

# 원격탐사 영상의 감독분류를 위한 개선된 하이브리드 c-Means 군집화 알고리즘

## Improved Algorithm of Hybrid c-Means Clustering for Supervised Classification of Remote Sensing Images

전영준\*, 김진일\*\*

Young-Joon Jeon\*, Jin-Il Kim\*\*

### 요약

원격탐사 영상은 파장대에 따라 나누어진 여러 개의 밴드로부터 수집된 다중분광 이미지 데이터이다. 위성영상 분류는 원격탐사 처리 과정에 있어서 가장 중요한 분석 기법으로써 영상을 구성하는 각각의 화소들 중 비슷한 분광 특성을 갖는 것끼리 집단화시켜 주는 방법이다. 본 논문에서는 PFCM 알고리즘을 응용한 원격탐사 영상의 패턴분류 방법에 관하여 연구하였다. PFCM 알고리즘은 각 데이터와 특정 클러스터 중심과의 거리에 대한 소속정도를 고려한 FCM 클러스터링 알고리즘과 데이터와 해당 클러스터 중심 간의 거리에 의존하여 패턴의 전형성(typicality)을 고려한 PCM 클러스터링 알고리즘을 결합한 방법이다. 본 연구에서는 분류 항목별 학습데이터를 선정한 후 이를 PFCM 알고리즘에 적용하여 감독분류를 수행하였다. Landsat TM과 IKONOS 원격탐사 위성영상을 이용하여 PFCM 알고리즘의 적용성을 검증하였다. PFCM 알고리즘을 이용한 감독분류는 PCM, FCM 분류방법보다 좋은 결과를 보여주었으며, 또한 전통적인 분류방법인 최대우도분류보다도 정확도가 더 높은 결과를 보여주었다.

### Abstract

Remote sensing images are multispectral image data collected from several band divided by wavelength ranges. The classification of remote sensing images is the method of classifying what has similar spectral characteristics together among each pixel composing an image as the important algorithm in this field. This paper presents a pattern classification method of remote sensing images by applying a possibilistic fuzzy c-means (PFCM) algorithm. The PFCM algorithm is a hybridization of a FCM algorithm, which adopts membership degree depending on the distance between data and the center of a certain cluster, combined with a PCM algorithm, which considers class typicality of the pattern sets. In this proposed method, we select the training data for each class and perform supervised classification using the PFCM algorithm with spectral signatures of the training data. The application of the PFCM algorithm is tested and verified by using Landsat TM and IKONOS remote sensing satellite images. As a result, the overall accuracy showed a better results than the FCM, PCM algorithm or conventional maximum likelihood classification(MLC) algorithm.

**Keyword** : possibilistic fuzzy c-means clustering, pattern classification, remote sensed images

## I. 서론

최근 1m급 해상도의 팬크로매틱 영상과 4m급 해상도의 다색 대역 영상을 촬영할 수 있는 고해상도 카메라(Multi-Spectral Camera, MSC)를 탑재한 지구관측용 다목적실용위성 2호(The Second Korea Multi-Purpose Satellite, KOMPSAT-2)가 발사에 성공한 이후 원격탐사 영상처리에 관련된 연구에 대한 관심이 더욱 높아지고 있다[1].

원격탐사 위성영상 처리에 대한 연구는 다양한 형태로 이루어지고 있으며, 특히 위성영상 데이터의 분류에 대한 연구는 다양하게 활용되고 있다. 원격탐사 위성영상의 분류는 영상을 구성하는 각각의 화소들 중 비슷한 분광 특성을 갖는 것끼리 집단화시켜 주는 방법이다. 위성영상의 분류는 지표면에 대한 정보 수집이나 환경변동 감시 등을 위한 다양한 자료로 활용할 수 있다[2].

본 논문에서는 기본적으로 퍼지 이론을 적용한 클러스터링 알고리즘에 대하여 연구하고 이를 위성영상의 분류에 활용하였다. 기존의 클러스터링 방법은 단순 클러스터링 방법(Hard clustering), FCM(Fuzzy c-Means) 방법 그리고 PCM(Possibilistic c-Means) 방법 등이 있다. 하드 클러스터링 기법은 주어진 데이터 상호간의 관계가 명확하다는 가정 하에 각 패턴을 분할하는 방식이다. 따라서 이 방법은 다루고자 하는 데이

\*대구대학교 컴퓨터-IT공학부 \*\*동의대학교 컴퓨터공학과

논문 번호 : 2007-3-3 접수 일자 : 2007. 5. 31

심사 완료 : 2007. 7. 23

\*본 논문은 2005학년도 동의대학교 교내연구비에 의해 연구되었음.(2005AA164)

터의 경계가 명확하지 않을 경우 실제 데이터 상호간의 군집성을 묘사하기에 부적절할 뿐만 아니라, 주어진 데이터 분포의 성질을 손실하는 결과를 초래 할 수도 있다. 이를 개선하기 위해 Bezdek 이 제안한 FCM 알고리즘은 각 데이터와 특정 클러스터 중심과의 거리에 소속 정도를 부여하는 알고리즘으로서 매 순간의 결정을 바로 이전에 결정된 결과에 근거하여 조정해 감으로서 영상의 분류에 좋은 결과를 주고 있다[3][4]. 그러나 FCM 알고리즘은 한 패턴 데이터의 각 클러스터에 관한 소속정도의 합이 1이 되는 확률적 제약 조건을 이용하므로 소속 함수의 값이 소속성 (belonging)이나 적합성(compatibility) 정도의 직관적인 개념과 항상 일치하지 않으며 클러스터 개수를 사전에 정해주어야 하는 문제점이 있다. 그리고 FCM의 경우에는 노이즈 패턴에 대해서도 다른 패턴과 같이 퍼지 소속정도(membership degree)를 할당하기 때문에 경계면의 위치가 영향을 받게 된다. 이는 FCM의 경우 소속정도가 다른 클러스터(클래스)와 관계가 있는 상대적인 값이기 때문이다. 이러한 문제점을 개선하기 위해 Krishnapuram은 PCM 클러스터링 알고리즘을 제안하였다[5][6].

PCM의 경우는 소속정도가 다른 클러스터와 관계가 없고, 패턴 데이터와 해당 클러스터 중심 간의 거리에만 의존하기 때문에 이러한 문제점을 개선할 수 있다. 즉 전체 패턴 데이터에 대해 해당 클래스의 전형성(typicality)을 고려한 소속정도를 할당하게 함으로써 노이즈 패턴 데이터에 대해서는 다른 패턴과 비교하여 상대적으로 작은 퍼지 소속정도를 할당하기 때문에 노이즈가 포함된 데이터에 대해서 더 뛰어난 클러스터링 수행이 가능하다. 그러나 PCM 방법을 적용하기 위해서는 각 영역에 해당하는 클러스터들의 크기, 평균내부거리, 초기 중심 값을 사전에 알아야 하므로 스케일 공간 필터링(Scale Space Filtering)과 같은 전처리 과정이 필요하다. 이에 Pal 등의 연구에서는 FCM과 PCM 클러스터링 알고리즘의 단점을 고려하여 두 알고리즘을 결합한 PFCM (Possibilistic Fuzzy c-Means) 클러스터링 알고리즘이 제안되었다[7][8]. 이 알고리즘은 FCM 클러스터링 알고리즘의 소속도와 전형성을 고려한 PCM 클러스터링 알고리즘의 소속정도를 결합한 알고리즘이다. 노이즈가 있는 상황에서 민감한 FCM 클러스터링 알고리즘과 클러스터에 부합하는 사전정보를 알아야 하는 PCM 클러스터링 알고리즘의 단점을 개선하였다.

본 논문에서는 다차원 다중분광 원격탐사 위성영상 데이터의 분류를 위하여 PFCM 알고리즘을 응용한 감독분류를 제안하였다. 제안된 분류방법은 분류 항목을 먼저 설정하고 영상에서 분석자가 학습데이터 영역을 선정하여 초기 중심 값을 구하고 이를 PFCM 분류 알고리즘에 적용하여 반복 또는 비반복 수행을 통하여 분류를 하였다.

## II. 퍼지 클러스터링 알고리즘

### 2.1 FCM 클러스터링 알고리즘

Bezdek에 의해 제안된 퍼지 c-Means 클러스터링 알고리즘은 퍼지 클러스터링 알고리즘 중에서 가장 폭넓게 사용되는 알고리즘이다[3][4]. 하드 클러스터링 알고리즘에서 경계가 불확실한 경우에 발생할 수 있는 데이터 분류 문제를 해결하기 위해 많이 사

용되고 있다. 각 데이터와 특정 클러스터 중심과의 거리에 소속정도를 부여하고 이 소속정도에 따라 패턴 데이터를 분류하는 알고리즘이다. 퍼지 c-Means 알고리즘은 각 데이터와 각 클러스터 중심과의 거리를 고려한 유사도 측정에 기초한 목적함수  $J_m(U, V; X)$ 을 최소화할 수 있도록 패턴 데이터 집합을 분류하는 알고리즘이다.

$$J_m(U, V; X) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m \|x_k - v_i\|_A^2 \quad (1)$$

여기서,  $c(1 \leq i \leq c)$ 는 클러스터의 개수,  $n$ 은 데이터의 개수,  $m$ 은 퍼지 정도를 나타내는 가중치이고  $1 < m < \infty$ 의 값을 사용하며 일반적으로  $m=2$ 값을 사용한다. 모든 데이터  $k$ 에 대하여  $\sum_{i=1}^c \mu_{ik} = 1$ 이다.  $X = \{x_1, x_2, \dots, x_n\}$ 인 데이터 벡터 집합과  $V = (v_1, v_2, \dots, v_c)$ 인 클러스터 중심들 사이의 소속정도를  $c \times n$ 인 행렬  $U = (\mu_{ik})$ 로 나타내면 식 (2)를 이용하여 계산하고, 이때의 클러스터 중심을  $V = (v_i)$ 로 나타내면 식 (3)과 같이 계산한다.

$$\mu_{ik} = \left( \sum_{j=1}^c \left( \frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad 1 \leq i \leq c; 1 \leq k \leq n; \quad (2)$$

$$v_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m}, \quad 1 \leq i \leq c \quad (3)$$

여기서, 클러스터 중심과 패턴 데이터와의 거리는 모든  $i, k$ 에 대해서  $D_{ik} = \|x_k - v_i\|_A > 0$ 으로 나타내며, 유클리디언 거리나 마하라노비스 거리 등을 이용한다.  $\mu_{ik}$ 는  $k$ 번째 데이터가  $i$ 번째 클러스터에 속하는 소속정도를 나타내고,  $v_i$ 는  $i$ 번째 클러스터 중심이다.  $m$ 이 1보다 큰 경우에 모든  $i, k$ 에 대해서  $v_i \neq x_k$ 를 만족한다고 가정하면 위의 식을 만족할 때만  $(U, V)$ 가 목적함수  $J_m(U, V; X)$ 의 최소화를 가능하게 한다. FCM 알고리즘은 식 (2)와 식 (3)의 과정을 반복하므로 목적함수  $J_m(U, V; X)$ 은 어떤 정해진 값에 수렴하게 된다.

### 2.2 PCM 클러스터링 알고리즘

PCM 클러스터링 알고리즘은 Krishnapuram과 Keller에 의해 제안되어졌다[5][6]. 기본적인 과정은 FCM과 유사하지만, 가중치를 보정할 때 사용하게 되는 소속정도를 할당하는데 있어서 패턴의 전형성을 고려한다는 것이 다르다. 이에 노이즈 패턴 데이터에 대해서는 다른 패턴과 비교하여 상대적으로 작은 퍼지 소속정도를 할당하기 때문에 노이즈가 포함된 데이터에 대해서 더 뛰어난 클러스터링 수행이 가능하다. 이러한 요구를 만족하는 PCM의 목적함수  $J_m(T, V; X, \gamma)$ 는 다음과 같다.

$$J_m(T, V; X, \gamma) = \sum_{k=1}^n \sum_{i=1}^c (t_{ik})^m \|x_k - v_i\|_A^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1 - t_{ik})^m \quad (4)$$

여기서 PCM에서는 FCM의 소속정도를 나타내는  $\mu_{ij}$  대신에 전형성을 고려한 소속정도를 나타내는  $t_{ik}$ 로 대체하여 나타낸다.  $\gamma_i$ 는 적절한 양수이다. 첫 번째 항은 패턴 데이터에서 클러스터 중심까지의 거리는 가능한 한 가까울 것을 요구한다. 이에 반해 두 번째 항에서는 거리가 가까울수록 소속정도  $t_{ik}$ 에게 가능한 한 큰 값을 갖도록 강요한다. 이와 같이 하여 패턴의 전형성을 고려할 수 있다. 목적함수  $J_m(T, V; X, \gamma)$ 의 전역 최소치를 만족하는  $T$ 는 식 (4)를  $t_{ik}$ 에 대해 미분하고 미분치를 0으로 놓을 때 다음 식을 얻을 수 있다.

$$t_{ik} = \frac{1}{1 + \left(\frac{D_{ikA}^2}{\gamma_i}\right)^{\frac{1}{m-1}}}, \forall i, k. \quad (5)$$

$$v_i = \frac{\sum_{k=1}^n t_{ik}^m x_k}{\sum_{k=1}^n t_{ik}^m}, \quad \forall i, k. \quad (6)$$

$t_{ik}$  값은  $k$ 번째 데이터가  $i$ 번째 클러스터에 속하는 소속정도를 나타낸다. 클러스터 중심 값( $v_i$ )의 적합성의 관점에서, 클러스터의 한 점의 소속정도는 해당 클러스터의 거리에 의해 결정되어야만 하고 다른 클러스터들과 겹치지 않아야 한다. 식 (5)에서,  $t_{ik}$ 의 값이 [0,1] 사이에 놓여 있다는 것을 알 수 있다.  $\gamma_i$  값은 클러스터에서 속성 데이터의 소속정도가 0.5가 되는 거리를 결정한다. 그러므로 이것은 각 클러스터의 가능성 분포의 적절한 대역폭에 의존하여 선택되어지는 것이 바람직하다.  $\gamma_i$ 는 목적함수에서 두 번째 항이 첫 번째 항과 비교하여 얼마나 중요한가를 나타내는 상대적인 정도를 결정한다. 여기서  $\gamma_i$ 는 클러스터의 전체 모양과 크기에 관계되며 Krishnapuram과 Keller는  $\gamma_i$ 를 식 (7)과 같이 추정하였다.

$$\gamma_i := K \frac{\sum_{k=1}^n \mu_{ik}^m D_{ikA}^2}{\sum_{k=1}^n \mu_{ik}^m}, K > 0 (\text{the most common choice is } K=1) \quad (7)$$

식 (7)에서  $n$ 은 속성 데이터의 개수이고,  $K=1$ 이다. PCM 알고리즘은 식 (5)와 (6)을 반복 수행하여 목적함수의 최소치를 만족하는 가능성 분포를 구할 수가 있다.

### III. PFCM 알고리즘을 이용한 감독분류 시스템

본 논문에서는 PFCM 알고리즘을 이용하여 다중분광 원격탐사 위성영상을 감독 분류하는 방법을 제안하였다. PFCM 알고리즘은 FCM의 소속정도와 PCM의 전형성 정도의 가중치를 모두 고려한 것이다. 분류 항목에 따른 학습 데이터를 사용하여 PFCM 알고리즘을 반복 혹은 비반복 수행함으로써 다양한 형태로 분포된 위성영상의 분광 데이터 값에 대하여 효과적인 패턴분류가 이루어질 수 있도록 하였다.

### 3.1 PFCM 클러스터링 알고리즘

PFCM 클러스터링 알고리즘은 FCM과 PCM 알고리즘을 결합한 알고리즘으로서 FCM 알고리즘의 소속정도와 PCM 알고리즘의 전형성에 대한 소속정도를 모두 고려한 알고리즘이다[7][8]. PFCM 알고리즘은 목적함수  $J_{m,\eta}(U, T, V; X)$ 을 최소화 할 수 있도록 데이터 집합을 분류한다.

$$J_{m,\eta}(U, T, V; X) = \sum_{k=1}^n \sum_{i=1}^c (a\mu_{ik}^m + bt_{ik}^m) \times \|x_k - v_i\|_A^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1 - t_{ik})^\eta \quad (8)$$

이 때 제약사항은 모든  $k$ 에 대하여  $\sum_{i=1}^c \mu_{ik} = 1$ 을 나타내고,  $0 \leq \mu_{ik}, t_{ik} \leq 1$ 이며,  $a > 0, b > 0, m > 1, \eta > 1, \gamma_i > 0$ 이다. 목적함수  $J_{m,\eta}(U, T, V; X)$ 에서  $a$ 와  $b$ 는 퍼지 소속정도와 전형성 정도에 대한 상대적인 중요성을 정의한다. 식(8)에서  $\mu_{ik}$ 는 FCM 알고리즘의 퍼지 소속정도와 같고  $t_{ik}$ 는 PCM의 전형성 정도를 나타내는 것과 같다. 소속정도인  $\mu_{ik}$  또는 전형성 정도  $t_{ik}$  각각을 이용하여 중심 값을 구하고 클러스터링을 수행 할 수 있다.

$$u_{ik} = \left( \sum_{j=1}^c \left( \frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad 1 \leq i \leq c; 1 \leq k \leq n \quad (9)$$

$$t_{ik} = \frac{1}{1 + \left(\frac{b}{\gamma_i} D_{ikA}^2\right)^{\frac{1}{\eta-1}}}, \quad 1 \leq i \leq c; 1 \leq k \leq n \quad (10)$$

PFCM 알고리즘의 중심 값은 식(11)을 이용하여 구한다.  $\mu_{ik}$ 와  $t_{ik}$ 의 가중치를 고려하여 중심 값을 구한다.

$$v_i = \frac{\sum_{k=1}^n (a\mu_{ik}^m + bt_{ik}^m)x_k}{\sum_{k=1}^n (a\mu_{ik}^m + bt_{ik}^m)}, 1 \leq i \leq c \quad (11)$$

PFCM 알고리즘은 식(9)에서 식(11)까지를 반복 수행하여 목적함수  $J_{m,\eta}(U, T, V; X)$ 의 최소치를 만족하는 값을 구함으로써 패턴 데이터 분류를 수행한다. PFCM 클러스터링 알고리즘에서  $a$ 의 값이 1이고  $b$ 의 값이 0이면 FCM결과와 동일하며,  $a$ 의 값이 0이고  $b$ 의 값이 1이면 PCM 결과와 같다. 즉  $a$ 와  $b$ 를 이용하여 가중치 즉, 중요성을 유연하게 조절하여 중심 값을 구하고 이를 이용한 소속정도 혹은 전형성 정도를 고려한 클러스터링 알고리즘을 수행한다.

### 3.2 위성영상 패턴분류 시스템

위성영상 패턴분류 방법에는 기본적으로 감독분류와 무감독분류로 나뉜다. 감독분류는 분석자가 영상 내에서 알고 있는 화소들의 영역을 지정하면, 각 군집의 중심, 밴드 간 공분산 등을 이용하여 자료내의 모든 화소를 분석하여 가장 유사한 분포특성을 가지는 군집에 할당시키는 과정이

다. 무감독분류 기법은 표본 집단과 같은 사전정보 없이 영상을 구성하는 화소 값의 공간적, 분광적 특성만을 이용하여 분류 작업을 수행하는 기법이다. 무감독분류의 대표적인 방법이 클러스터링을 이용한 방법이다. 일반적으로 클러스터링 알고리즘은 표본 집단과 같은 사전정보 없이 영상을 구성하는 화소 값의 공간적, 분광적 특성만을 이용하여 분류할 군집의 개수와 각 분광 군집 사이의 한계 거리만을 지정해 줌으로써 화소 값을 군집화 한다. 이때 각 군집은 분류 항목의 내용과는 상관없이 분광특성만을 나타내므로 분류 수행 후 항목의 정의를 내려준다.

본 논문에서는 PFCM 알고리즘을 응용하여 위성영상의 다중분광 원격탐사 위성영상을 감독 분류하는 방법을 제안하였다. 분석자가 분류할 항목을 먼저 설정하고 영상 내에서 알고 있는 화소들의 영역을 지정하여 분류항목별 학습데이터를 선정한 후 학습데이터의 모든 화소를 분석하여 PFCM 알고리즘에 적용하여 가장 유사한 분포특성을 가지는 군집에 할당시키는 감독분류를 수행하였다. 그림 1은 PFCM 알고리즘을 이용한 위성영상의 패턴분류과정을 나타낸 것이다.

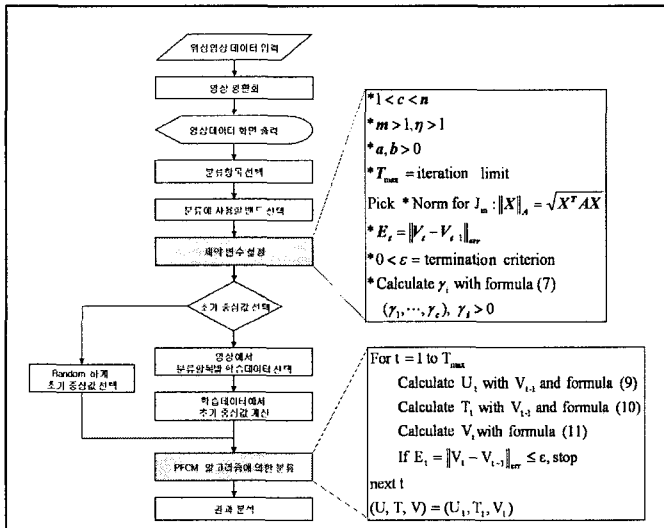


그림 1. PFCM 클러스터링 알고리즘을 이용한 위성영상의 패턴분류 과정

Fig. 1. Procedure of pattern classification by PFCM clustering algorithm.

제안된 방법의 패턴분류과정은 다음과 같이 각 단계별로 수행된다.

**[단계 1 : 분류하고자 하는 위성영상 데이터 입력]**

분류하고자 하는 위성영상을 불러들여 각각 밴드의 분광 데이터 값  $X = \{x_1, x_2, \dots, x_n\} \subset R^p$ 을 획득한다. 본 연구에서는 Landsat TM 위성영상과 IKONOS 위성영상을 입력데이터로 하고 있다.

**[단계 2 : 위성영상의 평활화]**

분광 데이터 값을 이용하여 각 밴드를 평활화 한 후 세

개의 밴드를 선택하여 RGB로 설정하고 화면에 보여준다.

**[단계 3 : 분류에 이용할 밴드 선택]**

위성영상은 파장대별 특성에 따라 여러 개의 밴드로 나누어진다. 분류 목적에 따른 필요한 위성영상의 밴드를 선택한다.

**[단계 4 : 분류항목 설정]**

분석자가 분류 주제에 따라 분류하고자 하는 군집 수 (분류항목 수) 및 분류 항목을 설정한다.

**[단계 5 : 제약 변수 설정]**

클러스터링 알고리즘을 수행하기 위한 제약사항, 즉 주요 변수들을 설정한다. 특히 소속정도  $\mu_{ik}$ 와 전형성 정도  $t_{ik}$ 에 대한 가중치를 설정하기 위해  $a > 0, b > 0$ 을 선택한다. 그리고 클러스터링의 반복과정의 중단 조건을 설정한다.

**[단계 6 : 초기 중심 값 설정]**

분류항목별 초기 중심 값  $V_0 = (v_{1,0}, \dots, v_{c,0}) \in R^p$ 을 설정한다. 초기 중심 값을 설정하는 방법은 분류 항목에 따른 학습데이터 영역들을 화면에 출력된 영상에서 분석자가 직접 분류항목별로 적절한 영역을 선택하고, 선택한 영역의 분광 데이터 값을 이용하여 각 분류항목별 중심 값을 구한다. 이때 군집에 따른 분류결과를 효과적으로 분석하기 위하여 분류항목별로 대응 색상을 선택한다.

**[단계 7 : PFCM 알고리즘 수행]**

중심 값을 이용하여 PFCM 알고리즘을 이용하여 분류를 수행한다. 클러스터링 알고리즘에 사용되는 놈 (norm)은 유클리디안 거리(Euclidean distance)를 일반적으로 사용하게 하였으며 필요에 따라서 마하라노비스 거리(Maharanobis distance)를 사용할 수 있도록 하였다. 소속정도  $\mu_{ik}$ 와 전형성 정도  $t_{ik}$ 를 이용하여 분류를 수행한다.

**[단계 8 : 결과분석]**

분류된 결과를 이용하여 주제도를 작성하고 정확도를 분석한다.

**IV. 실험 및 결과분석**

원격탐사 위성영상에 대하여 PFCM 알고리즘을 응용한 감독분류 방법의 효용성을 확인하기 위하여 1992년 6월에 서울 한강의 한 영역을 촬영한 Landsat TM 위성영상과 2001년 11월에 부산의 수영만 지역을 촬영한 IKONOS 위성영상을 사용하였다. 제안된 알고리즘의 테스트를 위하여 Visual C++ 6.0을 이용하여 알고리즘을 구현한 후 그 결과를 분석하였다.

Landsat TM 위성영상은 7개의 밴드로 구성된 다중분광 위성영상으로서 한 화소의 공간해상도는 30m×30m로 구성되어 있다. 본 논문에서는 실험을 단순화하기 위해서 7, 4,

2 밴드로 구성된 영상을 사용하였다. 영상의 크기는 136×136 화소로 구성되어 있다. 입력영상의 한 화소의 비트수는 8비트이다. IKONOS 위성영상은 고해상도 위성영상으로서 흑백 및 다중분광 영상 모두 11비트의 정보를 가지고 있으며, 1m 전정색(Panchromatic) 밴드의 흑백 영상 데이터는 가시광선에서 근적외선 파장대의 정보를 가지고 있다. 4m 다중분광밴드 영상은 적색(R), 녹색(G), 청색(B)과 근적외선(NIR)의 각 4개의 밴드로 구성되어 있다. 본 논문에서는 4m 다중분광 위성영상을 이용하여 실험하였다. 실험에 이용된 영상의 크기는 1000×1000개의 화소를 가지고 있으며, 4개의 밴드 모두를 실험에 사용하였다. 그림 2는 실험에 사용된 Landsat TM 위성과 IKONOS 위성의 영상으로서 평활화를 수행한 후의 영상이다.



그림 2. Landsat TM 위성영상(좌)과 IKONOS 위성영상(우)  
Fig. 2. Landsat TM satellite image(left) and IKONOS satellite image(right)

본 연구에서는 입력영상에서 각 분류항목별로 각각 몇 개의 학습영역을 선택하여 그 평균값을 계산하여 초기 중심 값으로 이용하였다. 실험의 결과를 얻기 위해 알고리즘을 30회 정도 수행하여 평균적인 값을 얻어서 해당되는 실험결과를 나타낸 것이다.

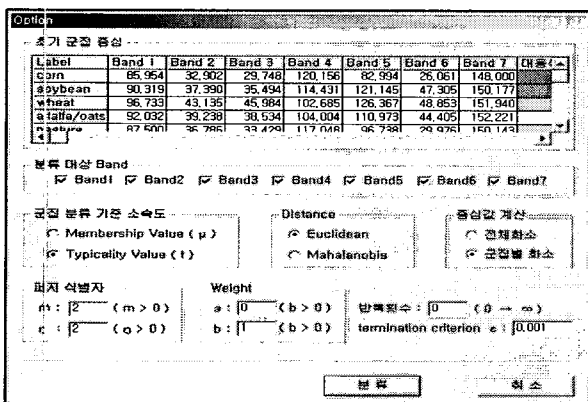


그림 3. 분류를 위한 제약사항 설정  
Fig. 3. Set of parameters for classification.

위성영상의 분류를 위한 실험을 위해서 Landsat TM 위성영상에서는 urban, water, crop, forest 등 4개의 분류 항목을 설정하였다. 그림 3은 PFCM 알고리즘을 적용하여 감독분류를 수행하기 위하여 밴드를 선택하고 반복, 비반복 횟수 등 여러 가지 제약 조건들을 설정하는 화면을 나타내었다.

표 1은 Landsat TM 위성영상에서 분류항목별 학습데이터에서 구한 초기 군집 중심 값과 이를 이용하여 반복 및 비반복 군집화 과정을 수행한 후의 밴드별 각 분류항목별 중심 값을 비교하였다. 실험에서 반복 수행을 하는 경우 중단 조건은 이전 클러스터링 과정을 수행한 중심 값과 비교하여 오차율이  $\epsilon=0.001$  이하가 되면 반복 과정을 중단한다. 실험 결과 초기 군집 중심 값과 비교하였을 때 비반복 수행하였을 경우와 반복수행 했을 경우의 중심 값의 변화는 크지 않고 비슷한 결과를 얻었다. 따라서 학습데이터를 초기 중심 값으로 이용하여 분류를 수행할 경우 반복수행을 하지 않아도 분류결과를 얻을 수 있음을 보여준다. 따라서 고해상도 대용량 영상의 경우에 있어서 반복 수행함으로써 많은 시간이 걸리는 단점을 해결할 수 있다.

표 1. Landsat TM 위성영상의 분류항목별 중심 값의 변화  
Table 1. Values of centroid by classification categories of IKONOS satellite image.

분류 항목	초기군집중심			PFCM 수행 후 군집중심					
	band7	band4	band2	비반복 수행			반복 수행 ( $\epsilon=0.001$ )		
Urban	88.95	80.18	62.03	83.93	80.81	60.70	86.19	81.21	61.22
Water	6.90	29.76	45.69	7.28	30.06	45.58	7.29	30.07	45.58
Crop	13.11	52.36	49.10	15.22	53.46	49.22	15.55	53.94	49.26
Forest	29.76	86.55	44.61	37.71	85.15	49.00	41.81	83.36	50.37

표 2. 제약변수에 변화에 따른 PFCM의 수행결과  
Table 2. Results produced by PFCM for different values of parameters.

a	b	m	$\eta$	$v_1$				$v_2=0.001$				반복 횟수
				Urban	Water	Crop	Forest	Urban	Water	Crop	Forest	
1	3	2	2	85.71	7.17	14.18	33.77	86.27	7.21	15.44	41.80	26
1	1	5	2	85.16	7.14	13.73	34.36	82.01	7.10	13.30	41.94	58
1	1	7	2	84.98	7.15	13.68	33.98	86.72	7.08	12.77	39.88	198
1	5	5	2	87.23	7.06	13.18	31.27	84.16	7.07	12.89	42.38	89
1	1	2	2	82.93	7.28	15.22	37.70	86.19	7.29	15.55	41.81	25
1	4	3	2	86.18	7.12	13.79	34.49	87.59	7.13	14.31	42.53	43
1	0	2	2	85.60	7.33	15.74	35.65	86.44	7.36	15.95	41.86	24
0.5	0.5	2	2	83.62	7.29	15.23	37.91	85.80	7.28	15.46	41.76	26
0	1	2	2	85.87	7.23	14.26	32.71	49.53	7.13	13.28	25.70	215

표 2에서는 제약변수들을 다양하게 변화시켰을 때 초기 중심 값과 반복 수행 후의 중심 값의 변화와 반복횟수를 확인하였다. a의 값은 소속정도의 가중치를 나타내며, b의 값은 전형성 정도의 가중치를 나타낸 것으로서 소속정도 a에 대한 전형성의 가중치 b의 변화에 따른 수행결과를 비교하였다. 이때 b의 값이 0이면 FCM 클러스터링 수행결과와 같고 a값이 0이면 PCM 클러스터링과 같은 결과를 얻는다.

표 3과 표 4는 Landsat TM 위성영상에서 학습데이터에 대한 PFCM 알고리즘을 비반복 수행했을 경우와 반복 수행한 후의 confusion matrix를 나타낸 것이다. 이것은 각 분류항목별로 학습데이터로 선정한 데이터들이 알고리즘

수행 후 처음에 선택된 분류 항목으로 얼마나 정확하게 분류되어졌는가를 나타낸 것이다. 분류결과를 보면 반복 수행한 경우와 비반복 수행한 경우의 결과는 비슷하지만 비반복 수행한 경우의 결과가 학습데이터의 분광 특성을 잘 반영한 분류결과 값을 나타냄을 보였다.

표 3. PFCM 비반복 수행 Confusion Matrix( $a=1, b=1, m=2, n=2$ )  
Table 3. Confusion matrix by the non-iterative execution of the PFCM algorithm, ( $a=1, b=1, m=2, n=2$ )

		학습데이터				Total
		Urban	Water	Crop	Forest	
분류 결과	Urban	1203	0	0	4	1207
	Water	0	1290	0	0	1290
	Crop	16	0	1187	5	1208
	Forest	86	0	13	1207	1306
Total		1305	1290	1200	1216	4887/5011 (97.5%)

표 4. PFCM 반복 수행 Confusion Matrix ( $a=1, b=1, m=2, n=2$ )  
Table 4. Confusion matrix by the iterative execution of the PFCM algorithm ( $a=1, b=1, m=2, n=2$ ).

		학습데이터				Total
		Urban	Water	Crop	Forest	
분류 결과	Urban	1160	0	0	8	1165
	Water	0	1290	0	0	1290
	Crop	14	0	1192	6	1211
	Forest	131	0	8	1202	1345
Total		1305	1290	1200	1216	4852/5011 (96.8%)

Landsat TM 영상에서와 같은 방법으로 IKONOS 위성영상을 이용한 실험을 수행하였으며 IKONOS 위성영상은 water, soil, urban, forest, shadow 등 5개의 분류 항목을 설정하였다. 분류항목별 초기 군집중심과 반복 및 비반복 수행하였을 경우에 Landsat TM 영상을 적용한 결과와 비슷한 결과를 얻을 수 있었다. 따라서 PFCM 클러스터링 알고리즘을 적용한 분류 기법은 IKONOS 위성과 같은 고해상도 위성영상에서도 적용 가능함을 확인하였다. 단 고해상도 IKONOS 위성영상은 파일이 대용량으로서 반복 과정을 적용함에 있어서 많은 시간이 소요되어짐을 보였으며, 그러므로 비반복 과정을 수행한 분류를 적용하는 것이 효과적이다.

표 5. IKONOS 위성영상의 분류항목별 중심 값의 변화  
Table 5. Values of centroid for classification categories of IKONOS satellite image.

분류 항목	초기군집중심				PFCM 수행 후 중심			
	band1	band2	band3	band4	band1	band2	band3	band4
Water	184.83	149.91	58.54	30.22	203.25	175.98	70.09	32.75
Soil	312.02	389.71	369.05	377.30	333.71	402.30	368.38	366.33
Urban	220.15	218.35	160.71	155.68	234.26	239.57	184.64	179.72
Forest	165.95	138.93	80.42	143.99	189.28	170.17	109.71	120.68
Shadow	170.48	133.91	67.89	44.83	176.18	144.87	78.30	59.22

표 6은 각 분류방법에 따른 위성영상의 분류결과를 비교하였다. FCM, PCM, PFCM 알고리즘은 각각  $a$ 와  $b$ 의 가중치를 조정하여 분류결과를 획득하였다. PFCM 알고리즘에서  $a$ 의 값이 1이고  $b$ 의 값이 0이면 FCM 결과와 동일하며,  $a$ 의 값이 0이고  $b$ 의 값이 1이면 PCM 결과와 같다. 최대우도분류(Maximum likelihood classification ; MLC) 방법은 기존에 구현된 알고리즘을 현재의 프로그램에 하나의 모듈로 삽입하여 그 결과를 얻었다.

표 6. 원격탐사 위성영상의 분류결과 비교  
Table 6. Comparison of classification results of remote sensing satellite images.

실험 위성영상		Landsat TM 영상		IKONOS 영상	
		반복	비반복	반복	비반복
분류 방법	FCM	91%	93%	89%	91%
	PCM	91%	93%	90%	91%
	MLC	N/A	95%	N/A	95%
	PFCM	96.8%	97.5%	93.5%	94.5%

표 6에서 PFCM 알고리즘을 적용한 방법은 제약조건을  $a=1, b=1, m=2, n=2$ 로 설정하고, 학습데이터에 의하여 계산된 초기 중심 값을 이용하여 위성영상의 분류를 수행하였다. PFCM 알고리즘을 이용한 분류방법에 대한 정확도는 제약변수의 설정에 따라서 차이가 있을 수도 있지만 대체적으로 FCM과 PCM 알고리즘에 비하여 향상되었다. 그리고 일반적으로 많이 사용되는 분류방법인 최대우도분류 방법보다도 실험에서 다소 좋은 분류결과를 얻을 수 있었다. 반복 수행한 경우와 비반복 수행한 경우를 결과를 비교해보면, 반복 수행한 경우 학습데이터의 분광특성을 고려하지 않고 각 화소들의 분광 값에 의한 분류가 수행되므로 분류의 정확도가 약간 감소되어지는 것을 볼 수 있다. PFCM 알고리즘의 경우에는 학습데이터를 선택하여 분류를 수행할 경우 반복수행한 경우와 비반복 수행한 경우는 차이가 크지 않았다. 그림 4는 Landsat TM과 IKONOS 위성영상에 대하여 PFCM 알고리즘을 이용한 분류를 수행한 후의 결과 영상을 나타낸 것이다.

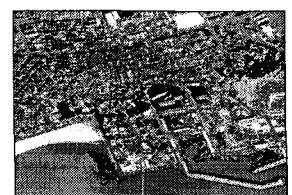
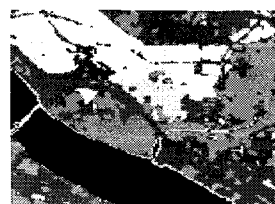


그림 4 PFCM 분류 후의 Landsat TM 위성영상(좌)과 IKONOS 위성영상(우)

Fig. 4. Result images of Landsat TM(left) and IKONOS(right) by the PFCM classification algorithm.

V. 결론

PFCM 클러스터링 알고리즘은 각 화소와 특정 클러스터 중심과의 거리에 대한 소속정도를 고려한 FCM 클러스터링 알고리즘과 화소와 해당 클러스터 중심 간의 거리에 의존한 전형성(typicality) 정도를 고려한 PCM 클러스터링 알고리즘을 결합한 방법이다. 본 연구에서는 원격탐사 영상을 분류하기 위하여 사전에 분류 항목을 선정하고 그에 따른 학습데이터를 영상에서 선택하여 초기 중심 값을 구한 후 PFCM 알고리즘을 적용하여 감독분류를 수행 하였다.

Landsat TM과 IKONOS 위성영상을 이용하여 학습데이터를 선정하고 PFCM 알고리즘을 적용한 감독분류를 수행한 결과에 따르면 제약 변수에 따라 약간의 차이가 있을 수도 있지만 대체적으로 비반복 수행한 경우와 반복 수행한 경우의 중심 값이 비슷한 결과를 보였으며, 또한 비반복 과정을 통한 분류결과는 반복 수행하는 것보다 분광특성을 잘 나타낸 분류결과를 얻었다. 따라서 학습데이터를 이용함으로써 PFCM 알고리즘의 반복적인 클러스터링 과정을 거치지 않고 한 번의 군집화 과정으로도 좋은 분류결과를 얻을 수 있으므로 반복 수행에 따른 시간을 줄일 수 있을 뿐만 아니라 IKONOS 위성영상과 같은 대용량의 고해상도 영상 분류에도 효과적으로 적용 가능하다. 특히 PFCM 알고리즘을 이용한 감독분류는 PCM, FCM 분류방법보다 좋은 결과를 보여주었으며, 또한 전통적인 분류방법인 최대우도분류보다도 정확도가 더 높은 결과를 보여주었다. 하지만 PCM, FCM 분류방법보다 분류시간이 단축되지만 최대우도분류보다는 분류시간이 더 많이 걸린다. 그리고 여러 가지 종류의 원격탐사 영상에서 학습데이터 및 가중치  $a, b$  에 따라 분류결과가 조금씩 차이가 있으므로 이에 대한 효과적인 선택이 요구된다.

향후 PFCM 알고리즘에서 적절한 가중치 및 제약변수의 자동적인 선택에 관한 연구가 뒤따라야 하며, 또한 대용량 위성영상 데이터 및 KOMPSAT-2 등의 다양한 종류의 위성영상의 분류에 본 알고리즘을 적용하여 그 효율성을 확인하는 연구가 뒤따라야 하겠다.

참고 문헌

[1] Moon-Gyu Kim et al., "Ground Receiving System for KOMPSAT-2," *The Korean Society of Remote Sensing, Korean Journal of Remote Sensing*, Vol.19 No.3, 2003.6, pp.191~200.  
 [2] John A. Richards, Xiuping Jia,, Remote Sensing

Digital Image Analysis : An Introduction, 4th ed., Springer Berlin Heidelberg, 2006.

[3] James C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NY, 1981.  
 [4] N.R. Pal and J.C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, Vol.3, No.3, 1995, pp.370~379  
 [5] R. Krishnapuram and J.M. Keller "A possibilistic approach to clustering," *Fuzzy Systems, IEEE Transactions*, Vol. 1, No. 2, 1993, pp.98~110.  
 [6] Jiang-She Zhang, Yiu-Wing Leung, "Improved possibilistic C-means clustering algorithms," *Fuzzy Systems, IEEE Transactions*, Vol 12, Issue 2, April 2004, pp.209~217.  
 [7] N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm," *Fuzzy Systems, IEEE Transactions*, Vol. 13, Issue 4, Aug. 2005, pp.517~530.  
 [8] N.R. Pal, K. Pal., J.M. Keller., J.C. Bezdek, "A new hybrid c-means clustering model," *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference*, Vol. 1, July 2004, pp.179~184.



전 영 준(Young-Joon Jeon)

1998년 2월 동의대 전산통계학과(이학사)  
 2000년 2월 동의대 전산통계학과(이학석사)  
 2005년 2월 동의대 컴퓨터공학과(공학박사)

2005년~현재 대구대학교 컴퓨터-IT공학부 초빙교수  
 ※주관심분야 : 패턴인식, 임베디드시스템, 센스네트워크



김 진 일(Jin-II Kim)

1976년~1982년 경희대학교 전자공학과  
 (학사, 공학석사)  
 1994년 서강대학교 컴퓨터학과(공학박사)

1982년~1984년 미국 Bon Scours 시스템,  
 맥도널다글라스 우주항공회사 등 연구원  
 1996년~1997년 미국 Purdue Univ.  
 전기및컴퓨터공학부 연구교수  
 1998년~2001년 위성탐사 영상처리 연구회(한국과학재단)  
 회장

1988년~현재 동의대학교 컴퓨터공학과 교수  
 ※ 주관심분야 : 패턴인식, 위성영상처리, 지능시스템