Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

ITERATIVE CLASSIFICATION USING AUTOMATIC TRAINING DATA SELECTION

H. SHIMODA, T. HOSOMURA, K. FUKUE
K. KOIDE, T. SAKATA

Tokai University/Research and
Information Center
Tokyo, Japan

I. ABSTRACT

Recently, remote sensing data analysis have been widely spreaded in many application fields. Especially, digital image processing facilities have been implemented in many institutes and Universities. In the image processing techniques, an automatic classification plays a main role in remote sensing data analysis.

A maximum likelihood method is one of the most popular algorithm in image classification techniques. However, it has several problems. In this algorithm, the probability density function of each training data are assumed to be a normal distribution, but it is not true in many cases, which results in much of misclassifications.

In order to decrease the miss classification, the probability density function of the samples in each class should be very similar to a normal distribution, and each class should be choosen as pure as possible from a spectral point of view. In this study, an algorithm which can select spectrally pure categories automatically in an iterative way is developed. This method is based on a method which automatically modify the inappropriate training data using a so called residual image.

Case studies using this algorithm showed significant improvements in their classification accuracies to a simple maximum likelihood method. Further, processing time can be shorted in 1/5 compared to a simple maximum likelihood method.

II. INTRODUCTION

Applications of remotely sensed images have been conducted in many fields. These images have been processed by computers in many institutes and Universities. In remotely sensed image processings, the maximum likelihood method is one of the most popular image classification technique. However, it has several problems. These problems can be divided in two categories. One is that a maximum likelihood method is a time consuming method. The other is that though probability density functions of each traning data are assumed to be a normal distribution, it is not true in many cases. Consequently, its classification accuracy becomes poor and sufficient results can not be obtained.

In actual cases, training data are modified interactively with a computer in order to decrease these misclassifications. But this kind of job is also a time consuming process. Authors have already proposed a maximum likelihood method with table look up method to eliminate the first defect[1]. Using this method, processing time have been decreased by a large margin.

In this study, a method to eliminate the second defect is proposed. The feature of this method is that a residual image[2] is used to modify the training data. The details of this method is described in the following section.

III. AUTOMATIC MODIFICATION ALGORITHM OF TRAINING DATA.

In the classification using maximum likelihood method, classified results are largely influenced by the training data. The probability density functions of training data usually does not show a normal distribution. It is because (1) number of samples is too small compared to the large numbers of the whole pixels, (2) each pixel contains several kind of ground targets and (3) predetermined categories do not correspond spectrally independent ground objects.

To avoid such problems, it is necessary that each class is chosen as pure as possible from a spectral point of view. The following algorithm were developed to select spectrally pure categories automatically in an iterative way. The whole process is made of 8 steps.

Step 1 M_0 categories are determined according to the purpose of the classification.

Step 2 Trainig data are taken from M_0 categories.

Step 3 Maximum likelihood classification is executed for the original image.

Step 4 The Euclidian distance between the mean vector for each category and the vector of each pixel in original image is calculated. The image which shows these distances is called a residual image.

Step 5 The pixels that have a larger distance than a certain value are extracted.

Step 6 If the number of pixels extracted in the step 5 is less than the value setted previously, the operation is terminated. If it is not so, next step is continued.

Step 7 Extracted pixels are classified to several classes M_k using a clustering.

Step 8 Jump to step 3 with categories $(M_0 + M_1 + \ldots + M_k)$
If this process terminates at the m-th iteration, the number of classes are $(M_0 + M_1 + \ldots + M_m)$

IV. EXPERIMENT

Two experiments using this algorithm have been executed in this study. One is the land cover classification of Osaka district. The other is the forest type clssification of Mt.Fuji area. If a conventional method is used, following defects would be observed in each experiment.

(1) The land cover classification; As many objects are mixed together in urban areas, extractions of appropriate training areas are very difficult. Therefore modifications of training data would have taken much time.

(2) The forest type classification; The extraction of appropriate training data is difficult too, because the features of spectrum for each class are very similar.

1. Land cover classification of Osaka district.

In this experiments, LANDSAT MSS image of Osaka district was used as the object image. Fig.1 shows the original object image. The size of this image is 512 pixels X 480 lines. At first, the training data of 9 classes were extracted as the initial training data. Fig.2 shows the residual image. At the 9-th iteration, the mean and variance of residual image came less than the threshold value setted previously. At this last classification stage, the number of classified categories were 78. The relation between iteration number and the mean or variance of residual images is shown in Fig.3.

It took 5 hours in the total to classify the image using it. On the contrary it took 25 hours to execute the same job using conventional method. The classified result is shown in Fig.4 To evaluate the classification accuracy of the result, the classified result using conventional method in which a skilled operator have managed the classification process was used.

Residual images were used for the comparison. It has been shown that the miss classifications of the result using this method was better than that of the result using conventional method.

2. Forest type classification of Mt. Fuji area.

A LANDSAT MSS image covering Mt. Fuji was used as the object image in this experiment. Fig. 5 shows the object image. At first, training data were extracted from 9 classes as the initial training data. Fig.6 shows the residual images. In this experiment, expected classification accuracy are obtained at the 6-th iteration. The number of the classified categories were 80. The classified result is shown in Fig. 7. In this result, conifer forests were classified into 5 forest groups, i.e. cryptomeria (Japan cedar) and Japanese cypress, pine tree, fir, larch, Japanese spruce.

Wide leaved forests were classified into two categories, high trees and low trees. Using conventional method, forest area of LANDSAT MSS images could be classified into only three groups, conifer forest, wide leaved forest and mixed forest. So, a much finer classification was realized using this proposed algorithm.

V. CONCLUSION

As mentioned above, two case studies using this algorithm showed a significant improve in classification accuracies compaired to a simple maximum likelihood method. Other advantages of this algorithm are ;

1. Even if the number of training data is not so many initially, it is increased to the whole pixels in the image from the second iteration.

2. Even if the initial training categories are not adequate, the final result is not so much influenced.

3. At the classification, each categories are composed of spectrally independent data, i.e. it gives the good result from the view point of spectral signatures.

VI. REFERENCE

1. K.Fukue, H.Shimoda and T.Sakata " Table Look up Maximum Likelihood Method", Jounal of the Japan Society of Photogrammetry and Remote Sensing, Vol.20, No.2, (1981)

2. D.L.B.Jupp and K.K.Mayo " The Use of Residual Images in Landsat Image Analysis", Photogrammetric Engineering and Remote Sensing, Vol.48, No.4, April (1982), pp.595-604.

AUTHOR BIOGRAPHICAL DATA

Haruhisa Shimoda recieved the Ph.D. degree in solid state physics of Organic semiconductor from the University of Tokyo, Tokyo, Japan, in 1972. Since 1972, he has been an Assistant Professor of the Department of Electro Photo-Optics Engineering at the Tokai University, Kanagawa, Japan. He is currently engaged in field of digital image processing, a development of image processing system, and application of digital image processing to remote sensing.

Tsukasa Hosomura recieved the ph.D. degree in physical information engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1978. Since 1978, he has been a lecturer of the Institute of Research and Development at Tokai University, Kanagawa, Japan. He is currently engaged in digital image processing, pattern recognition, and application of digital image processing to remote sensing.

Kiyonari Fukue recived the M.S. and ph.D. degrees from the Tokai University of Kanagawa, Japan, in 1978 and 1982, respectively. Currentry, he is an assistant at the Institute of Research and Development, Tokai University. His research interests include digital image processing and system, especially in remote sensing.

Kaoru Koide recived the M.S. degree from Tokai University of Kanagawa, Japan, in 1983. From April 1981 to March 1983 he was a graduated student of master cource in the Department of Electro Photo-Optics Engineering, Tokai University. He has been engaged in the digital image processing of remote sensing.

Toshibumi Sakata recived B.S. degree of Chemical Engineering from Chiba University. He took doctor in Engineering of Chemical Physics at the University of Tokyo and then joined to the Institute of Industrial Science there, as a research associate. He was a research scientist of Munich University during 1964 to 1966. In 1966 he moved to the Tokai University and he had a chair of professor in 1971. Presently he is the director of Tokai Research and Information Center, the Tokai University.

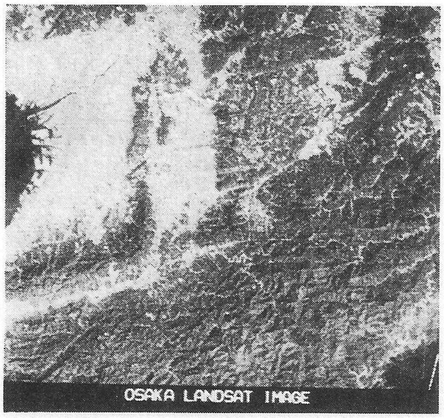

Figure 1. Original LANDSAT image of Osaka district.

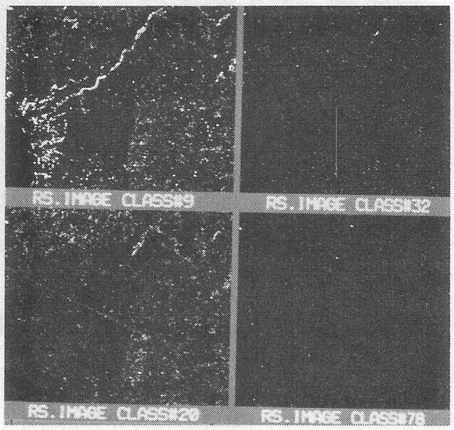

Figure 2. Residual images for Figure 1.

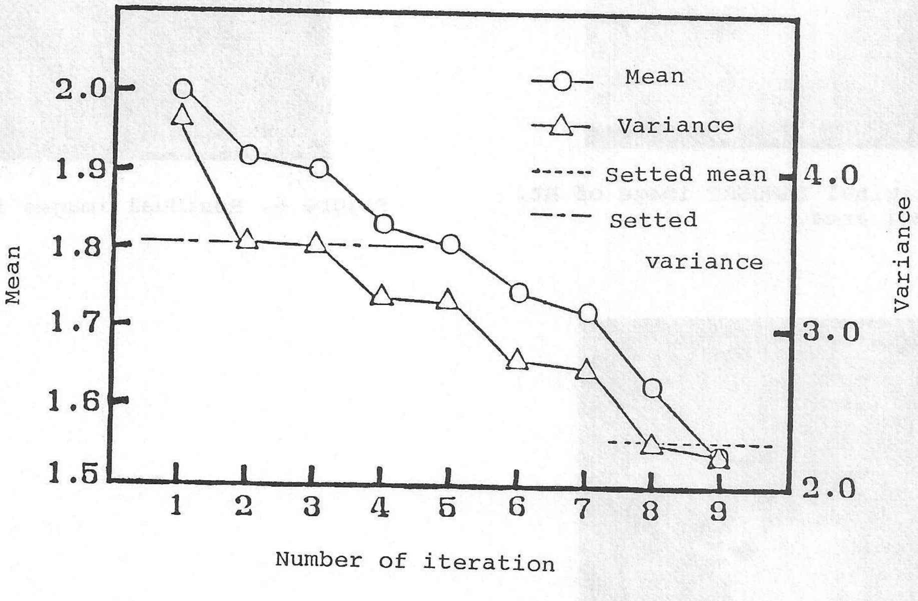

Figure 3. The relation between number of iteration and mean or variance of residual image.

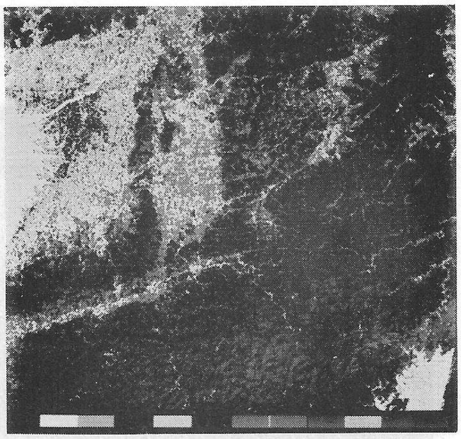

Figure 4. Classified result of Figure 1.

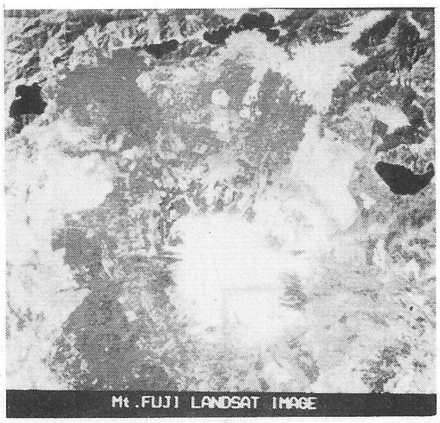

Figure 5. Original LANDSAT image of Mt. Fuji area.

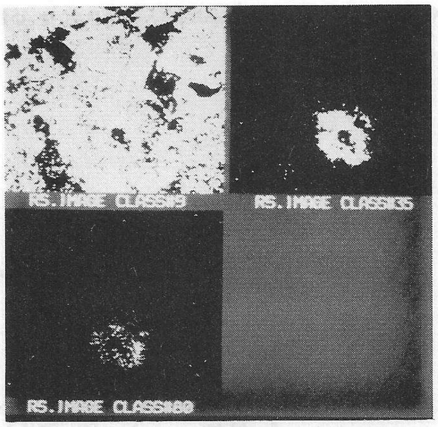

Figure 6. Residual images for Figure 5.

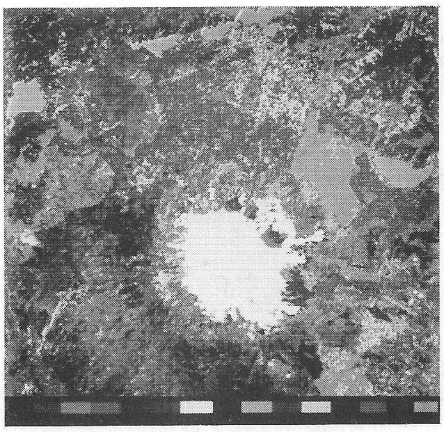

Figure 7. Classified result of Figure 5.