High Accuracy Classification Methods for Multi-Temporal Images

Sun Pyo Hong*, Dong Keun Jeon*

Abstract

Three new classification methods for multi temporal images are proposed. They are named as a likelihood addition method, a likelihood majority method and a Dempster-Shafer's rule method. Basic strategies using these methods are to calculate likelihoods for each temporal data and to combine obtained likelihoods for final classification. These three methods use different combining algorithms.

From classification experiments, following results were obtained. The method based on Dempster-Shafer's rule of combination showed about 12% improvement of classification accuracies compared to a conventional method. This method needed about 16% more processing times than that of a conventional method. The other two proposed method showed 1% to 5% increase of classification accuracies. However processing times of these two methods are almost the same with that of a conventional method. Among the newly proposed three methods, the Dempster-Shafer's rule method showed the highest classification accuracies with more processing time than those of other methods.

Ⅰ. Introduction

With the launch of second generation high resolution sensors like Landsat TM(Thematic Mapper) and SPOT HRV, many kinds of researches have been done to certificate the capability of these sensors for landcover classification[1-5]. Most of the results of these studies have shown that classification accuracies using these sensors are not so high as expected when applying conventional supervised maximum likelihood classifier using only spectral information.

One of promising methods, which can be through to increase classification accuracies, is to utilize multi-temporal data. And automatic method is desirable as it can be.

Using multi-temporal data, there are several studies on sophisticated classifier like expert systems or fuzzy classifiers[6, 7]. But these methods use some knowledge on the object area, so it is difficult to use practically.

The most popular method of combining multi-temporal data is to just increase the dimensions of classification feature space[8, 9]. In other words, multi temporal data are considered to be a set of multi channel data. This conventional method is called as a simple combination(SC) method in this paper. However, it has been known that this method does not necessarily show improvements on classification accuracies, because landcover on the object area is changing by time, and it usually increase variances of each training data. Variances of each training data will be increased as it use more multi-temporal data in order to obtain high accuracies. This phenomena make it difficult to use multi-temporal data. And the processing time should be increased exponentially as the number of channel is increasing.

The objective of this research is to try several new methods of utilizing multi-temporal data and find out the most useful method. Basic idea is as follow. That is, newly proposed methods do not process entire multi-temporal data entirely as SC method, but process each temporal data independently and combine the result of each temporal data. Basic strategies of these methods are to calculate likelihoods for each temporal data and to combine obtained likelihoods for final classification.

In this paper, three new methods using different combining algorithms are proposed and performance of four methods including the SC method have been examined.

Ⅱ. Proposed Methods

A pixel-wise maximum likelihood classifier based on spectral features is used as a basic classifier. Let LC be the likelihood of class-c derived from multi-temporal data set. In conventional simple combination method, LC is calculated as

* Department of Control & Instrumentation Engineering, Junior College of Inchon
Manuscript Received: September 16, 1996.

$$L_c = \frac{1}{(2\pi)^{nm/2}|S_c|^{1/2}} \exp\left\{-\frac{1}{2}(x-M_c)^t S_c^{-1}(x-M_c)\right\} \quad (1)$$

where

c : class,

n : number of spectral bands,

m : number of temporal data,

^{t} : transposed matrix,

| | : determinant,

$^{-1}$: inverse matrix,

S_c : variance-covariance matrix of class-c,

M_c : mean vector of class-c,

X : pixel vector shown as

$X = x_1(t_1), x_2(t_1), \ldots, x_n(t_1), x_1(t_2), x_2(t_2), \ldots, x_n(t_2), \ldots, x_1(t_j), x_2(t_j), \ldots, x_n(t_j), \ldots, x_1(t_m), x_2(t_m), \ldots, x_n(t_m)\}$

t : ID of observation date,

x_i : pixel value of spectral band-i

Then, a decision class(DC) is determined to the class showing the maximum likelihood as follows;

$$DC = c - \max, \text{ if } L_{c-\max} = \max_c [L_c] \quad (2)$$

In this method, temporal features are treated as the same feature with spectral features. That is, the dimension of feature space is equal to the product of the number of spectral bands(n) and the number of spectral bands (n) and the number of multi-temporal images(m). Usually separability of classes decreases, because the variance of each class usually increases compared to that of single temporal case. Consequently, the SC method does not always show improvement of classification accuracies.

Three new methods of utilizing multi-temporal data have been tested in this research. The first and second one are named likelihood addition(LA) method and likelihood majority(LM) method, respectively. The last one is based on Dempster-Shafer's rule of combination, and is named as a DR method. In these proposed methods, the likelihood of each class is calculated from each temporal image. That is, the likelihood of class-c obtained from temporal image-t is calculated for a pixel vector $X = \{x_1(t), x_2(t), \ldots, x_n(t)\}$. Let the likelihood calculated from temporal data-t be $L_c(t)$.

(1) likelihood adding(LA) method

A score of class-c, S(c), is calculated in the LA method by the following equation;

$$S(c) = \sum_t L_c(t), \quad c = 1, 2, \ldots, k \quad (3)$$

where k is the number of classes.

A decision class(DC) is determined to the class "c-max" if S(c-max) shows the maximum score, which can be written as follows;

$$DC = c - \max, \text{ if } S(c-\max) = \max_c [S(c)] \quad (4)$$

(2) likelihood majority(LM) method

In the LM method, scoring of likelihoods and decision of class are calculated using following eqns. (5) and (6), respectively.

$$S(c) = \sum_t f\{L_c(t)\}, \quad c = 1, 2, \ldots, k \quad (5)$$

where

$$f\{L_i(t)\} = \begin{cases} 1, & \text{if } L_i(t) = \max_c [L_c(t)] \\ 0, & \text{others} \end{cases} \quad (6)$$

Note that the function f converts the value of $L_c(t)$ to binary data.

(3) Dempster's rule(DR) method

Dempster's rule of combination is expressed by the following formula[10];

$$m_A(t_1, t_2) = \frac{\sum_{B \cap C = A} m_B(t_1) m_C(t_2)}{1 - k} \quad (7)$$

for $A \neq \phi$, where k is the normalization coefficient and are expressed as

$$k = \sum_{B \cap C = \phi} m_B(t_1) m_C(t_2) \quad (8)$$

Eqns. (7) and (8) show that the degree of evidence $m_B(t_1)$ from the first source which focuses on set B and the degree of evidence $m_C(t_2)$ from the second source which focuses on set C are combined by B and C. This is exactly the same way in which the joint probability distribution is calculated from two independent marginal distributions.

In this research $m_A(t_1, t_2)$ is treated as a score of each class. B and C are defined as a subset of 1st, 2nd and 3rd candidates of computational explosion. 1st, 2nd and 3rd candidates of decision classes correspond to classes having the largest, 2nd largest and 3rd largest likelihood. When it is assumed that C_1, C_2 and C_3 are the 1st, 2nd and 3rd candidate class, respectively, the subset B and C are expressed by

$$B=\{B_1, B_2, B_3, B_1\cup B_2, B_1\cup B_3, B_2\cup B_3, B_1\cup B_2\cup B_3\} \quad (9)$$

$$C=\{C_1, C_2, C_3, C_1\cup C_2, C_1\cup C_3, C_2\cup C_3, C_1\cup C_2\cup C_3\} \quad (10)$$

Then, $m_B(t_i)$ and $m_C(t_i)$ are calculated by the formula

$$m_B(t_i)\equiv L_B(t_i)=\sum_{k\in B} L_k(t_i) \quad (11)$$

$$m_C(t_i)\equiv L_C(t_i)=\sum_{k\in C} L_k(t_i) \quad (12)$$

Thus eq. (7) can be revised as follows;

$$S(c)=\frac{\sum_{B\cap C=c} L_B(t_1)L_C(t_2)}{1-k} \quad (13)$$

And a decision class is determined as follows:

$$DC=c-\max, \text{ if } S(c-\max)=\max_{c}[S(c)] \quad (14)$$

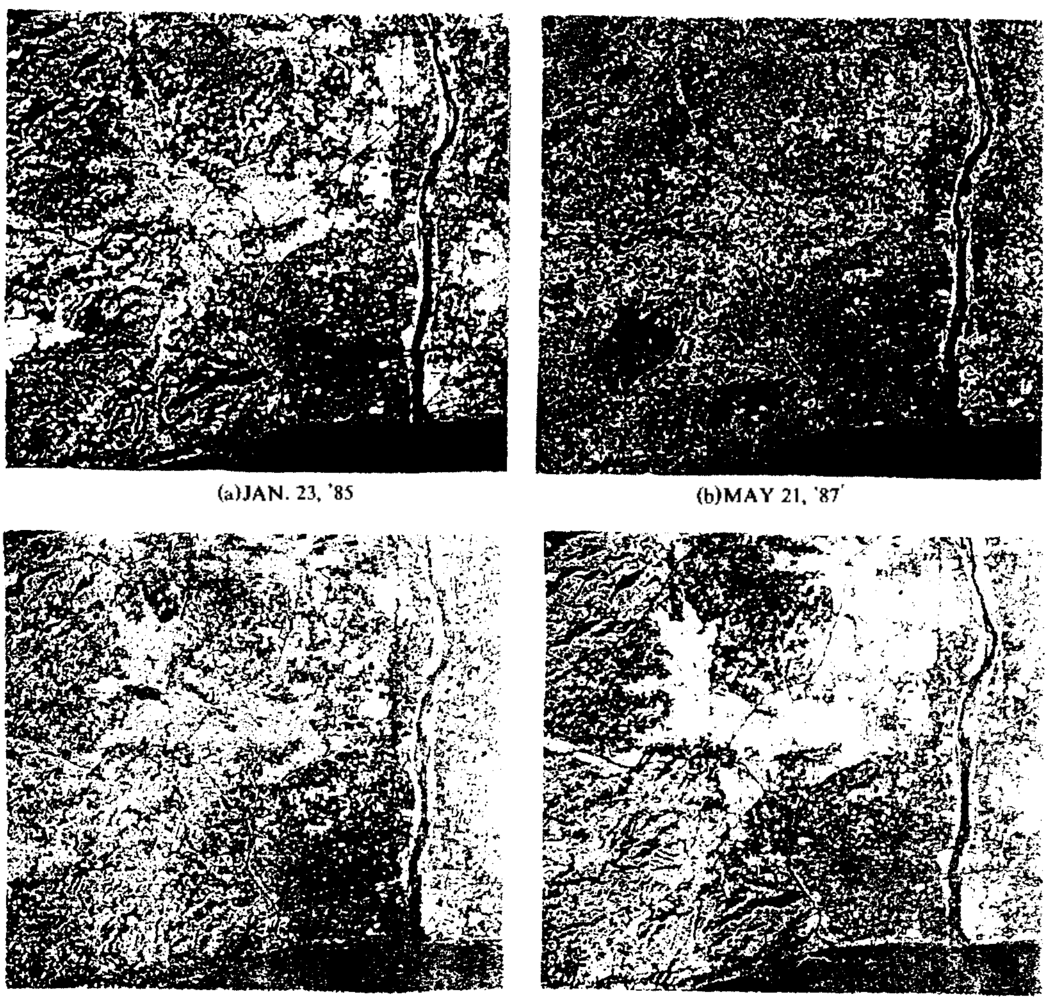
If i of "t_i" is greater than 2 such as $t=\{t_1, t_2, t_3, t_4, \ldots\}$, S(c) is calculated for all ti by applying eq. (13) iteratively.

Ⅲ. Experiments

3.1 Test Images and Test Site Data

In order to evaluate proposed methods described in Chapter 2, following four seasonal(spring, summer, fall, winter) Landsat TM data were classified by using three new methods and a conventional SC method. Sagami River basin was selected for object area to perform the quantitative evaluation. This area includes the test site area(width 2km, length 10km) which landcover is already investigated and categorized to test site data[11].

Figure 1 shows test images used in these experiments. These images were registered in the identical UTM(Universal Transverse Mercator) coordinate system.



(a)JAN. 23, '85 (b)MAY 21, '87 (c)AUG. 6, '86 (d)NOV. 4, '84

Figure 1. Test images

〔object area〕

Sagami River basin(in Japan)which has area of 12.8km × 12.0km.

〔observation date〕

Nov.4(1984), Jan 23(1985), Aug.6(1986) and May 21(1987)

〔image size〕

512 × 480 pixels, pixel size = 25m × 25m

〔used channels〕

TM Ch. 1, 2, 3, 4, 5 and 7 (Ch 6 was not used)

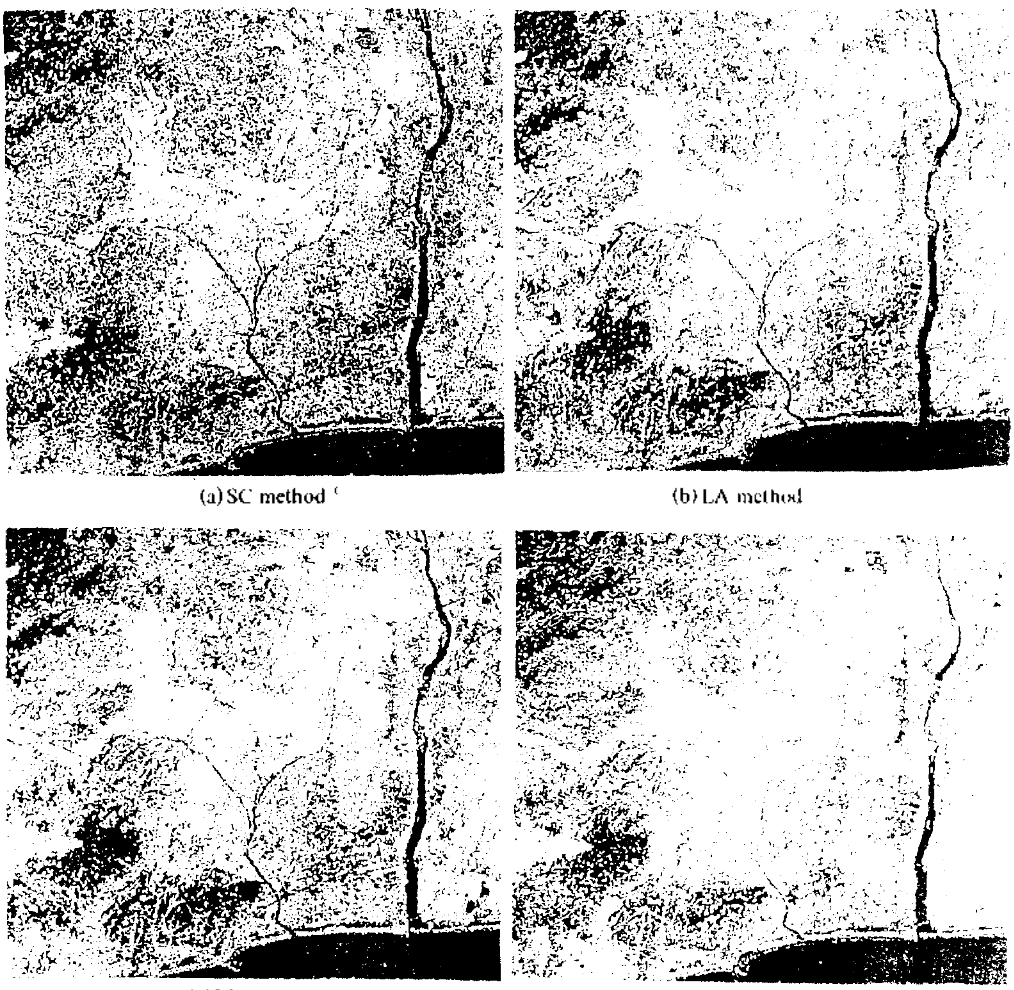
3.2 Classification and Results

Items in left hand side in Table 1 shows 15 classification categories used in the experiments. However the total number of classification classes were fifty-nine since each category has several sub-classes. Training data were obtained from the same training area for each seasonal image. Thus, training data set consists of four training data corresponding to four seasonal images.

Table 1. Classification categories

classification categories	number of classes	major categories
1. coniferous forest	3	vegetation
2. broad leaved forest	5	
3. mixed forest	3	
4. shadow of mountain	2	
5. paddy	9	paddy
6. high density urban	4	urban
7. low density urban	3	
8. housing area	3	
9. factories	5	
10. sea	2	water
11. river	2	
12. farm	6	other
13. grassland	5	
14. waste land	6	
15. sands	1	

At the first stage, likelihoods of each class, i.e. $L_C(t)$s, were calculated in all test images by using each training data. This calculation was done independently for each



(a) SC method (b) LA method (c) LM method (d) DR method

Figure 2. Classified results

test image. That is, $L_C(t_1)$, $L_C(t_2)$, $L_C(t_3)$ and $L_C(t_4)$ is calculated by using training data corresponding to the test image of t_1, t_2, t_3 and t_4, respectively.

At the second stage, three proposed methods(the LA, LM and DR methods) were applied to four seasonal likelihood data ($L_C(t)$s ($t = t_1$, t_2, t_3 and t_4)) obtained in the first stage. On the other hand, landcover classification using the conventional SC method was performed according to eqns. (1) and (2). In order to compare with a case of single temporal classification, conventional maximum likelihood classification(MLC) were conducted for each test images using the same training data set. Figure 2 shows classification results.

Since processing time depends on complexity of the algorithm, the processing time of DR method is expected longer than that of other methods. Processing times of MLC for a single image was about 15 minutes. Those of the SC, LA and LM methods are about 60 minutes, because of processing for four seasonal images. The DR methods needed processing times of about 70 minutes. Experiments were done by using HP9000/835 mini-computer system.

Finally, classification accuracies were estimated quantitatively by using digital test site data as shown in Figure 3. Test site data contain about 50 land-cover/use categories. As the categories used in test site data differ from those in landcover classification conducted in these experiments, fifteen classification categories were merged to five major categories as shown in Table 1. Accuracy evaluations were performed based on these five major categories.

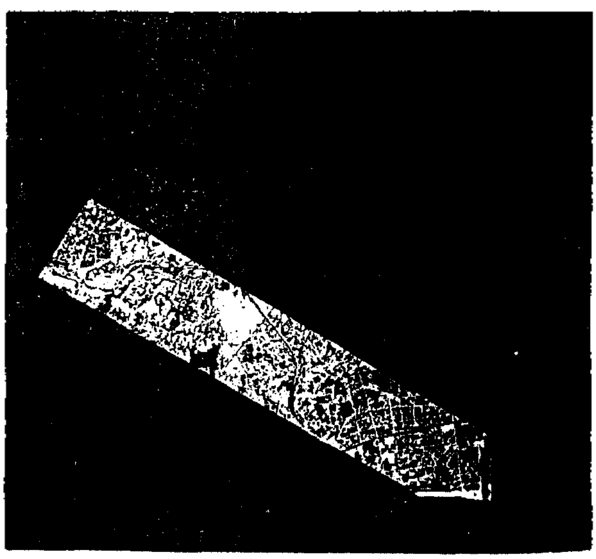


Figure 3. Digital test site data

Table 2. Classification accuracies

	forest	paddy	urban	water	other	mean
MLC JAN.	43.1	53.8	83.1	61.7	30.9	61.5
MAY.	42.7	38.2	82.9	61.6	39.5	62.3
AUG.	44.8	61.0	80.9	63.0	42.6	64.7
NOV.	43.8	63.5	81.8	64.2	39.0	64.3
SC method	49.8	50.7	88.5	62.0	33.9	65.1
LA method	45.6	69.3	84.2	63.9	38.1	65.9
LM method	47.3	72.3	88.2	66.8	43.6	70.1
DR method	50.0	77.5	93.7	75.3	44.8	76.9

Accuracy evaluations were performed by eq. (15) based on these five major categories.

$$mean = \frac{1}{5} \sum_{k=1}^{5} C_k \cdot A_k \qquad (15)$$

where C_k: accuracy of class k

A_k: occupation rate of class k

Table 2 shows estimated accuracies of classification results of the MLC applied to each test image, and the SC, LA, LM and DR method applied to the test image set. Average accuracies in right hand side of Table 2 were area weighted mean values for each category.

The MLC for each test image shows average accuracies from 62% to 65%. The SC method indicates an average accuracy of 65%, which is almost the same value to the largest one of the MLC result. LA, LM and DR method showed about 1%, 5% and 12% larger average accuracies, respectively compared to the results of the MLC and the SC method.

Ⅳ. Conclusions

Four classification methods for multi-temporal data were evaluated by experiments using four seasonal Landsat TM data. The first method is the conventional simple combination(SC) method, which combines spectral features and temporal features in the same feature space and performs a classification with the same manner as conventional maximum likelihood classification. Other three methods are newly proposed in this research.

The first proposed method is named as a likelihood addition(LA) method, in which method scores of each class are calculated by adding likelihoods obtained from each temporal data. The second method proposed is named as a likelihood majority(LM) method, in which method decision class is determined to the majority class in decision candidate class is determined to the majority class in decision candidate classes derived from each temporal data.

The last one is named as a Dempster-Shafer's rule(DR) method, which based on Dempster's rule of combination.

From results of classification experiments, following conclusions were obtained.

(1)The SC method and the LA method did not show large improvements of classification accuracies.

(2)The LM method can be conducted by relatively simple algorithm. However, it showed 5% higher accuracies compared to the SC method.

(3)The method which showed the highest classification accuracies is the DR method. Accuracies were improved about 12% compared to that of the SC method.

(4)As a conclusion, the DR method is recommended from the view point of classification accuracies. However the DR method needs about 16% more processing times than cases of the other methods.

References

1. S. Gorden, 1980 : Utilizing Landsat imagery to monitor land use change-A case study of Ohio, Remote Sensing of Environment, Vol. 9, 189-196.
2. J. P. Howarth, E. Boasson, 1983 : Landsat digital enhancement for change detection in urban environment, Remote Sensing of Environment, Vol. 13, 149-160.
3. A. Singh, 1989 : Digital change detection techniques using remotely sensed data, Inter. J. Remote Sensing, Vol. 10, No. 6, 989-1003.
4. J. R. Jenson, D. L. Toll, 1982 : Detecting residential land use development at the urban fringe, Photogrammetric Engineering and Remote Sensing, Vol. 48, 629-643.
5. R. F. Nelson, 1983 : Detecting forest canopy change due to insect activity using Landsat MSS, Photogrammetric Engineering and Remote Sensing, Vol. 49, 1303-1314.
6. P. N. Blonda, et al., 1991 : An experiment for the interpretation of multi-temporal remotely sensed images based on a fuzzy logic approach, Int. J. Remote sensing, Vol. 12, No. 3, 463-476.
7. D. G. Corr, et al., 1989 : Progress in automatic analysis of multi temporal remotely sensed data, Int. J. Remote sensing, Vol. 10, No. 7, 1175-1195.
8. N. Jewell, 1989 : An evaluation of multi-date SPOT data for agriculture and land use mapping in the United Kingdom, Int. J. Remote sensing, Vol. 10, No. 6, 939-951.
9. H. Thomas, et al., 1986 : Use of multi-temporal Spectral Profiles in Agricultural Land-cover Classification, Photogrammetric Engineering and Remote Sensing, Vol. 52, No. 4, 535-544.
10. George J. Klir and Tina A. Folger, 1988 : Fuzzy Sets, Uncertainty, And Information, Prentice Hall, 107-137.
11. 昭和60-62年度科學研究費補助金特定研究(1)研究成果報告書, 1988 : 陸地における衛星データの利用技術に關する研究.

▲Sun Pyo Hong

Sun Pyo Hong he received B.S. (1982) degrees in Electronics Engineering from A-jou University, M.S. (1986) degrees in Electronics Engineering from KAIST and Ph.D.(1993) degrees in Opto-electronics Engineering from Tokai University in Japan. He is an assistant professor at Junior College of Inchon, Department of control & Instrumentation.

▲Dong Keun Jeon

Dong Keun Jeon he received B.S. (1986), M.S.(1988), Ph.D.(1992) degrees in Electronics Engineering from Korea University. He is an assistant professor at Junior College of Inchon, Department of control & Instrumentation.