Comparing a Piecewise Linear Classifier with Gaussian Maximum-Likelihood and Parallelepiped Classifiers in Terms of Accuracy and Speed

Kai-Yi Huang and P. M. Mausal

Abstract
A piecewise linear classifier (PLC) was developed and tested to determine if it is superior to the Gaussian maximum likelihood classifier (GMLC) and parallelepiped classifier (PPC) for inventories of crop types in terms of classification accuracy and speed. The PLC was developed based upon the concepts of the single-sided decision surface, optimal weight vector, and seniority decision logic. These three classification algorithms were evaluated using multitemporal digitized video data. The PLC was much faster than the GMLC, and yet provided similar classification accuracy. Although the PLC was somewhat slower than the PPC, it provided much higher classification accuracy than did the PPC. The PLC was determined to be an optimal alternative to the GMLC or PPC for inventories of crop types in terms of classification accuracy and processing speed.

Introduction
The Gaussian maximum-likelihood classifier (GMLC) is commonly used to classify multispectral remotely sensed data for land-use/land-cover applications. Representative training samples must be available to estimate the mean vector and covariance matrix for each class to implement this parametric classifier (Swain and Davis, 1978). Moreover, the training samples for each class in each band must have a unimodal spectral signature distribution if an acceptable classification accuracy is to be attained (Lee and Richards, 1984).

The parallelepiped classifier (PPC), which utilizes a nonparametric approach, is an alternative to the GMLC in remote sensing applications. The PPC does not require its data to have a Gaussian distribution, but its implementation does require linear decision surfaces in feature space that are parallel to the feature axes. This limitation makes it difficult to separate overlapping classes in feature space (Lee and Richards, 1984). Also, determining the decision boundaries for the PPC is heavily dependent on human judgement. Therefore, it is more subject to human errors than many other classifiers.

A piecewise linear classifier (PLC), which is a form of a generalized PPC, was developed and tested in this study. The three classification algorithms (GMLC, PPC, and PLC) were evaluated according to their classification accuracy and speed using multitemporal digitized video data. This study attempted to determine if the PLC is best suited for inventories of crop types in terms of classification accuracy and speed of computation.

Study Area and Materials
The study area was located near Weslaco in Hidalgo county, Texas. It was a completely randomized block designed field experiment consisting of plots of the following surface features: (1) cotton, (2) cantaloupe, (3) sorghum, (4) johnsongrass, (5) pigweed, and (6) bare soil (Figure 1). Each of the 24 plots (six treatments and four replications) measured 7.11 by 9.14 m, making the total site dimension 42.67 by 36.56 m (Richardson *et al.*, 1985). However, the fourth replication (not drawn in Figure 1) was excluded from the study due to damage of this portion of the video data file.

The video imaging system was mounted in a single engine 182 Cessna aircraft to collect the data used for this study. The video data were collected on 31 May and 24 July 1983 near noon on moderately sunny days from an altitude of 900 m (3000 ft). The video system consisted of four black-and-white Sony AVC-3450 video cameras, each with a Sony SLO-340 Betamax I video cassette player/recorder (VCR). One of the four cameras was modified with an RCA Ultricon (TM) 4875/U camera tube to have a sensitivity in the 0.4- to 1.1-μm waveband. The other three cameras had a sensitivity in the 0.4- to 0.7-μm waveband (Richardson *et al.*, 1985). Visible and near-infrared narrowband filters were placed over the camera lenses to allow the video system to record any selected light segment within the visible/near-infrared region of the electromagnetic spectrum. The filters and camera aperture settings used were blue (420 to 430 nm), *f*1.8; yellow-green (520 to 550 nm), *f*2.8; red (640 to 670 nm), *f*1.8; and near infrared (850 to 890 nm), *f*8.0. Video images were digitized using a full frame grab to a quantization level of 256 using a Matrox MVP/AT board. Eight 512 by 512 data matrices were created. The eight video images were spatially regis-

Kai-Yi Huang is with the Energy & Resources Laboratories, Industrial Technology Research Institute, Bldg. 24, 195 Chung Hsing Rd., Section 4, Chutung, Hsinchu, Taiwan 310, Republic of China.

P. M. Mausal is with the Department of Geography and Geology, Indiana State University, Terre Haute, IN 47809.

Photogrammetric Engineering & Remote Sensing, Vol. 60, No. 11, November 1994, pp. 1333–1338.

0099-1112/94/6011-1333$3.00/0

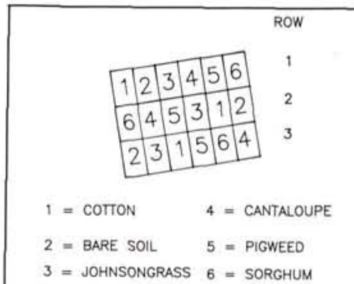

Figure 1. Plot identification map of study area (not drawn to scale).

tered using the CONTROL POINT and WARP modules of the International Imaging System (I^2S) software package (Richardson *et al.*, 1985).

The video data file (192 columns by 163 lines) was stored on a floppy diskette in band interleaved by line (BIL) format. There were eight channels in this file. Channels 1 to 4 were acquired on 24 July. Channels 5 to 8 were acquired on 31 May. Channels 1 and 5 were blue bands, channels 2 and 6 were red bands, channels 3 and 7 were yellow-green bands, and channels 4 and 8 were near-infrared bands (Huang, 1988). The accuracy assessment was performed over the entire study area, as shown in Figure 1. A digital ground truth mask was used to calculate classification accuracies of the three classifiers. Border areas of mask between plots were eliminated in order that incomplete pixels or overlapping plant and soil areas were excluded from the study.

Methodology

To appreciate the development of the piecewise linear classification algorithm, it is necessary to realize the theoretical basis of a simple linear classification. The family of line ar discriminant functions can be expressed in the form as follows (Nilsson, 1965):

$$G(\mathbf{X}) = W_1 * X_1 + W_2 * X_2 + \cdots + W_d * X_d + W_{d+1},$$

where $W_1, W_2, ..., W_d, W_{d+1}$ are weighting coefficients. G is a linear function of the components of $\mathbf{X}$. A simple linear classification is performed by a linear classifier. A linear classifier employs a linear discriminant function to partition the feature space into two regions. The linear discriminant function can be viewed as a separating surface in which the simplest form is a hyperplane. A hyperplane partitions the feature space into two: i.e.,

$$G(\mathbf{X}) = \mathbf{W} \cdot \mathbf{X} > 0 \quad \text{and} \quad G(\mathbf{X}) \;\; 0$$

where $\mathbf{W} = < W_1, W_2, ..., W_d, W_{d+1} >$,
$\mathbf{X} = < X_1, X_2, ..., X_d, 1 >$, and
d = dimension of the feature space.

Distinctive classes of data cannot always be separated by a single hyperplane (linear classifier). Therefore, a set (several sets) of linear classifiers is (are) generally used in feature classification. A committee classifier consists of a set of linear classifiers and decision logic (Lee and Richards, 1985). Hence, implementing a committee classifier, a set of linear classifiers and their sequence have to be specified. A variety of training algorithms (Nilsson, 1965; Osborne, 1977; Takiyama, 1978; Mizoguchi *et al.*, 1980; Byers and Richards, 1981) have been proposed for generating a committee classifier. These training algorithms, however, have their own weaknesses that limit their practicality (Lee and Richards, 1984). The training algorithm proposed by Lee and Richards (1984) is based upon the concepts of single-sided decision surfaces (Lee, 1982) and optimization method (Clark and Gonzalez, 1981). It seems to work well, but it may fail due to problems, an example of which is discussed later in this section (Huang, 1988).

The improved training algorithm used in this study was derived from the one proposed by Lee and Richards (1984). This algorithm is based upon the concepts of single-sided decision surfaces, optimal weight vectors, and seniority decision logic. An augmented pixel vector is defined as

$$\mathbf{A} = < \mathbf{U}^t, 1 >^t$$

where

$$\mathbf{U} = \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_n \end{bmatrix} \quad \mathbf{U}^t = < U_1, U_2, ..., U_n >$$

Assume that there are two classes of pixels in the feature space. A weight vector $\mathbf{W}$ is viewed to implement a single-sided decision surface if all augmented pixel vectors having the positive dot product values with $\mathbf{W}$ belong to the same class. Pixels lying on the other side of $\mathbf{W}$, that is, with negative dot product values, can be from either class. This is illustrated in Figure 2 and is defined as follows:

$$S_1 = \{\mathbf{A} \;\; S \mid \mathbf{W} \cdot \mathbf{A} > 0\}$$
$$S_2 = \{\mathbf{A} \;\; S \mid \mathbf{W} \cdot \mathbf{A} \;\; 0\}$$
$$S = S_1 + S_2$$

$\mathbf{W}$ is defined as a single-sided decision surface if the subset S_1 consists of pixels from one class only.

The second concept is to execute a loop to search for a weight vector that can correctly classify more training pixels than the previous iteration. If a better weight vector is found, then another iteration follows. This loop keeps executing until no better solution is found. This solution, placing the

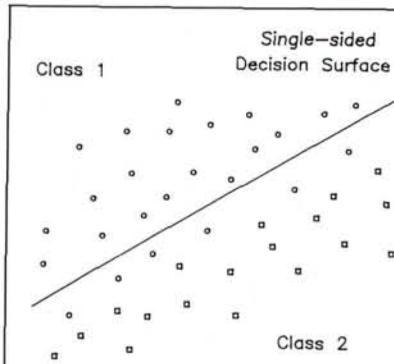

Figure 2. Concept of a single-sided decision surface in two-dimensional space (after Lee and Richards (1985)).

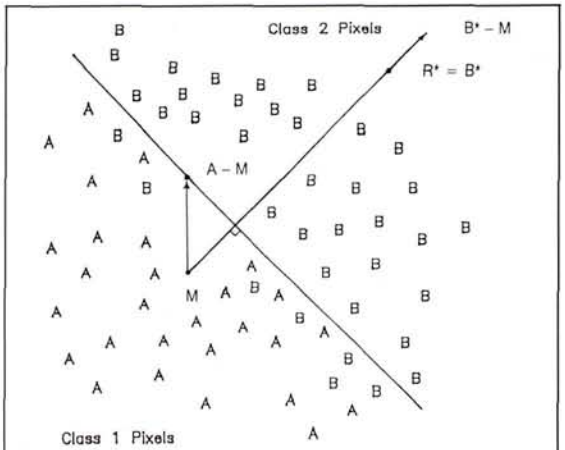

Figure 3. Initialization of a single-sided decision surface. The initial estimate passes through **A** — **M** (after Lee and Richards (1984)).

most pixels of the same class into S_1, is called an optimal weight vector. Before the optimization procedure is conducted, the training algorithm must be capable of generating an initial approximation to an optimal single-sided decision surface. The initialization approach is illustrated in Figure 3 and discussed in the following.

Let the pixels be augmented and expressed by the general column vector **X**. One pixel **R** is chosen from the training pixels so that $|\mathbf{R} - \mathbf{M}| > |\mathbf{X} - \mathbf{M}|$ for all **X**, where **M** is an arbitrarily chosen position vector in the feature space. Let **M** be the mean vector of all the training pixels. Under this condition, **R** and **M** define a line from the mean **M** to the farthest pixel (which could be from either class). Assume that there are two classes of training pixels represented by **A** and **B**, respectively, so that $\mathbf{R}\ \{ \mathbf{A} \} \neq \{ \mathbf{B} \}$. Assume further that $\mathbf{R}^*$ came from $\{ \mathbf{B} \}$ and now is denoted as $\mathbf{B}^*$. The next step is to find a pixel $\mathbf{A}^*\ \{ \mathbf{A} \}$ so that

$$(\mathbf{A}^* - \mathbf{M}) \cdot (\mathbf{B}^* - \mathbf{M}) > (\mathbf{A} - \mathbf{M}) \cdot (\mathbf{B}^* - \mathbf{M}).$$

$\mathbf{A}^*$ having the largest positive projection onto $(\mathbf{B}^* - \mathbf{M})$ is the pixel from $\{ \mathbf{A} \}$. It must be cautioned that the largest positive projection of $\mathbf{A}^*$ onto $(\mathbf{B}^* - \mathbf{M})$ cannot always be found because all of the inner product values may be negative. In this case, $\mathbf{A}^*$ having the biggest negative projection onto $(\mathbf{B}^* - \mathbf{M})$ is adopted. This is one of the weaknesses found in the training algorithm proposed by Lee and Richards (Huang, 1988). A line perpendicular to $(\mathbf{B}^* - \mathbf{M})$ and passing through $\mathbf{A}^*$ will be a good initial approximation to an optimal single-sided decision surface. It is defined by the weight vector:

$$\mathbf{W} = < (\mathbf{B}^* - \mathbf{M})^t, - \mathbf{A}^* \cdot (\mathbf{B}^* - \mathbf{M}) >^t$$

Because it is not necessarily optimal, an optimization procedure has to be performed to move the initial weight vector to a final effective position. Let **Y** be a matrix of row pixel vectors equal in number to the dimension of the feature space. Let one row of matrix **Y** be $\mathbf{A}^*$ and the remaining rows be a choice of the appropriate number of pixels from S_2. The decision surface passing through those pixels placed in **Y** serves as a solution to $\mathbf{YW} = 0$. This new **W** is evaluated to see if it places more pixels into S_1 than does the previous **W**. If it does so, then this new **W** replaces the old one. Obviously, this process has to be repeated by allowing all candidate pixels in S_2 to be placed in the matrix **Y** until no better solution is found. This approach has been found to work well (Clark and Gonzalez, 1981; Lee and Richards, 1984).

The third concept of the PLC is the seniority decision logic proposed by Osborne (1977). The linear classifiers in a committee are ranked in seniority. Each linear classifier can choose either to determine the class membership of a pixel when the dot product value is positive or to abstain when the dot product value is negative or zero. Should it abstain, then the next highest ranking classifier is evaluated. This process is continued until one of the classifiers does not abstain. A committee classifier consists of a set of linear classifiers of which seniority ranks have been specified at the same time when linear classifiers are generated from the training algorithm. Each classifier (single-sided decision surface) is represented by a weight vector along with its seniority rank. When an unknown pixel has to be classified, it is first assessed using the highest ranking weight vector. Should that pixel be found to belong to group S_1, then the process terminates. Alternatively, if that pixel falls into group S_2, then the next highest ranking weight vector is used to assess that pixel, and so on until it can be properly classified. Such a sequence of linear classifiers is implemented according to the seniority rank of each member in a committee classifier. The whole process for a hypothetical seniority committee classifier of three members is illustrated in Figure 4.

Results and Discussion

The seniority decision logic committee classifier used in the study was expanded from a two-class, two-channel linear classifier. Apparently, testing this special case was crucial before it could be expanded into a multi-class, two-channel PLC and eventually into a multi-class, multi-channel PLC. To examine the robustness of the training algorithm, many tests for different cases had been performed using dummy data

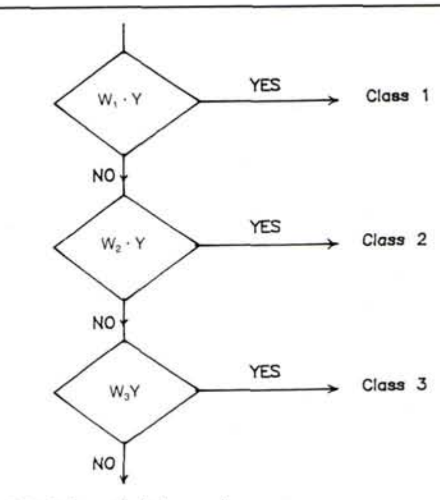

Figure 4. Classification procedure for a hypothetical seniority committee of three members (after Lee and Richards (1985)).

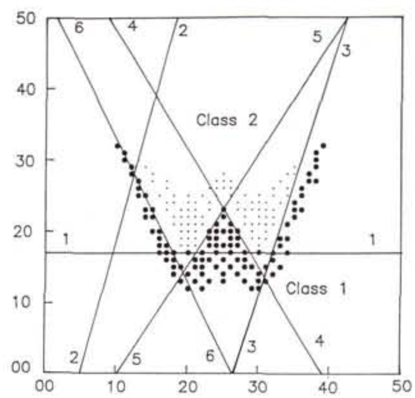

Figure 5. The butterfly diagram is computer output illustrating a two-class, two-channel case.

Table 1. Weight Vectors and Corresponding Line Equations and Seniority Ranks

Weight Vectors	Line Equations	Seniority Ranks
$\mathbf{W}_1$	$0.00X - 2.00Y + 34.00 = 0.00$	1
$\mathbf{W}_2$	$-30.00X + 8.01Y + 149.58 = 0.00$	2
$\mathbf{W}_3$	$114.19X - 36.00Y - 3030.65 = 0.00$	3
$\mathbf{W}_4$	$12.21X + 7.40Y - 477.79 = 0.00$	4
$\mathbf{W}_5$	$35.85X - 23.00Y - 364.60 = 0.00$	5
$\mathbf{W}_6$	$-2.00X - 1.00Y + 53.00 = 0.00$	6

Table 2. Weight Vectors and Corresponding Line Equations and Seniority Ranks

Weight Vectors	Line Equations	Seniority Ranks
$\mathbf{W}_1$	$-12.00X - 157.47Y + 2105.66 = 0.00$	1
$\mathbf{W}_2$	$8.88X - 13.13Y + 176.96 = 0.00$	2
$\mathbf{W}_3$	$10.90X + 2.49Y - 206.94 = 0.00$	3
$\mathbf{W}_4$	$-28.00X + 135.15Y - 3532.28 = 0.00$	4

and remotely sensed data. One example for a two-class, two-channel case and another example for a multi-class, two-channel case are presented in the following.

Testing a two-class, two-channel case is illustrated in Figure 5. As shown in the diagram, the graph looks like a butterfly consisting of two sets of dummy data. There was only one committee classifier. This committee classifier was composed of six linear classifiers represented by six straight lines, respectively. The equations of six straight lines along with their seniority ranks are given in Table 1.

Testing of a multi-class, two-channel training algorithm is illustrated in Figure 6. There were four seniority committee classifiers and each committee had only one member. The equations of four straight lines along with their seniority ranks are given in Table 2. Pixels in class 1 are separated from classes 5, 4, 3, and 2 by $\mathbf{W}_1$. Pixels in class 3 are separated from classes 5, 4, and 2 by $\mathbf{W}_2$. Pixels in class 4 are separated from classes 5 and 2 by $\mathbf{W}_3$. Finally, pixels in class 5 are separated from class 2 by $\mathbf{W}_4$.

The multi-class, multi-channel training algorithm was tested using six randomly selected channel combinations of the video image data. In order to be concise, only one of the classification maps, generated from channels 1, 2, 4 and 6 of the video image data, is shown in Figure 7. Because this classification map was oversize, every five lines and every five columns were displayed.

According to the test results, there seemed to be few problems with the PLC. In addition, six weight vectors were required to separate two clusters for the first example; four weight vectors were required to separate five clusters for the second example. The major difference is that the former training data were less compact than the latter training data in terms of their geometric shapes and densities (dots per unit area). Consequently, using compact training data may be helpful for reducing the number of weight vectors generated.

The GMLC and PPC coded in FORTRAN were executed on an IBM-4361 computer. The PLC coded in Turbo Pascal was executed on a 386 microcomputer. Because they were coded by different programmers, there might exist biases in programming style and language. Obviously, it is meaningless to compare their CPU time taken on different computer systems. It is desirable to count the number of multiplications and additions required to determine the potential membership of a pixel to one class for each classifier. Therefore, the number of computations was used as an alternative to computer CPU time evaluating the efficiency of each classifier.

Let N be the number of bands used for classification. The PPC required $2N$ comparisons of the spectral components of a pixel with decision boundaries to determine the potential membership of that pixel to a class. The PPC required no additions and multiplications to do that job. The PLC required N multiplications and N additions to do the same job stated above. The GMLC required $N^2 + N$ multiplications and $N^2 + 2N + 1$ additions to do the same job. Therefore, the PPC was the fastest among them without reference to an overlapping classes problem. Considering multiplications only, the GMLC took $N + 1$ times longer than the PLC to do that job. Thus, the PLC was second to the PPC in efficiency, but was far more efficient than the GMLC.

Accuracy assessment was conducted over the entire study area by comparing the classified video images with the ground-truth mask. Overall classification accuracies of the

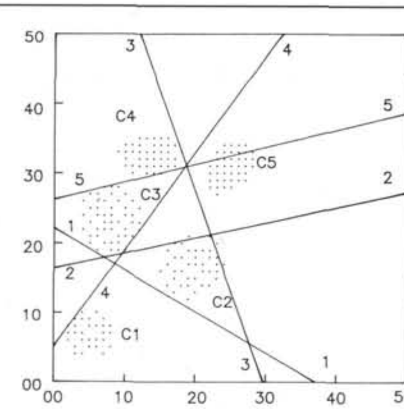

Figure 6. Diagram of the second example illustrates a five-class, two-channel case.

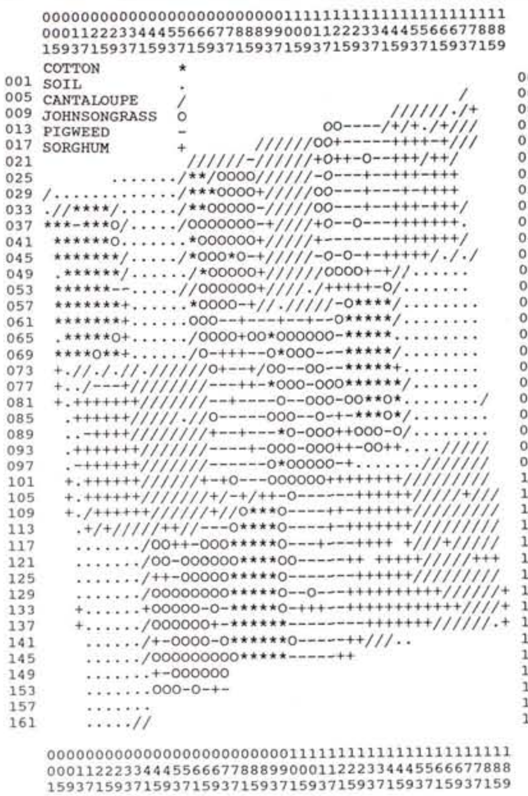

Figure 7. Classification map generated from channels 1, 2, 4, and 6 of the video image data using the piecewise linear classifier.

TABLE 3. OVERALL CLASSIFICATION ACCURACIES OF THE THREE CLASSIFIERS ON SIX CHANNEL COMBINATIONS

Channel Combinations	Overall Classification Accuracy (%)		
	PLC	GMLC	PPC
1,2,4	85.11	83.53	73.62
1,2,3,4	81.05	81.72	68.70
6,7,8	77.53	82.05	77.57
5,6,7,8	77.06	83.57	70.64
1,2,4,6	85.42	85.60	75.73
1,2,4,8	87.52	88.30	80.55
Average	82.28	84.13	74.47

three classifiers corresponding to six channel combinations were computed and are presented in Table 3. Through implementation of the Student's t-test, there was no significant difference in overall classification accuracies between the PLC and GMLC at the 0.05 rejection level. However, there was a significant difference in overall classification accuracies between the PLC and PPC at the 0.05 rejection level.

Conclusions

The piecewise linear classifier developed in the study was found to work effectively when applied to video image data. Meanwhile, the more compact the training data (convex data), the smaller was the number of single-sided decision surfaces generated. The PLC was far more efficient than the GMLC, and yet was similar in classification accuracy to the GMLC. Although the PLC was somewhat slower than the PPC without reference to an overlapping classes problem, the PLC was much more accurate than the PPC. Collectively, the PLC was considered as an optimal alternative to the GMLC or PPC when applied to inventories of crop types in terms of classification accuracy and processing speed.

Acknowledgments

The authors are grateful to William J. Kramber for his advice and suggestions. The authors appreciate the cooperation of the USDA-ARS Remote Sensing Unit in Weslaco, Texas for providing the multispectral video data and ground truth data in this research. The authors also want to thank Miss Chin-May Lu for her typing this manuscript.

References

Byers, R. S., and J. A. Richards, 1981. A Piecewise Linear Classifier for Remote Sensing Applications. *Proceedings of the Second Australian Remote Sensing Conference*, Canberra, P.6.2.1.

Clark, D. C., and R. C. Gonzalez, 1981. Optimal Solution of Linear Inequalities with Application to Pattern Recognition. *I.E.E.E. Trans. Pattern Anal. Machine Intell.*, PAMI-3(6):643-655.

Huang, K. Y., 1988. *The Use of Piecewise Linear Classifier in Video Image Analysis.* Doctoral Dissertation, Indiana State University, Terre Haute, Indiana.

Lee, T., 1982. *Committee Classification of Remote Sensing Data.* B.E. Thesis, University of New South Wales, Kensington.

Lee, T., and J. A. Richards, 1984. Piecewise Linear Classification Using Seniority Logic Committee Methods with Application of Remote Sensing. *Pattern Recognition*, 17:453-464.

———, 1985. A Low Cost Classifier for Multi-temporal Applications. *International Journal of Remote Sensing*, 6:1405-1417.

Mizoguchi, R., M. Shimura, and O. Kakusho, 1980. A New Algorithm for Constructing Piecewise Linear Discriminant Function. *Proceedings of IEEE Conference on Pattern Recognition*, Florida, pp. 666-670.

Nilsson, N. J., 1965. *Learning Machines.* McGraw Hill Inc., New York.

Osborne, M., 1977. Seniority Logic: A Logic for a Committee Machine. *IEEE Trans on Computers*, C-26(12):1302-1306.

Richardson, A. J., R. M. Menges, and P. R. Nixon, 1985. Distinguishing Weed from Crop Plants Using Video Remote Sensing. *Photogrammetric Engineering & Remote Sensing*, 51(11):1785-1790.

Swain, P. H., and S. M. Davis (editors), 1978. *Remote Sensing, The Quantitative Approach.* McGraw Hill Inc., New York.

Takiyama, R., 1978. A General Method for Training the Committee Machine. *Pattern Recognition*, 10:255-259.

(Received 21 January 1992; revised and accepted 14 April 1993)

Kai-Yi Huang

Kai-Yi Huang is a Researcher at the Energy and Resources Laboratories, Industrial Technology of Research Institute in Taiwan, Republic of China. He specializes in applications of remote sensing and GIS to natural resources management and development and in digital image processing. He received his B.S. in Geography from National Taiwan University, and his M.A. and Ph.D. from Indiana State University. He is currently directing some projects on unsupervised/supervised classification algorithms and GIS applications.

Paul W. Mausel
Paul W. Mausel developed the Indiana State University Remote Sensing Lab (ISURSL) in 1974 and served as its director during most of its 20 year history. He has published approximately 100 referred articles and book chapters on remote sensing, soils, and GIS topics. Research areas include Amazon deforestation and land-use changes; videography applications; sea level rise impact modeling; and integration of satellite data into GIS systems. Funds from EPA, NSF, NIGEC (DOE), and the Indiana Dept. of Transportation have supported these research interests.