

# Modified Highest Confidence First Classification

Paul C. Van Deusen

## Abstract

A classification procedure is developed that distinguishes between pixels that are clearly associated with a given class versus those where class assignment is uncertain. Alternatively, most commonly used classification algorithms force each pixel into a single class without regard to certainty. By classifying pixels in order of certainty and considering spatial context, pixels with weak observational evidence for classification are prevented from contributing to their neighbor's decisions. Subsequently, a better decision is made for the uncertain pixels by considering the previously classified neighbors. Degrees of certainty measures can assist in later map accuracy assessment by allowing for stratified sampling of zones having similar certainty levels.

## Introduction

The objective of this work was to develop a classification procedure that recognizes that pixels have varying degrees of observational evidence favoring membership in a single class. A degree of certainty (DOC) measure is defined for each pixel and utilized in determining the manner and order in which the pixel is assigned to a class. The DOC is then available for other uses, such as creation of certainty strata for map accuracy assessment. Related work (Wood and Foody, 1989; Foody *et al.*, 1992; Trodd *et al.*, 1989) has discussed utilizing information on strength of class membership derived from maximum-likelihood calculations to identify potentially misclassified cases and for directing ground surveys.

The algorithm developed here is called the modified highest confidence first (MHCF) procedure, because it is motivated by the highest confidence first (HCF) algorithm of Chou and Brown (1990). Both HCF and MHCF can be viewed as modifications of the iterated conditional modes (ICM) procedure developed by Besag (1986). ICM initializes all pixels by committing them to the class determined by application of the standard maximum-likelihood (ML) classifier. ML utilizes observed data associated with each pixel to make the classification without reference to information from neighbors. Subsequent passes of ICM utilize neighborhood information within a Bayesian theoretical framework to refine the classification.

Klein and Press (1992) criticize ICM for its final outcome being heavily dependent on the initial classification and they propose modifications to all stages of ICM. Likewise, ICM has been criticized for the way it handles neighborhood information in later stages. Specifically, Owen (1986) noted that if a pixel has eight A neighbors all with *a-posteriori* A-probabilities of 0.99, then ICM handles it the same way as the situation where three pixels were 0.99 and five were 0.51. The point of these criticisms is that ICM and ML in their usual forms fail to consider the level of certainty with which a pixel is associated with a given class.

A review is given of the ML and ICM classification algorithms in the next two sections. Then the DOC measure and the MHCF algorithm are described. After a discussion on choosing the DOC cutoff value for MHCF and a critique of MHCF versus HCF, applications are made to classifying a human face image and a Landsat TM image.

## Maximum-Likelihood Classifier

One of the most widely employed methods for pixel classification is the ML procedure. A brief review is given to clarify the details of the implementation used here. Denote a pixel location by  $i$  and its observed data vector by  $\mathbf{y}_i$ , which contains the spectral information for the pixel. The unobserved class,  $x_i$ , may be one of  $K$  unordered class values. Although a pixel may contain more than one class in reality, we seek here a unique assignment to the most likely class.

The ML procedure consists of maximizing the likelihood function

$$L(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n f(\mathbf{y}_i|x_i) \quad (1)$$

which assumes that the observational evidence,  $\mathbf{y}_i$ , given  $x_i$  is independently distributed for each pixel. If a multivariate normal distribution is assumed, this is equivalent to minimizing, for each pixel, the function

$$\mathbf{D}(x_i) = \frac{1}{2} (\mathbf{y}_i - \mathbf{u}(x_i))' \Sigma_x^{-1} (\mathbf{y}_i - \mathbf{u}(x_i)) + \frac{1}{2} \log |\Sigma_x| \quad (2)$$

where  $\mathbf{u}(x_i)$  is the mean vector for the class  $x_i$ , and  $\Sigma_x$  is the covariance matrix for each class. The mean and covariance for each class are derived from a reference (training) set of observations. In the case where each class has the same covariance matrix,  $\mathbf{D}(x_i)$  is equivalent to linear discriminant analysis.

## Iterated Conditional Modes Classifier

One feature of the standard ML classifier that may be undesirable is that no spatial information is included. In recognition of this deficiency, Switzer (1980) has developed a simple modification that incorporates average  $y$  values from neighbors into the ML approach. Haslett (1985) developed a more formal ML approach to incorporate observed  $y$  values of certain neighbors of pixel  $i$ . An approach developed in the seminal paper of Geman and Geman (1984) utilized Gibbs distributions within a Bayesian framework to incorporate spatial information. Besag (1986) developed a procedure that

Photogrammetric Engineering & Remote Sensing,  
Vol. 61, No. 4, April 1995, pp. 419-425.

0099-1112/95/6104-419\$3.00/0

© 1995 American Society for Photogrammetry  
and Remote Sensing

NCASI, Department of Civil Engineering, Anderson Hall,  
Tufts University, Medford, MA 02155.

is motivated by Geman and Geman's work, but is computationally more tractable.

Formally, Besag's ICM procedure begins with Bayes rule

$$p(x|y) \propto L(y|x)p(x) \quad (3)$$

where,  $p(x)$  represents *a priori* knowledge about the unobserved true classification,  $L(y|x)$  is the likelihood function discussed above, and  $p(x|y)$  is the *a posteriori* distribution of  $x$  given the observations. The *a priori* knowledge,  $p(x)$ , in Equation 3 is represented by a Markov random field (Mrf) with a locally dependent conditional distribution (a Gibbs distribution). Thus, if we let  $X_{s/i}$  denote the class labels everywhere except pixel  $i$  and let  $X_{\delta_i}$  be the neighbors of  $i$ , we have

$$p(x_i|X_{s/i}) = p(x_i|X_{\delta_i}) \quad (4)$$

which simply states that the conditional *a priori* distribution of the pixel label depends only on the neighboring pixel labels. The neighborhood being considered here consists of the eight nearest neighbors.

Besag's ICM procedure is expressed in a form similar to Equation 3 as

$$p(x_i|y, \hat{x}_{s/i}) \propto f(y_i|x_i)p(x_i|\hat{x}_{\delta_i}) \quad (5)$$

The "hats" imply the use of estimated class assignments from the previous iterations in the current iteration of ICM. The specific conditional Mrf considered here involves a K-class distribution with pairwise interaction between neighbors: i.e.,

$$p(x_i=k|X_{\delta_i}) \propto \exp\left(\alpha_k - \sum_{j \neq k} \beta_{jk}u_i(j)\right) \quad (6)$$

where  $\beta_{jk} = \beta_{kj}$  and  $u_i(j)$  is the number of neighbors with label (color)  $j$ . Besag (1986) demonstrated Equation 6 with  $\alpha_k$  and  $\beta_{jk}$  being the same for all classes. Thus,  $\alpha_k$  is absorbed into the proportionality constant to give the following ICM formulation:

$$E(x_i) = D(x_i) - \beta u(x_i) \quad (7)$$

where  $D(x_i)$  is given in Equation 2. This implementation of ICM assumes a normal likelihood and a simple pairwise interaction Mrf to describe the *a priori* knowledge. The ICM classification assigns the label to  $x_i$  that minimizes the energy  $E(x_i)$ . The energy analogy comes from Geman and Geman's (1984) use of simulated annealing to find a maximum *a posteriori* solution to Equation 3 (annealing refers to the schedule for cooling metals to obtain good crystalline structure).

The method of implementing ICM used here is to iteratively increase  $\beta$  from 0 to 0.5 to 1.0. This initializes the procedure with the standard ML classifier and no use of spatial information. As  $\beta$  increases, the spatial information is given more weight.

### Modified HCF Algorithm

Chou and Brown (1990) present HCF as a modification of ICM. Rather than force every pixel into a class immediately, HCF classifies the pixels in order according to a stability measure that is related to differences of the energy function (Equation 7). MHCF classifies all pixels at once that exceed a user defined DOC cutoff value. Thus, the formation of certainty strata is more natural with MHCF than with HCF, and MHCF approaches ICM as the DOC cutoff approaches 0. Likewise, MHCF

approximates HCF as the DOC cutoff approaches a value large enough to allow only one pixel to be classified at each iteration. Unfortunately, as discussed below, the HCF procedure is prone to allowing the more certain classes to dominate the final results to the point that other classes may be totally suppressed.

The HCF and MHCF procedures are only meaningful within a Bayesian context because, without spatial information, they would lead to the same result as the ML procedure. Unlike ICM or ML, HCF and MHCF require a null class into which all pixels are initially assigned. Pixels in the null class do not contribute to neighbors decisions, but committed neighbors help uncertain pixels make their decisions later in the process.

The MHCF procedure is similar in spirit to HCF, but uses a slightly different measure of certainty for each pixel,  $G_i$ , and classifies all pixels with  $G_i \geq G_c$ , where  $G_c$  is a user defined cutoff value. MHCF is identical to ICM when  $G_c = 0$ . Consider the sorted energy values (Equation 7)  $E^{(1)}, \dots, E^{(N)}$ , where  $E^{(1)}$  is the minimum and the associated  $x_i$  value represents the class with minimum energy. With ICM, the class assignment for pixel  $i$  would be determined solely by  $E^{(1)}$ . Now define the DOC measure to be the likelihood ratio

$$G_i = E_i^{(2)} - E_i^{(1)} \quad (8)$$

which by definition has  $G_i \geq 0$ . This is the difference between the energy of the classes that are closest and second closest to pixel  $i$ . Clearly, the closer  $G_i$  is to 0, the less certain the class assignment.

### Choosing the Cutoff Value

Selecting the initial  $G_c$  value needs careful consideration. If at least one pixel from each class doesn't exceed the  $G_c$  value, that class will never appear in the final results. This is because, when  $D_i^{(2)} - D_i^{(1)} < G_c$  for the first ML pass of MHCF, the only way that  $G_i$  can later be made to exceed  $G_c$  is based on neighborhood information. But if class  $k$  initially gets no committed pixels, then clearly no  $G_i$  value can later exceed  $G_c$  based on neighboring class  $k$  pixels. Two methods are given for choosing  $G_c$ . The first allows the user close control over how many pixels are classified in the ML pass, and the second method is based on analytical considerations.

The first method begins by taking a sample of  $G_i$  values from the image to estimate percentiles of the  $G$ -distribution. For example, if a sample of size 1000 is taken and  $G^{(1)}, \dots, G^{(1000)}$  represent the ordered  $G$ -values, then letting  $G_c = G^{(750)}$  sets the cutoff value at the 75th percentile. This would lead to approximately 25 percent of the pixels being classified by the first iteration of MHCF. The  $\beta$ -parameter is set to 0.5 for the second MHCF iteration and pixels with initial  $G$ -values slightly below  $G_c$  would have the potential to rise above  $G_c$  and be classified based on neighborhood information. This process continues until the  $\beta=1.0$  iterations are complete. The  $G_c$  value can be moved toward zero in subsequent iterations holding  $\beta = 1.0$  until all pixels are committed. In the example applications in the next section,  $G_c$  initially corresponds to the 30th percentile and is moved to zero with two intermediate steps:  $G_c/2$  and  $G_c/4$ . Iteration can continue with  $\beta = 1.0$  and  $G_c = 0$  until no further changes occur. In practice, this seems to require only one or two iterations.

Now consider the components of the DOC measure:

$$G_i = E_i^{(2)} - E_i^{(1)} = D_i^{(2)} - D_i^{(1)} - \beta (u_i^{(2)} - u_i^{(1)}) \quad (9)$$

where  $D_i^{(1)}$  and  $D_i^{(2)}$  are the ML components and  $u_i^{(1)}$  and  $u_i^{(2)}$

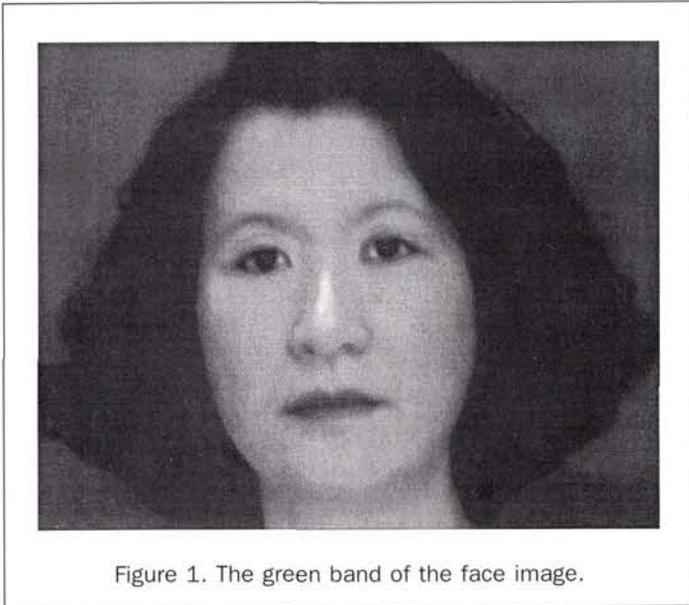


Figure 1. The green band of the face image.

are the associated counts of neighboring pixels. In the later stages of MHCF, uncommitted pixels will have  $D_i^{(2)} \approx D_i^{(1)}$  so we can assume that

$$G_i \approx -\beta (u_i^{(2)} - u_i^{(1)}) \quad (10)$$

Therefore, a pixel with eight neighbors of the same color in the closest class has  $G_i \approx \beta * 8$  and with seven neighbors in the closest class and one in the second closest has  $G_i \approx \beta * (7 - 1)$ . However, pixels with the weakest evidence in favor of a class might have  $u_i^{(1)} - u_i^{(2)} = 1$ , so setting  $G_c = \beta_{\max} * 1.0$  could be considered a minimum starting value, where  $\beta_{\max} = 1.0$  is suggested here. This will allow pixels with quite weak observational and spatial evidence to commit in the final stages of MHCF. Then  $\beta$  could be moved towards zero with two intermediate steps, say  $G_c = 0.5$  and  $G_c = 0.25$ . Any pixels that haven't committed before  $G_c = 0$  can be considered very uncertain indeed.

The second method for choosing  $G_c$  follows from the above analytical discussions. Simply set  $G_c = \beta_{\max}$  initially and in the final iterations move  $G_c$  toward zero according to the schedule discussed above. The first method allows the user complete control and for the opportunity to locate the pixels with the highest DOC values. The second method is more expedient and requires no user input.

To summarize, there are two major differences between HCF and MHCF. First, the  $G_i$  values for committed pixels in HCF are based on the difference between the current class assignment and the next closest class, rather than on the difference between the two closest classes as with MHCF. However, in most cases these will be the same. The second difference is more substantial and is related to choosing the cutoff value,  $G_c$ . HCF always classifies the pixel with the largest  $G_i$  value, which leads to the possibility that some classes may never appear in the final result. Heuristically, this is because the most certain classes commit first and then have an undue spatial influence on uncommitted and less certain neighbors. As discussed above, the MHCF user controls this undesirable situation when selecting a  $G_c$  value. MHCF would be essentially identical to HCF if at each iteration  $G_c$  were set equal to

the maximum  $G_i$  value so that only one pixel could commit. This extreme situation would clearly be undesirable, and this is why the HCF algorithm is not compared in the examples below.

### Example Applications

Comparing alternative classification procedures in a meaningful way is difficult. Image summarization is more subtle than it first appears, as Ripley (1986) demonstrates with two classifications of heather that visually appear quite different, but are indistinguishable according to several standard summary statistics. Because of this difficulty, the first application uses a face as opposed to a natural scene because we know what a face is supposed to look like and can therefore judge the results of the classification algorithms quite easily. The second application uses Landsat TM data of a natural scene.

The specific implementation of MHCF recommended and used here is now outlined. First, the user selects a cutoff value,  $G_c$ , as discussed above. For both applications,  $G_c$  corresponds to the 30th percentile so that approximately 70 percent of the pixels are classified in the first ML pass of MHCF. Then the following steps are taken: (1) Initialize the classification by accepting the ML results for all pixels where  $G_i \geq G_c$ ; (2) accept the ICM results with  $\beta = 0.5$  for all pixels with recomputed  $G_i$ 's such that  $G_i \geq G_c$ ; (3) repeat step 2 with  $\beta = 1.0$ ; and (4) repeat step 2 with  $\beta = 1.0$  and with each of three  $G_c$  values, i.e.,  $G_c/2$ ,  $G_c/4$ ,  $G_c = 0$ . This particular implementation of MHCF assures that all pixels will be classified and seems to work well, but it is by no means the only possible implementation. One could iterate to convergence at each step and move  $G_c$  toward zero either more quickly or more slowly.

The ICM implementation used here involves just three iterations. First, initialize with the ML classification. Second, implement ICM as defined in Equation 7 with  $\beta = 0.5$ . Third, iterate on the results of the second step with  $\beta = 1.0$ . Software for these applications was programmed with MATLAB<sup>®</sup> (The Math Works, Inc., Natick, Mass.).

### Application to a Face

For the first application, the MHCF procedure is compared to ML and ICM by classifying a human face picture (Figure 1) into four classes: hair, background, skin, and eyes, which are denoted as classes 1, 2, 3, and 4. This image was captured with a digital camera in 24-bit RGB format consisting of 595 rows and 800 columns of pixels with the green band displayed in Figure 1.

The ML result (Figure 2a) can be compared with ICM (Figure 2b) and MHCF (Figure 2d). The unclassified pixels from the first MHCF pass are shown in white (Figure 2c). Notice that the "eye" class is rather generously distributed around the perimeter of the face for all procedures. This is because the training data for this class were obtained by drawing a polygon around the eye to include the pupil and the white of the eye, which leads to a highly variable class. The "hair" class is also rather variable. The unclassified pixels (Figure 2c) show much of the uncertainty to be in the transition zones between the hair and background classes and the skin and erroneous eye class bordering the face. Thus, although the DOC measure has correctly picked up some problem areas, it has missed much of the erroneous "eye" area.

The DOC  $G$ -values (Figure 3b) are shown over their associated sample number. A 1 percent sample was taken from the image systematically within rows with random starts for

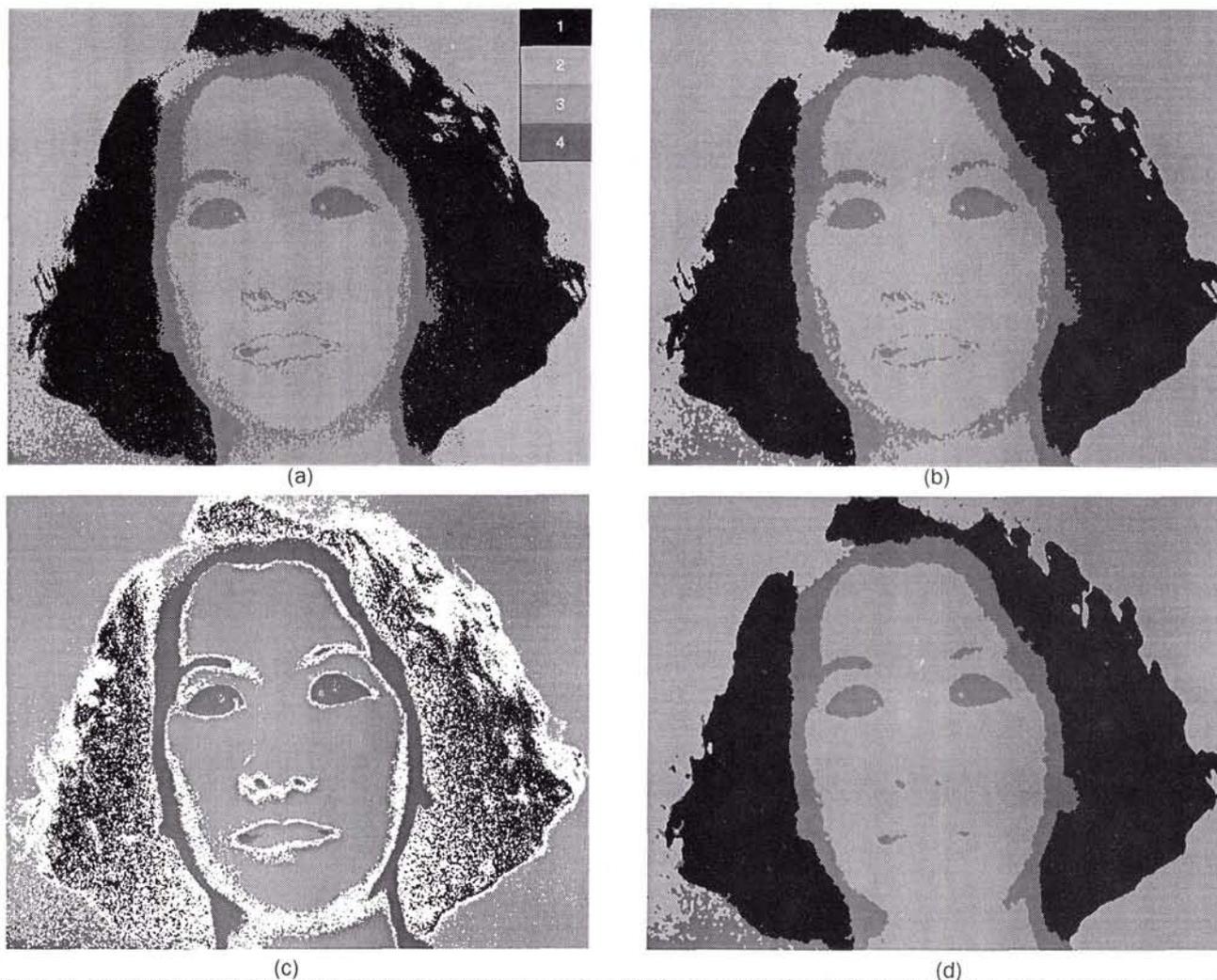


Figure 2. (a) Maximum-likelihood classification of the face image into four classes: hair, background, skin, and eyes; (b) ICM classification; (c) Result of the first pass of MHCf showing the pixels in white that did not exceed the  $g$ -cutoff value; (d) MHCf classification.

each new row. The corresponding  $G_c$  was slightly greater than 2.0 and was quite consistent over repeated 1 percent samples. The  $G$ -value plot (Figure 3b) gives an indication of the overall distribution of  $G$ -values as well as the spatial distribution, because they are obtained by row scanning the image beginning in row 1. The number of pixels, by class, that have  $G$ -values greater than  $G_c$  (significant  $G$ -values) in the sample (Figure 3a) is also important to consider. As stated above, there must be some pixels with significant  $G$ -values in each class before MHCf begins, or that class will not appear in the final result. This restriction places some limits on the choice of  $G_c$ . With this particular example, letting  $G_c$  correspond to the 50th percentile left none of the hair class significant.

Table 1 shows which classes made up the  $G$ -values of the 1 percent sample. The rows are the closest class and the columns represent the second closest class. For example, 569 pixels were closest to the hair class and second closest to the background class. The diagonals of Table 1 must be empty

by definition. Table 2 gives the same information for the pixels that exceeded  $G_c$ . Tables 1 and 2 indicate which classes are likely to be confused, but should not be interpreted in the same way as an actual confusion matrix. Because skin only appears as the second closest class to the eye class, one can assume that the skin spectral emissions make it reasonably distinct. On the contrary, the eye class appears frequently as second closest to every other class. Thus, one can assume that the eye class will be frequently confused, and this is evident in the final classification (Figure 2).

The ML classification (Figure 2a) shows its characteristic salt-and-pepper appearance and shows the greatest error in the eye class. ICM (Figure 2b) smooths over much of the salt-and-pepper appearance from ML, but otherwise is quite similar to ML. The first iteration of MHCf (Figure 2c) indicates uncertainty in the hair class and transition zones between face and hair. The final MHCf result (Figure 2d) has more skin in the chin area correctly classified than do either ML or ICM. This is due to the fact that these areas were initially left

unclassified and benefitted by neighboring pixels that were correctly classified as skin. Otherwise, there appears to be no major difference between the results of MHCF and either ML or ICM.

**Application to TM Data**

A Landsat Thematic Mapper image from September, 1984 (Figure 4) of the area around Bay St. Louis, Mississippi is put into four classes for this example: (1) water, (2) marsh, (3) forest, and (4) non-forest. TM bands 1 to 4 are used for this analysis, and Figure 4 is based on a composite of bands 4, 3, and 1. The results of the classification by ML, ICM, and MHCF are given in Figure 5. The ML classification (Figure 5a) exhibits somewhat of a salt-and-pepper appearance, although the bay and the Jourdan River that enters it are classified with little error. The ICM classification (Figure 5b) has a



Figure 4. Landsat TM image of Bay St. Louis, Mississippi derived from bands 1, 3, and 4.

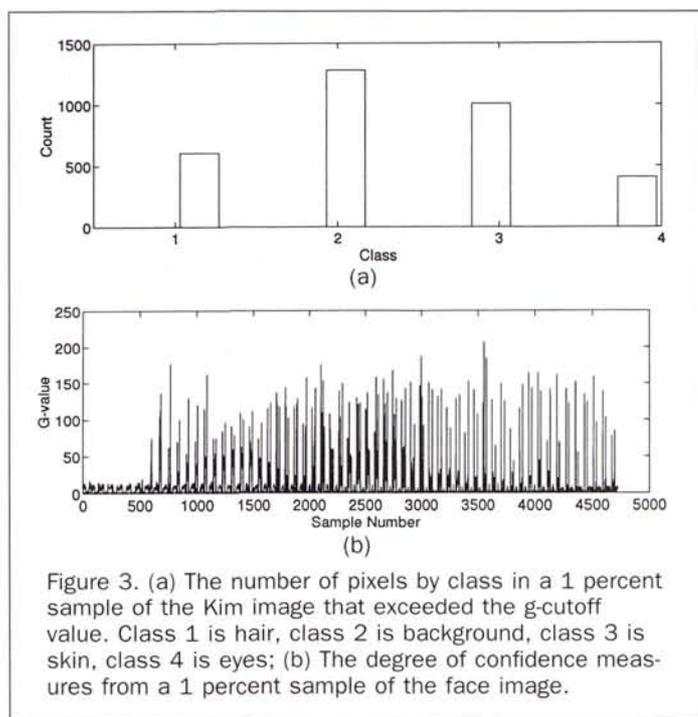


TABLE 1. ELEMENT R,C SHOWS THE NUMBER OF TIMES CLASS R WAS CLOSEST AND CLASS C WAS SECOND CLOSEST TO EACH PIXEL IN THE INITIAL 1 PERCENT SAMPLE FOR APPLICATION 1.

	Hair	Background	Skin	Eye
Hair	0	569	0	869
Background	171	0	0	1292
Skin	0	51	0	1119
Eye	263	253	173	0

TABLE 2. ONLY THE PIXELS FROM TABLE 1 THAT EXCEED  $G_c$  ARE INCLUDED

	Hair	Background	Skin	Eye
Hair	0	253	0	357
Background	60	0	0	1218
Skin	0	51	0	947
Eye	197	195	56	0

much smoother appearance and has clarified some of the well defined components of the image. For example, Interstate 10 runs diagonally through the scene north of the bay and is better defined by ICM than by ML.

The pixels remaining unclassified after the first MHCF pass (Figure 5c) are shown in white and include much of the water area. However, unlike with the first example, ML correctly classified most of these uncertain pixels so that the final MHCF (Figure 5d) result shows fewer noticeable differences from the ML result. The final MHCF result is somewhat smoother than the ICM result, which may be due as much to the additional iterations involved as to the fact that only the most certain pixels can commit after each iteration. ICM would also result in some further smoothing if another iteration were allowed with  $\beta = 1$ .

Tables 3 and 4 show the closest and second closest classes for the sample pixels for Application 2. Table 3 indicates that water would most frequently have marsh as the second closest class with 645 entries. However, marsh only had water as the second closest class 246 times. The most confused class appears to be non-forest, which is second closest to forest 1006 times and has all other classes as second closest a number of times.

Figure 6a shows that all classes will be present after the first iteration of MHCF, which is necessary for them to appear in the final result. Figure 6b shows the distribution of the  $G_i$  values for the initial 1 percent sample. In this case,  $G_c = 0.48$ , which indicates that many of the  $G_i$  values were quite small.

**Summary and Conclusions**

A modified highest confidence first classifier has been developed. The MHCF procedure is an extension of the ICM procedure of Besag (1986), and is the same as ICM when the cutoff value ( $G_c$ ) is set to zero. The user of MHCF must select a  $G_c$  value that allows commitment of some pixels to each class in the first iteration of MHCF. Suggestions were given for doing this along with demonstrations of diagnostic aids in the example applications. Spatial information is incorporated in

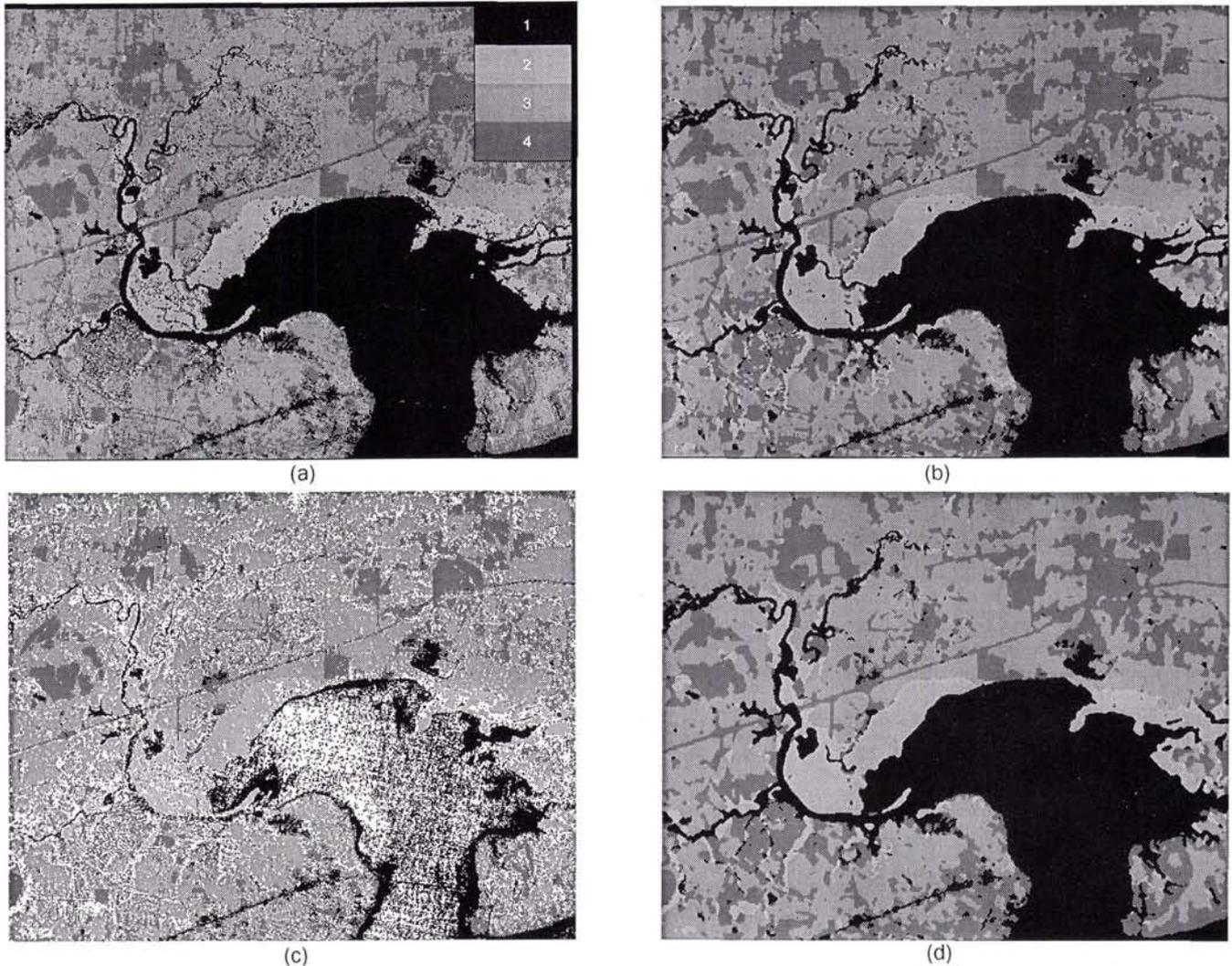


Figure 5. (a) Maximum-likelihood classification of the TM image into four classes: water, marsh, forest, and non-forest; (b) ICM classification; (c) Result of the first pass of MHCN showing the pixels in white that did not exceed the  $G_c$ -cutoff value; (d) MHCN classification.

MHCN after the first iteration to allow previously classified pixels to influence neighbors decisions.

The motivation for leaving the most uncertain pixels unclassified in early iterations is clear; only the more certain pixels should be initially influencing the decisions of neighbors. However, the choice of  $G_c$  for MHCN allows the user to prevent highly certain classes from completely overwhelming

less certain classes as may happen with the HCF procedure of Chou and Brown (1990). In fact, MHCN would be nearly equivalent to HCF if  $G_c$  were reset after each iteration to be large enough to allow only one pixel to be classified, but it was shown that this would be generally undesirable.

MHCN is also very amenable to the formation of strata indicating the certainty of classification. These strata could be used to guide sampling for the determination of map accuracy. One approach to strata formation would be to choose

TABLE 3. ELEMENT R,C SHOWS THE NUMBER OF TIMES CLASS R WAS CLOSEST AND CLASS C WAS SECOND CLOSEST TO EACH PIXEL IN THE INITIAL 1 PERCENT SAMPLE FOR APPLICATION 2

	Water	Marsh	Forest	Non-Forest
Water	0	645	0	48
Marsh	246	0	2	48
Forest	0	11	0	1006
Non-Forest	134	137	421	0

TABLE 4. ONLY THE PIXELS FROM TABLE 3 THAT EXCEED  $G_c$  ARE INCLUDED

	Water	Marsh	Forest	Non-Forest
Water	0	341	0	42
Marsh	126	0	0	13
Forest	0	51	0	844
Non-Forest	115	109	300	0

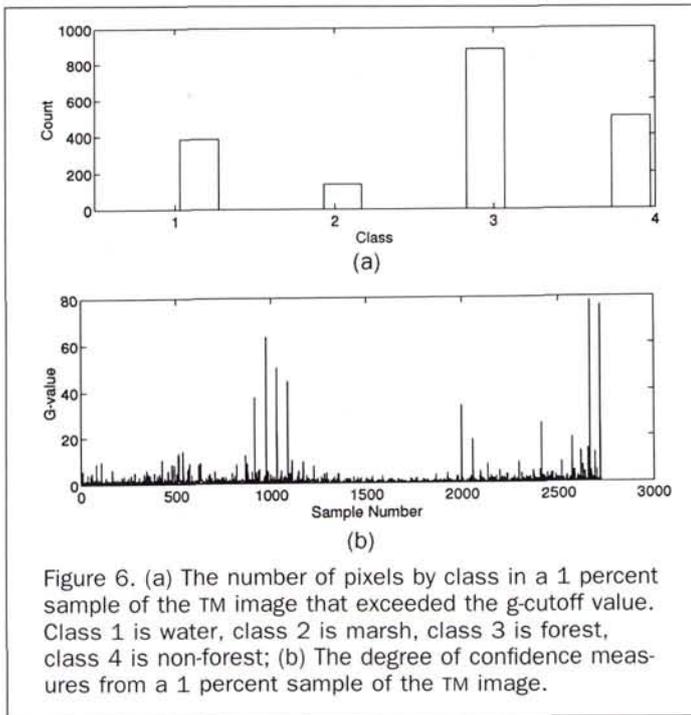


Figure 6. (a) The number of pixels by class in a 1 percent sample of the TM image that exceeded the  $g$ -cutoff value. Class 1 is water, class 2 is marsh, class 3 is forest, class 4 is non-forest; (b) The degree of confidence measures from a 1 percent sample of the TM image.

percentile-based strata boundaries. For example, pixels having  $G_i$  values greater than or equal to the 75th percentile could go in the first strata, pixels with  $G_i$  between the 50th and 75th percentiles in the second strata, etc. Another approach would be to have strata correspond to the iteration when the pixel was first committed to a class.

Classification results are also heavily influenced by the choice of training data. The act of selecting the  $G_c$  value and being made aware of the location of uncertain pixels may make the user realize that the training data for a particular class are inadequate and revisit this stage of the process. The DOC measures defined by Equation 8 can also be viewed as useful diagnostic aids like the probability measures discussed in Foody *et al.* (1992), but there is only one DOC value per pixel rather than  $K$  probability measures to store.

#### Acknowledgments

This work was funded by two USDA Forest Service programs: the Forest Health Monitoring Program and the Southern Global Change Program.

#### References

- Besag, J., 1986. On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Association*, B 48(3):259-302.
- Chou, P. B., and C. M. Brown, 1990. The theory and practice of Bayesian image labeling, *International Journal of Computer Vision*, 4:185-210.
- Foody, G. M., N. A. Campbell, N. M. Trodd, and T. F. Wood, 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification, *Photogrammetric Engineering & Remote Sensing*, 58(9):1335-1341.
- Geman, S., and D. Geman, 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721-741.
- Haslett, J., 1985. Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context, *Pattern Recognition*, 18:287-296.
- Klein, R., and S. J. Press, 1992. Adaptive Bayesian classification of spatial data, *Journal of the American Statistical Association*, 87(419):844-851.
- Owen, A., 1986. Discussion of Statistics, Images, and Pattern recognition (Ripley, 1986), *Canadian Journal of Statistics*, 14(2): 106-110.
- Ripley, B. D., 1986. Statistics, images, and pattern recognition, *Canadian Journal of Statistics*, 14(2):83-111.
- Switzer, P., 1980. Extensions of linear discriminant analysis for statistical classification of remotely sensed imagery, *Mathematical Geology*, 12(4):367-376.
- Trodd, N. M., G. M. Foody, and T. F. Wood, 1989. Maximum likelihood and maximum information: Mapping heathland with the aid of probabilities derived from remotely sensed data, *Remote Sensing for Operational Applications*, Remote Sensing Society, Nottingham, pp. 421-426.
- Wood, T. F., and G. M. Foody, 1989. Analysis and representation of vegetation continua from Landsat Thematic Mapper data for lowland heaths, *International Journal of Remote Sensing*, 10: 181-191.

(Received 2 July 1993; revised and accepted 28 October 1993)



#### Paul C. Van Deusen

Paul C. Van Deusen received the B.S. degree in Forest Management in 1975 from the University of Massachusetts, the M.S. degree in Forest Biometrics in 1979 from Mississippi State University, and the Ph.D. degree in Forest Biometrics in 1984 from the University of California. Formerly Project Leader at the Institute for Quantitative Studies in the USDA Forest Service's Southern Forest Experiment Station. Currently, Quantitative Forest Ecologist/GIS program leader with NCASI, the National Council of the Paper Industry for Air and Stream Improvement.

**Would you like to see your company's imagery  
on the cover of PE&RS?**

Cost is \$2,000.

Call Joann Treadwell, ASPRS Director of Publications  
301-493-0290; fax 301-493-0208; email [jtread@asprs.org](mailto:jtread@asprs.org)