A PRIOR PROBABILITY METHOD FOR SMOOTHING DISCRIMINANT ANALYSIS CLASSIFICATION MAPS

BY

PAUL SWITZER, WILLIAM S. KOWALIK and RONALD J.P. LYON

TECHNICAL REPORT NO. 1
JULY 15, 1981

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

A PRIOR PROBABILITY METHOD FOR SMOOTHING DISCRIMINANT ANALYSIS CLASSIFICATION MAPS

BY

PAUL SWITZER, WILLIAM S. KOWALIK and RONALD J.P. LYON

TECHNICAL REPORT NO. 1
JULY 15, 1981

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

A PRIOR PROBABILITY METHOD FOR SMOOTHING DISCRIMINANT ANALYSIS CLASSIFICATION MAPS

by

Paul Switzer
Department of Statistics
Stanford University

and

William S. Kowalik* and Ronald J. P. Lyon
Department of Applied Earth Sciences
Stanford University

Introduction

It has been widely recognized that digital classification maps which are based on information about the adjacent pixels, or which include information about the probability of occurrence of a class, rather than just the per-point information, have an improved appearance, and, perhaps, improved classification accuracy.

Various smoothing techniques have previously been described. Ballew (1977) passed a simple, weighted 3×3 filter across Landsat data (Maxwell (1976) before using the data to create a discriminant analysis classification map of rock alteration patterns in the Goldfield District, Nevada. Davis and Peet (1977) described a post-processing algorithm which replaced all areas below an empirically chosen, minimum size with the class of the surroundings. Letts (1979) used an algorithm for empirically replacing small, isolated areas with their weighted, most frequently occuring neighbors. Also, Thomas (1980) described an empirical $1/r^2$ distance rule to reevaluate the classification of a pixel based upon the initial classifi-

*Now a Postdoctoral Fellow with the Branch of Petrophysics and Remote Sensing, USGS, Denver, Colorado 80225

cation and that of the surrounding 3×3 pixel area. Richards, et al., have developed the implementation of spatial context (e.g. wheat fields are unlikely in an urban area) in an iterative, probabilistic relaxation labeling scheme which takes the initial label of a pixel into account at each iteration of the relaxation procedure. Strahler (1979, 1980) described the application of prior probability estimates, derived from auxiliary data, for the likelihood of occurrence of each class in the map area.

This paper presents a systematic method for the inclusion of intrinsically estimated prior probability information in maximum likelihood classification with application to linear discriminant analysis. The prior probability information in our method is derived from the frequencies of tentative class assignments in a window moving across an initial per-point classification map. Prior probability estimates are pixel specific and they employ the confusion matrix based on training set data. The class of the pixel at the center of the window is reevaluated using the spectral data for that pixel and the spatial information inherent in the classes of the neighborhood pixels. This work is an outgrowth of previous research (Switzer, 1980). We give an example of application of the technique to classification of rock types in western Nevada using Landsat spectral data.

All computer processing described herin was performed on a PDP 11/34 minicomputer in the Stanford Remote Sensing Lab.

Theory

This section develops the proposed procedure as a consequence of rather straightforward probabilistic reasoning. The procedure is illustrated step-by-step in the following section.

When geographic continuity is on a scale substantially larger than the dimensions of a single pixel, then classification of a pixel should exploit the increased likelihood that a pixel and its neighbors belong to the same class. One approach is to pass a moving average filter over the satellite data; an example and a mathematical justification are described in Switzer (1980). The approach of this paper uses the Bayesian concept of a prior probability distribution over the m possible classes. We differ from conventional Bayesian treatments in that the prior distribution is allowed to be pixel specific and that these priors are estimated directly from the satellite data.

Such a program is made possible by the anticipated geographic continuity. Over a small area containing the pixel in question the prior distribution is assumed to be approximately constant. Operationally, we take a $k \times k$ moving window containing k^2 pixels. The pixel-specific prior class probabilities for the pixel at position t are denoted $\pi_1(t)$, $\pi_2(2)$, ..., $\pi_m(t)$; they are the object of estimation which will be done in two stages.

(i) Use the available training data for identified pixels to establish a per-pixel classification algorithm such as an ordinary linear discriminant analysis. Estimate the confusion matrix of probabilities for this algorithm, viz.

$f_{ij} \equiv$ probability that the algorithm assigns a pixel to class j given that its true class is i.

The training data will provide direct estimates of the f_{ij} but they may be biased. Jackknifed estimates will serve to reduce bias. The confusion matrix is taken as fixed over the whole application area. For example, taking the distribution of the satellite data to be class-dependent but

not pixel-dependent gives a geographically constant confusion matrix. This is the common working hypothesis in statistically based classification.

(ii) Apply the per-pixel classification algorithm separately to each pixel of the application area, thereby obtaining the rough classification map. For a fixed k×k window centered at position t, let $P_j(t)$ denote the relative frequency of class j assignments in that window, i.e.,

$$P_j(t) \equiv \frac{1}{k^2} \text{ (number of class } j \text{ assignments in window centered at } t) .$$

These locally estimated relative frequencies are related to the sought after pixel specific prior probabilities via the confusion matrix. Specifically, the law of total probability gives:

$$P_j(t) \overset{e}{=} \sum_i \pi_i(t)\, f_{ij} \quad \text{for each class } j ,$$

where $\overset{e}{=}$ denotes equality of expected value. This is a system of linear equations in the m unknown probabilities $\pi_i(t)$. Assuming a nonsingular confusion matrix, this system can be solved to obtain moment estimates of the required prior distribution at position t.

Each pixel has its own window and its own local class frequencies derived from the rough classification map. Therefore, in principal, there is a new system of linear equations to solve at each pixel location. However, since the confusion matrix is fixed, it can be inverted once and for all, and the estimates of the prior distributions are obtained by multiplying the fixed inverse matrix by the locally estimated relative frequency vectors.

While the sum of the estimated prior class probabilities is guaranteed to be unity, it may occasionally happen that the solution of the linear system gives some estimated probabilities which are outside the unit interval. There are several possible fix-ups of which the simplest might be to set any negative estimates to zero and to rescale the remaining positive probabilities to add up to one. If large negative estimates occur persistently in the application area, one might question the appropriateness of the confusion matrix derived from the training data.

We now reclassify each pixel according to the posterior most probable class, i.e., using Bayes' Theorem. Bayes' Theorem requires, for each pixel, its estimated prior class distribution, which is obtained by the method described above. Bayes' Theorem also requires estimates of the class-specific probability density functions of the satellite data. In the context of linear discriminant analysis, these density functions are taken to be approximately multivariate Gaussian with class-specific mean vectors $\underset{\sim}{\bar{x}}_i$ and common covariance $\underset{\sim}{S}$. It follows that the logarithm of the posterior probability for class i at pixel position t, given the satellite data vector $\underset{\sim}{x}(t)$ at position t, is given by:

$$L_i(t) + \ln \pi_i(t) + \text{constant}$$

where

$$L_i(t) \equiv \text{usual per-pixel linear discriminant function for class } i$$
$$= \underset{\sim}{x}'(t)\ \underset{\sim}{S}^{-1}\ \underset{\sim}{\bar{x}}_i - 1/2\ \underset{\sim}{\bar{x}}_i'\ \underset{\sim}{S}^{-1}\ \underset{\sim}{\bar{x}}_i\ .$$

So, whereas the per-pixel "rough" classification chooses that class which maximizes $L_i(t)$, the "smooth" classification chooses that class which

maximizes $L_i(t) + \ln \pi_i(t)$ where $\pi_i(t)$ is estimated locally as described above. An application and an example of the calculations are given in the next section.

It is important to note that this proposal is not equilavent to prefiltering of the satellite data, nor is it equivalent to reclassification of a rough map based on local "majority voting", nor is it equivalent to Bayes classification using an overall estimate of prior probabilities for the application area as a whole.

Example of Application

An area of known rock type in the Yerington District of western Nevada adjacent to the MacArthur porphyry copper prospect was used to assess the usefulness of the processing methodology described above. The Yerington District has sparse vegetation cover (mean about 14%), and consequently, the rock and soil cover at the surface is well exposed to aerial view. The application area is a rectangular array of 23 Landsat pixels by 14 scan lines (1.3 × 1.1km) and includes 3 major rock types: 1) Oligocene volcanic tuffs, 2) Limonitically stained Jurassic quartz monzonite, and 3) Non-limonitic Jurassic quartz monzonite. The data array is from lines 1287 to 1300 and pixels 634 to 656 of Landsat-1 image 1307-18-62. This is a high sun angle image (61 degrees) recorded on May 26, 1973.

Color, stereo, aerial photographs of the application area were studied, and coupled with field visits to problem areas, each pixel of the test area was assigned to one of the 3 rock types. This digital geologic map (Figure 1) was used to assess the classification accuracy obtainable w th the 4-band Landsat data.

Pixels from every other scan line across the test area were used as training set data. Discriminant classification functions were developed in nonhierarchical fashion (all 3 groups together) from those training pixels with the BMDP7M Stepwise Linear Discriminant Analysis Program (Dixon and Brown, 1977). These linear classification functions were then used to classify a large array of Landsat data centered on the Yerington District. Classification mpas of the application area showing the 3 rock types were printed on a matrix line printer at 1:15,000 scale. These maps are presented here as Figures 2 through 6.

Figure 2 is the initial, per-point classification map. The classification of each pixel was then reevaluated with the prior probability information coming from neighborhoods of 3×3, 5×5, 7×7, and 9×9 pixels. Figures 3, 4, 5, and 6 show the classification maps derived in those respective cases. Notice that the smoothing operation tends to eliminate stray pixels and thereby produces a more coherent map pattern. Figure 7 shows a sample calculation using a 5×5 neighborhood for one pixel of the application area.

Table 1 shows the overall classification accuracy of each Landsat-derived map. The classification accuracy is the number of correctly classified pixels divided by the total number of pixels in the area (322).

The classification accuracies are low, and indicate that these 3 rock types are not spectrally very distinct in the Landsat data. Each increase in window size from the per-point classification, however, caused an increase in classification accuracy of 2% to 3%. The 7×7 and 9×9 windows gave the same accuracy, 8.4% larger than that obtained by the per-point classification. We conclude that this smoothing algorithm can provide improved map appearance and increased classification accuracy. The best window size in any case is, no doubt, a function of surface terrain variability and the data noise characteristics.

References

Ballew, G.I., (1977), Alteration Mapping at Goldfield, Nevada by Cluster and Discriminant Analysis of Landsat Digital Data, in Proc. of 11th Int. Symp. on Remote Sensing of Env., p783-790.

Davis, W.A., and Peet, F.G., (1977), A Method of Smoothing Digital Thematic Maps, in Remote Sensing of Environme t, Elsevier Press, v6, p45-49.

Dixon, W.J., and Brown, M.B., (1977), BMDP-77 Biomedical Computer Programs, P Series, Univ. of Calif. Press, Berkeley, Calif. 880 p.

Letts, P.J., (1979), Small Area Replacement in Digital Thematic Maps, in Proc. of 1979 Machine Processing of Remotely Sensed Data Symp., p431, (abstract).

Maxwell, E.L., (1976), Multivariate System Analysis of Multispectral Imagery, in Photog. Eng. and Remote Sensing, v42, #9, p1173-1186.

Richards, J.A., Landgrebe, D.A., and Swain, P.H., (1980), Pixel Labeling by Supervised Probabilistic Relaxation, LARS Technical Report 022580, Purdue University, West Lafayette, Indiana, 15 p.

Strahler, A.H., (1979), The Use of Prior Probabilities in Maximum Likelihood Classification, in Proc. of 1979 Machine Processing of Remotely Sensed Data Symposium, p356, (abstract).

Strahler, A.H., (1980), The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data, in Re ote Sensing of Environment, Elsevier Press, v10, p135-163.

Switzer, P., (1980), Extensions of Linear Discriminant Analysis for Statistical Classification of Remotely Sensed Satellite Imagery, in Proc. of Int. Geol. Congress, Paris, July, p .

Thomas, I.L., (1980), Spatial Post Processing of Spectrally Classified Landsat Data, in Photog. Eng. and Remote Sensing, v49, #9, p1201-1206.

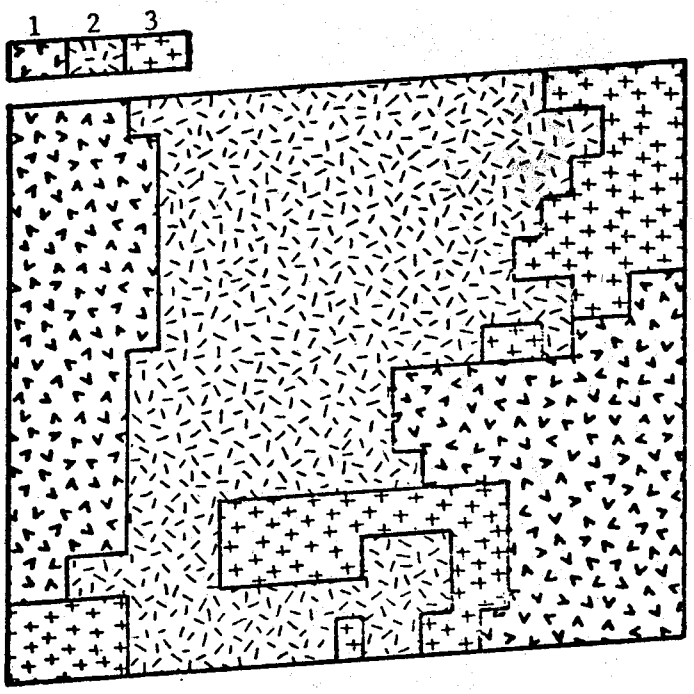

Figure 1. True geologic map of 3 rock types interpreted from color aerial photography. V = volcanic rock group; L = Limonitic rock group; N = non-limonitic rock group.

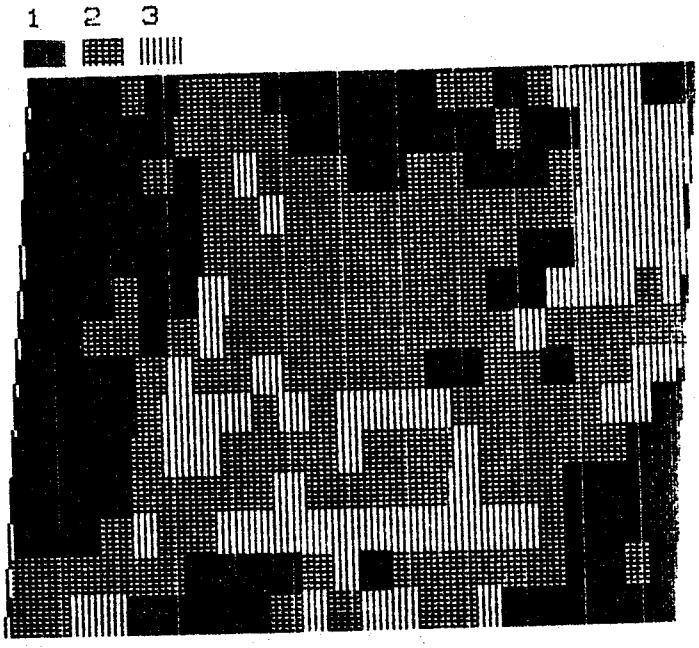

Figure 2. Original, unsmoothed, discriminant classification map of the application area. Symbol 1 - Volcanic rock group, Symbol 2 - Limonitic rock group, Symbol 3 - Non-limonitic group.

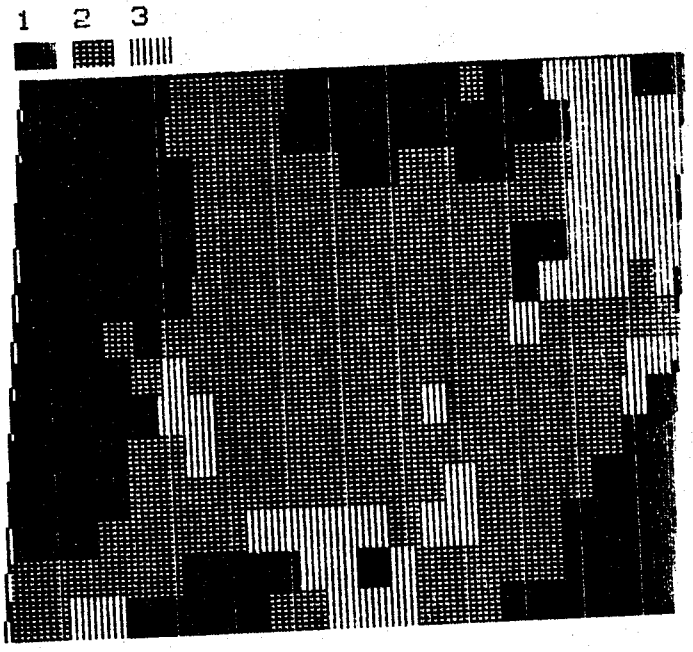

Figure 3. Discriminant classification map after prior probability smoothing with a 3×3 pixel kernel. The symbols are the same as in Figure 2.

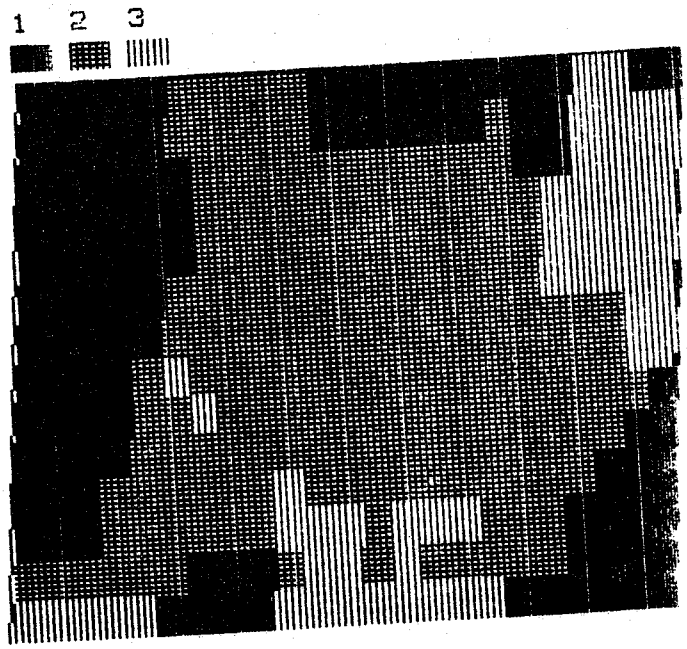

Figure 4. Discriminant classification map after prior probability smoothing with a 5×5 pixel kernel. The symbols are the same as in Figure 2.

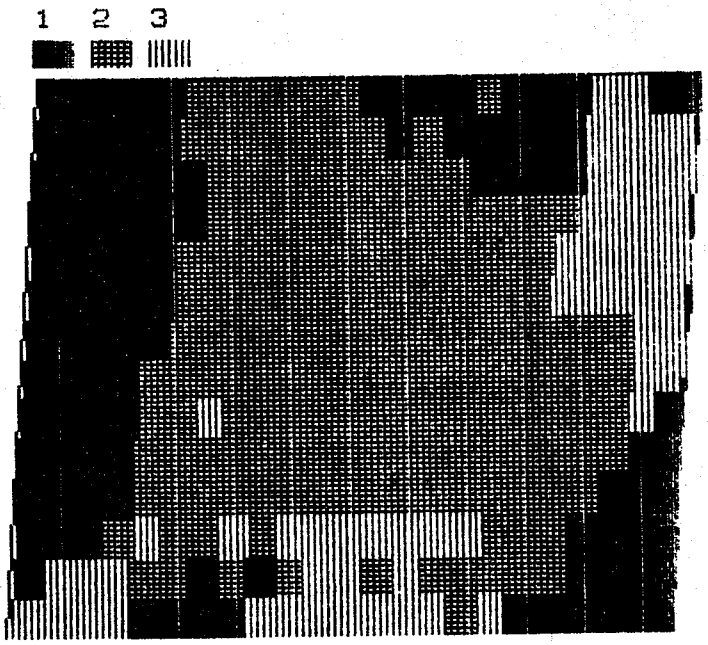

Figure 5. Discriminant classification map after prior probability smoothing with a 7×7 pixel kernel. The symbols are the same as in Figure 2.

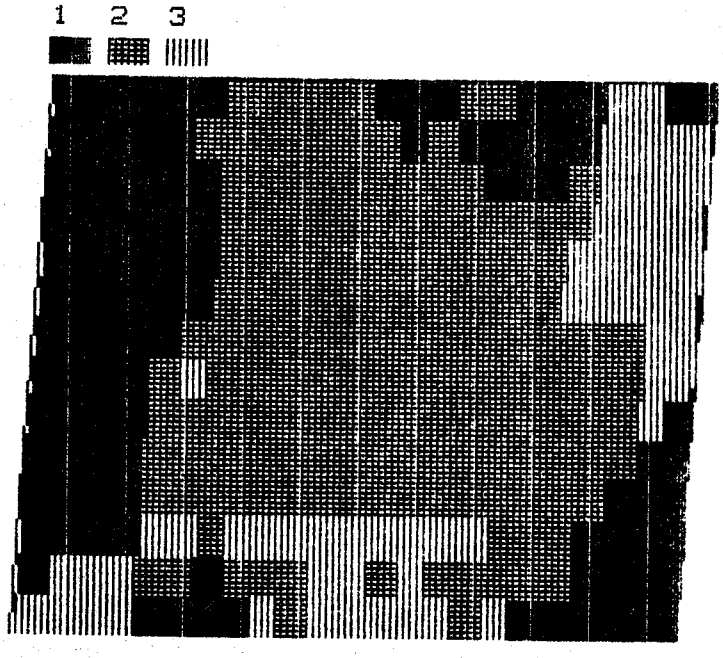

Figure 6. Discriminant classification map after prior probability smoothing with a 9×9 pixel kernel. The symbols are the same as in Figure 2.

Figure 7. Sample calculation for one pixel of the application area. The case shown is for the pixel at scanline 1293 and column 637 where a 5×5 pixel neighborhood is considered.

The original per point classification functions from the BMDP7M program are:

V: Volcanic rock group = L_1 = 2.87(Band 4) + 1.72(Band 5) - 119.53

L: Limonitic rock group = L_2 = 2.69(Band 4) + 2.09(Band 5) - 131.47

N: Non-Limonitic group = L_3 = 3.36(Band 4) + 1.77(Band 5) - 147.75

The pixel at line 1293 and column 637 has the values: Band 4 = 48
Band 5 = 57.

The above classification functions yield the following values for this pixel: L_1 = 116.3
L_2 = 116.8
L_3 = 114.4.

Therefore, the pixel is classified as L (Limonitic) because the limonitic group function yields the largest value.

Now, we take the spatial information from the surrounding 5×5 neighborhood into account. The original classification map of that 5×5 pixel area is:

V V V V V
V V L V V
V L L V L
V V V L N
V V V L N

The frequency of each class in the area is: (V-17, L-6, N-2). These frequency values are converted to percentages by dividing by 25 = (5×5): P_1 = .680, p_2 = .240, P_3 = .080.

The prior probabilities π_1, π_2, and π_3 are given by: $\underset{\sim}{P} = \underset{\sim}{\pi}\ \underset{\sim}{f}^T$, where $\underset{\sim}{f}$ is the confusion matrix estimated from the training pixels (after conversion to percentages). That is:

The Confusion Matrix Frequencies (from BMDP7M):

	V	L	N	Row Total
True V	39	14	6	59
True L	19	43	16	78
True N	0	6	18	24
			Total	161

The Confusion Matrix; Percentage of Row Total (Table 1A of Table 1):

V	L	N
.66	.24	.10
.24	.55	.21
.00	.25	.75

The system of 3 linear equations to be solved for is:

$$.680 = \pi_1(.66) + \pi_2(.24) + \pi_3(.10)$$

$$.240 = \pi_1(.24) + \pi_2(.55) + \pi_3(.21)$$

$$.080 = \pi_1(.00) + \pi_2(.25) + \pi_3(.75) \ .$$

The solution is: $\pi_1 = 1.03$

$\pi_2 = -0.07$

$\pi_3 = 0.13$.

The negative probability value for π_1 requires a fixup. We use a rule which sets negative probabilities arbitrarily to zero.

Normalize all positive probabilities to 1.0 by dividing by their sum. The normalized values are: $\pi_1 = .89$, $\ln \pi_1 = -.12$

$\pi_2 = .00$, $\ln \pi_2 = -\infty$

$\pi_3 = .11$, $\ln \pi_3 = -2.21$.

The new classification values are:

$$L_1^* = L_1 + \ln \pi_1 = 116.3 - .12 = 116.2$$

$$L_2^* = L_2 + \ln \pi_2 = 116.8 - \infty = -\infty$$

$$L_3^* = L_3 + \ln \pi_3 = 114.4 - 2.21 = 112.19 \ .$$

The pixel is now assigned to the first class, V, its correct class.

Table 1. Confusion matrices which correspond to Figures 2 through 6. The matrices give the frequency of correctly classified and misclassified pixels for each of the 5 figures. Matrix 1A is the confusion matrix for the training pixels only. Matrices 1B, 2, 3, 4, and 5 are based on the training pixels and the other half of the pixels in the application area.

Matrix Number	Frequency			Total (Frequency)	Total Correct (Frequency)	Overall Percentage Correct
	V	L	N			
1A. Original Per-Point						
True V	39	14	6			
True L	19	43	16	161	100	62.1
True N	0	6	18			
1B. Original Per-Point (all pixels)						
True V	69	32	10			
True L	40	89	25	322	190	59.0
True N	4	21	32			
2. Smoothed 3×3						
True V	70	36	5			
True L	41	100	13	322	200	62.1
True N	4	23	30			
3. Smoothed 5×5						
True V	70	34	7			
True L	31	109	14	322	212	65.8
True N	3	21	33			
4. Smoothed 7×7						
True V	71	32	8			
True L	27	112	15	322	217	67.4
True N	4	19	34			
5. Smoothed 9×9						
True V	71	32	8			
True L	28	111	15	322	217	67.4
True N	4	18	35			