Evaluations of Unsupervised Methods for Land-cover/use Classifications of Landsat TM Data

Kiyonari FUKUE, Haruhisa SHIMODA, Yoshiaki MATUMAE, Ryouji YAMAGUCHI and Toshibumi SAKATA

Tokai University Research and Information Center
2-28-4 Tomigaya, Shibuya-ku, Tokyo 151, Japan

ABSTRACT

Supervised classification methods have been mainly used for land-cover/use classifications from the view point of classification accuracy, especially in the area where detailed land use dominates like in Japan. However, for high ground resolution image data such as Landsat TM and SPOT HRV data, it has been clarified that the classification accuracy using supervised classifications is lower than what have been expected. One of the major reason of this phenomena may be caused by the difficulty with selecting sufficient training data. There is a possibility to solve this problem by using an unsupervised learning method because of its independent sampling characteristics. However, quantitative evaluations of performances of unsupervised classification methods for high resolution satellite data are not yet established.

In this study, classification accuracies of unsupervised classification methods were evaluated for Landsat TM data with comparison to a conventional supervised maximum likelihood classification. The evaluated unsupervised methods are six kinds of hierarchical clustering, and a minimum residual clustering. Classification accuracies were estimated quantitatively by using digital land-cover/use test site data which were created by the authors.

As a result, most of clustering methods showed higher classification accuracies than a conventional supervised maximum likelihood classification, especially for urban and agricultural areas.

1. INTRODUCTION

Applications of Landsat TM and SPOT HRV data have been just started in the last several years in the field of land-cover /use classifications. However, classification accuracy using supervised pixel-wise method is lower than what was expected for those high resolution satellite data. Following two major reasons may have caused this phenomena[1),2)]. The first reason is that each pixel does not correspond to one landcover category. For instance, low density urban area as a classification category can be resolved to many kinds landcover components such as house roofs, lawns, trees, concretes, asphalts, bare grounds,etc. The same landcover components can be contained in a pixel of other classification categories such as high density urban areas, agricultural areas, forests,etc. These kinds of mixtures of landcover components in a pixel may cause a large misclassification. The second reason is that it is difficult to extract sufficient training data using supervised method because of large image level variances in each region. For instance, values of pixels in low density urban areas have very different values by the same reason above. Therefore, it is very difficult to select sufficient homogeneous areas for each category.

For the first problem, Zhan et.al.[2)] proposed a two step classification procedure consisted of pixel-wise landcover classification and spatial land-cover/use recognition. As for the second problem, there is a possibility to solve the problem by using unsupervised learning method. However, quantitative performances of unsupervised classification methods for high resolution satellite data are not yet clarified. In this study, classification accuracies of unsupervised classification methods were evaluated for Landsat TM data.

2.Evaluated Unsupervised Classification Methods

The evaluated unsupervised methods are six kinds of hierarchical clusterings and a minimum residual clustering[(3)]. Nearest neighbor, farthest neighbor, median centroid, group-average method and Ward methods were used for hierarchical clustering. The outline of these clustering procedure are follows.

1)hierarchical clustering

Figure 1 shows the procedure of hierarchical clusterings. Distances between clusters are defined as in Table 1. As a distance, d_{fg} in equation(1) is used for nearest neighbor, farthest neighbor and median methods. On the other hand, d_{fg}^2 in equation(2) is used for centroid, group-average and Ward methods.

$$d_{fg} = C_h d_{fh} + C_1 d_{fl} + C_1 d_{hl} + C_2 |d_{fh} - d_{fl}| \quad .....(1)$$

$$d_{fg}^2 = C_h d_{fh}^2 + C_1 d_{fl}^2 + C_1 d_{hl}^2 + C_2 |d_{fh}^2 - d_{fl}^2| \quad .....(2)$$

cluster-h and l : merged cluster
cluster-g : new cluster(cluster-h+l)
d_{ab} : distance between cluster-a and b

C_h, C_l, C_1, C_2 are coefficients and the values of these coefficients are shown in Table 2.

2)Minimum Residual Clustering

Figure 2 shows the procedure of the minimum residual clustering. Data for clustering are extracted adaptively from the object image in this method. At first, a maximum likelihood classification is conducted by using initial training data sets selected by a supervised method. At the second step, residual image R(x,y) is calculated following the equation(3).

$$R^2(x,y) = \{M(x,y)-O(x,y)\}^2 \quad .....(3)$$

where $O(x,y)$: original image
$M(x,y)$: mean image

A mean image M(x,y) can be created as follows. At first, mean vectors for each category are calculated from the classified result and the original image. Each pixel of the mean image has the same mean vector corresponding to the classified category. Then a clustering is performed for pixels in the original image which has large residual values. In this study, a hierarchical clustering with group-average distance was used as a clustering method in this stage. Finery, the generated clusters are added to the training data set and the procedure jumps to the first step,i.e. a maximum likelihood classifier. This procedure is iterated until 1)the mean value of a residual image is smaller than a certain threshold value, or 2)the total number of training classes is larger than a predefined number, or 3)the iteration number is larger than a predefined value.

3 Experiments

3.1 Test image data

TM data used in the experiments are show in Table 3. In order to reduce the amount of data, a principal component analysis were conducted and the first four components were used for classifications. The accumulated proportion of 4th component is about 99.3%. Fig.3 shows the object area. This area contains urban, agricultural areas, forests, rivers, a sea, etc.

3.2 Digital test site data

The test site covers 2km x 10km area which contains city areas, agricultural areas, forest, rivers and a sea. It locates about 50km west form central Tokyo and Includes the main campus of our university.

The test site data was generated as follows. First, aerial infra-red color photographs over the test site area were taken. A land-cover/use thematic map with scale of 1:2,500 was generated by photo interpretations of these photographs. This thematic map was than improved by highly intensive ground investigations which was conducted from October to November in 1983. The improved thematic map was digitized with ground resolution of 1m. After several preprocessings, each polygons of thematic map were automatically recognized.

Finally, the test site data was resampled with 10m x 10m pixel size. It contains 52 land-cover/use categories. In order to assess classification results of large pixel size images, this basic 10m pixel data were further resampled to 20m, 25m, 50m and 75m using majority law. Figure 3 shows the digital land-cover/use test site data.

3.3 Hierarchical clustering

For Hierarchical clusterings, samples were sampled from 11x11 grid points. Eleven iterations were done for this sampling by moving the grid, hence 121x11 pixels were sampled and used for clusterings.

In each clustering experiments, 40 categories were generated and the clustered data were used as training data for a classifier using a minimum distance method with Euclidian distance.

Classified results are shown in Fig. 4. In this figure, 40 categories are merged to 16 categories for easier discrimination as shown in Table 4. Classification accuracies calculated by using the digital test site data are shown in Table 5. In this table, categories are further reduced to 5 in order to coincide the classified categories and test site categories. Accuracies in the table are area weighted means for each category.

3.4 Minimum residual clustering

In the case of minimum residual clustering, 80 categories were generated. These 80 clusters were used as training data for a maximum likelihood classifier. Classified results are shown in Fig. 5 with the same 16 categories defined in Table 4. Classification accuracies are shown in Table 5 together with the results of hierarchical clusterings.

3.5 Supervised maximum likelihood classification

For the comparison with clustering methods, a conventional supervised maximum likelihood classification was conducted to the same data. In this case, 90 categories were chosen as training data by a skilled operator. Classified results and classification accuracies are shown in Fig.6 and Table 5, respectively.

3.6 Generate categories

Table 6 shown the number of classes generated by each classification method for each 5 major categories. As the total number of classes are different between classification methods, normalized number of classes are shown in Table 7.

4. Evaluations

1) Hierarchical Clustering

The performances of hierarchical clusterings can be divided into two groups from the view point of mean classification accuracies. The first group is composed of Ward method, group average method and centroid method while the second group is composed of the remaining 3 methods which have showed 5 to 11% lower accuracies than the first group.

Hierarchical clusterings showed better mean classification accuracies than supervised maximum likelihood method except nearest neighbor method. These results contradicts the expectation which has originated from experiences for MSS data analyses. It is natural that the accuracy of nearest neighbor method is so low, because is will generate statistically biased training data. This bias is caused by the fact that clusters tends to be merged in a chain shape in this method. However, the result that several clusterings showed more than 6% higher accuracy than MLM forces us to reconsider our common knowledge.

The largest reason for the fact that the second group showed lower accuracies than the first group is that this group showed more than 10% lower accuracies in urban areas which have 49.3% area within the digital test site area. The numbers of clusters generated for the second group are less than those for the first group as shown in table 5. This can be thought to be one reason for the lower accuracies. Another reason can be thought as follows. Substantially, centroids of each cluster in the second group largely more when clusters merge, because distances between clusters are calculated without any influence from number of elements in each cluster. Therefore, clusters generated by the second group may not be so reliable as training data.

The reason why the first group showed about 7% higher accuracies than MLM is that those methods have showed about 10% higher accuracies for urban areas as well as about 20 to 30% higher accuracies for paddies. From the confusion matrix for MLM, 46.9% paddies are misclassified to others. Further, number of classification classes for MLM is larger than those of the first group as shown in Table 5. It means that the variances of training data used in the MLM are far smaller than those of the populations.

These kinds of errors concerning training data selection may occur generally considering the fact that these training data were selected by a skilled operator on the basis of detailed group truth. It may be concluded that the classification accuracy problem concerning with paddies is not a special case. Rather, it means the effectiveness of clusterings for training data selections.

Extractions of sufficient training data from high ground resolution images like TM are very difficult for human being, especially in the areas like urban where so many ground cover materials exist in each pixel. From table 5, it can be shown that the numbers of classes selected by a human operator is about 1/4 of those generated by the first group clusterings in urban areas. This fact supports the above discussions.

2) Minimum residual clustering

The mean classification accuracy of the minimum residual clustering (MRC) is 64.3% which is almost the same as the results of Ward method and group average method. MRC should have shown a better result than Hierarchical clusterings which are substantially simple algorithms. Our experiments, however, showed a different result.

The reason for this rather low accuracy can be considered as follows. In the MRC, clusters tends to be generated around pixels which have exceptionally different values compared to the most of pixels. In order to eliminate this effect, the number of classes generated by the MRC should be far larger than that which is necessary for classification. In the case of MSS data where around 20 categories ate usually selected by human operators, 80 classes were sufficient for MRC. However, for TM, 80 classes may have been too small to eliminate the above effect.

5.Conclusion

Evaluations of classifier capabilities for land use/cover classifications of two clustering algorithms, i.e. hierarchical clustering and minimum residual clustering were conducted based upon a digital test site data. The former algorithm is composed of 6 distance measures, i.e. nearest neighbor method, farthest neighbor method median method, centroid method, group average method and Ward method. The target image used in the evaluation experiments was a TM image. In order to compare these clusterings with supervised method, a supervised maximum likelihood classification was also performed to the same data. In this study, clustering were used to generate training data sets for classifiers.

Three hierarchical clusterings, i.e. nearest neighbor method, farthest neighbor method and median method showed the lowest classification accuracies because of the used distance measures which are not weighted by numbers of elements within clusters. Especially, the nearest neighbor method showed about 5% lower accuracy than MLM.

On the contrary, 3 clusterings, i.e. Ward method, group average method and minimum residual clustering showed 6 to 7% higher accuracies than MLM. The MRC should have shown a higher accuracy. An experiment with a larger number of classification classes should be tried in the future.

As a whole, statistics of automatically generated training data by clusterings have been nearer to those of the population than those of training data extracted by human operators. This fact suggests that the human extractions of training data may be not appropriate for high ground resolution images.

References

1) H. Shimoda, et. al, "Accuracy of land-cover/use classification of TM data", Proc. 8th Asian Conf. Remote Sensing, pp. B.4.1-B.4.9(1987)

2) Zi-jue Zhang, et. al, "Spatial information processings of TM data", Proc. 8th Asian Conf. Remote Sensing, pp. F.5.1-F.5.6(1987)

Table 1 Distances between clusters.

nearest neighbor method	: minimum distance of between cluster components
farthest neighbor method	: maximum distance between cluster components
median method	: median distance between cluster components
centroid method	: mean distance between cluster components
group-average method	: root square mean distance between cluster components
Ward method	: variance of the merged cluster

Table 2 Coefficients of the equation(1) and (2)

	Ch	Cl	Cl	C2	eq.
nearest neighbor method	1/2	1/2	0	-1/2	(1)
farthest neighbor method	1/2	1/2	0	1/2	(1)
median method	1/2	1/2	-1/4	0	(1)
centroid method	Nh/Ng	Nl/Ng	-NhNl/NgNg	0	(2)
group-average method	Nh/Ng	Nl/Ng	0	0	(2)
Ward method	(Nf+Nh) / (Nf+Ng)	(Nf+Nl) / (Nf+Nl)	-Nf / (Nf+Ng)	0	(2)

Table 3 Object TM data.

channels	1,2,3,4,5,7 (6channels)
path-row	107-35
date	Nov. 4, 1984
level	bulk processed data

Table 4 Classification Categories.

Color Coding Categories	(number of training classes)	Major Categories
1. coniferous forest 1	(8)	trees
2. broad leaved forest 1	(7)	
3. coniferous forest 2	(2)	
4. broad leaved forest 2	(7)	
5. bright shadow	(4)	
6. dark shadow	(5)	
7. urban	(3)	urban
8. low density urban area	(3)	
9. factories	(3)	
10. paddy	(6)	paddy
11. sea, river	(10)	water
12. waste lands,grounds,sandy beach	(10)	other
13. rock	(2)	
14. farm	(7)	
15. grasslands	(10)	
16. golf courses	(2)	

Table 5 Classification accuracies.

	TREES	PADDY	URBAN	WATER	OTHER	mean
hierarchical clustering						
1)nearest neighbor method	57.0	60.4	52.7	74.2	47.0	53.8
2)farthest neighbor method	62.2	53.1	71.3	77.3	35.2	59.6
3)median method	42.6	67.0	67.2	75.0	40.7	59.0
4)centroid method	54.9	70.8	83.5	76.6	22.4	63.3
5)group-average method	43.2	71.8	80.6	69.9	36.4	64.5
6)Ward method	54.9	57.4	80.4	76.5	37.6	64.6
minimum residual clustering	55.6	56.1	76.1	72.3	45.5	64.3
maximum likelihood method	51.3	41.0	70.5	74.4	38.1	57.9
area coverage	6.3	9.5	49.3	7.6	27.3	

Table 6 The number of classes.

	TREES	PADDY	URBAN	WATER	OTHER	total
hierarchical clustering						
1)nearest neighbor method	9	2	12	2	15	40
2)farthest neighbor method	11	2	12	3	12	40
3)median method	8	2	13	2	15	40
4)centroid method	9	1	16	2	12	40
5)group-average method	7	1	17	2	13	40
6)Ward method	7	2	17	2	12	40
minimum residual clustering	19	9	19	6	27	80
maximum likelihood method	33	6	10	10	31	90

Table 7 The normarized number of classes.

	TREES	PADDY	URBAN	WATER	OTHER
hierarchical clustering					
1)nearest neighbor method	22.5	5.0	30.0	5.0	37.5
2)farthest neighbor method	27.5	5.0	30.0	7.5	30.0
3)median method	20.0	5.0	32.5	5.0	37.5
4)centroid method	22.5	2.5	40.0	5.0	30.0
5)group-average method	17.5	2.5	42.5	5.0	32.5
6)Ward method	17.5	5.0	42.5	5.0	30.0
minimum residual clustering	23.8	11.3	23.8	7.5	33.8
maximum likelihood method	36.7	6.7	11.1	11.1	34.4

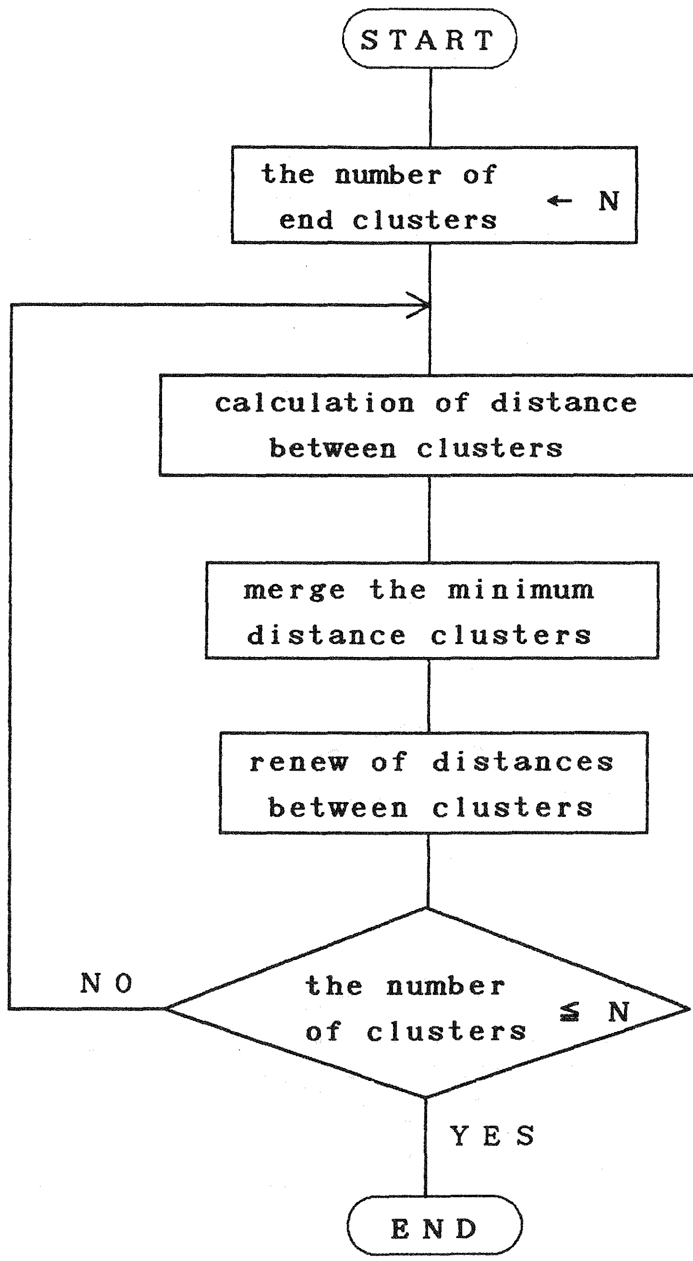


Fig.1 The procedure of hierarchical clustering.

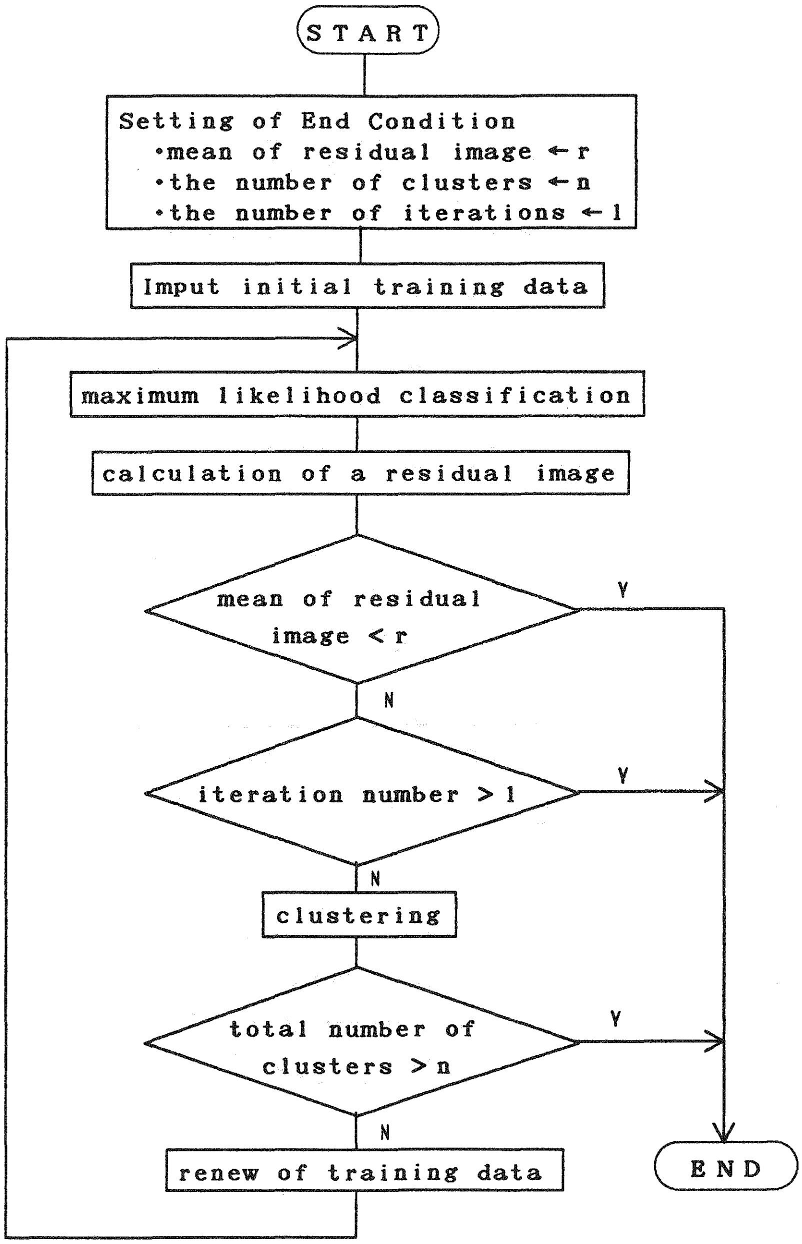
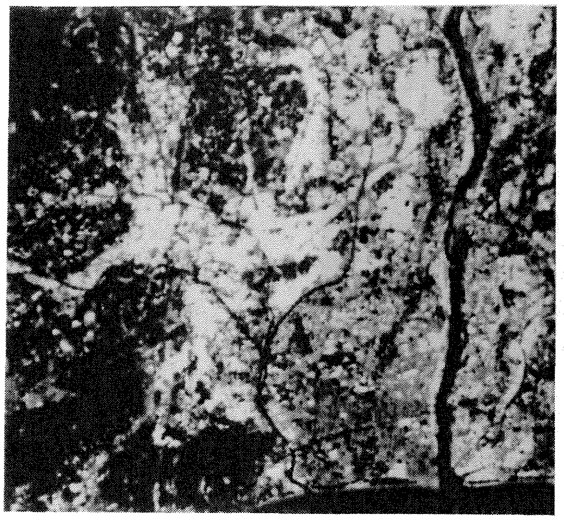
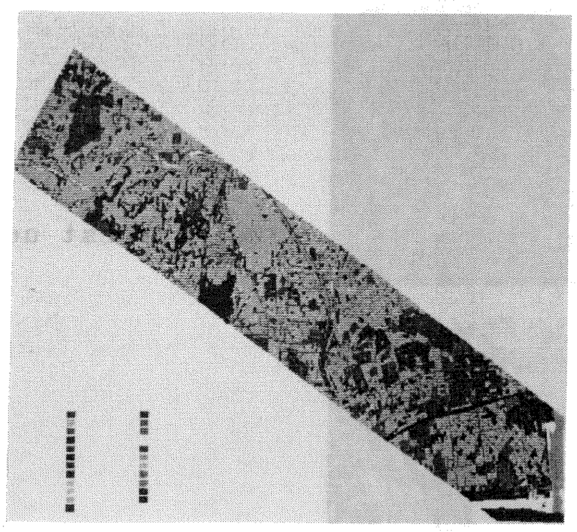


Fig.2 The procedure of minimum residual clustering.

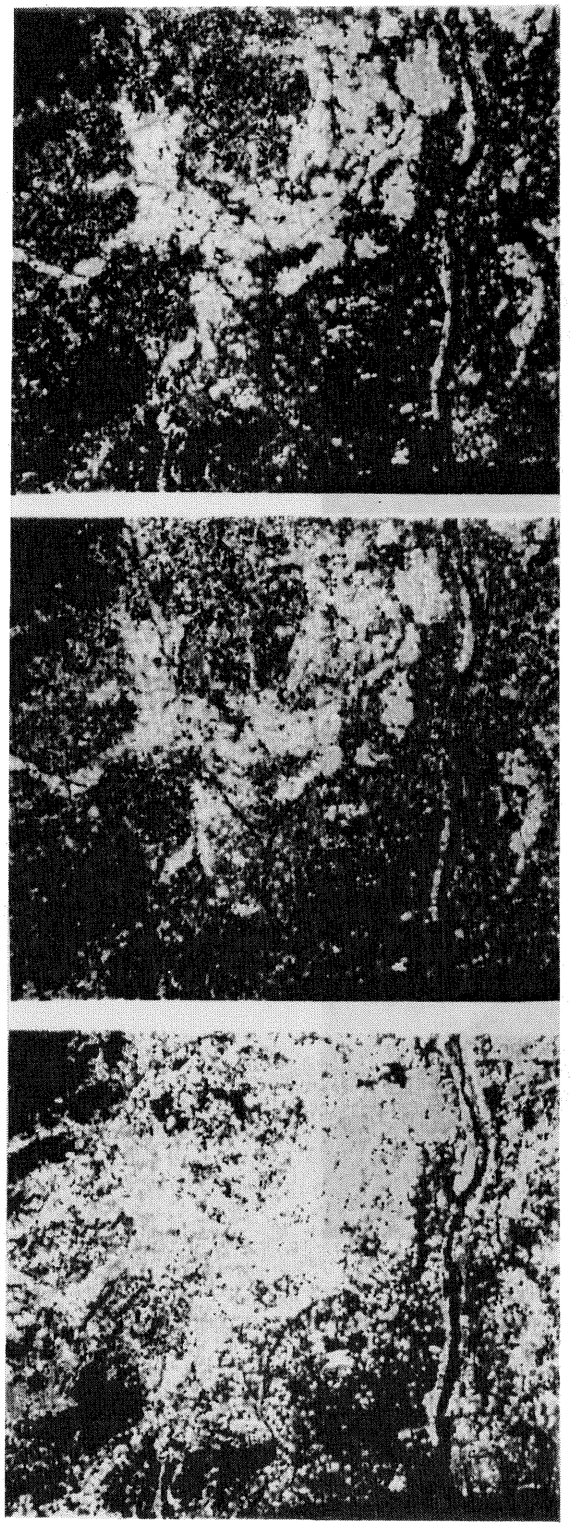


(a)Object TM data.



(b)Digital test site data.

Fig.3 Object TM image and digital test site data.



(a)nearest neighbor method

(b)farthest neighbor method

(c)median method

Fig.4 Classification results of hierarchical clusterings.

(d)centroid method

(e)group-average method

(f)Ward method

Fig.4 (continue)

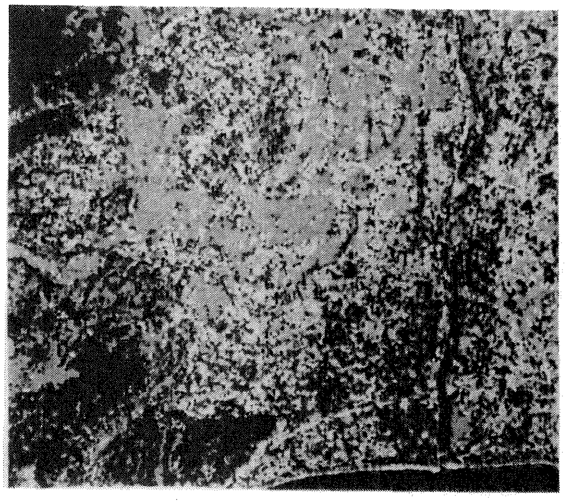


Fig.5
Classification results of the minimum residual clustering.

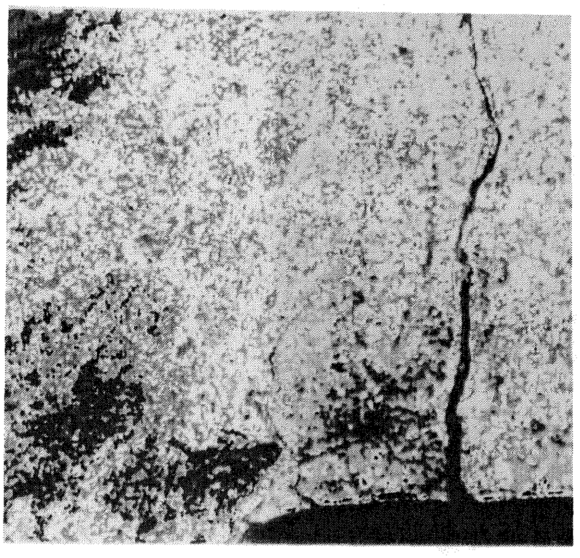


Fig.6
Classification results of the maximum likelihood classification.