MAXIMUM LIKELIHOOD METHOD MODIFIED IN ESTIMATING A PRIOR PROBABILITY AND IN IMPROVING MISCLASSIFICATION ERRORS

Junichi SUSAKI[*], **Ryosuke SHIBASAKI**[**]
[*]Institute of Industrial Science, University of Tokyo
Meguro-ku, Tokyo, 153-8505 Japan, e-mail: suzaki@skl.iis.u-tokyo.ac.jp
[**]Center for Spatial Information Science
University of Tokyo
Meguro-ku, Tokyo 153-8505 Japan, e-mail: shiba@skl.iis.u-tokyo.ac.jp

KEY WORDS: Maximum Likelihood Method (MLM), Mixel, Data Fusion, Decision Tree

ABSTRACT

Maximum Likelihood Method (MLM) has been one of the most traditional classification methods in remote sensing field, but its disadvantages have been also pointed out. While a prior occurrence probability gives a crucial effect to classification results, most of classifications have been conducted on an assumption that each *a prior* probability of land cover is equal because of insufficient a priori information. And as long as the class showing the highest likelihood is allocated to a pixel, misclassification errors are unavoidable. Authors modified method can estimate *a prior* probability through EM (Expectation Maximization) algorithm, applying a finite mixture model for a target image histogram. And misclassification errors can be overcome by data fusion model. Validation results demonstrate that data fusion model is effective to improve misclassification errors.

1 INTRODUCTION

The environmental issues have been focused on not only local level but also global level, which requires that land cover data from satellite images have both wideness and high accuracy. On the other hand, even now that multi-resolution and multi-spectral satellite images can be obtained, most of traditional classification methods deal with just single source of data, and an effective method of multi-data fusion has not been reported so far in the remote sensing field.

This paper focuses on land cover classification method using satellite images. Firstly, authors explain the limitation of MLM, one of the most traditional classification method. Then, authors propose a classification method based on data fusion technique in order to overcome the MLM limitation.

2 PROBLEMS IN CLASSIFICATION BY MLM

2.1 A priori probability

A priori probability in MLM is originated in the context of Bayesian estimation, expressed in Equation (1).

$$P(H_i \mid A) = \frac{P(A \mid H_i) \cdot P(H_i)}{\sum_i P(A \mid H_i) \cdot P(H_i)} \tag{1}$$

Here, H_i : hypothesis, A : experimental data, $P(H_i \mid A)$: *a posterior* probability, $P(A \mid H_i)$: probability density, $P(H_i)$: *a priori* probability. Equation (1) means that likelihood of hypothesis H_i when data A is observed is proportional to the multiplication of *a priori* probability $P(H_i)$ with likelihood A given H_i ($P(A \mid H_i)$). Most of classification methods used in remote sensing field assumed that *a priori* probability for each class is identical because the information is difficult to be estimated. The assumption gives a crucial influence to the classification results. On the other hand, some methods extracted *a priori* probability by examining a specific area and apply it for the whole of the classification area. But, this application is effective only when there is no reasonable difference between occupying ratio of land covers in the specific region and the whole of area. But, in fact, the similarity between the local area and the whole is not always true. Figure 1 shows how a priori probability gives classification result. If *a priori* probabilities are assumed identical of

all the classes, even when they are not so, there'll be a discrepancy between real parameter for classification and estimated one. The discrepancy gives a mis-classification.

2.2 Overlapping of probability distributions

When a probability distribution is not clearly discriminated from another, the situation is expressed in Figure 2. The distributions intersection means each likelihood for each class is equal, and is utilized as a threshold for classification. If the whole of probabilities in the overlapping section is bigger, i.e. the gray area in Figure 2 is larger, more misclassification tends to be caused. Even if the overlapping area is limited on the training data level, it is not always true on the whole level, which is essentially same as 2.1. Especially, wide-area classification requires to examine the difference between features extracted on the local level and ones on the whole level.

2.3 Mixel

The overlapping section, mentioned in sub-section 2.2, has a high possibility that probability distribution of mixel class is existing. This tendency is dominant especially when low-spatial resolution satellite images are used [1]. Figure 3 shows the mixture of both pure class distributions and mixel class distribution. Even if the whole of probabilities in the overlapping section is very small, the classification accuracy will be limited without consideration of mixel existence.

The points are, 1) even if there is no mixel in a satellite image, overlapping of probability distributions in the feature space causes misclassification, and 2) if a low-spatial resolution satellite image is used for classification, mixel class distribution exists and the classification accuracy is dependent on how mixel is dealt with. If traditional MLM is simply applied for classification, it cannot overcome the problems described in this section, and high classification accuracy cannot be always expected.
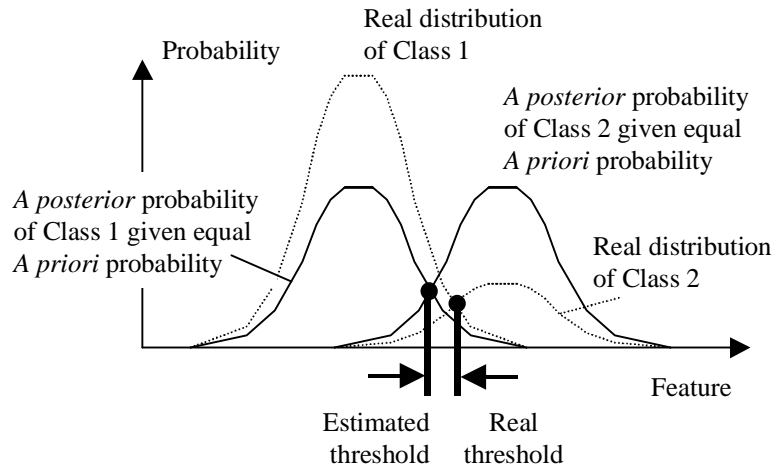

Figure 1. Discrepancy between real threshold and estimated one for classification

Figure 2. Misclassification due to overlapping distributions

Figure 3. Mixel class distribution

3 METHODOLOGY

As for the problems of overlapping distribution and mixel, authors consider those problems cannot be overcome completely just by use a one-level spatial resolution image. For example, NOAA AVHRR images, low-spatial resolution satellite images, have limitation to give the information which are more detailed and accurate than 1.1km-level information. But, it is possible to detect the feature space where possibility of misclassification from overlapping distribution and mixel problems is quite high. In authors method, histogram decomposition method (Susaki *et al.* 1999) is used for the detection. The method assumes that an image histogram is composed of pure class distributions and
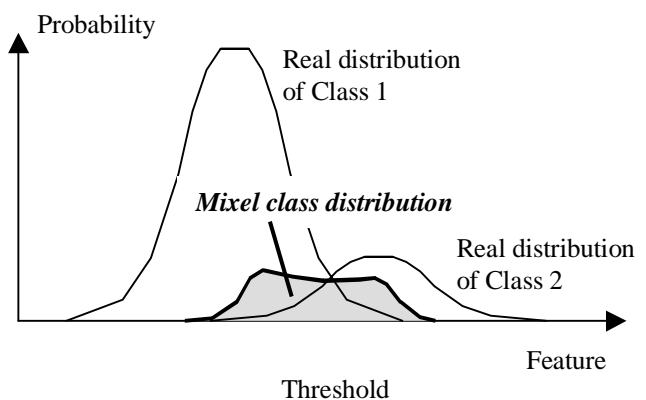
mixel class distributions. Initial parameters are given automatically and iteration for optimization is performed by EM (Expectation Maximization) algorithm (Redner *et al*. 1984). The method can decompose the image histogram into each class distribution, including mixel class. The distributions for all classes reflect a priori probability. For example, in the case of 2 pure classes, the problem is to decompose an image histogram into 2 pure classes distributions and 1 mixel class one.

And after the whole of histogram is decomposed into each class distribution, firstly feature spaces are divided into two, feature space classified with high reliability and one classified with low reliability. Pixels belonging to the low-reliability feature space are classified by use of higher-spatial resolution satellite images.

These procedures are based on the decision tree, one of the traditional classification methods in remote sensing field. The principle is to discriminate firstly the class with the highest discrepancy. In order to quantify the discrepancy of one class from others, training data are used and KL (Kulllback-Leibler) entropy can be an index for the discrepancy. The procedures are described below.

1) Determine candidates of land cover classes temporarily
2) Consider all the combinations from the class candidates, and calculate the discrepancy of the combination by use of training data set
3) Architect a decision tree based on the calculated discrepancy
4) (As for low-reliability feature space) classify through traditional MLM by use of a higher-spatial resolution satellite image

4   EXPERIMENT

As low-spatial resolution satellite images, AVHRR NDVI 10-day composite images were downloaded from the web site EROS (The Earth Resources Observation Systems) and monthly composite images (92.April to 93 March) calculated from them were used. The classification was targeted on Indo-China peninsula, and the images coordinate were N^o-$N25^o$, $E90^o$-$E110^o$. On the other hand, Landsat TM NDVI images were used as high-spatial resolution satellite images. The decision tree for classification was constructed by comparing KL entropy, which reflects the distance between probability distributions, and Susaki method (Susaki *et al*. 1999) was used to determine the feature space classified with high reliability from the one with low reliability.

As following 1)-4) described in section 3, the process to construct the classifier is briefly explained. In the step 1), land use map of Thailand in 1990 and temporal AVHRR NDVI images were examined. In fact, there are lots of land uses, but AVHRR NDVI images could not identify all of them. Authors temporarily determined four class candidates, evergreen forest, deciduous forest, crop and paddy.

In the step 2), firstly training area for four class candidates were selected by comparing land use map of Thailand and TM NDVI images. Then, KL entropy for all the combinations were calculated by referring to the training area in the temporal AVHRR NDVI images.

This procedure is based on the background below. This research focused on Indo-china peninsula, but the monthly composite AVHRR NDVI images on the region have quite high temporal correlation. Principal component analysis (PCA) gave the first principal component whose ration is more than 98%. Therefore, the first principal component can be regarded as an index reflecting temporal information, and authors judged that the first principal component can be used for classification considering temporal feature changes. And the discrepancy between the feature distributions of land covers is not always maximized when use of the first principal component calculated from 12-month NDVI data set. So, we have to examine which data set gives the biggest discrepancy before constructing decision tree. If the data set giving the biggest discrepancy is used for construction of the decision tree, the most optimal can be constructed. The discrepancies were calculated by use of KL entropy. In this experiment, four class candidates were firstly assumed, then the KL entropies in all the combinations were examined. Examples of the combinations are below; (1) Evergreen vs Others (Deciduous + Crop + Paddy), (2) Evergreen + Deciduous (called as "Forest") vs Crop + Paddy (called as "Non-Forest"), (3) Paddy vs Others (Evergreen + Deciduous + Crop). Authors calculated KL entropy for such an each combination by changing the dimension of input data set. The decision tree was constructed like Figure 4. The phrase "the first principal component from 12-month NDVIs" in Figure 4 means the KL entropy calculated from it was maximum of all the data set and all the combination. In the first stage of the decision tree, the combination, Forest vs Non-Forest, and the data set, the first principal component from 12-month NDVIs, were adopted. Next, in the second stage to classify Evergreen from Deciduous, the data set, the first principal component from 5-month NDVIs gave the
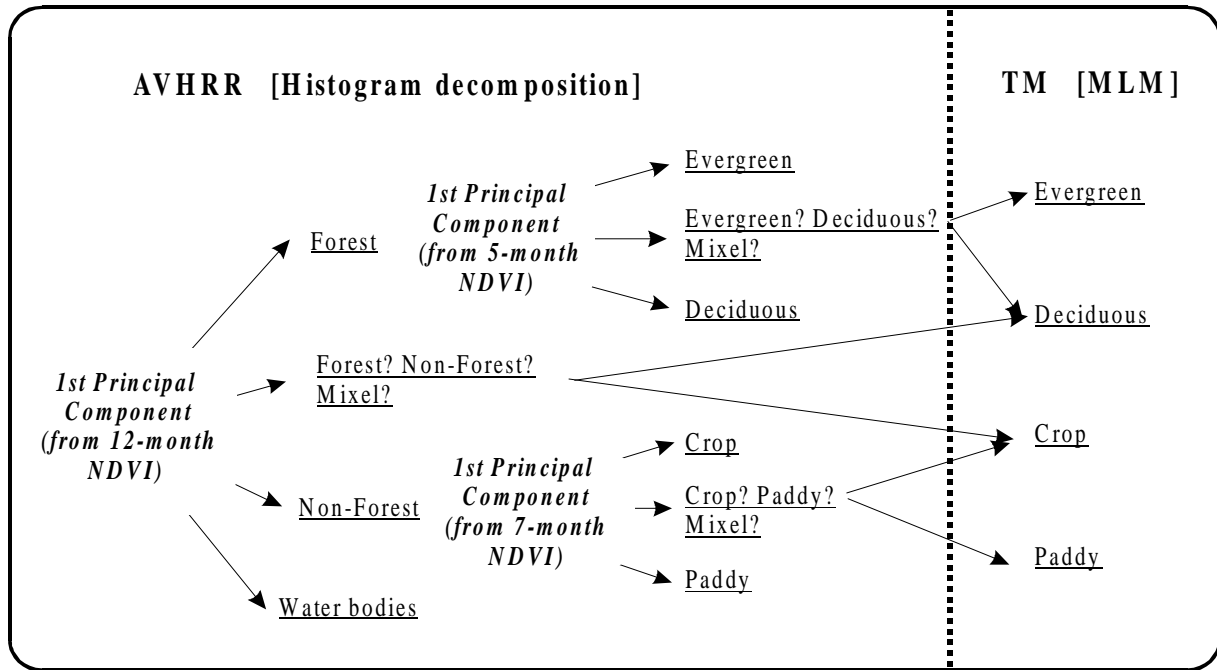
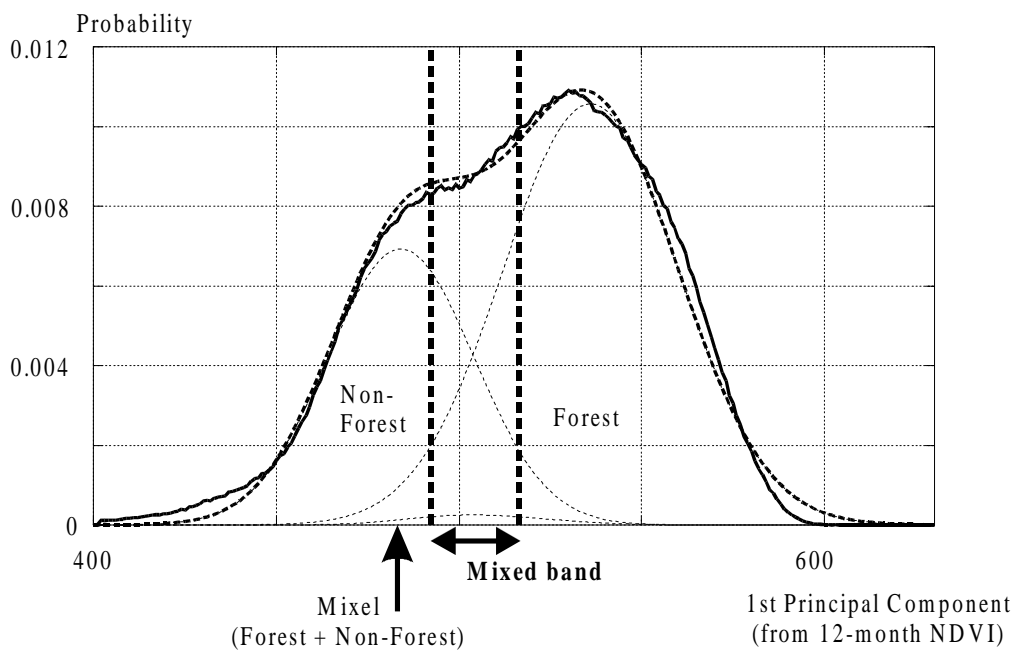Figure 4. Classification flow fusing multi-resolution satellite images

Figure 5. Example of histogram decomposition (Forest vs Non-Forest)

maximum KL entropy. Here, "5-month NDVIs" means NDVIs during dry season, i.e. 92 November, December, 93 January, February, March. Then, in the third stage to classify Crop from Paddy, the data set, the first principal component from 7-month NDVIs gave the maximum KL entropy. Here, "7-month NDVIs" means NDVIs during dry season, i.e. 92 April, May, November, December, 93 January, February, March. In the step 1) described in section 3, land cover class candidates were temporarily determined as four, and the classification feasibility was also examined by KL entropy. In the third stage of the decision tree, the maximum KL entropy was 5.92 given the first principal component from 7-month NDVIs. Authors judged the KL entropy was big enough to be used for classification, then four class candidates were adopted as four classes for classification.

In the step 1) to 3), classification was executed by use of low-spatial resolution satellite images. Finally, in the step 4), pixels not classified because of low reliability were classified through MLM by use of high-spatial resolution satellite images, TM images. The used TM images were two scenes; Cambodia (Path 127, Row 051:1996. Jan. 18 observed) and Thailand (Path 128, Row 049:1995. Dec. 5 observed).

Figure 4 shows histogram decomposition result in the first stage of the decision tree. The horizontal axis means the first principal component from 12-month NDVIs, and the vertical axis means probability. In this research, a condition for pure feature space was defined as more than 80% of occupancy by one class. If the occupancy was below 80%, the pixels belonging to the feature space were classified by use of a Landsat TM image.

Classified results were validated by comparison of an image visually interpreted by experts. Traditional MLM was used as a comparative method. It dealt with 13 dimensional data, 12-month monthly composite AVHRR NDVIs and TM NDVI, and classified based on the likelihood calculated from the 13 dimensional data. Validation results of two scenes are shown in Table 1. Figure 5 shows visually interpreted image, the image classified by authors proposed method, and the image classified by traditional MLM.

			Land covers in visually interpreted images					
			Evergreen	Deciduous	Crop	Paddy	Sum	Ratio
R E S U L T	Evergreen	Proposed	3.26%	0.661%	0.344%	6.90×10^{-6}%	4.28%	76.4%
		MLM	4.03%	4.07%	1.71%	2.64×10^{-4}%	9.84%	41.0%
	Deciduous	Proposed	1.61%	12.3%	4.19%	0.969%	19.0%	64.4%
		MLM	0.750%	10.6%	5.79%	2.05%	19.2%	55.1%
	Crop	Proposed	3.42×10^{-4}%	2.65%	12.8%	15.1%	30.6%	41.8%
		MLM	1.18×10^{-4}%	0.437%	11.2%	14.7%	28.2%	39.7%
	Paddy	Proposed	2.02×10^{-6}%	0.406%	6.00%	39.7%	46.1%	86.1%
		MLM	4.60×10^{-6}%	0.286%	5.02%	37.5%	42.3%	87.6%
	Sum	Proposed	4.90%	16.0%	23.3%	55.8%	100.0%	
		MLM	4.79%	17.2%	23.7%	54.2%	100.0%	
	Ratio	Proposed	66.5%	76.7%	54.8%	71.1%		68.0%
		MLM	84.1%	61.2%	47.2%	69.1%		63.3%

Table 1. Validation result

5 DISCUSSION

Histogram decomposition method used for parameter estimation is based on just 2-class mixture model. The method can be extended into multi-class mixture model. But obtaining stable parameters requires multi-dimensional data. If the mixture class number is much larger than the dimension of input data, the estimated parameters will be unstable because there'll be lots of solutions which satisfy conditions. In this research, 12-month monthly composite AVHRR NDVI images were used. If channel 1 and 2 data are used, we can obtain 24 dimensional data. But, those data are affected by cloud noise, and it was found that NDVI is more appropriate for classification than those data because NDVI is less affected by cloud noise. Moreover, as authors mentioned above, the temporal NDVI data have high correlation. In fact, the first principal component proportion calculated from 12-month monthly composite AVHRR NDVIs of Indo-china peninsula was more than 98%. This means that we cannot expect to deal with temporal 12-month data as the same dimensional data substantially, and that just 1 dimensional data is effective for classification. And dimensional compression like PCA is effective in the view of noise removal and it gives less dimensional data which are easy to be dealt.

The situation is even if there are multi-dimensional data, we can substantially use 1 dimensional data for land cover classification. On the other hand, examining training data set revealed that feature distributions of land covers are not clearly separated from other distributions. If the area under such a situation is classified into several land cover classes simultaneously, like traditional MLM, classification accuracy will be low. In order to overcome those problems, authors adopted the principle of decision tree. It is based on the classifier to discriminate with priority the class which is the easiest to be classified. Even if there are multi-classes and there is just 1 dimensional data for classification, and if the class which is the easiest to be classified is classified with priority, it is possible to classify with high accuracy and with consideration of mixel existence. But, such a method has limitations that mixel mixture can be considered up to two classes. In fact, there are lots of mixels mixed of several land covers on the AVHRR data level. In this research, it was found that classification considering multi-class mixels requires substantially multi-dimensional data.

Visually interpreted image

Classified result by proposed method
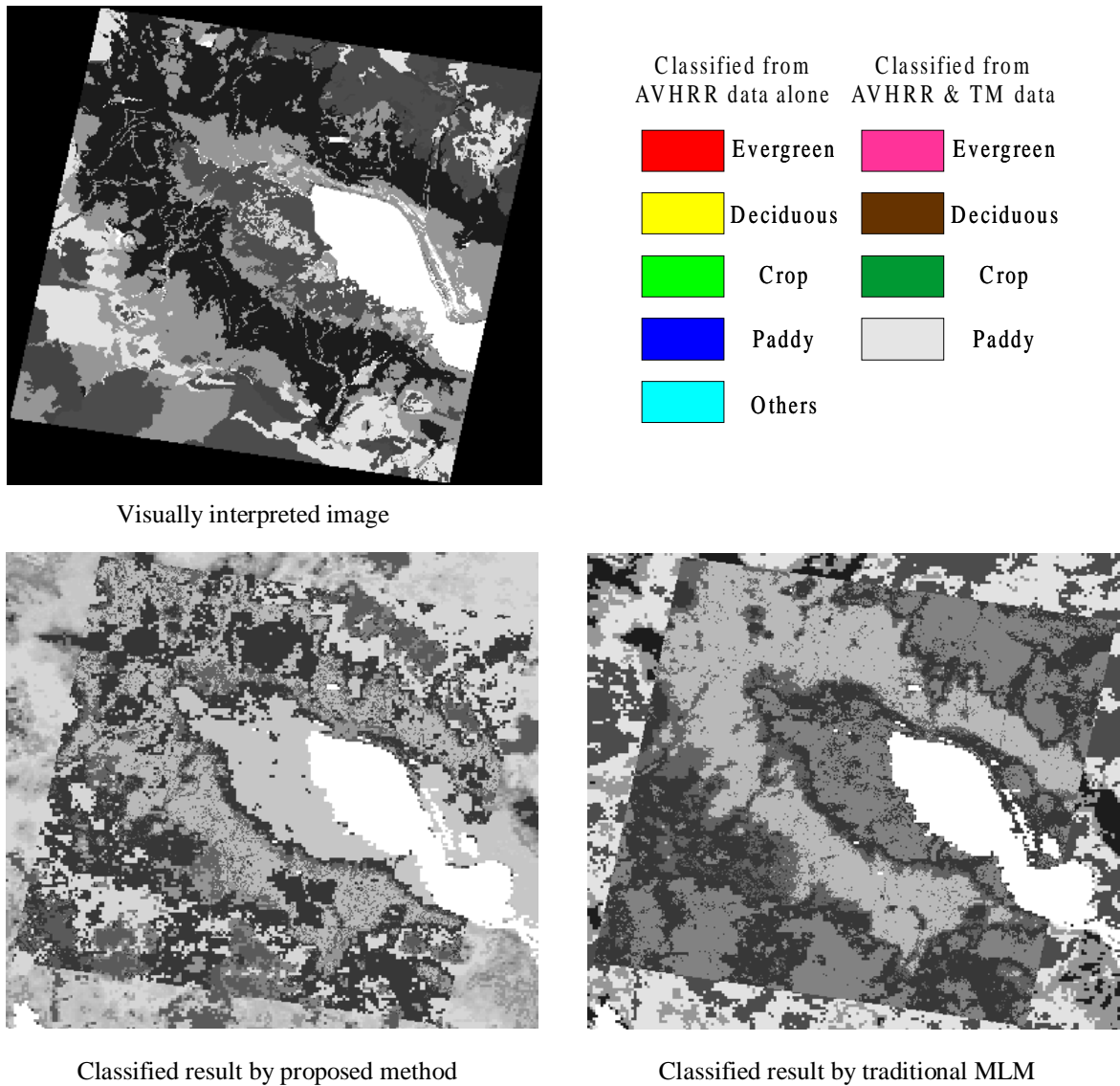

Classified result by traditional MLM

Figure 6. Classified results fusing AVHRR and TM data (Cambodia)

6 CONCLUSION

Authors pointed out limitations of classification by MLM in the first part of this paper, then proposed a classification method fusing multi-resolution satellite images. The method is mixed of decision tree and MLM. And it considers mixels especially existing in low-resolution satellite images and give a severe effect to classification accuracy. The experiment used Landsat TM and NOAA AVHRR images as high- and low-resolution satellite images, and the validation showed that the proposed method gave higher classification accuracy than traditional MLM

REFERRENCES

Redner R. A., Walker H. F., 1984. Mixture densities, maximum likelihood and the EM algorithm, SIAM Review, Vol.26, No.2, pp.195-239

Susaki J., Shibasaki R., 1999. Estimation method of spectral characteristic and area proportion of land cover based on probabilistic mixel model , Journal of Japan Society of Photogrammetry and Remote Sensing, Vol.38, No.3, pp.44-59