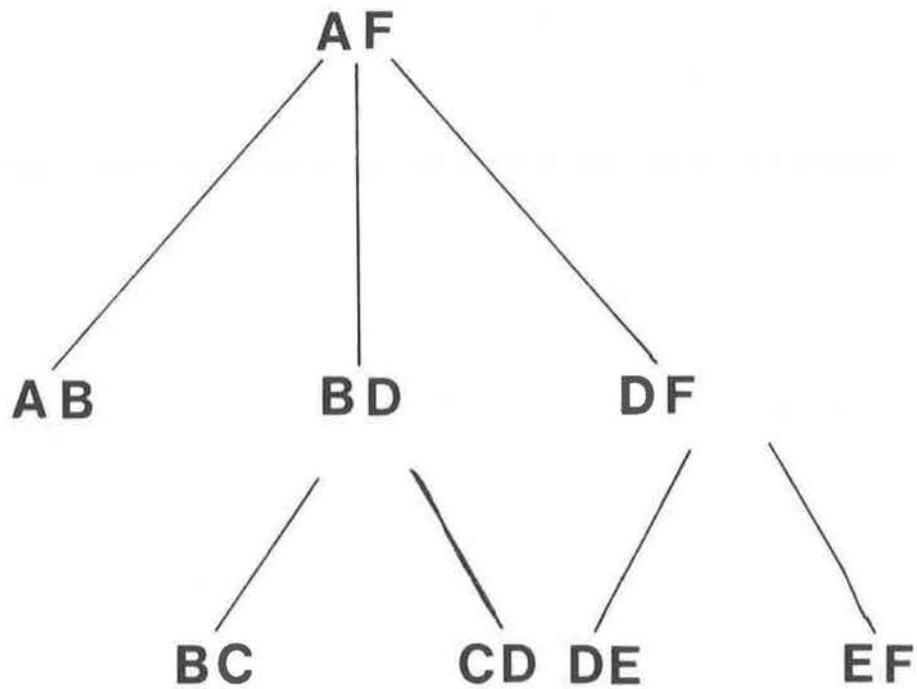
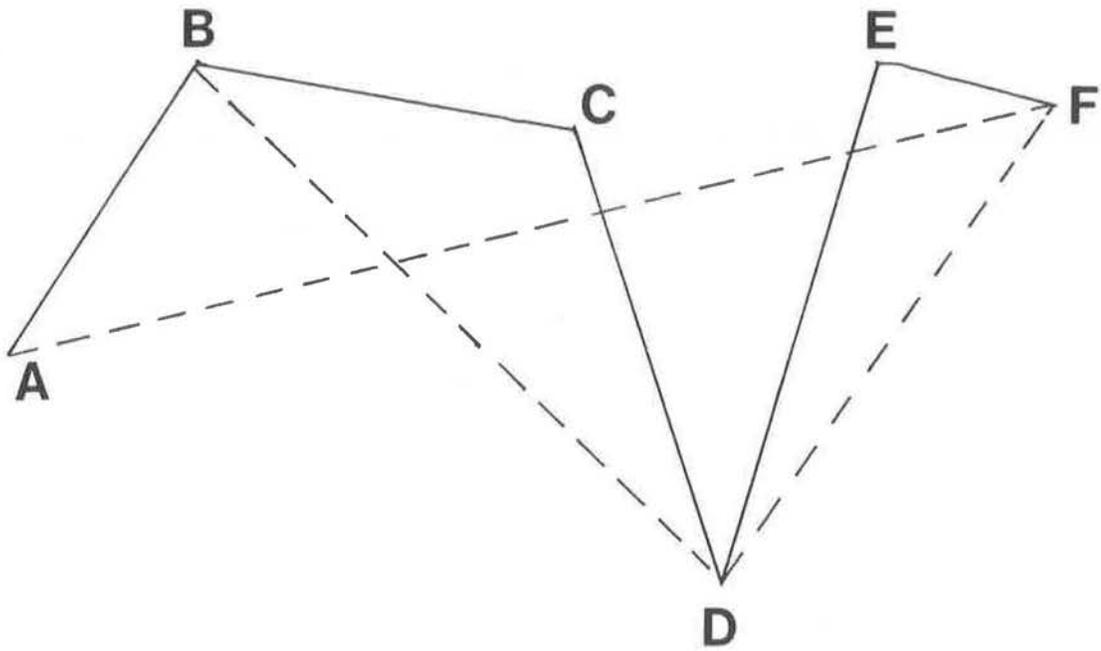


Figure 17

A curve and the tree representation of its generalized form.

subsets of the line are processed recursively in a similar fashion. In figure 17, these subsets are AB, BCD and DEF. If the point farthest from the segment in a particular subset is within the threshold distance, then processing of that subset of the line is stopped, and only the endpoints of the line segment retained. The process of finding such an approximation is termed 'generalization'. The tree-like structure derived in this fashion is useful in cartographic computations (Ballard,1979). A tree for a generalized form of a curve is also shown in figure 17.

By varying the detail level used in computing the curve approximation, a family of approximations is generated (figure 18) . Alternatively, the perpendicular distance of a point in the curve from the next highest level segment can be recorded in the approximation. The distance so found is a measure of the significance of that point in the approximation of the curve.

By constructing two lines, parallel to and offset from a given segment of a curve, and at a given perpendicular distance, a region in the plane is described which is called the 'band' of that segment of the curve. When the segment connects the endpoints of the curve, the region is the band of the curve (figure 19). The detail at which the curve is examined can be varied by altering the perpendicular distance at which the band is constructed. The band of a curve will be used to determine whether a curve overlaps another, in curve comparison and in testing of the registration.

Since most of the lines in the Landsat image contain many colinear points, the line approximations contain significantly fewer points than the the original lines represented as connected sets of pixels. Using the approximations, directional decisions involving the orientation of line segments are less affected by perturbations at the end of curves caused by quantization. The output of the feature-detector is the set of lines in

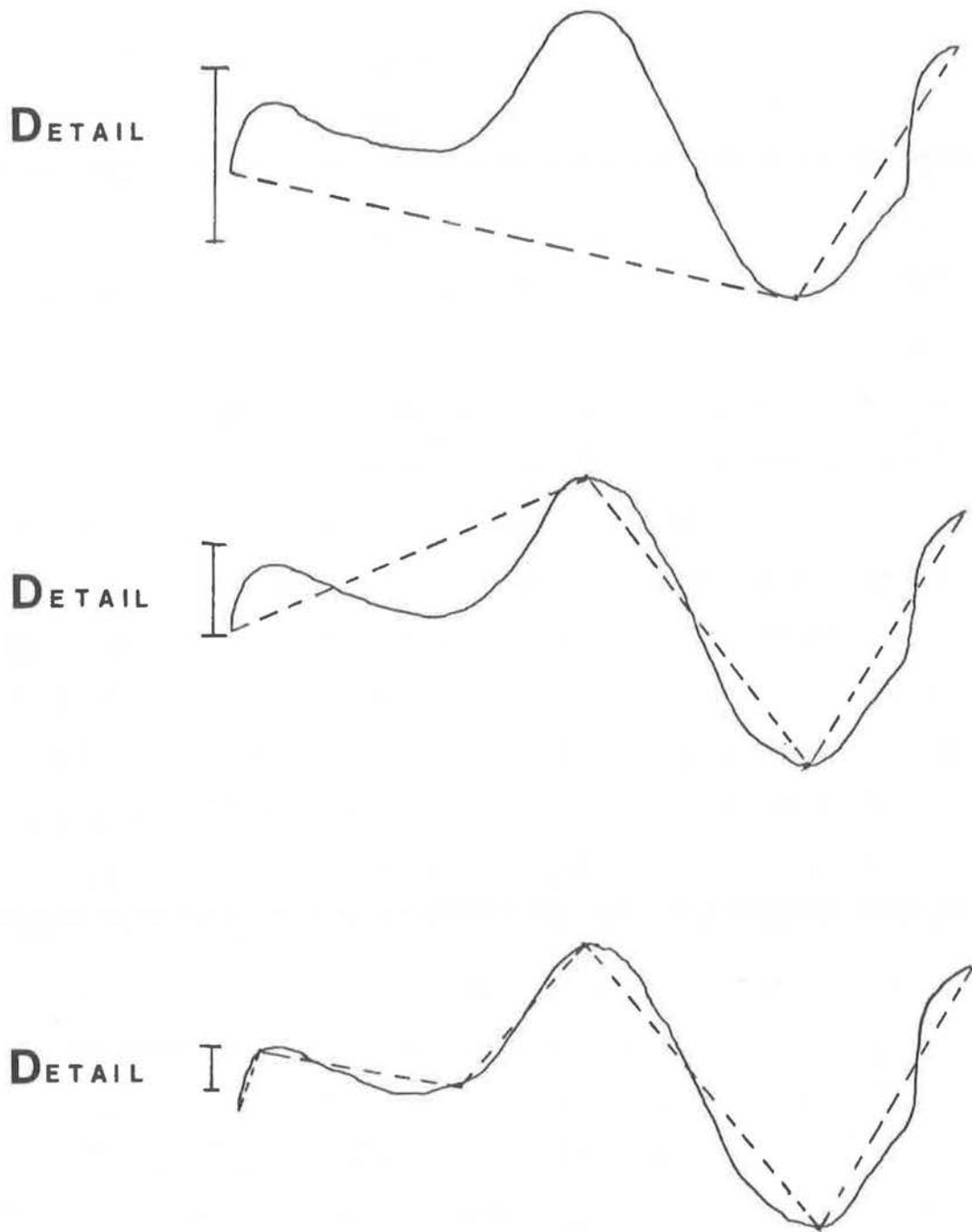


Figure 18

Varying the detail level in generalization. The dotted lines represent the generalized form of the curve at each level.

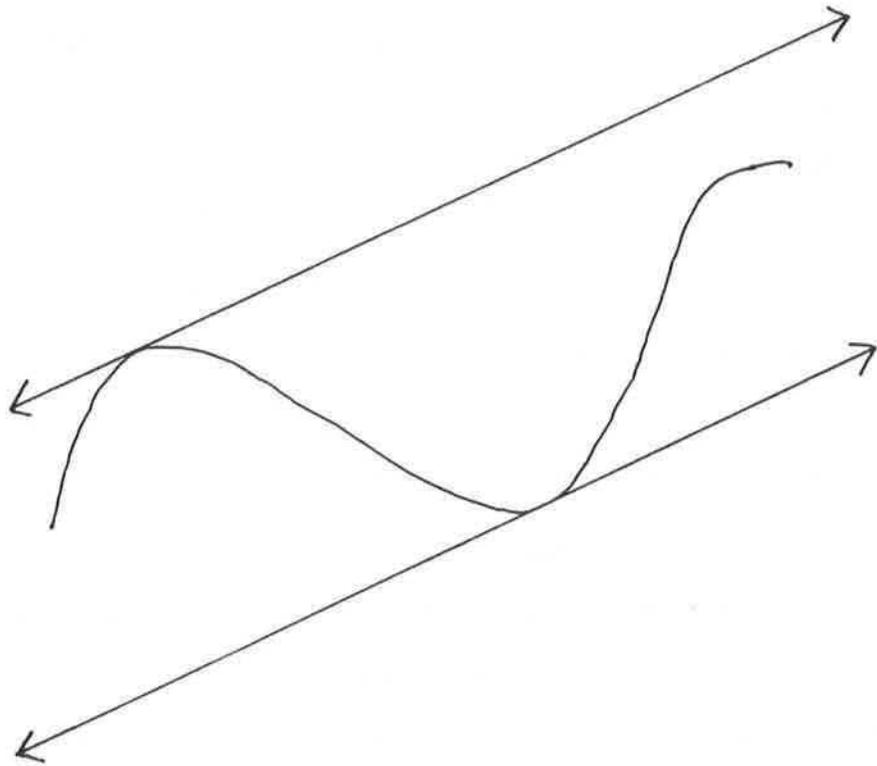


Figure 19
The band of a curve.

generalized form, which are longer than a specified minimum length (figure 20). These curvilinear features derived from a Landsat image will be termed "l-edges".

Since the ridges are derived from the TIN representation, they are represented as a string of contiguous line segments. It is natural to convert this form to the piecewise linear approximations used for curves found in the Landsat image. If a DTM feature is represented by a single straight line segment, then it is 'simple'. Any curve whose representation includes interior points is said to be 'structured'. The ridges and l-edges are input to the second stage.

3.2 The Second Stage: Matching

The basic approach in the second stage is to locate the known features, the ridges, in the image. Features are located sequentially, and the location of a feature in an image will constrain search for other features. Ridges will be located by structurally matching ridges with l-edges. A ridge matched to an l-edge is a "pairing". A pairing locates a ridge in the image and provides a pair of matched points which are used as ground control points (section 3.4). Because of the sequential nature of the algorithm, it is useful to introduce an ordering.

The ridges are considered in the order of their structural complexity. This is because it is assumed that the more strongly an element differs from a straight line the less likely it is to be matched incorrectly. The goal of the matching process is to pair a sufficient number of ridges and l-edges to compute transform parameters. Features in the DTM are ordered as follows:

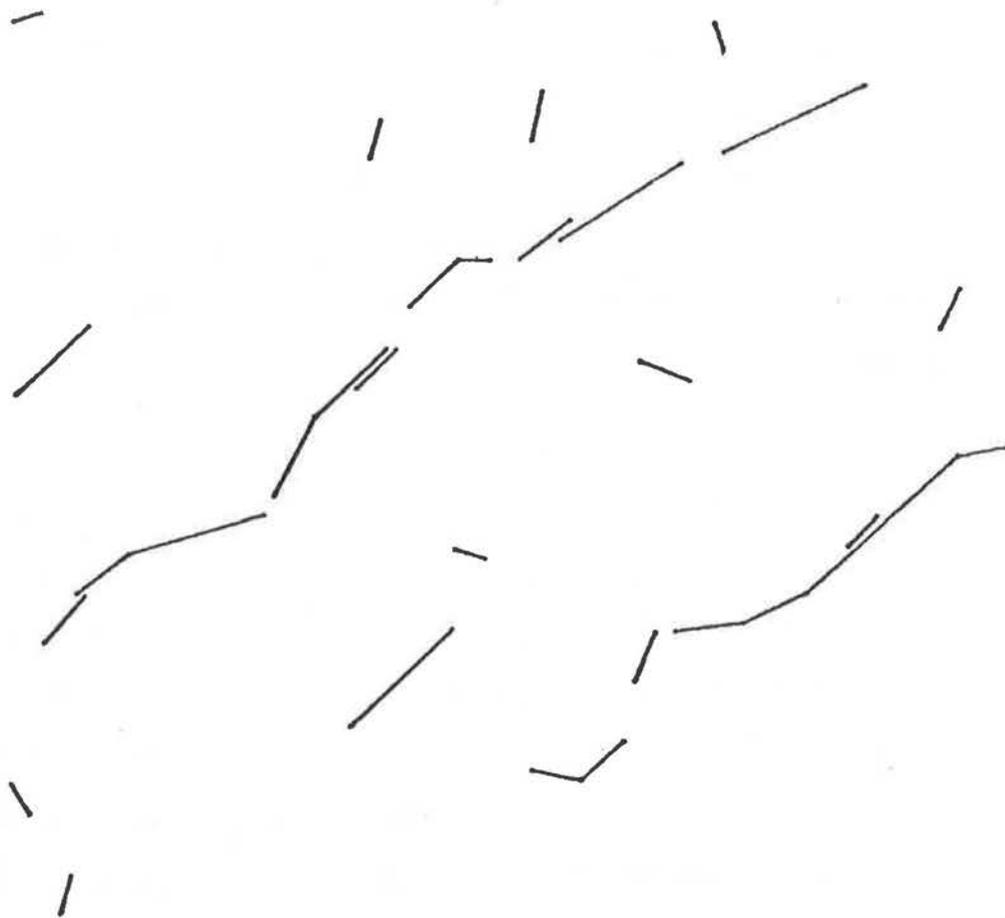


Figure 20

Landsat image features for subsection in figures 11-16.

1. All structured ridges are considered before simple ridges (figure 21).
2. The strength of a ridge is the product of its length and the estimated brightness discontinuity across it. Structured ridges and simple ridges are each ordered according to the ridge strength.

The relation of structured feature to structured feature is potentially a many-to-many relation. A portion of a ridge may be represented by two or more line segments while the corresponding 1-edge is represented as a single line segment, or vice versa. To consider all possible relations between two structured curves means examining the relations between all powersets of both. Representing and manipulating such relations significantly complicates curve matching. Consequently, all structured features in the Landsat image are broken down into simple elements, represented by single line segments.

3.3 The Affine Transformation

Horn and Woodham (1979) demonstrate that, if small, second-order effects are ignored, an affine transformation is sufficient to register small subsections of a Landsat image to a plane tangent to the earth's surface. The parameters of this transformation can be expressed in terms of the parameters of the satellite's orbit and other fixed quantities. An affine transform has 6 degrees of freedom and can be written as:

$$x' = a x + b y + c$$

$$y' = d x + e y + f$$

where x, y are image coordinates and x', y' are DTM coordinates. A subset of

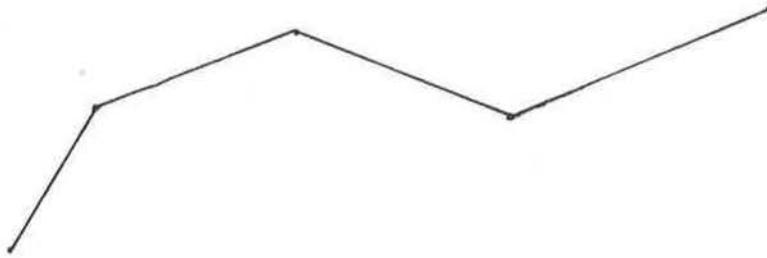
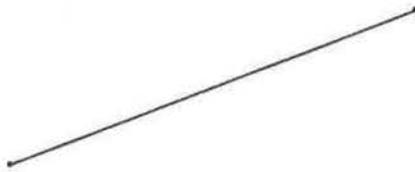
**A****B**

Figure 21
Structured (A) and simple (B) ridges.

affine transformations can be expressed as a composition of rotation, translation and scaling of the coordinate axes. However, the fully general affine transform does not admit this simple decomposition.

Finding the transformation parameters requires at least three matched points. These can be determined manually by identifying ground control points in both the DIM and the image. Let the coordinates of the image points be (x_1, y_1) , (x_2, y_2) and (x_3, y_3) and the DIM coordinates be (x_1', y_1') , (x_2', y_2') and (x_3', y_3') . Then

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} a & d \\ b & e \\ c & f \end{vmatrix} = \begin{vmatrix} x_1' & y_1' \\ x_2' & y_2' \\ x_3' & y_3' \end{vmatrix}$$

so

$$\begin{vmatrix} a & d \\ b & e \\ c & f \end{vmatrix} = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}^{-1} \cdot \begin{vmatrix} x_1' & y_1' \\ x_2' & y_2' \\ x_3' & y_3' \end{vmatrix}$$

If more than three GCP's are supplied, a least-squares estimate of the transform can be computed. If

$$M = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_n & y_n & 1 \end{vmatrix}$$

Then the least-squares estimate with equal weighting to all points for the transform parameters is:

$$\begin{vmatrix} a & d \\ b & e \\ c & f \end{vmatrix} = (M \cdot M)^{-1} \cdot M^T \begin{vmatrix} x_1' & y_1' \\ x_2' & y_2' \\ \cdot & \cdot \\ \cdot & \cdot \\ x_n' & y_n' \end{vmatrix}$$

Once three feature pairings have been established, the affine transform can be estimated. A set of three pairings will be termed a 'matching' and a set of more than three pairings an 'extended matching'. Exhaustive

examination of all matchings is too expensive. The number of triples grows as n^3 . Hence the number of matchings grows as n^3 , where each of the feature sets is of cardinality n . Knowledge of the constraints imposed on the problem is used to limit the search space. An estimate of the affine transform is derived (following Horn and Woodham, 1979) from orbital parameters included in the image annotation, and other fixed parameters of the scanner. Section 4.4 presents the analytic expressions for the parameters of the affine transform and describes the parameters of the satellite orbit. Because this estimate of the transform is available, it can be used to eliminate the generation of some incorrect matchings.

3.4 Construction of a Pairing

Initially, the location of an image feature in the DIM is known only to within 10 kilometers. This delimits a search region for a feature. The system begins by selecting a ridge and finding the l-edges in its search region. Each l-edge is transformed according to the 'a priori' transform estimate, and compared with the candidate ridge. The basis of the comparison is the representation of a curve as a piecewise linear curve. The construction of this representation, as described in section 3.1.5, involves determination of the direction of the curve, which is the direction of the vector connecting its endpoints. The band about the curve is a region in the plane bounded by two lines parallel to the direction vector and offset from it by a fixed amount. The width of the band is the distance between the parallel lines. Assessment of a pairing of features proceeds as follows:

a) If the ridge is simple, the transformed l-edge is translated so that one of its endpoints coincides with an endpoint of the ridge segment. The perpendicular distance D from the ridge to the other endpoint of the l-edge is computed. There are three cases:

- 1) D is less than or equal to the band width. The positions at which the l-edge can be matched include all points in the ridge segment. In practice, three positions are used (figure 22): at either endpoint, or at the centerpoints, the averages of the endpoints of the segments.
- 2) D is less than twice the band width. The l-edge is constrained to match its centerpoint to the centerpoint of the ridge (figure 23).
- 3) Otherwise, the pairing is rejected.

In cases 1 and 2, the measure of goodness of the match is the cosine of the angle between the two curves. In case 3, the measure of the match is arbitrarily set to 0. The translation vector for the pairing is constructed from the difference of the matched points.

b) If the ridge feature is structured, then the l-edge is compared with each of the line segments in the ridge as above. The result of the comparison is a list of matchings of the l-edge with each of the segments in the ridge.

If structured l-edges were used in pairing development, endpoint matches could be confirmed on the basis of the relative orientation of the segments meeting at that endpoint. Local context is provided by the adjacency of segments. In the present system, local support for a pairing serves this purpose.

The result of a match determines a point-to-point correspondence which is used in estimating the affine transform. The list of pairings of ridge and l-edges, sorted by value, is associated with the ridge. Pairings whose value is too small are not allowed to enter into the construction of a matching. This acts to eliminate pairings which can only arise from combinations of rotation, scaling and skewing inconsistent with the known imaging geometry.

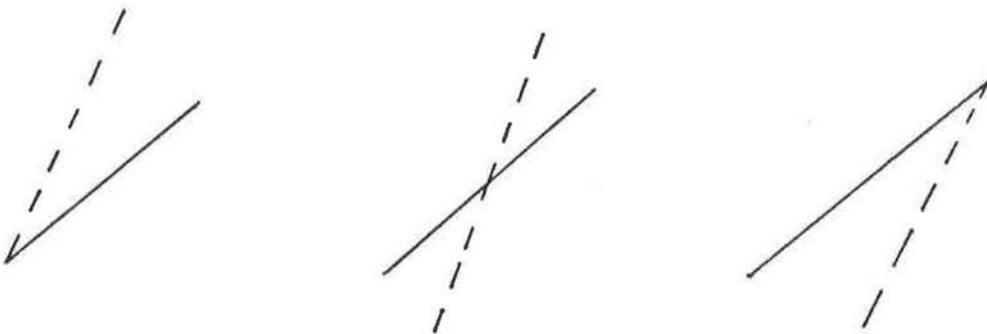
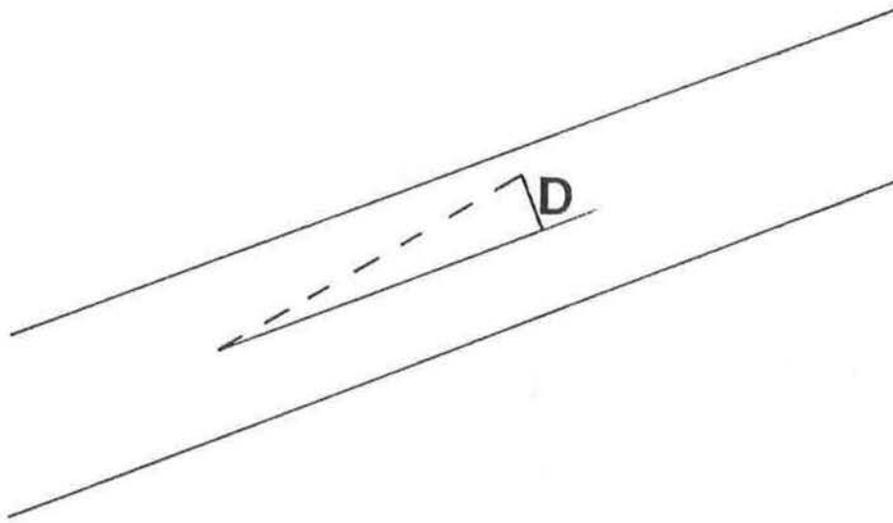


Figure 22

Close fit in a pairing, with the three matching positions.

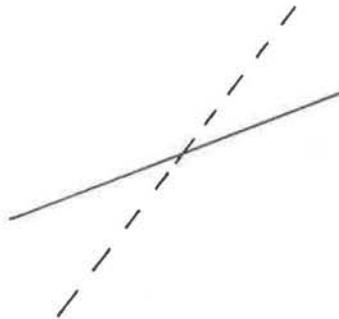
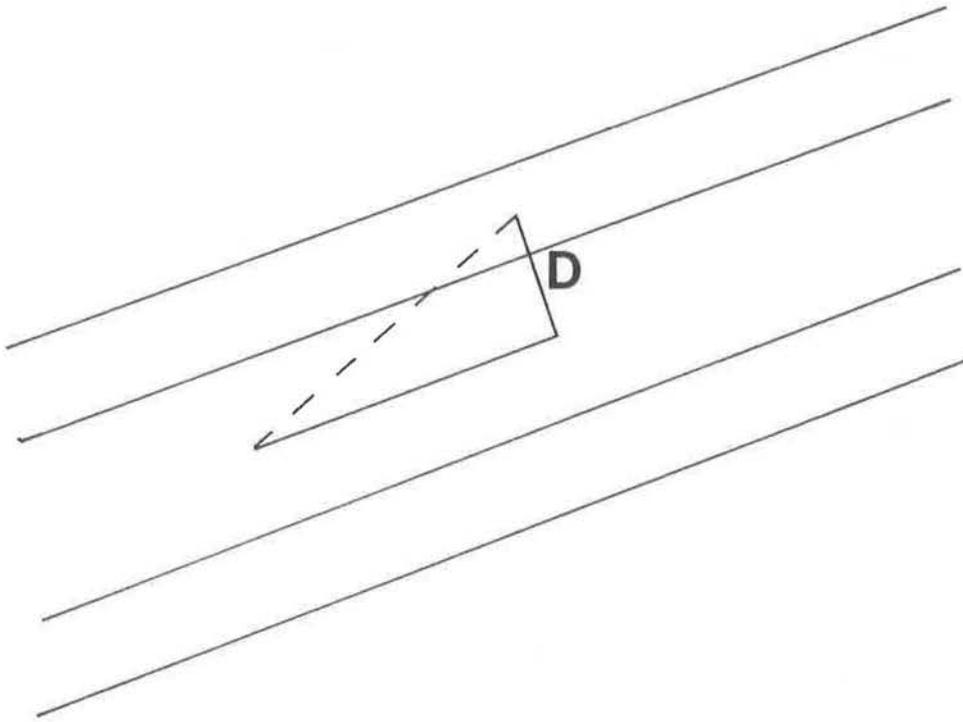


Figure 23

Loose fit in a pairing. Matching is permitted only at centerpoints, the average of segment endpoints.

3.5 Support for a Pairing

A pairing specifies a translation vector. The residual error in position of an image feature after transformation can be modeled as a translation. Hence the translation needed to bring a transformed l-edge into correspondence with a ridge is used to guide the development of matchings. Each subsequent pairing must be consistent with the previous pairings, that is, the translation required to construct the pairing must be similar to those previous. Similarity between translation vectors is measured by treating each translation as a point in the plane and finding the distance between the points. If the distance is too large, the translations are incompatible. Otherwise, the translations are considered consistent. Testing translation consistency eliminates the generation of many incorrect matchings.

Experimentation with the feature sets has shown that translation alone is not a sufficient constraint. For example, the location and orientation of ridges is often controlled by the underlying geological structure of the region. Ridges are often parallel or nearly so, and the spacing between ridges can be very regular. Hence, a pairing of a ridge to an image feature may be correct in orientation, but offset by the inter-ridge spacing.

Consider the following example: The problem is to register the image resembling the numeral 4 represented in figure 24 to a model of the numeral. Image segments are referred to by their endpoints, ABCDEF, and the model segments by the same letters, with quotes, A'B'C'D'E'F'. In this example, if segment A-C is compared with B-D, the match will be high in value, assuming an identity a priori transformation. Orientation and length of the

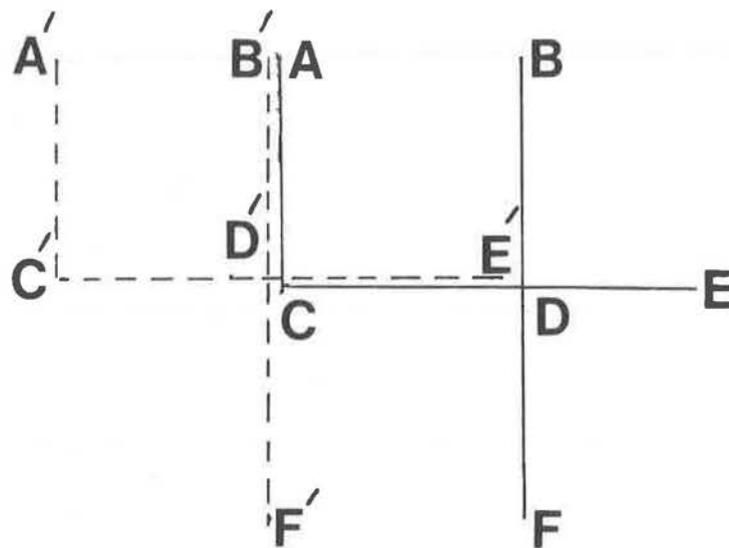
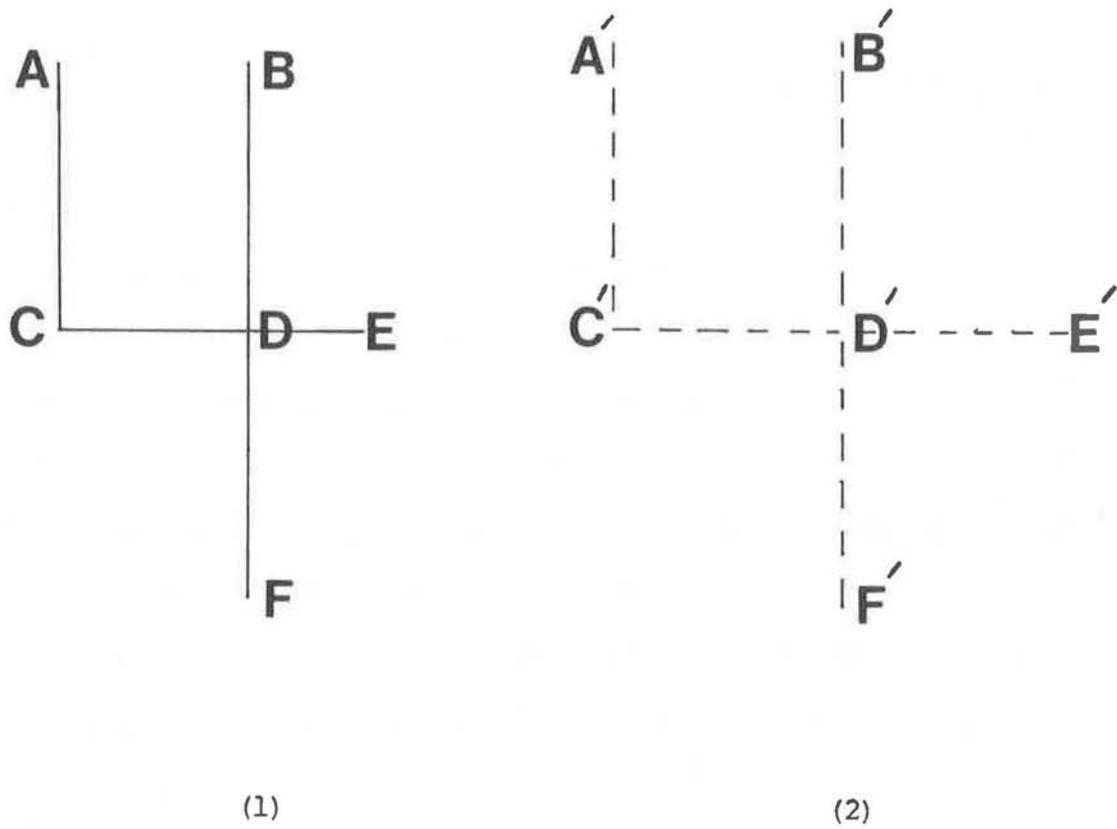


Figure 24

An image (1), its model (2), and a mis-match to demonstrate support for a matching.

segment do little to distinguish A-C and B-D. To eliminate incorrect matchings cause by this phenomenon, the local spatial structure of both the ridges and the l-edges must be used to guide the matching. In terms of the example, note that when A-C is registered to B'-D', C-D overlaps D'-E', partially confirming the match with B'-D'. But when A-C is paired with A'-C', all segments in the image will participate in a pairing consistent with that involving A-C.

When an initial pairing of features is made, nearby ridges are examined and a tally is kept of the number of nearby ridges which can be paired with l-edges in a matching consistent with that under construction. Consistency here is again measured by comparing the translations necessary to bring a feature into alignment, under the a priori transform, with a given l-edge. If a structured ridge is being considered, the tally is formed by counting the number of segments in the ridge which can be paired with l-edges under mutually compatible translations. Developing a pairing of the elements of a structured ridge with several simple l-edges compensates somewhat for the decomposition of structured l-edges into separate simple features. The simple l-edges can be paired with the elements of a structured ridge as they would have been had they still been joined in a structured l-edge. The pairings of ridge and l-edges are ordered by the number of supporting pairings, (i.e., by the extent to which they can be locally extended). This strategy can be understood as a generalization of the scheme of determining local support for linear features employed in the SRI system (Bolles et al., 1979).

At this point, a feature (whole or part) is matched to a l-edge, and a set of compatible pairings has been generated. Each of the elements in this set is in turn selected as the second pairing for the matching. By

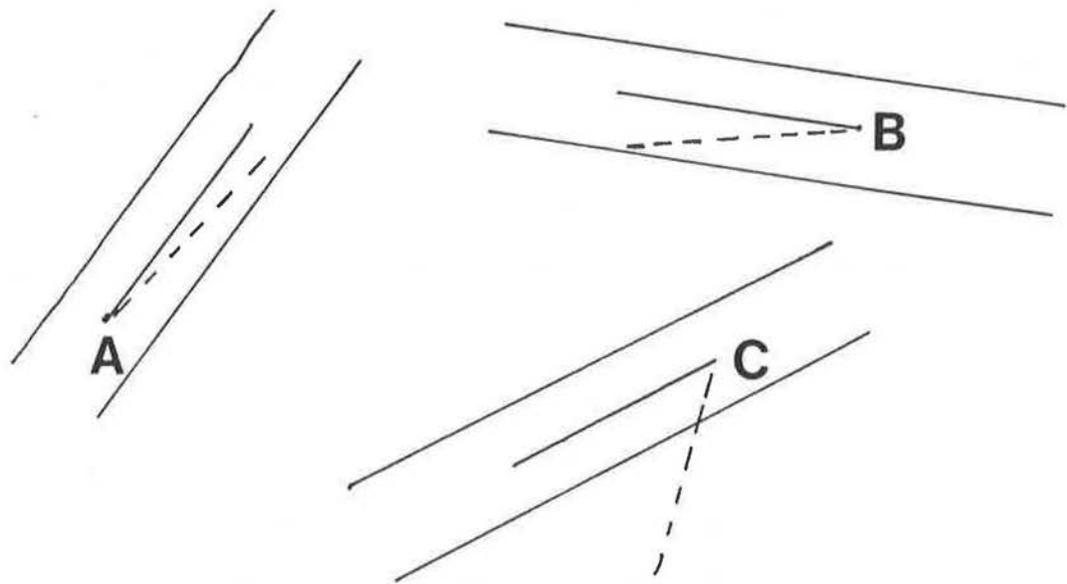
selecting other ridges with translation-compatible pairings as the third part of the match, the matching is extended to include three mutually translation-consistent pairings of features. With the six values from the matching, an affine transform can be determined. Each pairing of a ridge and an l-edge provides a point-to-point match for the parameter determination.

3.6 Consistency of the Transform

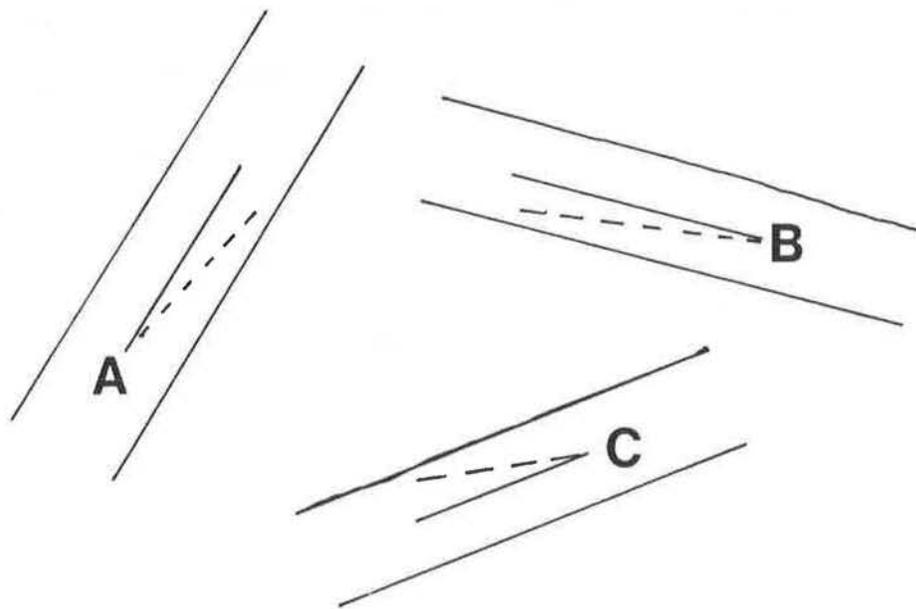
When an affine transform is determined for a set of three point pairings, the resulting transform will predict the points with no error. Hence it is necessary also to test how the transform predicts the segments passing through the matched points. This first estimate of the transform computed from the three pairings is tested for self-consistency. The transform is self-consistent if, using the new transform, the transformed l-edges and their matching ridges overlap (figure 25). Many matchings yield inconsistent transforms, which are rejected. This avoids the relatively expensive procedure of predicting and verifying feature locations.

3.7 Verification

If the transform is self-consistent, it is then used to predict the the location of the remaining l-edges in the terrain model. When a l-edge is compared to a ridge, it is examined at the positions for matching as described above (3.4). If the l-edge lies within a narrow band of the appropriate segment on the ridge, the points corresponding to that match are entered as the match points, and the features are matched. The number of



(1)



(2)

Figure 25

In a self-consistent transform (2), the bands of the ridges overlap the transformed l-edges (dotted lines). When the transform is not self-consistent (1), the l-edges extend outside the bands. A-C are the matched points.

l-edges which overlap existing ridges is used as the measure of the quality of the matching. Also the average and root-mean-square of the differences between points in the DTM and their matched l-edge points are calculated. If enough features can be matched in this way, the set of pairings is used to form an extended matching, from which a least-squares estimate of the affine transform parameters is computed. In this extended matching, any point pairs whose associated error is larger than the average error are rejected. In a good matching most of the point pairs produce errors less than the average. Removing pairs with large errors and re-computing the transformation is a heuristic for improving the registration. The new transform is computed from the remaining pairs. This transform, in turn, is used to predict the location of the l-edges in the DTM. If the matching improves, a new least-squares estimate of the transform is computed. This iterative process terminates when error terms are sufficiently small and the number of features predicted is sufficiently high. Indeed, if the average and RMS errors are less than a pixel, searching stops and success is indicated.

4. Implementation and Testing

4.1 The Input

To test the method, a 100x100 pixel subsection of a Landsat image (figure 1) is registered to a digital terrain model. Band 7 of the Landsat image was used because the effects of terrain relief are most apparent in that band. The Landsat image was acquired on September 14, 1976 (frame ID 11514-17153). The digital terrain model was digitized from the 1:50,000 series contour map, Canadian National Topographic System (NTS) sheet 82 F/9, (St. Mary Lake), covering an area from latitude 49 degrees, 30 minutes to latitude 49 degrees, 45 minutes and in longitude from 116 degrees to 116 degrees, 30 minutes. This area is northwest of Cranbrook, British Columbia. An area, 30 kilometers by 23 kilometers, is represented in the TIN digital terrain model by approximately 5500 points. This terrain model was utilized in other research on modeling image formation in remote sensing (Woodham, 1980).

The DIM was prepared manually by the author. The ridges and channels of the area were digitized. Additional points were included to shape the terrain surface between the ridges and channels. The complete set of points was triangulated, and automatically edited to include the edges joining points along the ridges and channels. When a TIN format DIM is not available, there exist automatic procedures for converting a DIM in grid format to a TIN (Fowler and Little, 1979).

4.2 Programming Languages

Implementation of the various parts of the system has been accomplished in several different programming languages. The procedures to extract features from a DTM and brightness discontinuities from a Landsat image were both written in PASCAL-UBC (Jolliffe and Pollack, 1979). Brightness discontinuities are not derived on demand during registration, but are determined in a preprocessing step. The DTM and the image are not available during registration matching. Registration matching uses ridge features and Landsat features written to ancillary files during the preprocessing steps. The registration system is written in LISP-MTS (Wilcox and Hafner, 1976) and reads the files from the Pascal procedures. LISP was chosen for the major component of the implementation because of the ease of experimentation with control structures and the simplicity of dynamic storage allocation.

4.3 Data Structures

The curvilinear features of the DTM and the Landsat image are represented in the piecewise linear approximation described in section 3.1.5. They are structured as lists when written to the feature files. A curve is represented as a 3 element list, as follows:

```
( first-point
  last-point
  internal-structure )
```

where internal structure is a 5 element list defined recursively as:

```
( left-point
  right-point
  internal-structure between first-point and left-point
  internal-structure between left-point and right-point
  internal-structure between right-point and last-point )
```

or

```
NIL when the perpendicular distance of all points between
  first-point and last-point is less than the detail level.
```

Points are represented as a list of the two coordinates. At lower levels in the structure, internal-structure is expanded using the endpoints of the enclosing segment as the first- and last-point. For example the line A-F in figure 17 is represented in the list structure for a generalized curve as:

```
( A F ( B D NIL (C NIL NIL NIL NIL) (E NIL NIL NIL NIL)))
```

The dotted lines in figure 17 show the approximating segments for various portions of the curve. Figure 17 also shows a tree representation of the generalized curve. This representation can easily be converted into the original line structure, a list of points, by a pre-order traversal of the tree.

Because it is necessary to search the area around a feature, a data structure was added to the system which would succinctly represent spatial relations. A coarse mesh is placed over the region in the plane containing the features. Each cell defined by this mesh is termed a bucket. On input, the features are compared with this mesh and the names of all features passing through a given bucket are added to the list of features in the bucket. When it is necessary to find all features within a certain distance

of a feature, a region of the appropriate shape around the feature is generated, and the list of buckets which this shape overlaps is derived. By merging the lists of feature names associated with this list of buckets, it is possible to determine the names of all features which may lie within the correct region.

4.4 Estimating the Affine Transform

If the change in the satellite's attitude during image acquisition is ignored, the parameters of the transform are:

$$\begin{aligned} a &= (M z_0 S) \cos (H + y) \\ b &= (O R L) \sin H + (E R L) \cos G \\ c &= x_0 - (r \cos H + p \sin H) z_0 \\ d &= -(M z_0 S) \sin (H + y) \\ e &= (O R L) \cos H \\ f &= y_0 - (-r \sin(H) + p \cos(H)) z_0 \end{aligned}$$

where

M is the angular velocity of the scanning mirror
 z_0 is the distance of the satellite from the surface of the earth at reference time t_0
 S is the sampling interval along the scan
 O is the angular velocity of the satellite in its orbit
 R is the radius of the satellite's orbit
 L is the time interval between scan lines
 G is the geocentric latitude at the sub-satellite point
 H is the heading of the satellite - the angle its orbit makes with a meridian
 r, p, y are the roll, pitch and yaw angle of the satellite platform measured with respect to x, y, z axes
 E is the angular velocity of the earth
 x_0, y_0 are the image coordinates of the point directly below the satellite at reference time t_0

The a priori estimate of the affine transform used in registration was computed using the following substitutions to the equations for the transform:

$M = 6.21 \text{ rad/sec}$
 $z_0 = 900 \text{ kilometers (a more accurate altitude is contained in the image annotation)}$
 $S = 9.958e-6 \text{ sec}$
 $O = 1.014e-3 \text{ rad/sec}$
 $R = 6370 \text{ kilometers}$
 $L = 12.237e-3 \text{ sec}$
 $G = 49 \text{ deg } 35 \text{ min}$
 $H = 0.246 \text{ rad}$
 $r, p, y = 0, 0, 0 \text{ rad}$
 $E = 72.722e-6 \text{ rad/sec}$
 $x_0, y_0 = 0, 0$

Using these parameters, the resulting affine transform is:

$a = 0.539797$
 $b = 0.236592$
 $c = 0.0$
 $d = -0.13550$
 $e = 0.766666$
 $f = 0.0$

4.5 Examples of Registration and Results

The position of the sun for the September 14, 1976 image was determined using a version of the method of (Horn, 1977) implemented by R.J. Woodham. The sun's position so determined was azimuth 134.5 degrees, elevation 34.4 degrees. Figure 26 shows a synthetic image generated using the calculated position of the sun. The DIM features selected using this sun position are depicted in figure 27. All of these ridge lines are longer than 250 meters. The curves are generalized using a detail level of 80 meters.

For the Landsat image, the same length and generalization parameters were used. The top 20 percent of the feature cells were used in the construction of the l-edges. The position of the sun was input as well, so that shadow edges could be rejected.

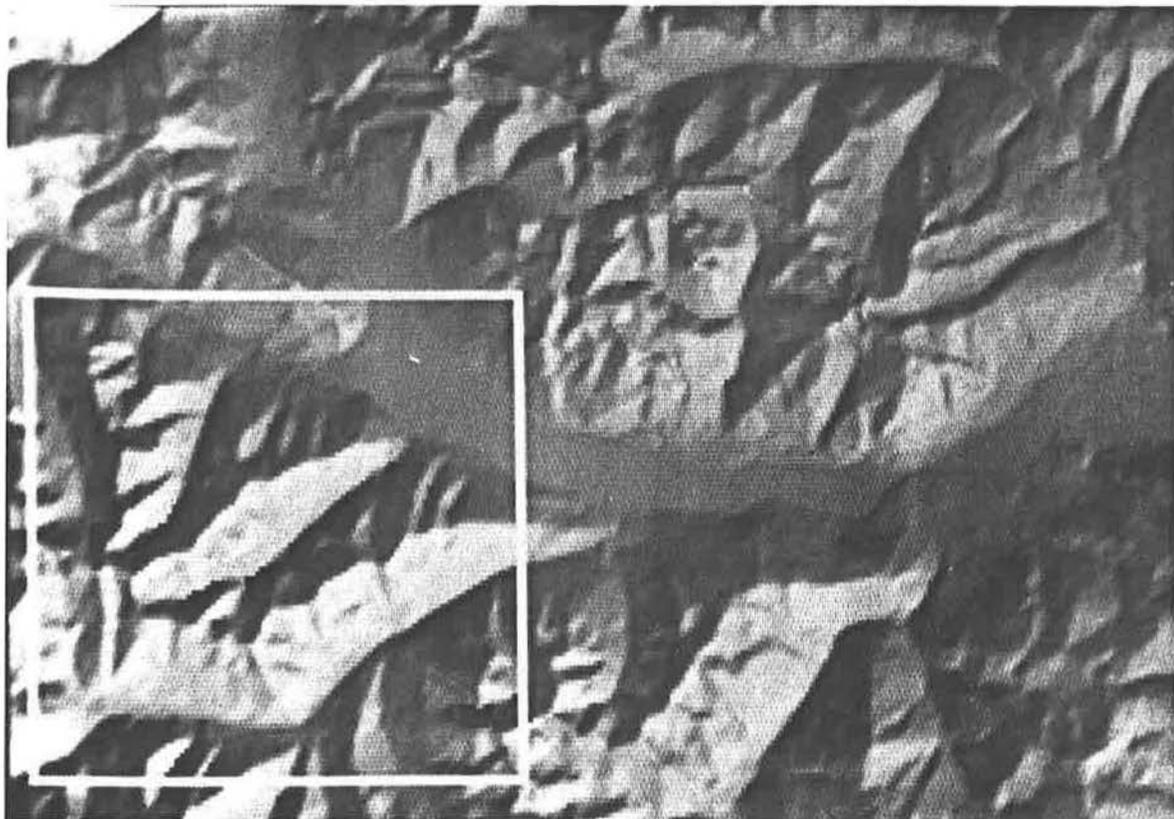


Figure 26

Synthetic image for September 14, 1976. The sun's position is azimuth 134.5 degrees, elevation 34.4 degrees. The white box outlines the portion of the DIM used in feature selection. Photographed from the screen of the COMTAL Vision 1.



Figure 27

DTM features for September 14, 1976, with the matched points (1-6).

In the test case, the location of the Landsat image in the DIM was estimated by hand to within 0.75 kilometers, or approximately 10 pixels. This reduced the search region size so as to reduce the expense in developing the system. The affine transform for this image was determined to be:

$$\begin{aligned} a &= 0.555292 \\ b &= 0.131612 \\ c &= 1.944259 \\ d &= -0.143495 \\ e &= 0.773612 \\ f &= 2.197464 \end{aligned}$$

The errors associated with a transformation are determined by comparing the positions of transformed Landsat points with the positions of the corresponding DIM points. The point pairs are derived from the features which overlap using the transformation being evaluated. The registration determined from the matching found by the system resulted in the following errors:

$$\begin{aligned} \text{Average error} &= 30.9 \text{ meters or } 0.3862 \text{ pixels} \\ \text{Root mean square error} &= 53.5 \text{ meters or } 0.66875 \text{ pixels} \end{aligned}$$

Figures 27 and 28 show the DIM and Landsat features with the matched points. There are 33 ridges and 18 l-edges in this example. Twenty-five matchings (three pairings each) were examined before a matching was accepted. Of these matchings, 14 produced transforms which were self-consistent. The remaining were rejected on the grounds of the inconsistency of the transform. Two of the six pairings in this matching are at junctions between features in the DIM.

A second image of the same region (figure 29) obtained January 8, 1979, (frame ID 30309-17575), was registered. The a priori estimate of the affine transform used for this case was the same as that used for the first image. The position of the sun for this image was calculated as above to be azimuth



Figure 28

Features from Landsat image, September 14, 1976, with matched points (1-6).

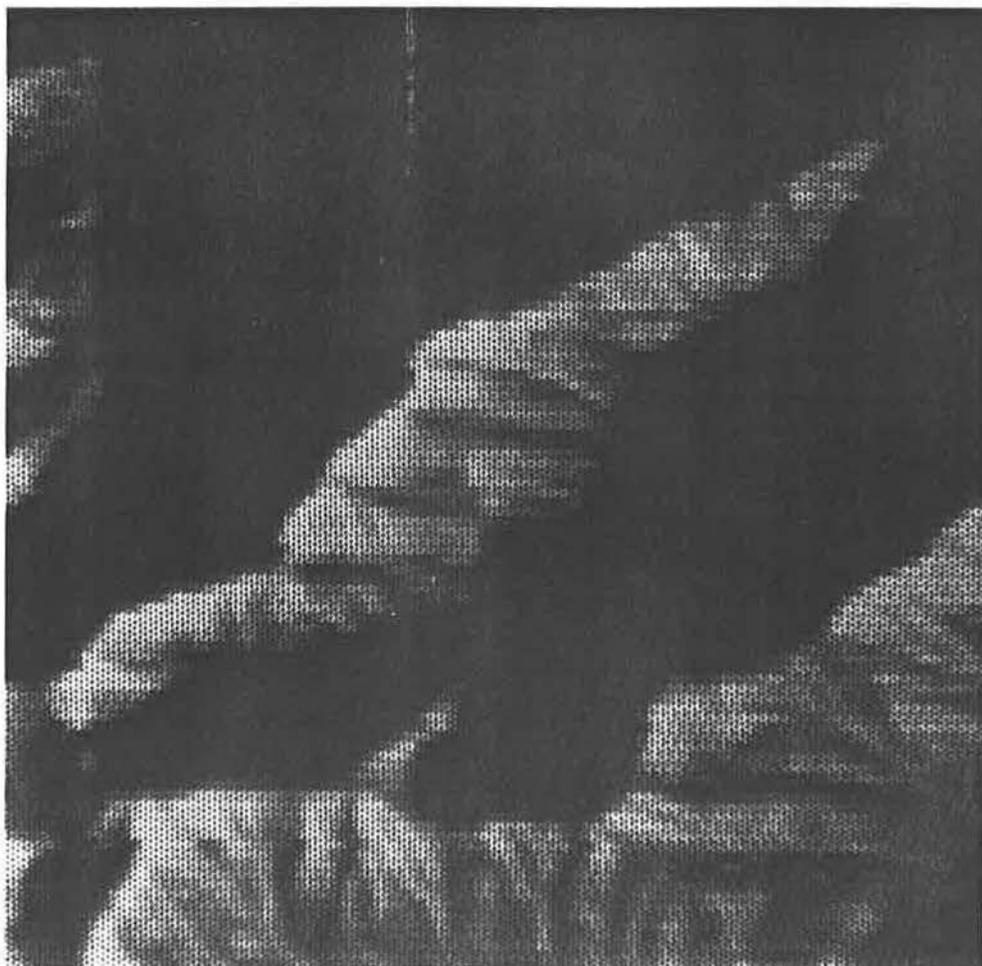


Figure 29

100 X 100 pixel subsection of Landsat image (band 7) from January 8, 1979, frame ID 30309-17575. Photographed from the screen of the COMTAL Vision 1.

153.1 degrees, elevation 13.8 degrees. Many of the ridges selected from the DTM using surface orientation alone were in shadow. The portion of the terrain in shadow can be detected by using a standard 'hidden-surface' algorithm (Woodham, 1980) in which the viewing point is located at the position of the light source. The portion of the surface which is invisible to an observer thus situated is in shadow. Any part of a ridge which is in shadow is 'clipped' to the boundaries of the shadow. Shadow calculation was not implemented for feature selection. Instead, the program of R.J. Woodham for producing synthetic images (figure 30) was used to determine the locations of regions in shadow. Features lying in those regions were removed manually.

The affine transform for the January 8, 1979 image was determined to be:

a = 0.537147
 b = 0.150337
 c = 1.620489
 d = -0.132218
 e = 0.694722
 f = 2.330885

The error terms were:

Average error = 38.5 meters or 0.48125 pixels
 Root mean square error = 56.9 meters or 0.71125 pixels

Figures 31 and 32 show the DTM and Landsat features with the matched points indicated. There are 15 ridges and 22 l-edges. In this example, the first matching developed yielded this good set of matches with an acceptable error. Four of the seven pairings in this matching occur at ridge junctions. Figure 33 shows the registered Landsat image.

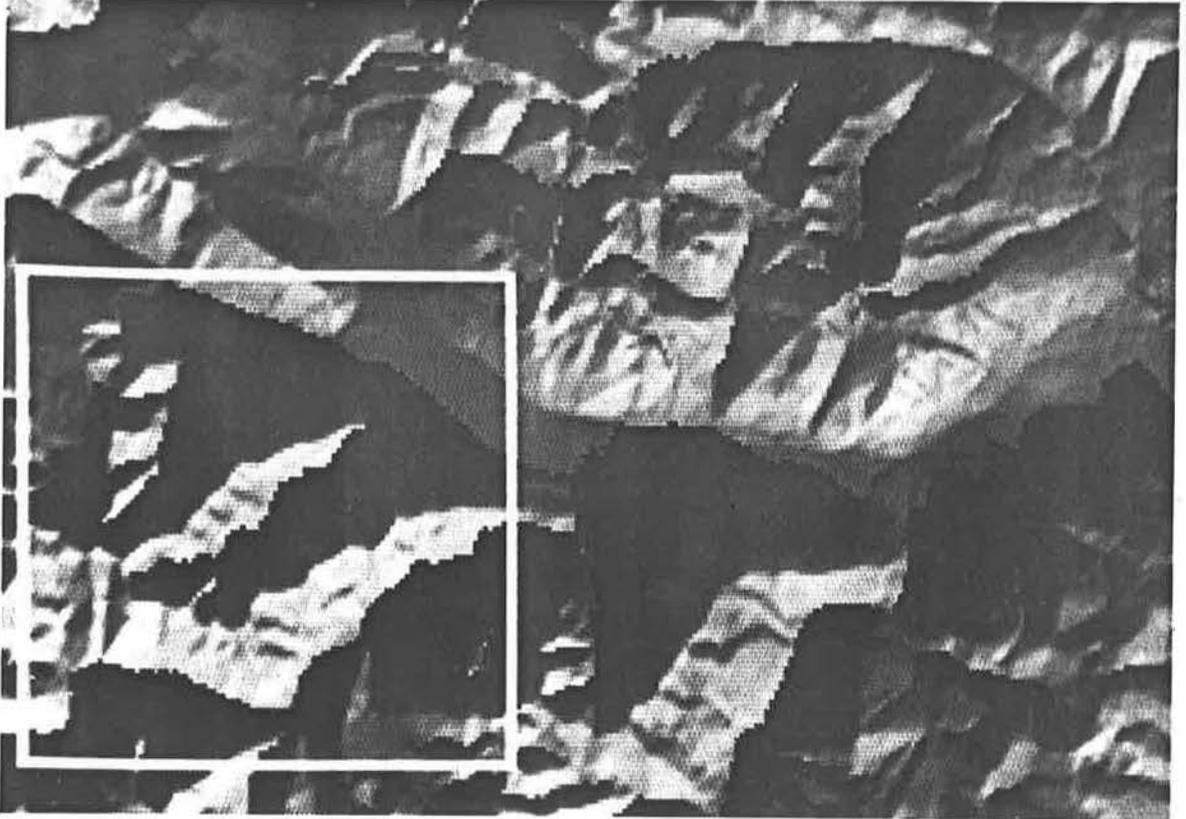


Figure 30

Synthetic image for January 8, 1979. The sun's position is azimuth 153.1 degrees, elevation 13.8 degrees. The white box outlines the portion of the DIM used in feature selection. Photographed from the screen of the COMTAL Vision 1.

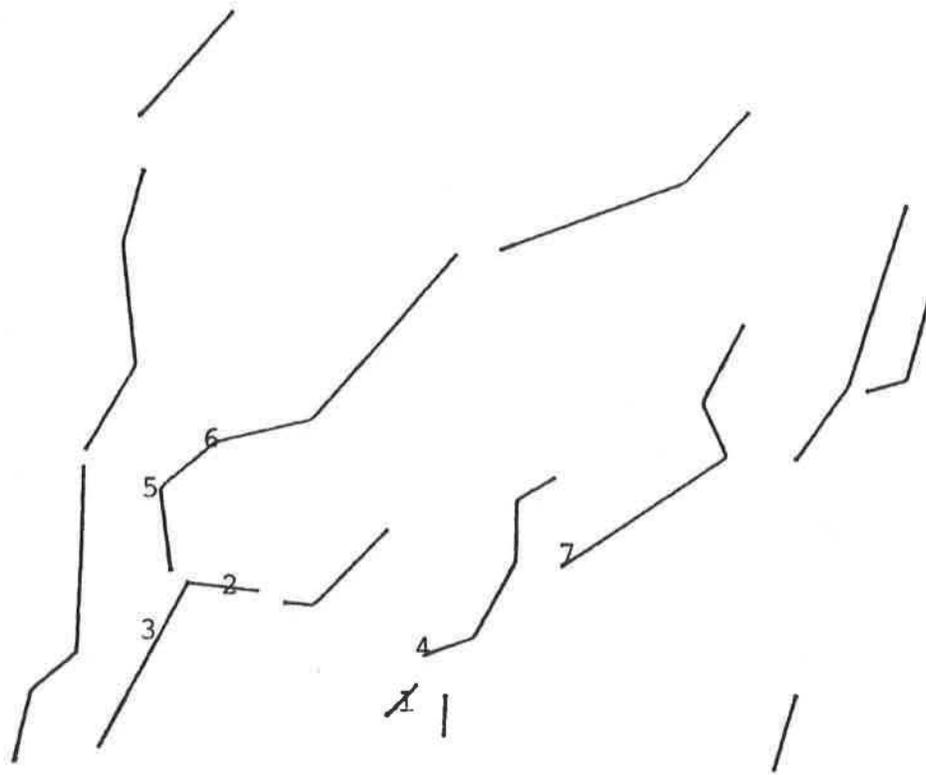


Figure 31

DTM features for January 8, 1979, with matched points (1-7).

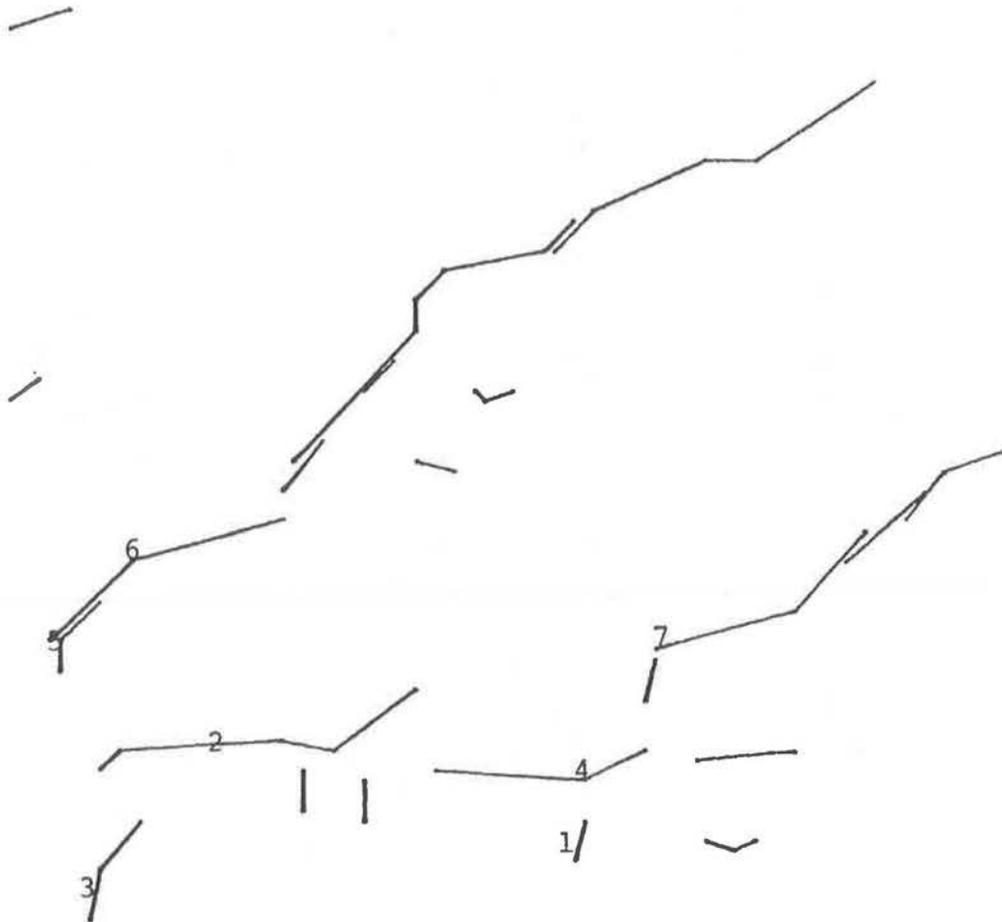


Figure 32

Landsat features for January 8, 1979 with matched points (1-7).

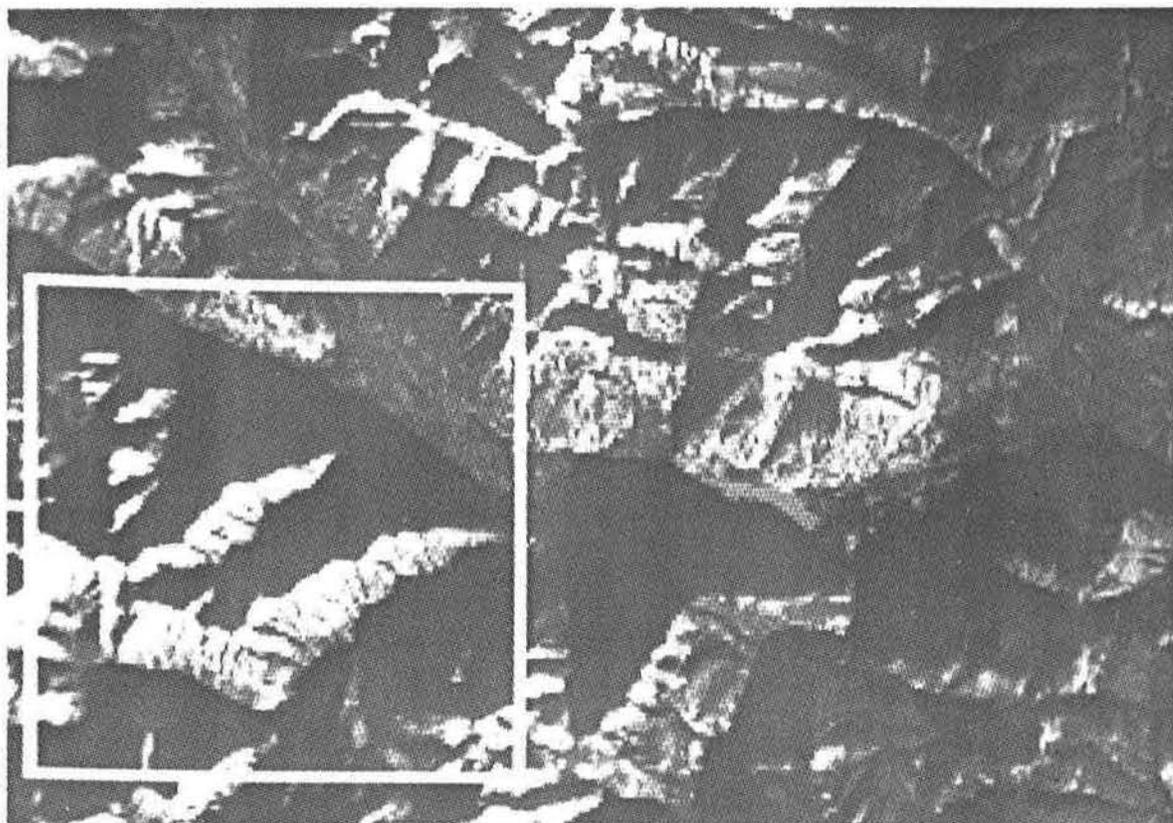


Figure 33

Registered Landsat image for January 8, 1979. The white square outlines the 10 km² area covered by the DIM. Photographed from the screen of the COMTAL Vision 1.

5. Discussion and Conclusions

5.1 Discussion

The results of the tests indicate that feature matching can be an effective procedure for registering images. Registration errors for the examples are well within accepted standards. The automatic registration procedure presented relies upon the existence of a detailed terrain model for its use. Currently, such DTM's are not generally available. However, in Canada, the Department of Energy, Mines and Resources is committed to production of such DTM's for most of Canada. It has been shown (Woodham, 1980) that registration of an image to a digital terrain model is helpful in determining shading effects which affect image analysis. Benefits such as this can sometimes justify manual generation of a DTM for a particular study area.

Insofar as the method is based upon determining terrain features which will appear distinctly in an image, the method is restricted to application in areas of mountainous terrain. There is little possibility that the method as it stands would be useful in registering images of prairie land. The benefits of registering an image to a DTM in such a situation are minimal also. Nevertheless, the principle of using known illumination conditions and a scene model can find applicability elsewhere. The features selected can be water-land boundaries, roads and other distinctive scene elements. The application to DTM's and Landsat images is particularly appealing since the imaging geometry is simple. There is no need to solve hidden surface problems. Funt (1980) has proposed using synthetic images in

interpreting indoor scenes. Features extracted from any scene model containing information on surface orientation and position can be used with the method presented.

The representation of curves by generalization simplifies the process of analysing the relations between curves. By matching portions of curves to each other in varying positions, distinctive matchings are determined. Although the notion of representing a curve by its band has existed for some time, its use in curve matching is new in this application. By permitting looser matching between curves and segments of curves, the band representation facilitates curve matching.

Determining local support for a match appears to disambiguate false matches readily. In images of mountainous terrain, it is unlikely that local support will be insufficient for detecting correct matches. However, in scenes of urban landscapes, or in industrial applications, regularity is intrinsic. Local support will be very necessary in distinguishing false and true matches. In addition, these situations will require more careful selection of distinct subsets of matching features. The representations for curves advanced in this thesis is advantageous for such feature selection.

Difficulties with the method will occur in areas of low relief or strongly regular terrain, what geomorphologists call "strongly controlled" terrain. Clouds, depending upon the sun's position, can be problematic. The boundaries of shadows of clouds will appear in images as strong brightness discontinuities and will not be discriminated from the images of ridges. Clouds themselves will generate brightness discontinuities. The method may prove itself robust enough to meet this challenge.

5.2 Further Work

The handling of curve matching where both curves are structured was eliminated in this implementation of the registration method. By subdividing l-edges into simple segments, some of the power of the representation is lost. The junctions at which the l-edge segments meet are then unavailable. However, the implementation is simplified. By including a facility for manipulating and assessing structured-to-structured feature matching a significant improvement could be made. Davis (1979) has developed one such method.

The control of matching development is very simple. All translation-consistent triples are examined for self-consistency. If a matching is self-consistent with respect to its derived transformation, the transform is used to predict the location of image features in the DTM. The result of this test of the transform is binary: either accept or fail. When a set of predicted l-edge to ridge overlaps is generated, the structural relations between the overlapping l-edges and ridges could be used to guide further adjustment of the parent matching. This upward flow of information is very important in image analysis in general.

5.3 Conclusions

This work demonstrates the effectiveness of matching features derived from digital terrain models with image features for solving the registration problem. It is hoped that the techniques presented here and the principles underlying them can find application elsewhere.

Bibliography

- Agin, G.J., "Computer Vision Systems for Industrial Inspection and Assembly", *Computer* 13, 5(May, 1980), 11-20.
- Aho, A.V., Hopcroft, J.E. and J.D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Ma., 1974.
- Ambler, A.P. and R.S. Popplestone, "Inferring the positions of bodies from specified spatial relationships", *Artificial Intelligence*, 6, 2 (1975), 157-174.
- Bajcsy, R. and P. Chance, "Picture recognition applications in biochemistry", Technical Report, University of Pennsylvania, Moore School of Electrical Engineering, (May, 1975).
- Bajcsy, R., and M. Tavakoli, "Computer recognition of roads from satellite pictures", *International Joint Conference for Pattern Recognition*, (1976).
- Ballard, D.H., C.M. Brown and J.A. Feldman, "An Approach to Knowledge-Directed Image Analysis", *International Joint Conference on Artificial Intelligence*, (1979).
- Ballard, D.H. "Strip Trees: A Hierarchical Representation for Map Features", *Proceedings of IEEE Computer Society Conference on Pattern Recognition and Image Processing*, Chicago, Illinois, August 1979, 278-285.
- Barnea, D.T. and H. Silverman, "A Class of Algorithms for Fast Digital Image Registration", *IEEE Transactions on Computers*, C-21, 179, (1972).
- Barrow, H.G. and R.M. Burstall, "Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques", *Information Processing Letters*, 4, 4, (Jan. 1976), 83-84.
- Barrow, H.G. and R.S. Popplestone, "Relational descriptions in picture processing", in *Machine Intelligence 6*, Meltzer, B., Michie, D. (Eds.), 377-396.
- Barrow, H.G., Tenenbaum, J.M., Bolles, R.C. and H.C. Wolf, "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching", Technical Note 153, SRI International, Menlo Park, California.
- Bernstein, R. "Digital Image Processing of Earth Observation Sensor Data", *IBM Journal of Research and Development*, (Jan. 1976), 40-57.
- Bolles, R.C., "Robust feature matching through maximal cliques", *Proceedings of Society of Photo-optical Instrumentation Engineers*, 182 *Imaging Applications for Automated Industrial Inspection and Assembly*, (1979), 140-149.

- Bolles, R.C. et al., "Automatic Determination of Image-to-Database Correspondences", International Joint Conference on Artificial Intelligence, (1979).
- Davis, L.S. "Shape Matching Using Relaxation Techniques". IEEE Transactions on Pattern Analysis and Machine Intelligence, 1,1 (Jan. 1979), 60-72.
- Davis, L.S. "A Survey of Edge Detection Techniques", Computer Graphics and Image Processing 14, 3 (1975), 248-270.
- Davis, W.A. and S.K. Kenue, "On the Detection of Changes in Digital Images", Computer Science Technical Report, University of Alberta, Edmonton, (Sept. 1977).
- Fischler, M.A. and R.A. Elschlager, "The Representation and Matching of Pictorial Structures", IEEE Transactions on Computers, C-22, 1, (Jan. 1973), 67-92.
- Fowler, R.J. and J.J. Little, "Automatic Extraction of Irregular Network Digital Terrain Models", Proceedings of SIGGRAPH-79, Chicago, Illinois, (Aug. 1979), 199-207.
- Funt, B.V. "Towards Synthetic Images in Scene Analysis", Proceedings of the CSCSI/SCEIO Conference, Victoria, B.C., (1980), 158-165.
- Horn, B.K.P., "The Position of the Sun", MIT AI Lab Working Paper 162.
- Horn, B.K.P. and B.L. Bachman, "Using Synthetic Images to Register Real Images with Surface Models", Comm. ACM 21, 11 (Nov. 1978), 914-924.
- Horn, B.K.P. and R.J. Woodham, "Landsat MSS Coordinate Transformations", Proceedings of the Fifth Annual Symposium on Machine Processing of Remotely Sensed Data, Purdue University, 1979.
- Hsieh, Y.Y. and K.S. Fu, "A Method for Automatic IC Chip Alignment and Wire Bonding", Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing, Chicago, Illinois, (August 1979), 101-108.
- Iannino, A. and S. Shapiro, "An Iterative Generalization of the Sobel Edge Detection Operator", Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing, Chicago, Illinois, (August, 1979), 130-137.
- Itai, A., Rodeh, M. and S.L. Tanimoto, "Some matching problems for bipartite graphs", Journal of the Association for Computing Machinery 25, 4 (Oct. 1978), 517-525.
- Jolliffe, B. and B.W. Pollack, "UBC PASCAL, PASCAL/UBC Programmer's Guide", Computer Center, University of British Columbia, (Sept. 1979).
- Kaneko, T., "Evaluation of LANDSAT Image Registration Accuracy", Photogrammetric Engineering and Remote Sensing XLIII, 1976, 1285-1299.

- Myers, W., "Industry Begins to Use Visual Pattern Recognition", *Computer* 13, 5 (May, 1980), 21-31.
- Nevatia, R. and K.R. Babu, "Linear Feature Extraction and Description", *International Joint Conference on Artificial Intelligence*, (August, 1979).
- Pavlidis, T., Structural Pattern Recognition, Springer-Verlag, New York, 1977.
- Peucker, T.K. and D.H. Douglas, "Detection of surface-specific points by local parallel processing of discrete terrain elevation data", *Computer Graphic and Image Processing* 4, (1975), 375-387.
- Peucker, T.K., R.J. Fowler, J.J. Little, and D.M. Mark, "The Triangulated Irregular Network", *Proc. of the Digital Terrain Models Symposium*, May 9-11, 1978.
- Ramer, U. "An iterative procedure for the polygonal approximation of planar curves". *Computer Graphics and Image Processing* 1,3 (1972).
- Sutherland, I.E., Sproull, R.F. and R.A. Schumaker, "A Characterization of Ten Hidden-Surface Algorithms", *Computing Surveys* 6, 1 (March 1974), 1-55.
- Tanimoto, S.L. "Analysis of biomedical images using maximal matching", *Proceedings of 1976 IEEE Conference on Decision and Control Adaptive Processes*, Clearwater Beach, Fla., (Dec. 1976), 171-176.
- Toriwaki, J. and T. Fukumaru, "Extraction of Structural Information from Grey Pictures", *Computer Graphics and Image Processing* 7, (1978), 30-51.
- Wilcox, B. and C. Hafner, "LISP/MIS User's Guide", Department of Computer Science, University of British Columbia, (July, 1976).
- Woodham, R.J. "Using Digital Terrain Data to Model Image Formation in Remote Sensing", *Proc. Soc. Photo-Optical Instrumentation Engineers*, Vol. 238, Bellingham, Wa., (July, 1980).