Extracting Road Networks from Natural Terrain for Visualization

Edward Riseman, Xiaoguang Wang
Brian Hanechak, Howard Schultz, Allen Hanson
{riseman, xwang, hanechak, hschultz, hanson}@cs.umass.edu

University of Massachusetts
Computer Vision Laboratory
Amherst MA 10003
USA

Abstract

The ability to effectively generate and assimilate terrain information is central to many computer vision (CV) and computer graphics (CG) applications. In particular, effective interactive visualization applications require techniques for generating a digital elevation map (DEM) and ortho-image, classifying ground cover, and extracting symbolic information. In this paper we describe a program underway at the University of Massachusetts for generating complex geospatial databases using photogrammetric techniques and context-sensitive strategies. We present a specific example in which the input are two monochromatic aerial images of a rural area of Fort Hood, Texas, and the results are a DEM, ortho-image, classified ground cover and a symbolic representation of a road network.

1. Introduction

Effectively generating and applying terrain information is crucial for both civilian and military applications. Advances in computer vision and graphics, including sensors, sensor platforms, computing platforms, and storage capacity, can facilitate the automated classification of terrain and the extraction of features for both natural terrain and cultural sites.

This paper focuses on a subset of work that forms part of a major DARPA (Defense Advanced Research Projects Agency) program encompassing a wide range of image understanding techniques. Our emphasis is on providing more accurate three-dimensional reconstruction, extracting symbolic information, and enhanced visualization techniques. To demonstrate our approach, we analyzed a rural scene containing trees, bare ground, a dry river bed, a partially occluded rural road network, and a bridge. The goals of this investigation were to determine the three-dimensional structure of the terrain, the type of ground cover, and to extract the road network. We obtained a semantic interpretation of aerial images to classify the terrain, and then applied context-sensitive strategies to extract the road network. These results provide an enhanced graphical model for use by high level applications such as fly-through visualization in support of logistics, training, and navigability investigations.

The results shown in this paper were produced by integrating a number of systems. The processing sequence of these systems operates in the following manner. Given a pair of aerial images of rural terrain, the UMass terrain reconstruction system (Terrest) produces a digital elevation map. The next step is to classify pixels according to the type of ground cover. The classified pixels, together with the ortho-image, become the input for a context-sensitive procedure to extract the road network. This procedure takes into account reasoning about where the trees occlude the roads, and applies geometric grouping criteria to extract a symbolic road network.

2. Generation of Digital Elevation Maps

With the advent of unmanned air vehicles (UAV) and high resolution digital imaging devices, it has become feasible to collect high quality digital images over a wide area. Applications that analyze these image sources must be able to efficiently process images taken from widely spaced locations. However, most terrain reconstruction systems perform well only when the two views are close to nadir (i.e., down-looking), and are not too widely separated. Widely spaced images present unique problems for terrain reconstruction systems because of the perspective distortion resulting from oblique viewing geometry and large base-to-height ratios. The UMass terrain reconstruction system (Terrest) employs techniques specifically designed to reconstruct terrain from images with a significant amount of perspective distortion. These techniques include iterative

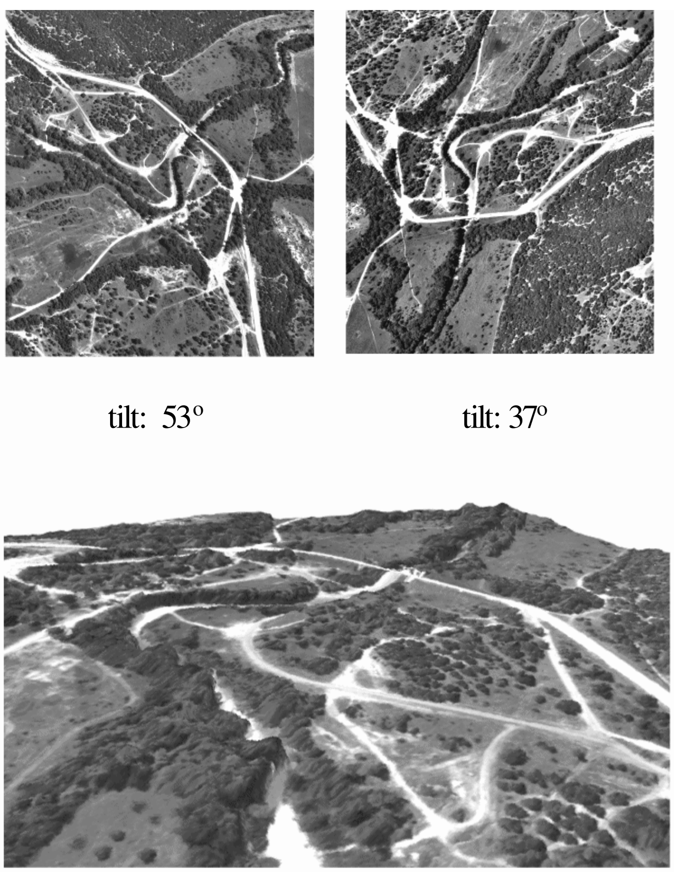


Figure 1. (top) Two aerial images of a rural section of Fort Hood, TX taken at oblique viewing angles. (bottom) A rendered view of the reconstructed terrain.

hierarchical matching, a weighted correlation mask, and subpixel registration. Terrest requires intrinsic and extrinsic camera parameters and assumes that the surfaces are textured. For a more detailed treatment see [Schultz, 1994 and 1995]. We used Terrest to analyze a pair of oblique views of rural terrain in Fort Hood, Texas. Figure 1 shows the $2k \times 2k$ pixel input images, which were taken at an off nadir tilt of 53° and 37° respectively, and a rendered view of the test area created by draping the ortho-image over the DEM.

3. Terrain Classification

Image classification and segmentation is a key stage in terrain interpretation tasks. The goal of terrain classification is to label the pixels of a scene according to the type of ground cover; this procedure allows us to extract objects of semantic interest for future use. The test area was divided into four classes, foliage (which included trees and shrubs), grass covered ground, bare ground (which included exposed earth, dirt roads and riverbeds), and shadows. These classes were selected based on observations of the test area. Training data were generated by manually segmenting small sub-regions into irregular patches in which each patch belonged to one of the four classes

3.1 The feature set

The choice of features is critical to the performance of an image classifier. Because of the rural nature of the test area, we selected texture sensitive features. Several investigators [Ohanian and Dubes, 1992; Conners and Harlow, 1980; and duBuf et al., 1990] have demonstrated that co-occurrence features are well-suited for classifying images characterized by varying textures. Haralick et al. (1973), Weszka et al. (1976), and du Buf et al. (1990) showed that the co-occurrence features were generally superior to other types of features when applied to classifying terrain from aerial images. A later theoretical study by Conners and Harlow (1980) supported this finding by demonstrating the advantages of co-occurrence features for classifying terrain. Ohanian and Dubes (1992) studied four types of features, including Markov random field parameters, multi-channel filtering features, fractal based features, and co-occurrence features, and concluded that co-occurrence features yield the best results.

The co-occurrence features describe the texture within a local image window based on gray-level spatial dependencies. The (i,j) matrix element $P_{ij\theta}$ represents the relative frequency of occurrence of two pixels in the window, separated by distance d and angle θ, one with gray-level i and the other with gray-level j. In their original work, Haralick et al. (1973) provided 14 types of features. However, later studies by Conners and Harlow (1980), and Ohanian and Dubes (1992) suggest that three of these features, ASM (the angular second-moment), CON (the contrast), and ENT (the entropy), are sufficient for classifying natural scenes. The definition of these features are

$$f_\theta^{ASM} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{ij\theta}^2 \quad (1)$$

$$f_\theta^{CON} = \sum_{n=0}^{G-1} n^2 \sum_{|i-j|=n} P_{ij\theta} \quad (2)$$

$$f_\theta^{ENT} = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{ij\theta} \ln P_{ij\theta} \quad (3)$$

The ASM feature is a measure of the homogeneity of the texture intensity. The CON feature measures the local variation. The ENT feature measures the complexity. Based on experiments conducted by Ohanian and Dubes (1992) and our preliminary experiments, we applied the following settings: The window size is set to 17×17, d=1 (i.e., the co-occurrence matrix reflects neighbor pixel dependencies), θ = (0°, 45°, 90°, 135°), and G=8 (i.e., the number of gray levels is 8). In a pre-processing step, histogram equalization was applied to the images to reduce the total number of gray-levels to G.

3.2 The classifier

Reducing the dimensionality of feature space is an important consideration in classification. In the example presented here, there are thirteen features consisting of the original gray-scale intensities plus the twelve co-occurrence features generated by using Equations 1,2 and 3 with θ=(0°,45°,90°,135°). The intensity feature provides additional discriminability, especially for road and shadow classes which have a wide range of gray levels. The complex thirteen-dimensional feature space was reduced to a much simpler three-dimensional algebraic feature space via a linear transformation given by

$$\mathbf{G}_j = \sum_{i=0}^{12} \mathbf{a}_{ij}\mathbf{F}_i \qquad j = 0,1,2$$

where $\mathbf{F}_i$ are the old features, $\mathbf{G}_j$ are the new features, and $\mathbf{a}_{ij}$ are the coefficients.

The classification system is based on the Foley-Sammon transform (FST) which is a linear discriminant method that employs the Fisher criterion [Foley and Sammon, 1975; Tian 1986]. Our implementation includes extensions to the FST proposed by Liu et al. (1992 and 1993) and by Weszka et al. (1976) who demonstrated the utility of the FST for classifying texture features.

We further modified the FST to account for the fact that some classes have a significantly greater variance in feature space than others. Instead of using the Euclidean distance between a feature vector and a class centroid in feature space, we used the Mahalanobis distance (the Euclidean distance weighted by the class variance). The class variance is estimated from the training data. The advantage of using the Mahalanobis distance over Euclidean distance is that it takes into account the within class scatter. For example, if a feature vector is equidistant between two classes, *A* and *B*. Furthermore, if class *A* has a greater variance than class *B*, then class *A* occupies a greater volume in feature space than class *B*. Thus, when the inter-class Euclidean distances are scaled by the class variances, the feature vector is closer to *A* than to *B*. As a result, the feature vector would be assigned to class *A*.

For the Fort Hood test area, the dimensionality of the feature space was reduced from 13 to 3; Table 1 shows the coefficients $\mathbf{a}_{ij}$ of the linear transformation which were determined from the training data. The training data were extracted from 14 image chips ranging in size from 20×20 to 99×99 that were randomly distributed over the ortho-image. The results of the classification for a 590×350 sub-region of the test area are shown in Figure 2; Figure 2A shows the ortho-image, and Figure 2B shows the classification results. A summary showing the classification results versus ground truth is shown in Table 2. A perfect classifier should result in unity along the diagonal and zeros in the off diagonal terms.

Table 1. Coefficients $\mathbf{a}_{ij}$ used to generate the new features from the old features

***i* (Old Features)**	***j* (New Features)**		
	1	**2**	**3**
1	0.203851	0.242146	0.061315
2	-0.277844	0.529578	-0.443270
3	0.268151	0.209577	-0.088036
4	0.092234	0.730391	0.134553
5	0.153504	-0.034972	-0.374619
6	0.220371	0.035253	0.389429
7	-0.220631	-0.034401	-0.050839
8	0.069955	0.023174	0.240010
9	-0.197589	-0.022663	-0.003371
10	-0.546828	-0.045067	-0.256101
11	-0.234117	-0.087290	0.297282
12	-0.423391	0.254793	0.503026
13	-0.320942	-0.065184	0.121256

Table 2. Classification results versus ground truth.

Truth	Classified: shadow	grass covered ground	foliage	bare ground
shadow	0.544	0.056	0.397	0.003
grass covered ground	0.004	0.842	0.057	0.097
foliage	0.011	0.244	0.707	0.038
bare ground	0.000	0.020	0.001	0.979

4. Extraction of Road Networks

In the next phase, higher level symbolic information is extracted from the classified images. We demonstrate this process using the previous results and a series of context-sensitive strategies to extract a road network. These procedures take into account reasoning about where the trees occlude the roads, and apply geometric grouping criteria to extract a symbolic road network (Fischler et al., 1981).

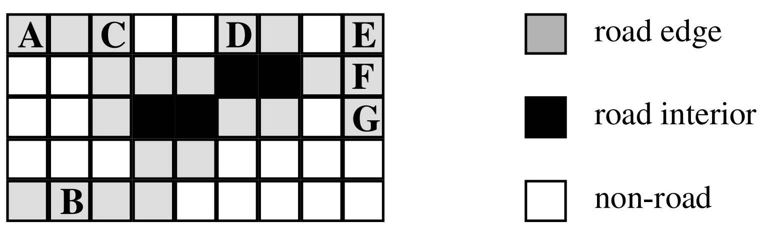
4.1 Forming and thinning road segments

The first step is to reclassify the image into road and non-road pixels by combining the grass covered ground, foliage and shadow classes into one, non-road class. Next, road segments are formed from the clusters of road pixels by smoothing the boundaries and closing interior gaps. The road segments are then thinned to produce a sequence of line segments that lie approximately along the center of the roads. Finally, the line segments are grouped to form a continuous road network.

The smoothing procedure involves the following steps. A small window is passed over the image. At each position, the total number of road pixels in the window, and the total number along the perimeter of the window are counted. If both of these sums are greater than or equal to a specified threshold, all of the pixels within the window are classified as belonging to the road interior. In the example presented here, the window size was 5×5 (which has a perimeter of 16 and an area of 25), and the perimeter and area thresholds were set to 10 and 14. Next, small holes in the road segments that arise from irregularities in the road surface and occlusions from surrounding trees are closed using a connected components algorithm that fills any holes that are smaller than 12 pixels across.

Once the road segments have been closed and their boundaries smoothed, an iterative thinning process is applied to reduce the road segments to line elements. At each step in the thinning process, pixels are removed from the borders of the road segments until the road is represented by a single line. At each iteration the thinning process takes three steps: (1) All road pixels are marked as belonging to the interior of a road or lying on the edge of a road. An edge pixel is a road pixel that has a non-road pixel either above or below it, or to its left or right. (2) All road edge pixels are marked as *essential* or *non-essential*. (3) All non-essential road edge pixels are reclassified as non-road. The iterations continue until no non-essential pixels are found.

The method for defining a non-essential road edge pixel is as follows. Starting and ending at the pixel directly above a road edge pixel, the algorithm scans the eight adjacent pixels and counts the number of times the pixel class crosses a road boundary. This count divided by two represents the number of roads connected to the central pixel (N_c). If $N_c = 1$ and the number of road pixels adjacent to the central pixel is greater than 1, then the central pixel is considered non-essential and reclassified as non-road. Consider the following example of a 9×5 image with 25 road pixels, of which 4 are road interior and 21 are road edge pixels.



For the pixels labeled **A-G**, Table 3 shows the number of roads connect to the pixel (N_c), the number of adjacent road pixels, and whether the pixel was classified as essential or non-essential. Figure 2C shows the road and non-road pixel classes, and the thinned road segments for the sub-area of the ortho-image shown in Figure 2A.

Table 3. Road thinning example.

Pixel	**Connect road segments** N_c	**Number of adjacent road pixels**	**essential**
A	1	1	Yes
B	2	2	Yes
C	2	3	Yes
D	1	4	No
E	1	2	No
F	3	3	Yes
G	1	2	No

4.2 Perceptual grouping

After the thinning process is completed, road vertices are marked. A road vertex is defined as any road pixel for which N_c=1 or N_c>2. If N_c=1, the pixel is located at a dead end; if N_c=2, the pixel lies on a road, or a bend in a road; and if N_c>2, the pixel lies at the junction of two or more roads. Because of occlusions, shadows and irregularities in the road surface there are many gaps in the thinned road elements. To generate a connect road network, the next task is to convert the collection thinned road elements to a network. The first step in this process is to convert the pixel representation to a graph/vector representation, where each intersection, dead end, and turning point in the road is a vertex in the graph.

The algorithm starts by constructing a graph using the vertices described above, and connects any pair of vertices with a straight line, if there is a road between them. It determines this condition using a depth-first search in the thinned image from one vertex pixel along road pixels. The depth-first search ends when another vertex pixel is encountered.

The program attempts to refine each line in the graph by replacing it with a polyline (a connected sequence of shorter line segments) that more accurately correspond to the original road. For each straight line segment, the algorithm finds the pixel on the original thinned road segment that is furthest away. If this distance is above a threshold (set here at 1.5 pixels), a new vertex is inserted at this pixel, and the straight line segment is replaced with two shorter ones. The algorithm repeats until no straight line segments are replaced. After adding the new segments, very short dead ends (less than 12 pixels long) are removed from the graph.

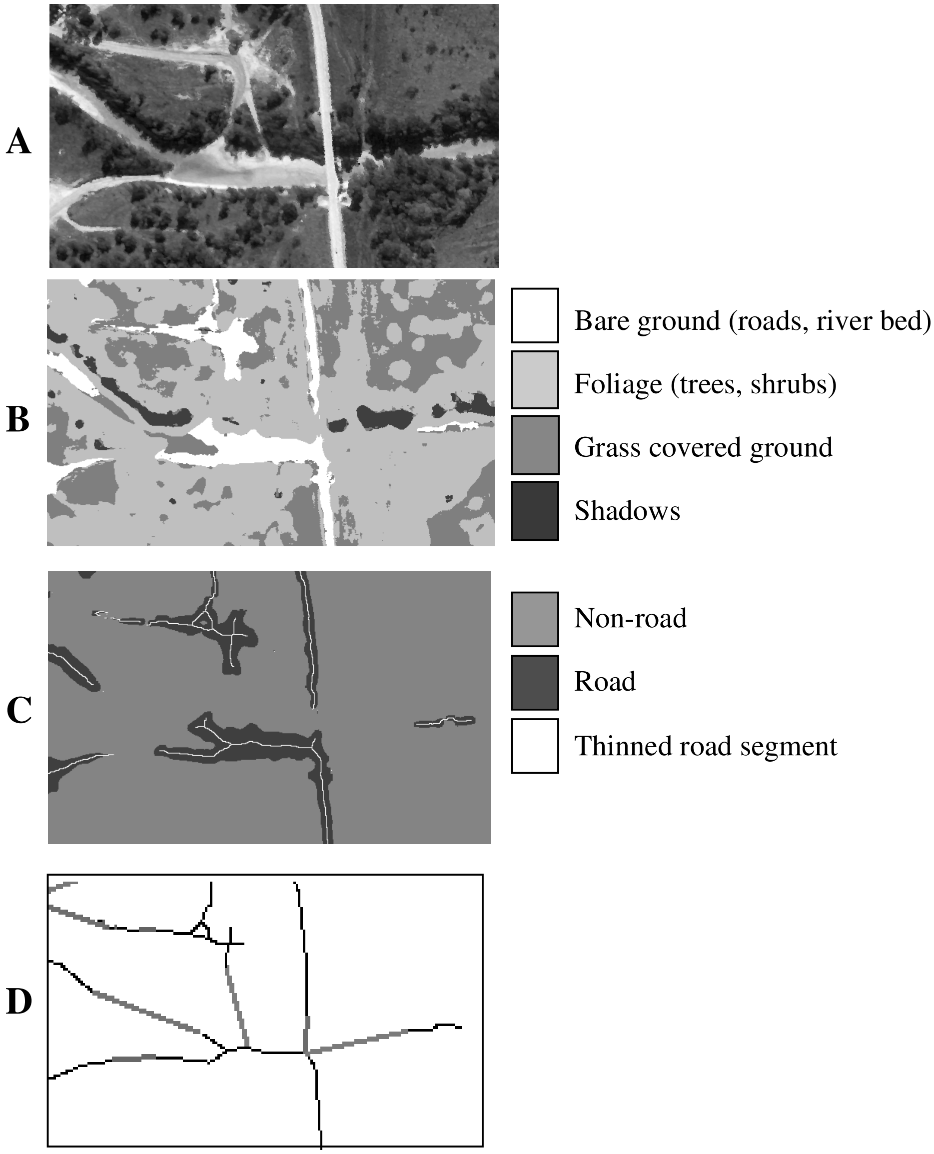


Figure 2. Four 590 by 350 pixel sub-areas of the Fort Hood test area. (A) Ortho-image, (B) Classified image, (C) Thinned road segments, and (D) Final road network.

A perceptual grouping algorithm was employed to close the gaps in the road network. The algorithm attempts to create new road segment between any two vertices on the graph provided that there is evidence for a connecting road segment. Evidence for inserting a road segment included: (1) Close proximity between two vertices, (2) collinearity between the hypothetical segment and an existing one, and (3) lack of a nearly parallel road segment between two vertices.

These factors are taken into account in a heuristic formula that determines the certainty of a hypothetical road segment. The segments where then in order of certainty by a greedy algorithm. After each road segment is added, the certainties are recomputed for the remaining hypothetical road segments. When there is no evidence for adding a new segment the algorithm terminates, and the network is considered complete. Figure 2D shows the final road network for the sub-area shown in Figure 2A. Figure 3 shows a rendered view of the final road network overlaid on the reconstructed terrain. We successfully extracted the road network, except for a few cases where the roads were very thin (a dirt path) or very wide (a broad patch of bare ground). Generally, the system successfully extracted a significant portion of the road network. Future work will concentrate on improvements in grouping strategies, especially near complex junctions, and classification techniques.

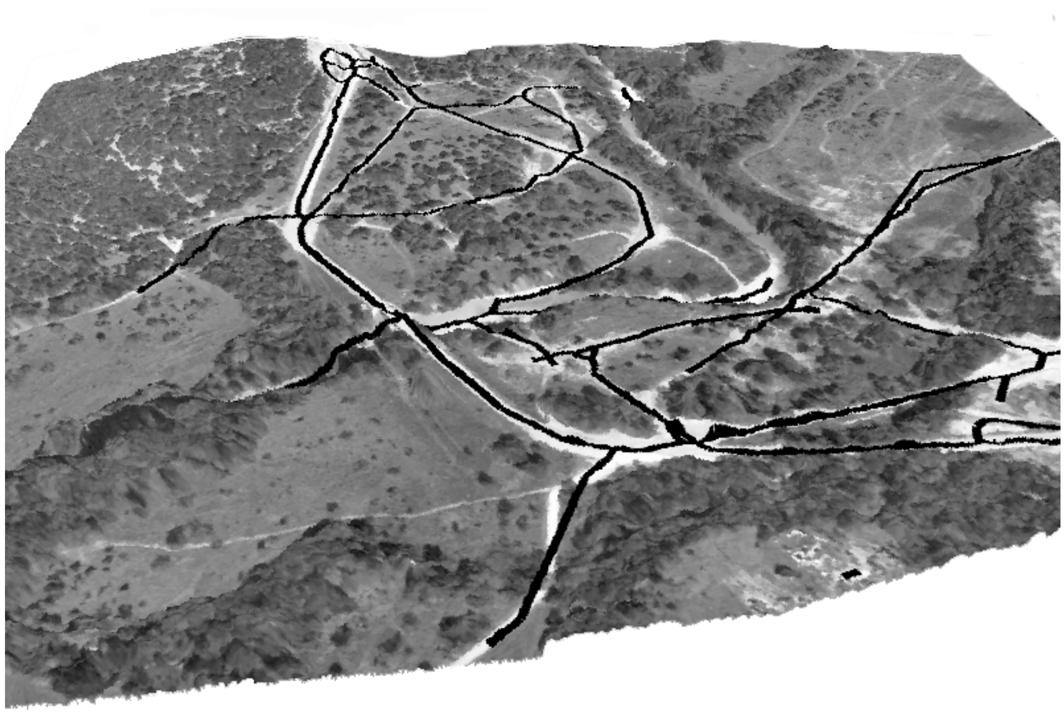


Figure 3. A rendered view of the reconstructed terrain with the extracted road network overlaid in black.

5. Conclusion

This paper demonstrates an effective method for extracting high level symbolic information from aerial images. Starting with a pair of aerial images that may have been taken from oblique viewing angles, the UMass terrain reconstruction system produces a DEM and ortho-image. The ortho-image is classified by terrain type, and a road network extracted using context-sensitive strategies. These strategies apply expected geometric properties of road and perceptual grouping methods to extract a connect road network. The resulting DEM, ortho-image, pixel classifications, and symbolic information greatly enhance the capability of visualization applications to accurately represent complex terrain information.

6. References

R. W. Conners and C.A. Harlow, "A Theoretical Comparison of Texture Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 3, pp. 204-222, 1980.

J. M. H. duBuf, M. Kardan, and M. Spann, "Texture Feature Performance for Image Segmentation," *Pattern Recognition,* Vol. 24, No. 3/4, pp. 291-309, 1990.

M. Fischler, J. Tenenbaum, H. Wolf, "Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique," *Computer Graphics and Image Processing*, Vol. 15, pp. 201-223, 1981.

D. H. Foley and J.W. Sammon, Jr., "An Optimal Set of Discriminant Vectors," *IEEE Tans. of Computers,* Vol. 24, No. 3, pp. 281-289, 1975.

R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 3, No. 6, pp. 610-621, 1973.

K. Liu, Y.Q. Cheng, and J.Y. Yang, "Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion," *Pattern Recognition,* Vol. 26, No. 6, pp. 903-911, 1993.

K. Liu, Y.Q. Cheng, J.Y.Yang, and X. Liu, "An Efficient Algorithm for Foley-Sammon Optimal Set of Discriminant Vectors by Algebraic Method," *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 6, No. 5, pp. 817-829, 1992.

P. P. Ohanian and R.C. Dubes, "Performance Evaluation for Four Classes of Textural Features," *Pattern Recognition,* Vol. 25, No. 8, pp. 819-833, 1992.

H. Schultz, "Terrain Reconstruction from Widely Separated Images", *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II,* SPIE Proceedings Vol. 2486, pp. 113-123, Orlando. FL, April 1995.

Q. Tian, M. Barbero, Z.H. Gu, and S.H. Lee, "Image Classification by the Foley-Sammon Transform," *Optical Engineering,* Vol. 25, No. 7, pp. 834-840, 1986.

J. S. Weszka, C.R. Dyer, and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. 6, No. 4, pp. 269-285, 1976.