Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects

R. L. KETTIG AND D. A. LANDGREBE, SENIOR MEMBER, IEEE

This paper appears in: Geoscience Electronics, IEEE Transactions on
Issue Date: Jan. 1976
Volume: 14 Issue:1
On page(s): 19 - 26
ISSN: 0018-9413

Digital Object Identifier: 10.1109/TGE.1976.294460
Date of Current Version: 29 enero 2007

Abstract—A method of classification of digitized multispectral image data is described. It is designed to exploit a particular type of dependence between adjacent states of nature that is characteristic of the data. The advantages of this, as opposed to the conventional "per point" approach, are greater accuracy and efficiency, and the results are in a more desirable form for most purposes. Experimental results from both aircraft and satellite data are included.

ECHO (extraction and classification of homogeneous objects)

II. Sample Classification

Each pixel in an object is a q-dimensional random variable,

$p(\boldsymbol{x}|W_i)$,

$\boldsymbol{x} \in R^q$, denotes this class-conditional density function for the ith class. Another common assumption is that the classes can be defined such that $p(\boldsymbol{x}|W_i)$ is approximately multi-variate normal (MVN); i.e.,

$$p(\boldsymbol{x}|W_i) = N(\boldsymbol{x}; \boldsymbol{M}_i, \tilde{C}_i)$$
$$\triangleq (|2\pi\tilde{C}_i| \exp((\boldsymbol{x} - \boldsymbol{M}_i)^t \tilde{C}_i^{-1}(\boldsymbol{x} - \boldsymbol{M}_i)))^{-1/2}$$

for some q-dimensional positive-definite, covariance matrix $\tilde{C}_i$ and some mean vector $\boldsymbol{M}_i \in R^q$. Parametric estimates of these

II. Sample Classification

Two pixels in spatial proximity to one-another are unconditionally correlated, with the degree of correlation decreasing as the distance between them increases.

we assume class-conditional independence

$X = (X_1, \cdots, X_n) \in R^{nq}$ represents a set of pixels in some object

One popular approach is the "minimum distance (MD) strategy

Bhattacharyya distance, which for $N(\boldsymbol{x}; \boldsymbol{M}_i, \tilde{C}_i)$ and $N(\boldsymbol{x}; \boldsymbol{M}, \tilde{C})$ is given by:

$$B = \frac{1}{4}\left(\ln \frac{|(\tilde{C}_i + \tilde{C})/2|^2}{|\tilde{C}_i|\,|\tilde{C}|} + \operatorname{tr}\left((\tilde{C}_i + \tilde{C})^{-1}(\boldsymbol{M}_i - \boldsymbol{M})(\boldsymbol{M}_i - \boldsymbol{M})^t\right)\right). \quad (1)$$

II. Sample Classification

Our preference is the maximum likelihood (ML) strategy which assigns X to class i if

$$\ln p(X|W_i) = \max_j \ln p(X|W_j).$$

Due to the assumption of class-conditional independence, these quantities can be computed as:

$$\ln p(X|W_i) = -\frac{1}{2}\operatorname{tr}(\tilde{C}_i^{-1}\tilde{S}_2) + \boldsymbol{M}_i^t\tilde{C}_i^{-1}\boldsymbol{S}_1 - \frac{1}{2}n(\boldsymbol{M}_i^t\tilde{C}_i^{-1}\boldsymbol{M}_i + \ln|2\pi\tilde{C}_i|),$$

$$\boldsymbol{S}_1 = \sum_{i=1}^{n}\boldsymbol{X}_i, \qquad \tilde{S}_2 = \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^t. \tag{2}$$

Of course: $\boldsymbol{M} = \boldsymbol{S}_1/n$ and $\tilde{C} = \tilde{S}_2/n - \boldsymbol{M}\boldsymbol{M}^t$. Formula (2) is much faster to compute than formula (1) for each $(\boldsymbol{S}_1, \tilde{S}_2)$

II. Sample Classification

the ML strategy is computationally efficient.

Chernoff bound for ML no-memory classification ($n = 1$) can be extended to provide an error bound for ML sample classifi-

III. Image Partitioning

At the first level of testing, each group becomes a unit called a "cell," provided that it satisfies a relatively mild criterion of homogeneity.

At the second level, an individual cell is compared to an adjacent "field," which is simply a group of one or more connected cells that have previously been merged. If the two samples appear statistically similar by some appropriate criterion, then they too are merged.

III. Image Partitioning

Annexation Criterion

Let $X = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$ represent the pixels in a group of one or more cells which have been merged by successive annexations. Let $Y = (\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m)$ represent the pixels in an adjacent, non-singular cell. Since both X and Y have satisfied certain criteria of homogeneity, we assume that each is a sample from a MVN population. Let f and g represent the corresponding density functions. It is desired to test the (null) hypothesis that $f = g$. This is a composite hypothesis, since it does not specify f and g. The "likelihood ratio procedure" [10] provides an effective statistic for testing this hypothesis. Van Trees [11] refers to it as the "generalized likelihood ratio."

III. Image Partitioning

Annexation Criterion

Let

$$H_0(x, y) = \{p(x, y|f, g): \ g = f, \qquad f \in \Omega\}$$

$$H_1(x, y) = \{p(x, y|f, g): \ f \in \Omega, \qquad g \in \Omega\}$$

where $p(x, y|f, g)$ is the conditional joint density of X and Y evaluated at $x \in R^{nq}$ and $y \in R^{mq}$, and Ω is a set of MVN density functions. The assumption of class-conditional independence enables us to express the joint-density of pixels as the product of their marginal densities. Thus:

$$p(x, y|f, g) = p(x|f)\, p(y|g) = \left(\prod_{i=1}^{n} f(\boldsymbol{x}_i)\right)\left(\prod_{i=1}^{m} g(\boldsymbol{y}_i)\right).$$

III. Image Partitioning

Annexation Criterion

The generalized likelihood ratio is given by:

$$\Lambda = \frac{\sup H_0(X, Y)}{\sup H_1(X, Y)} = \frac{\max_{f \in \Omega} \; p(X|f)\, p(Y|f)}{\max_{f \in \Omega} \; p(X|f) \max_{g \in \Omega} p(Y|g)}.$$

For an "unsupervised" approach to partitioning we take Ω to be the following set of functions of $\boldsymbol{x} \in R^q$:

$$\Omega = \{N(\boldsymbol{x}; \boldsymbol{M}, \tilde{C}):\ \boldsymbol{M} \in R^q,$$

$$\tilde{C} = \text{symmetric and positive-definite}\}$$

Annexation Criterion

For a "supervised" approach to partitioning we take Ω to be:

$$\Omega = \{p(\mathbf{x}|W_i):\ i = 1, \cdots, K\}.$$

This greatly simplifies each hypothesis, but paradoxically the resultant test criterion is much more complicated:

$$\Lambda = \frac{\max_i p(X|W_i)\, p(Y|W_i)}{\max_i p(X|W_i)\, \max_j p(Y|W_j)}. \tag{6}$$

III. Image Partitioning

Annexation Criterion

Anderson [12] shows that:

$$\Lambda = \Lambda_1 \cdot \Lambda_2$$

where

$$\Lambda_1 = (|\tilde{A}|/|\tilde{B}|)^{N/2}$$

$$\Lambda_2 = (|\tilde{A}_x/n|^n \, |\tilde{A}_y/m|^m / |\tilde{A}/N|^N)^{1/2}$$

we can construct a significance test of the null hypothesis. Λ_1 and Λ_2 are independent under the null hypothesis [12], so the procedure we use is to test Λ_1 at significance level α_1 and Λ_2 at level α_2, and reject the null hypothesis if either test produces a rejection.

III. Image Partitioning

Annexation Criterion

For a "supervised" approach to partitioning we take Ω to be:

$$\Omega = \{p(\mathbf{x}|W_i):\ i = 1, \cdots, K\}.$$

This greatly simplifies each hypothesis, but paradoxically the resultant test criterion is much more complicated:

$$\Lambda = \frac{\max_i p(X|W_i)\, p(Y|W_i)}{\max_i p(X|W_i)\, \max_j p(Y|W_j)}. \tag{6}$$

$$T = 10^{-t}, \qquad t \geqslant 0.$$

In other words, we reject the null hypothesis if $\Lambda < T$ or equivalently $-\log \Lambda > t$. Otherwise we accept it. Experimentally we investigate the effect of different values of t on performance.

III. Image Partitioning

Cell Selection Criterion

III. Image Partitioning

Cell Selection Criterion

"Cell selection" refers to the Level-1 test which is used to detect cells that overlap boundaries. Such cells frequently exhibit abnormally large variances. Thus, in the unsupervised mode, we say that a cell is singular if the ratio of the square root of the sample variance to the sample mean falls above some threshold, c, in any channel.

III. Image Partitioning

Cell Selection Criterion

In the supervised mode we call a cell singular if $Q_j(Y) > c$, where:

$$Q_j(Y) = \mathrm{tr}\left(\tilde{C}_j^{-1} \sum_{i=1}^{m} Y_i Y_i^t\right) - 2M_j^t \tilde{C}_j^{-1} \sum_{i=1}^{m} Y_i + m \cdot M_j^t \tilde{C}_j^{-1} M_j$$

where j is such that:

$$\ln p(Y|W_j) = \max_i \ln p(Y|W_i)$$

$$= \max_i -\frac{1}{2}\left(m \cdot \ln|2\pi\tilde{C}_i| + Q_i(Y)\right).$$

The decision rule is to accept the hypothesis that Y is homogeneous if $Q_j(Y) < c$, where c is a prespecified threshold. Otherwise the hypothesis is rejected. This criterion has the particu-

IV. Experimental Results

Two aircraft and two LANDSAT-1 data sets, for which large amounts of training and test data are available, were classified by the following six methods:

1. Conventional ML No-Memory Classification [14]
2. Supervised Cell Selection only ($t = 0$); ML Sample Classification
3. "Optimized" MUV Unsupervised Partitioning; ML Sample Classification
4. Supervised Partitioning ($t = 4$); ML Sample Classification
5. ML Sample Classification of Test Areas Only
6. MD (Bhattacharyya) Sample Classification of Test Areas Only [14].

IV. Experimental Results

method #1.　　　　　　　　　　method #4

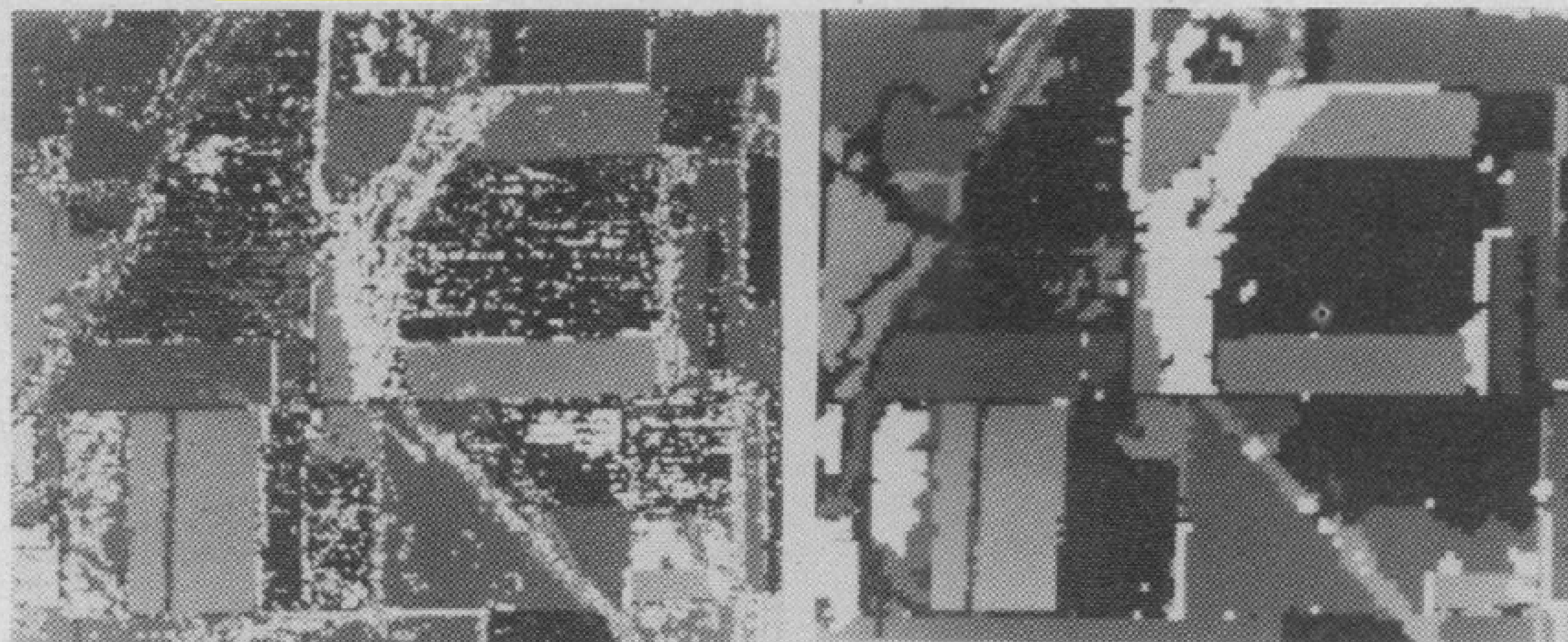


Fig. 2. Gray-scale-coded classification maps produced by no-memory classifier (left) and sample classifier (right).

IV. Experimental Results

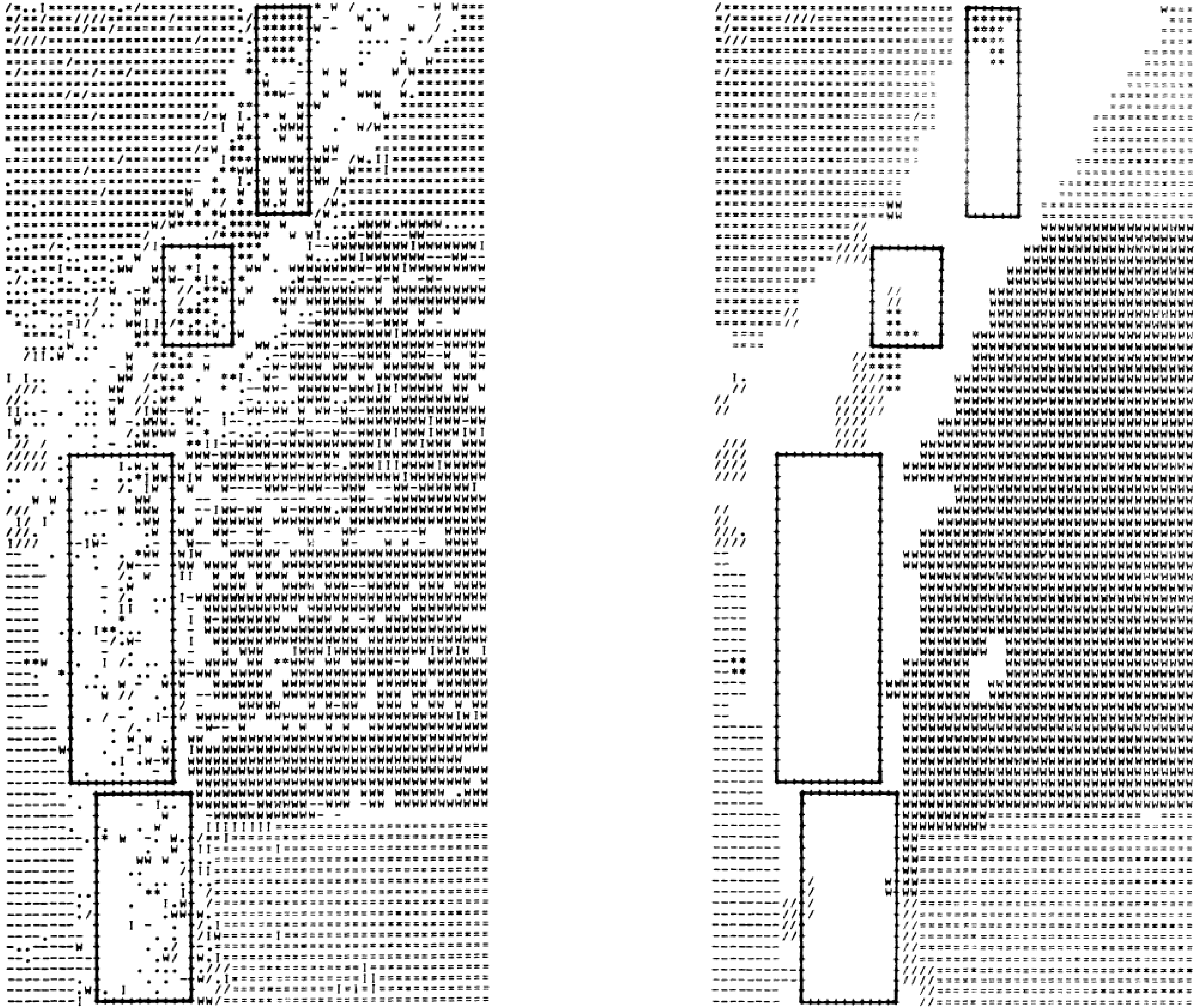
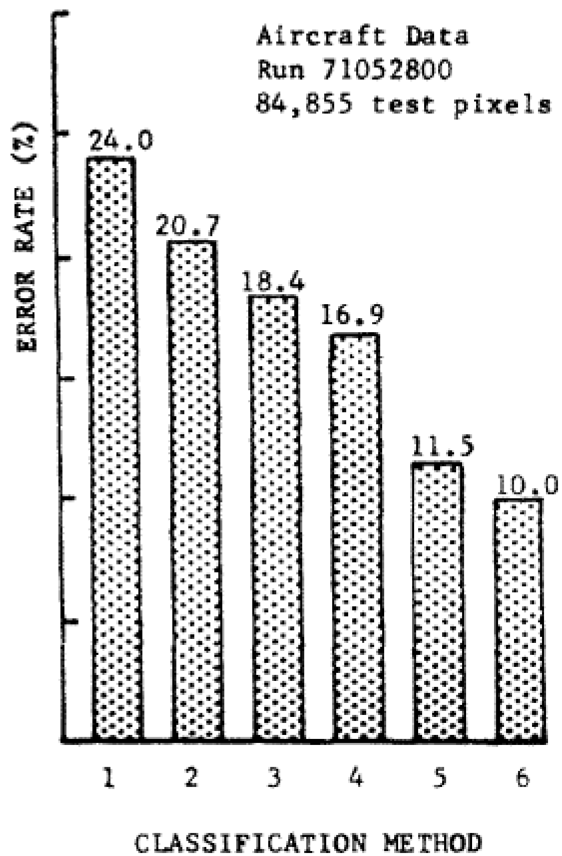


Fig. 3. Logogrammatic classification maps produced by no-memory classifier (left) and sample classifier (right).

IV. Experimental Results

Figure 3 shows the centers of these two maps in greater detail. Each class is represented by an assigned symbol and each symbol represents one pixel. The four rectangular areas are test areas designated as wooded pasture (displayed as a blank).

IV. Experimental Results



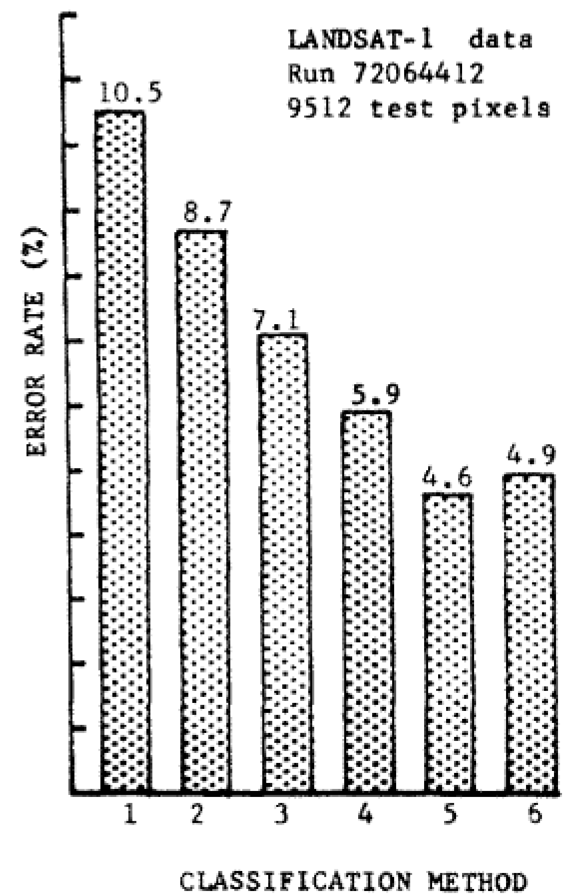


Fig. 4. Classification performance of six different methods applied to four different data sets.

IV. Experimental Results

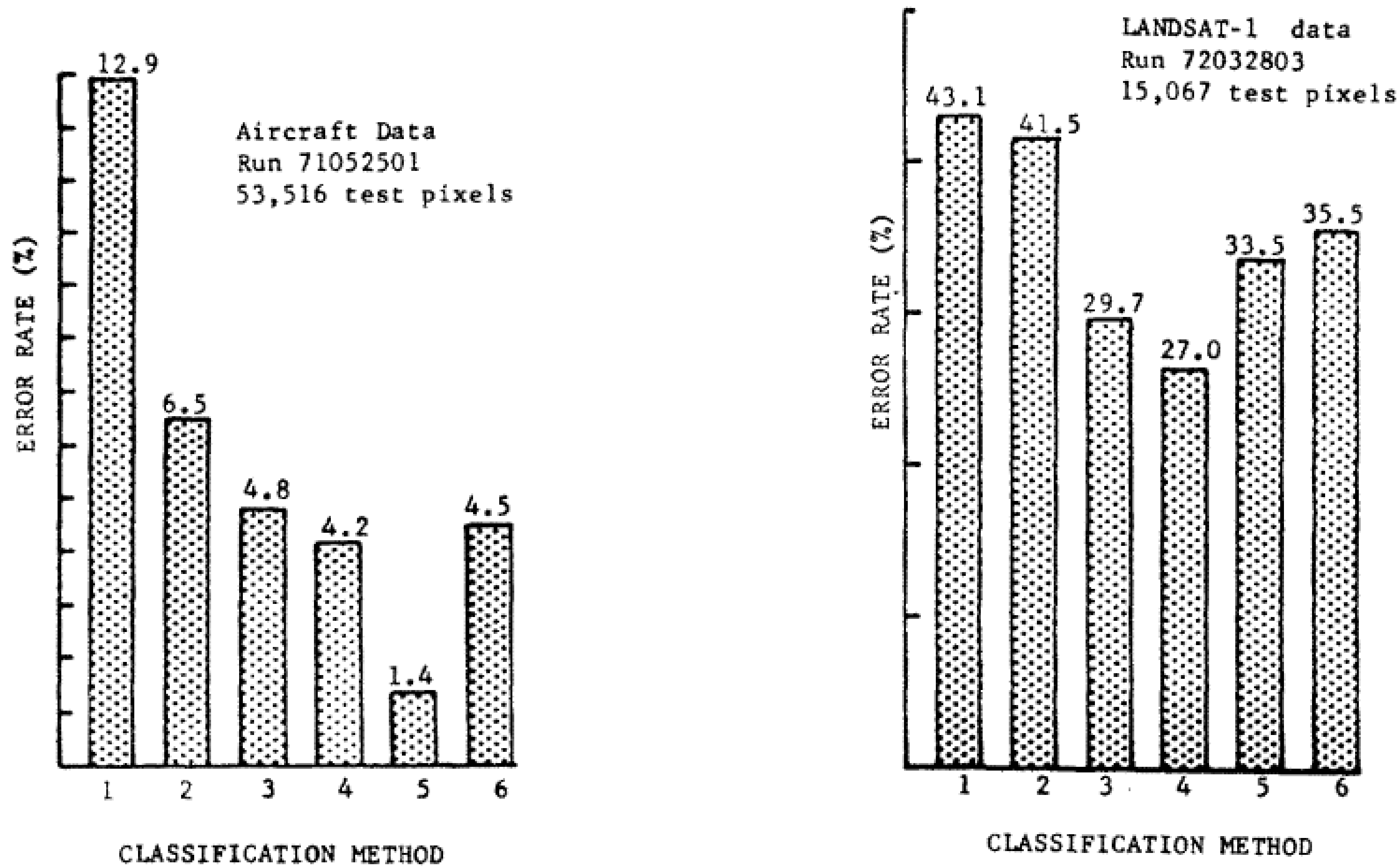
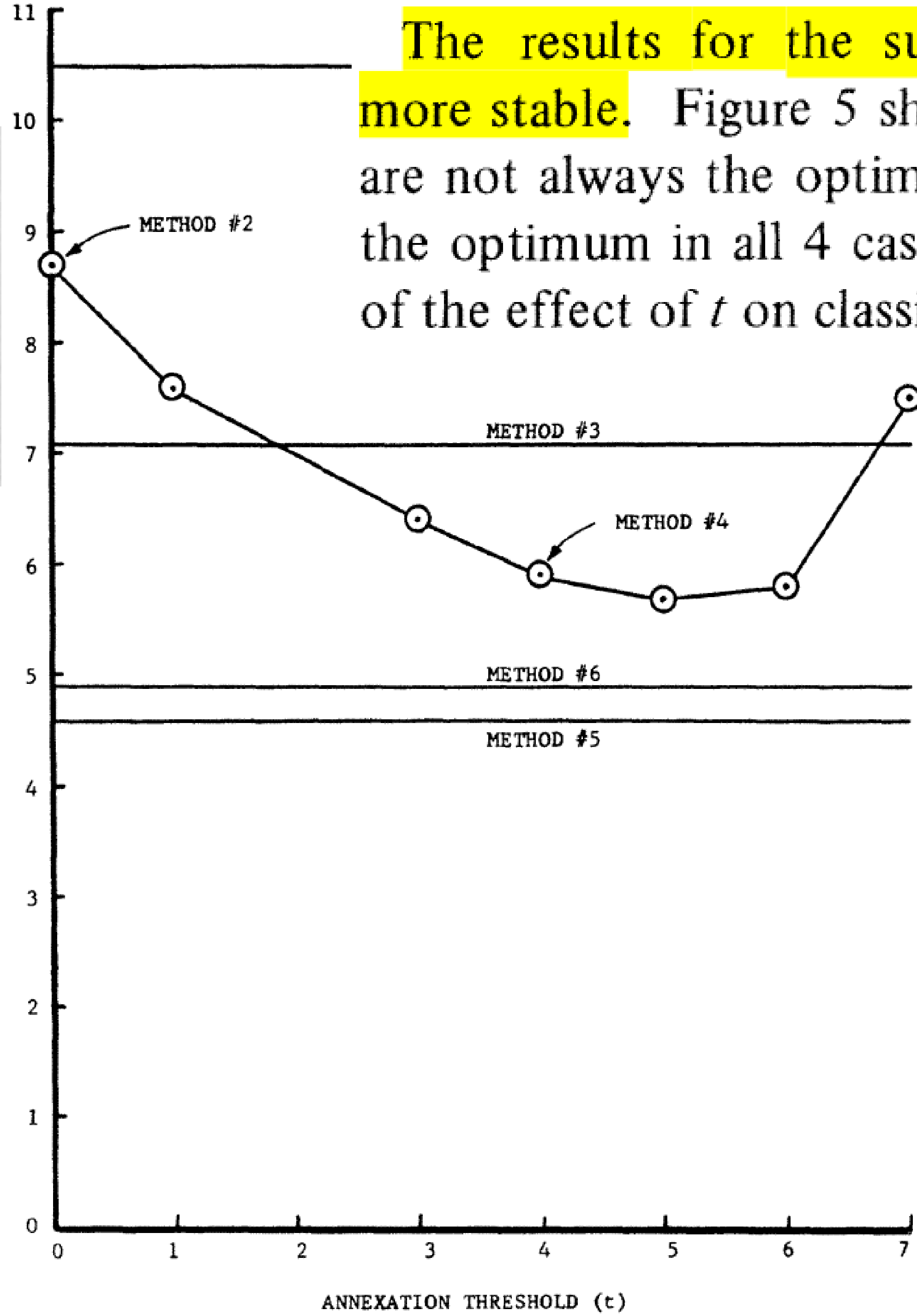


Fig. 4. Classification performance of six different methods applied to four different data sets.

The results for the supervised mode, however, are much more stable. Figure 5 shows only the results for $t = 4$, which are not always the optimum results, but they are within 1% of the optimum in all 4 cases. Figure 5 shows a typical example of the effect of t on classification error rate.



g. 5. Effect of annexation threshold (t) on classification performance–run 72064412.

V. Conclusion

We have successfully exploited the redundancy of states that is characteristic of sampled imagery of ground scenes to achieve better accuracy and reduce the number of actual classifications required. The only training used is the same as that required by a conventional maximum likelihood, no-memory classifier; i.e., estimates of the class-conditional, marginal densities for a single pixel. Thus we have not relied on specific spatial features, textural information (class-conditional spatial correlation), or on the contextual information associated with spatial relationships of objects.