Effect of spatial transformations on the accuracy of supervised classification

Chi-Chuan Cheng
Taiwan Forestry Research Institute,
53 Nan-Hai Road, Taipei,
Taiwan, R.O.C.

Abstract: The effect of spatial transformations on the accuracy of supervised classification was investigated. The objective was to use the proposed spatial transformations, the residual analysis and the principle component analysis, to yield transformed training statistics and then to investigate whether the performance of supervised classification would be improved by these transformed training statistics.

The nine-band airborne multispectral scanner data were classified by the maximum likelihood classifier based on untransformed (biased) and transformed (unbiased) training statistics. The classification results indicated that the incorporation of spatial transformation for defining training statistics resulted in better classification accuracy than the conventional supervised approach which was based on biased training statistics. Improved classification resulted from the reduced spatial correlation and the change in the covariance matrix.

1 INTRODUCTION

Spatial correlation between pixels exists in remote sensed data (Basu and Odell, 1974; Craig & Labovitz, 1980; Tubbs, 1978-1980), particularly in high spatial resolution aircraft scanner data (Mobasseri et al, 1978). The spatial correlation functions approximately follow a decaying exponential which means that the adjacent pixel correlation is high at a low distance lag and will decrease to zero at a high distance lag.

The sources of spatial correlation are due to the physical properties of the sensor (Kalayeh, 1982) and the target (Tubbs, 1979). Spatial correlation between training pixels affects the estimation of population mean vectors and covariance matrices and then affects the classification performance of likelihood ratio classifiers which require a mean vector and a covariance matrix(Campbell, 1981; Basu & Odell, 1974; Tubbs & Coberly, 1978). Therefore, classification results are suboptimal, not optimal.

To remove or adjust for this effect, spatial transformations are proposed to eliminate the errors caused by spatial correlation. The objectives were to investigate the effects of spatial transformations on likelihood ratio classifiers and to investigate whether the classification performance can be improved.

2 MATERIAL AND STUDY AREA

A high resolution nine-band airborne MSS data set, which was operated at an altitude approximately 4572 meters (15,000 feet) above sea level and collected on October 2, 1984, was chosen for this investigation. The study area is located in Chi-Tou Experimental forest of The National Taiwan University, Taiwan, Republic of China. The scanner has a 2.5 milliradian instantaneous-field-of-view (IFOV). The chosen data represents the forested area. The ground cover types include forest species such as Taiwania cryptomerioides, Phyllostachys nuda, Cunnighamia lanceolata, Chamaecyparis formosensis, and large areas of Cryptomeria japonica.

The test area with size 150x150 each as shown in Figures 1 was selected for this study. Forest cover types in this test area include:

(1) Bamboo (Phyllostachys nuda),
(2) Peacock pine (Cryptomeria japonica).

Where Cryptomeria japonica consists of two different classes with stand volume 330+20 cubic meter/ha and 200+70 cubic meter/ha. In following discussions, these two stand volume classes are simply referred to Cryptomeria-1 and Cryptomeria-2.

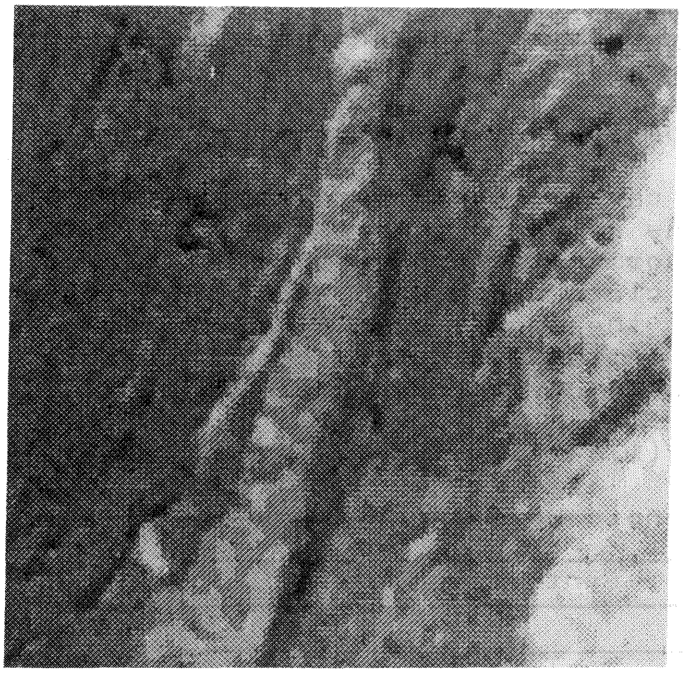

Figure 1. Band 8 brightness map of the experimental data (150 rows x 150 columns) from Chi-Tou test area 1.

3 RESEARCH METHODS

Supervised approach is utilized much of the time for defining training statistics in remote sensing applications. However, from previous work this approach is the most ineffective technique (Fleming & Hoffer, 1977) because spatial correlation between adjacent pixels may lead to biased estimates of parameters and degrade the classification performance. This study was to investigate the effects of untransformed and proposed (transformed) supervised approaches. Figure 2 is an analytic flowchart of these two approaches.

3.1 Untransformed (Biased) Training Statistics in Supervised Classification

Supervised classification consists of a training stage and a classification stage. In the training stage, about 1% of the entire image was selected as training classes based on a contiguous block selection. The selected training classes had biased estimates for the parameters because of spatial correlation and then allowed for examining the effect of biased training statistics on supervised classification.

3.2 Transformed (Unbiased) Training Statistics in Supervised Classification

Two proposed spatial transformations, the residual analysis and the principal component analysis, were used to yield transformed training statistics for supervised classification. The objective was to investigate whether these transformed training statistics could improve the classification performance and if so how much improvement could be obtained.

The whole process of yielding transformed statistics consisted of three steps: the extraction of spatial correlation, the generation of spatially uncorrelated training data, and the synthesis of spatially uncorrelated training data.

3.2.1 Extraction of Spatial Correlation

A two-dimensional causal model represented by partial differential equations as shown in equation 1 and Figure 3 was utilized to extract horizontal and vertical spatial correlations for each training class under each band. The reason is that this model can fit the image, with a nonseparable, isotropic covariance model, better than other common covariance models such as the separable covariance model (Jain, 1977).

$$Y(i,j)=r_h \cdot Y(i,j-1)+r_v \cdot Y(i-1,j)-r_h \cdot r_v \cdot Y(i-1,j-1)+W(i,j) \quad (1)$$

where r_h and r_v denote horizontal and vertical correlations.

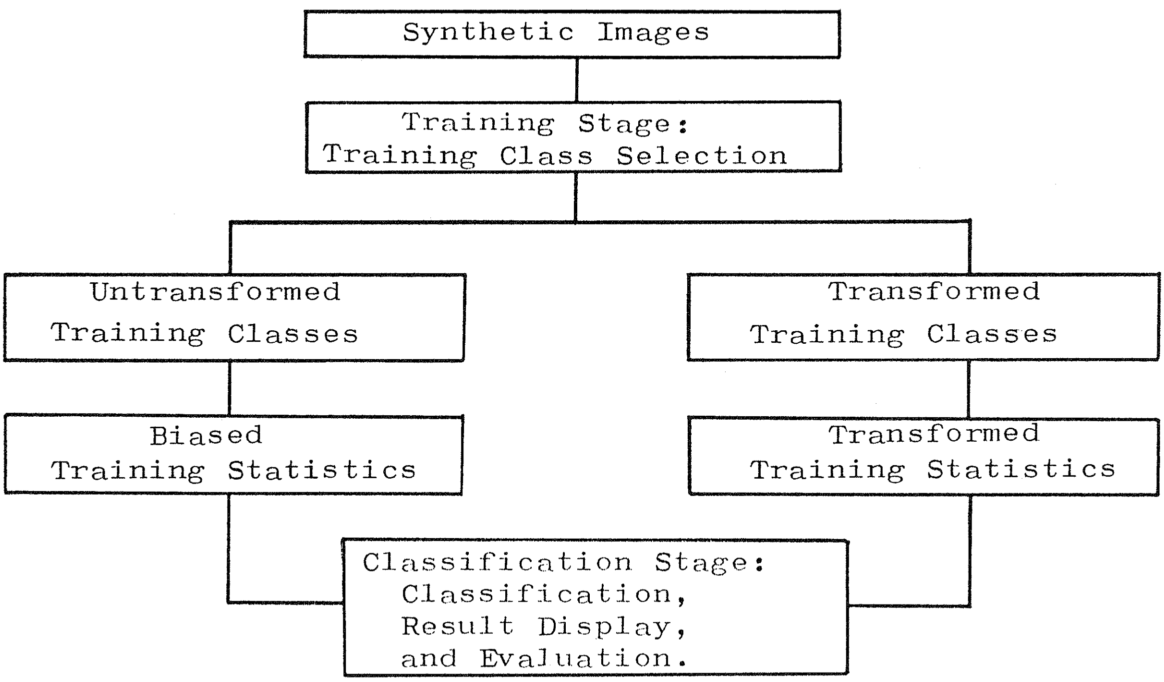

Figure 2. An analytic flowchart of using biased and transformed training statistics for supervised classification in forest applications.

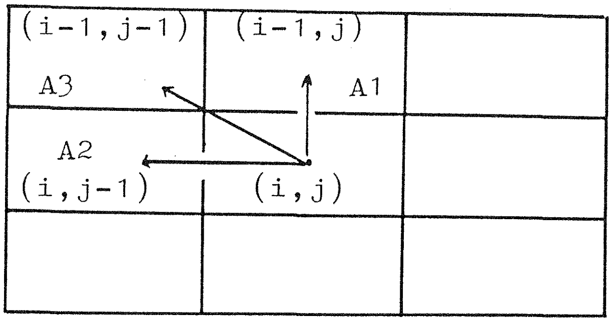

Figure 3. Random field models by partial differential equations: a causal model representation.

3.2.2 Generation of Spatially Uncorrelated Training Data

Trainin classes with spatial correlations were linearly transformed based on the following two proposed spatial transformations.

3.2.2.1 Residual Analysis

The training class used for defining training statistics is assumed to be homogeneous. It is commonly described by a spatially stationary mean and spatially stationary covariance, and modeled by

$$F(i,j)=M+W(i,j) \tag{2}$$

where M is the mean of a class, and W(i,j) is the white noise with

$$E\ W(i,j)=0 \tag{3}$$

$$E\ W(i,k).W(j,l) = \begin{cases} V & i=j \\ 0 & i\neq j \end{cases} \tag{4}$$

and V is the variance. Since the training data obtained from the supervised approach are spatially dependent, to yield the spatially uncorrelated training data for a training class, the residual analysis as in equation (5) was implemented once the vertical and horizontal spatial correlations were extracted.

$$W'(i,j)=G(i,j)-r_v.G(i-1,j)-r_h.G(i,j-1)$$
$$+r_h.r_v.G(i-1,j-1) \tag{5}$$

where G(i,j)=F(i,j)-M. The white noise W'(i,j), assumed to be spatially uncorrelated, was then used to replace the original white noise W(i,j).

3.2.2.2 Principal Component Analysis Based on a Two-Dimensional Markov Model

The white noise W'(i,j) produced by the residual analysis was assumed to be uncorrelated. However, statistically it was still somewhat spatially correlated (Jain, 1976). This kind of residual correlation can be eliminated by using the principal component transformation which will yield spatially uncorrelated white noise W"(i,j).

If a training class is represented by a two-dimensional separable covariance model with zero mean, the covariance matrix can be expressed in direct product.

$$S_x=S_h \otimes S_v \ . \tag{6}$$

If the covariance function is separable, then the transform matrix T is also separable and can be written by (Pratt, 1978)

$$T_x=T_h \otimes T_v \tag{7}$$

where T_h and T_v are the transform matrices of the horizontal and vertical directions respectively. They satisfy the following relations given in equations (8) and (9).

$$T_h \cdot S_h = D_h \cdot T_h \tag{8}$$

$$T_v \cdot S_v = D_v \cdot T_v \ . \tag{9}$$

As a result, the principal component transform of a training class may be given by

$$G=T_h \cdot F \cdot T_v \, . \tag{10}$$

It is known that the 2-D separable covariance model can be obtained using 1-D separate transformations in the horizontal and vertical directions. The elements of the transform matrix for each direction are defined by the sine transform.

$$T(i,k)=\sqrt{2/(N+1)} \ \sin\frac{ik\pi}{N+1} \tag{11}$$

$$T(j,\ell)=\sqrt{2/(N+1)} \ \sin\frac{j\ell\pi}{N+1} \tag{12}$$

The principal component transform of a 2-D separable covariance model is then given by the 2-D sine transform.

$$G(i,j)=\frac{2}{N+1}\sum_{k=1}^{N}\sum_{\ell=1}^{N} F(k,\ell)\cdot\sin\left[\frac{ik\pi}{N+1}\right]\cdot\sin\left[\frac{j\ell\pi}{N+1}\right] \ . \tag{13}$$

This 2-D sine transform is also related to the FFT (Jain, 1976; Cheng, 1987). To enhance computational efficiency, the two-dimensional sine transformation can be implemented via the two-dimensional FFT.

3.2.3 Synthesis of Spatially Uncorrelated Training Data

After the spatial transformations, the white noises $W'(i,j)$ and $W''(i,j)$, obtained from the residual analysis and the principal component analysis, respectively, can be used to replace the original training class because they are spatially uncorrelated. However, the means for both white noises are zero. Thus, translation to the origin is needed. Merembeck (1979) showed that this translation has no effect on the analysis. Both white noise are translated to the original feature space through the following relations.

$$F'(i,j)=M+W'(i,j) \tag{14}$$

$$F''(i,j)=M+W''(i,j) \tag{15}$$

The steps of both spatial transformations were summarized as follows.

The steps for the residual analysis were:

(1) Generate a zero mean training class by subtracting the class mean on a pixel-by-pixel basis.

(2) Based on the spatial correlation, generate the white noise $W'(i,j)$ by using a simple subtraction as in equation (5).

(3) Translate $W'(i,j)$ to the original feature space by using $F'(i,j)=M+W'(i,j)$. The $F'(i,j)$ training data are then used in the classification.

The steps for the principal component analysis were:

(1) Perform steps similar to the first two steps of the residual analysis and obtain the white noise $W'(i,j)$.

(2) Transform the $W'(i,j)$ to yield the $W''(i,j)$ by using the 2-D sine transform or the 2-D discrete Fourier transform.

(3) Translate $W''(i,j)$ to the original feature space by using $F''(i,j)=M+W''(i,j)$. The $F''(i,j)$ training data sets was then used in the classification.

3.3 Classification and Evaluation

Based on these biased and transformed training statistics, the maximum likelihood classifier was then used to classify the test area under the use of nine bands or a subset of nine bands selected by the stepwise discriminant analysis (BMDP7M program). Thematic maps of this test area were produced using the LMAP program of the ORSER (office for Remote Sensing of Earth Resources) package. Comparison of these thematic maps were made by visual comparison, by the tabular summaries of the mapping results, and by pixel-by-pixel comparison of the mapping results. A program entitled PERFORM, which was written for this study, was used to compare the classification map with the forest type map. A summary of pixel comparisons was then produced.

4. RESULTS

The residual analysis and the principal component analysis were used to transform biased training classes into transformed training classes. However, according to the comparison of classification results between these two transformations described in Cheng's research (1987), the residual analysis appears to be better. The reason is because the principal component analysis used more computer time, but the improved effect on the classification performance is small. Therefore, the following description of classification results is based on the residual analysis.

4.1 Classification Results Using Biased and Transformed Training Statistics with Nine-Band Data

Figures 4 and 5 are the LMAP maps produced by biased and transformed training statistics. Visual comparison of these two maps reveals differences. When compared with the ground truth information, the map using the transformed training statistics gives better classification results. The major exception of these two maps using biased and transformed training statistics is the

Cryptomeria-1 in the upper right hand portion and the middle bottom portion of the image. For the map using biased training statistics, this area is classified into bamboo. In addition, the map based on transformed training statistics is less noisy than that using biased training statistics. This is particularly noticeable in the right hand portion, the upper right portion, and the center bottom portion of the image.

In the PERFORM results, shown in Table 1(a), the map using biased training statistics differed on 2565 of 22500 pixels (11.4 %) while the map using transformed training statistics differed on 539 pixels (2.4%).

4.2 Classification Results Using Biased and Transformed Training Statistics with Four-Band Data

Four bands (Bands 1, 3, 5, and 8) were selected from the original nine bands based on the stepwise discriminant analysis (BMDP7M program). The classification maps using biased and transformed training statistics are listed in Figures 6 and 7. Visual comparison of these two maps reveals some differences. The differences are noticeable in upper center portion and the lower right hand portion of the image. Also, the map using transformed training statistics is less noisy than the use of biased training statistics. When comparing this with the nine bands, the results show that when using biased training statistics, the use of four bands gives better results than the use of nine bands. The differences are particularly noticeable in upper right portion of the image. As for the use of transformed training statistics, the use of nine bands gives less noisy map than the use of four bands.

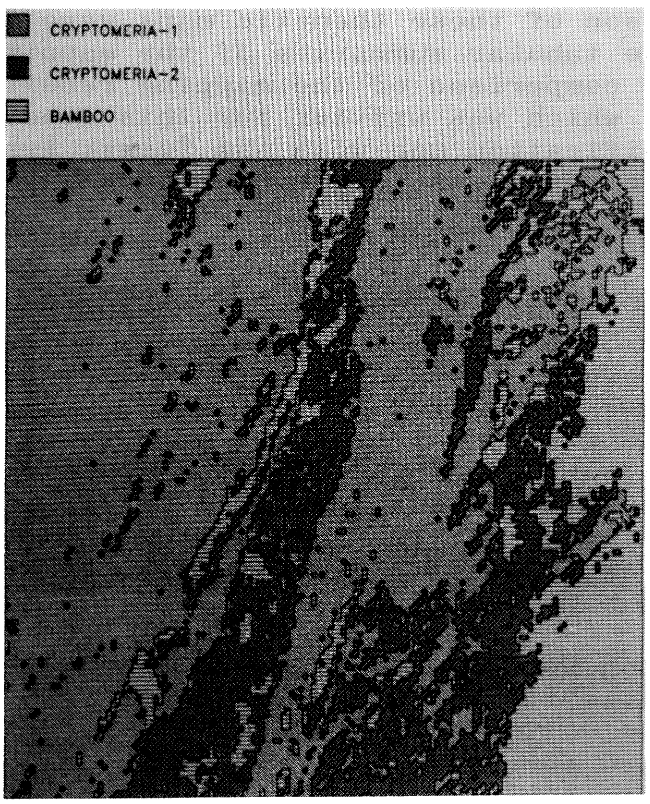

Figure 4 Thematic map of Chi-Tou test area 1 using biased training statistics with nine bands.

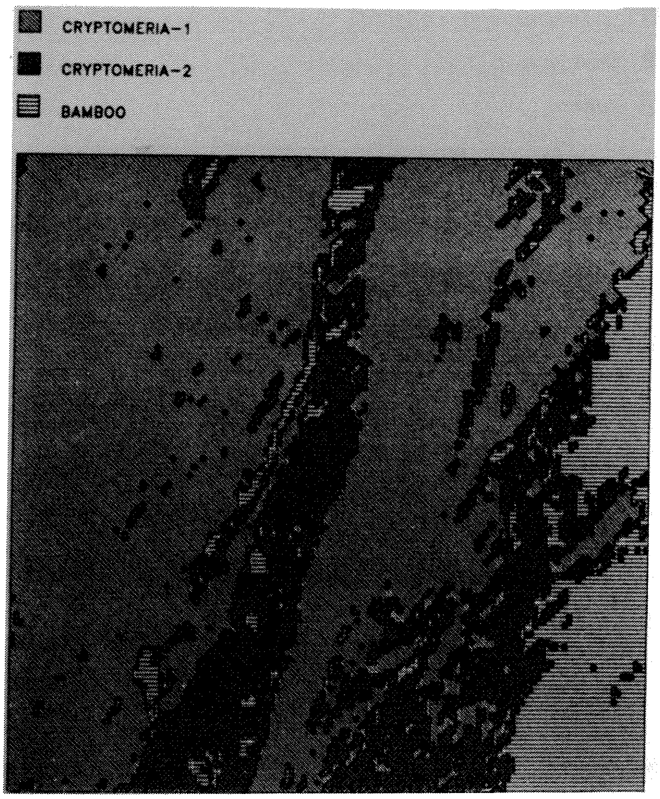

Figure 5 Thematic map of Chi-Tou test area 1 using transformed training. statistics with nine bands.

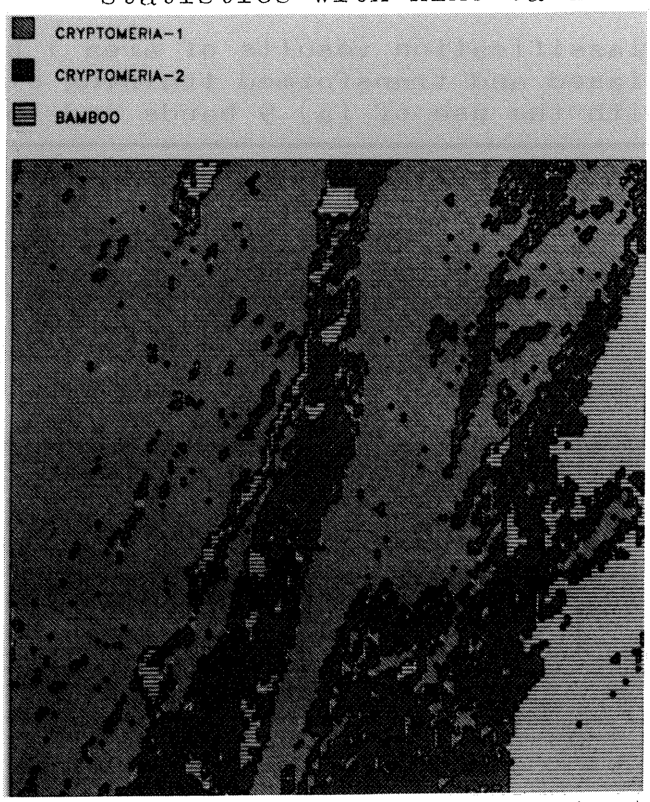

Figure 6 Thematic map of Chi-Tou test area 1 using biased training statisitcs with four bands.

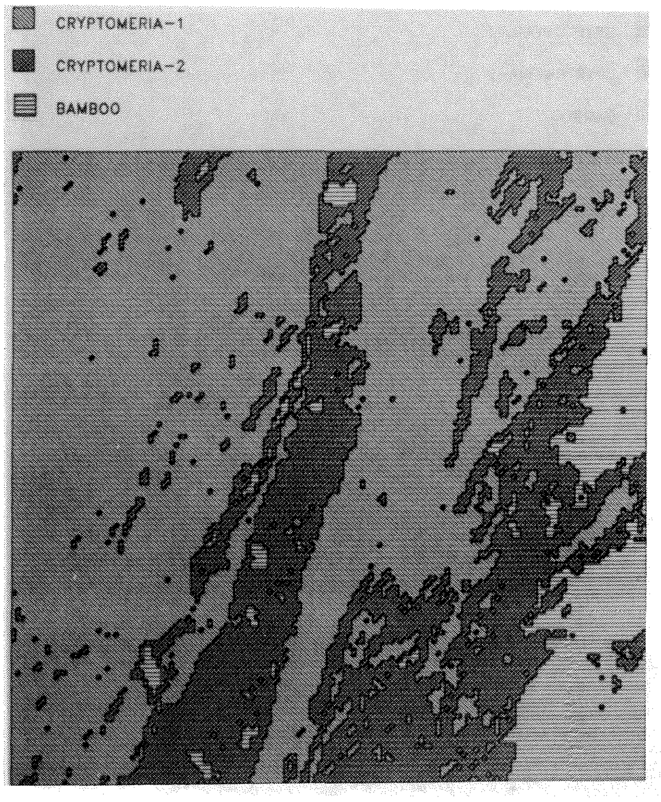

Figure 7 Thematic map of Chi-Tou test area 1 using transformed training statistics with four bands.

Table 1. Classification results of area 1 based on biased and transformed training statistics with the use of (a) 9 bands and (b) 4 bands.

	Number of samples	Biased training statistics			Transformed training statistics		
		C1	C2	C3	C1	C2	C3
C1	14471	12991	645	834	14225	170	43
C2	5132	59	4086	987	93	4914	125
C3	2897	2	37	2858	11	64	2822

Note: C1 to C3 represent Cryptomeria-1, Crytomeria-2, and bamboo.

(a) Use of nine bands

	Number of samples	Biased training statistics			Transformed training statistics		
		C1	C2	C3	C1	C2	C3
C1	14471	13033	1363	75	13427	1041	3
C2	5132	60	4662	410	125	4967	40
C3	2897	1	292	2604	1	497	2399

(b) Use of four bands

Table 1(b) shows the PERFORM results using four bands. The map using biased training statistics differs on 2165 of 22500 pixels (9.6%) while the map using transformed training statistics differs on 1707 of 22500 pixels (7.6%). Comparing these results with Table 1(a) indicates that, with the use of transformed training statistics, about 2% improvement (from 90.4% to 92.4%) is obtained when using four bands and about 9% improvement (from 88.6% to 97.6%) for the use of nine bands. The comparison also shows that when using biased training statistics, the classification results using nine bands were worse than those using four bands. However, the use of nine bands resulted in better classification performance when using transformed training statistics.

5. CONCLUSIONS AND RECOMMENDATIONS

The classification results may be concluded that low classification accuracy produced by conventional supervised approach was due to violation of the assumption of independence. However, in case of assumption violation, the proposed spatial transformations could eliminate certain errors, make the data more consistent with the assumptions, and improve the performance of classification.

The incorporation of spatial transformations for defining training statistics resulted in better classification accuracy than the conventional supervised approach. Improved classification resulted from the reduced spatial correlation and the change of the covariance matrix. From these results, it may be concluded that the incorporation of spatial transformations into conventional supervised approach may be helpful in computer-assisted classification of remotely sensed data for forestry purposes.

For practical applications in classifying forest areas from remotely sensed data, the incorporation of spatial processes into conventional supervised classification is recommended for further research. Figure 8 illustrates an analytic flowchart to be followed in incorporating spatial processes into conventional supervised classification.

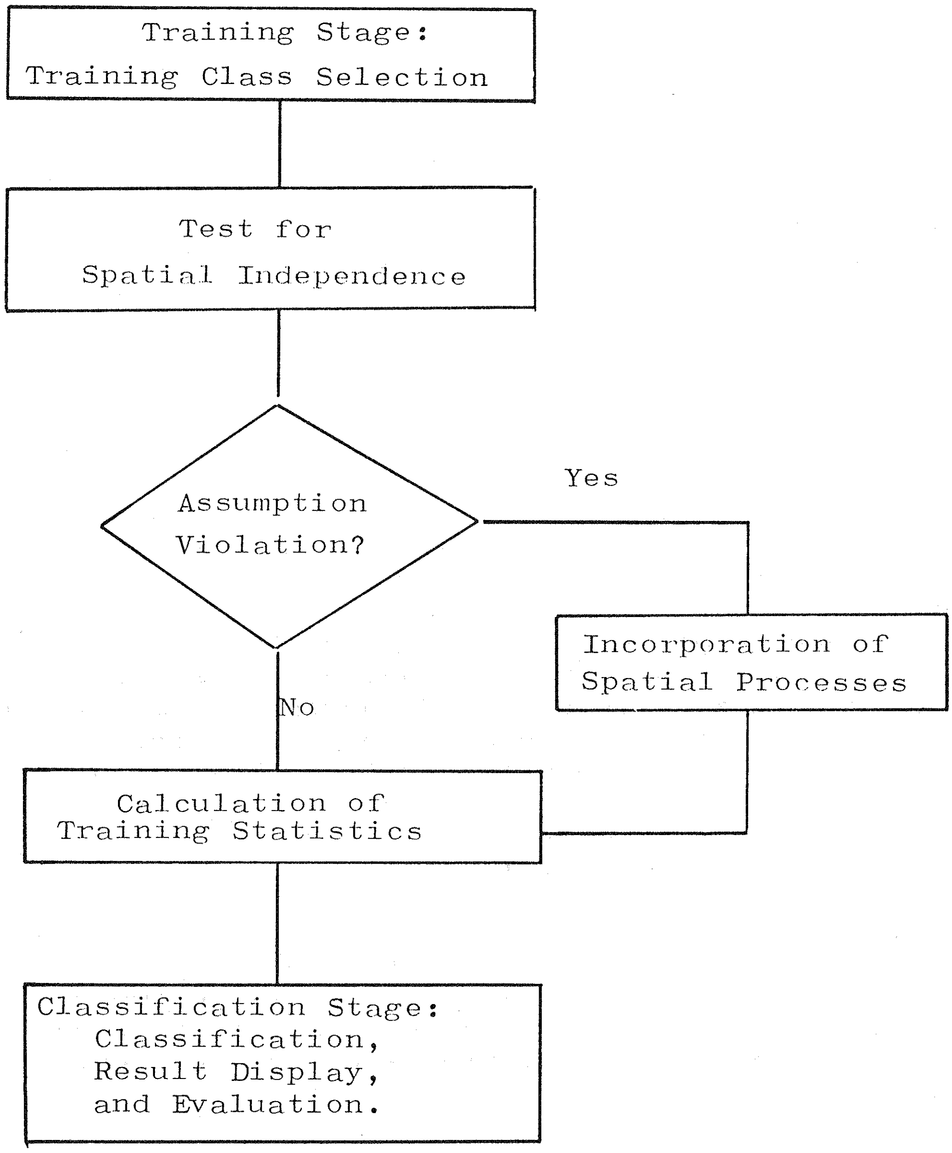

Figure 8. An analytic flowchart to incorporate spatial processes into conventional supervised classification in forest applications.

REFERENCES

Anderson, T. W. 1984. An introduction to multivariate statistical analysis. John Wiley & Sons Inc., New York.

Basu, J. P. and P. L. Odell. 1974. Effect of intraclass correlation among training samples on the misclassification probabilities of Bayes' procedure. Pattern Recognition, Vol. 6, pp. 13-16.

Campbell, J. B. 1981. Spatial correlation effects upon accuracy of supervised classification of Landsat cover. Photogrammetric Engineering and Remote Sensing, 47, pp. 335-336.

Cheng, C. C. 1987. Effects of non-normality and spatial correlation on computer-assisted classification in forest thematic mapping. Ph. D. thesis. The Pennsylvania State Univ., University Park.

Craig, R. G. and M. L. Labovitz. 1980. Sources of variation in Landsat autocorrelation. Proceedings of 14th International Symposium on Remote Sensing of Environment, pp. 1755-1767.

Jain, A. K. 1976. A fast Karhunen-Loeve transform for a class of random process. IEEE Transactions on Communication, pp. 1023-1029.

Jain, A. K. 1974. A fast Karhunen-Loeve transform for finite discrete images. Proc. National Electronics Conference, Chicago, Illinois, pp. 323-328.

Jain, A. K. and E. Angel. 1974. Image restoration, modelling, and reduction of dimensionality. IEEE Transactions on Computers, Vol. C-23, No. 5, pp. 470-476.

Jain, A. K. 1976. Some new techniques in image processing. Image Science Mathematics, pp. 201-223.

Jain, A. K. 1977. Partial differential equation and finite difference methods in image processing, Part1: Image representation. Journal of Optiminization Theory and Application, Vol. 23, No. 1, pp. 65-91.

Jain, A. K. 1981. Advances in mathematical model for image processing. Proceedings of the IEEE, Vol. 69, No. 5, pp. 502-528.

Jain, A. K. 1975. Image coding via a nearest neighbors image model. IEEE Transactions on Communications, Vol. BOM-23, No. 3, pp. 318-331.

Jennrich, R. and P. Sampson. 1977. P7M stepwise discriminant analysis in biomedical computer programs, pp. 711-740.

Kalayeh, H. M. 1982. Classification of remotely sensed image data using multitype information. Ph. D. thesis. Purdue Univ., 139 pages.

Mobasseri, B. G., C. D. McGillem and P. E. Anuta. 1978. A parametric multispectral scanner spatial model performance evaluation. LARS Technical Report 061578, Purdue Univ., 272 pages.

Pratt, W. K. 1978. Digital image processing. Wiley-Interscience Publishing, N. Y., 750 pages.

Tubbs, J. D. and W. A. Coberly. 1978. Spatial correlation and its effect upon classification results in Landsat. Proceedings of 12th International Symposium on Remote Sensing of Environment, pp. 775-780.

Tubbs, J. D. 1979. Classification results using spatially correlated Landsat data. Proceedings of 13th International Symposium on Remote Sensing of Environment, pp. 1499-1505.

Tubbs, J. D. 1980. Effect of autocorrelated training samples on Bayes' probabilities of misclassification. Pattern Recognition, Vol. 12, pp. 351-354.