

하이브리드 분류기법을 이용한 위성영상의 분류에 관한 연구

전 영 준[†] · 김 진 일^{††}

요 약

본 논문에서는 위성영상의 분류에 대한 성능 개선을 위하여 ISODATA 클러스터링, 퍼지 C-Means 알고리즘, 베이시안 최대우도 분류기법을 통합한 하이브리드 분류기법을 제안하였다. 본 연구에서는 분석자에 의하여 분류항목별 학습 데이터를 선정한 후 이를 ISODATA 클러스터링을 이용하여 각각의 분류항목별로 분광특성에 따라 학습 데이터를 세분화하여 새로운 학습 데이터를 선정하였다. 새롭게 선정된 학습 데이터를 이용하여 퍼지 C-Means 알고리즘을 이용하여 분류를 수행하고 그 결과를 베이시안 최대우도 분류기의 사전확률로 적용하여 분류를 수행하였다. 그 결과 분석자가 선정한 분류항목별 훈련데이터의 분광적인 특징에 관계없이 분류를 수행할 수 있었으며 위성영상의 분류의 성능을 개선할 수 있었다. 제안된 기법은 Landsat TM 위성영상을 이용하여 그 적용성을 시험하였다.

A Study on the Classification for Satellite Images using Hybrid Method

Young Joon Jeon[†] · Jin Il Kim^{††}

ABSTRACT

This paper presents hybrid classification method to improve the performance of satellite images classification by combining Bayesian maximum likelihood classifier, ISODATA clustering and fuzzy C-Means algorithm. In this paper, the training data of each class were generated by separating the spectral signature using ISODATA clustering. We can classify according to pixel's membership grade followed by cluster center of fuzzy C-Means algorithm as the mean value of training data for each class. Bayesian maximum likelihood classifier is performed with prior probability by result of fuzzy C-Means classification. The results shows that proposed method could improve performance of classification method and also perform classification with no concern about spectral signature of the training data. The proposed method is applied to a Landsat TM satellite image for the verifying test.

키워드 : 하이브리드 분류기법(Hybrid Classification Method), ISODATA 클러스터링(ISODATA Clustering), 퍼지 C-Means(Fuzzy C-Means), 베이시안 최대우도 분류기법(Bayesian Maximum Likelihood Classifier)

1. 서 론

원격탐사 기술은 컴퓨터, 전자 기술의 발전에 의하여 1m의 공간해상도를 가지는 위성영상과 200채널을 가지는 하이퍼스펙트럴 위성영상의 획득이 가능하다. 따라서 원격탐사 기술은 새로운 시대를 맞이하고 있다. 원격탐사(remote sensing)란 멀리 떨어져 있는 대상물을 관측하는 기술로서 측정하고자 하는 목표물에 직접 접촉하지 않고 목표물에서 복사되어 나오는 전자파를 감지하여 그 물리적 성질을 측정하는 과학이다[1-2]. 지구로부터 수집된 지상 물체에 대한 다양한 정보를 얻어내기 위하여 여러가지 파장대에서 획득한 영상 데이터를 분석하여 유용한 주제정보를 추출한다. 원격탐사에 의한 정보는 토지의 체계적인 활용, 환경오염의 감시 및 통제, 지도제작 등에 요구되는 복잡 다양한

정보를 신속하고 정확하게 분류, 해석하는데 있어서 커다란 잠재력을 가지고 있는 최첨단 과학기술로 각광받고 있으며, 그 활용 분야는 환경에서부터 토목, 농업, 해양, 지질, 입암, 수산업 등 각 분야에 걸쳐 급속도로 확산되고 있다. 최근 위성 탐측기의 급속한 발전에 의해 해상력이 향상되고 자료획득 시간이 단축됨에 따라 이용분야가 확대되고 있다. 이러한 정보의 추출을 위해 가장 자주 이용되는 방법은 패턴인식을 이용한 다중분광 분류이다. 패턴인식은 데이터로부터 중요한 특징이나 속성을 추출하여 입력 데이터를 식별할 수 있는 부류로 분류하는 것이다.

인공위성으로부터 얻어진 원격탐사 위성영상은 다양한 보정 작업을 거친 후 분류를 위한 자료로 사용되어진다. 영상의 분류는 영상을 구성하는 각각의 화소들 중 비슷한 분광 특성을 갖는 것끼리 집단화시켜 주는 방법으로서 일반적으로 이와 같이 구성된 각각의 집단을 클래스(class), 분류항목이라고 한다. 영상의 각 화소는 분류를 수행한 후에 하나의 분류항목 또는 값을 갖게 된다. 이것은 영상으로부터

* 이 논문은 2003년도 BB21 사업에 의해 수행되었습니다.

† 정 회 원 : 동의대학교 컴퓨터공학과 교수

†† 정 회 원 : 동의대학교 대학원 컴퓨터공학과

논문접수 : 2003년 9월 21일, 심사완료 : 2004년 2월 5일

터 관심 대상인 정보를 추출하기 위해 수행되는 것으로 주로 논, 밭, 나대지 등의 토지피복, 표면 온도, 대기 상태 등의 환경상태 등 광범위한 지역을 몇 종류 또는 몇 단계로 분류할 수 있는 정보를 추출하는 데 사용된다[3, 4].

위성영상 분류의 최근 연구를 살펴보면 베이지안 분류와 같은 기존의 통계적 이론을 확장하여 보다 개선된 내용을 제안한 연구[5, 6], 원격탐사 영상의 분류에 있어 클래스의 경계를 확실히 정하기 어려운 경우에 효과적으로 이용되는 퍼지 분류기법을 적용한 연구[7, 8], BP(Backpropagation), BPNN(BP neural network), Kohonen의 Self-organizing Map (SOM), Self-Organizing Neural Networks(SONN) 등의 신경망 이론을 적용한 연구[9-11]가 이루어지고 있다. 또한 Markov Random Field(MRF)를 이용하여 분류를 수행하는 방법에 대한 연구[12], 퍼지 이론과 신경망을 융합한 기법과 같이 여러 개의 분류기법을 혼합하여 분류 정확도를 향상시키는 연구[13] 등 다양한 연구가 이루어지고 있다.

본 논문에서는 위성영상의 분류의 정확성을 개선시키기 위하여 감독 분류를 기반으로 부분적으로 무감독 분류기법을 적용한 하이브리드 분류방법을 제안하였다. 제안된 방법은 ISODATA 클러스터링을 이용하여 분류항목별 학습데이터를 분광특성에 따라 세부항목으로 분리하여 분석자가 선정한 분류항목별 학습 데이터의 분광적인 특징에 관계없이 분류를 수행할 수 있도록 하였다. 이렇게 선정된 학습데이터를 이용하여 퍼지 C-Means 알고리즘을 수행하여 애매모호한 지역의 분류문제를 고려한 분류를 수행하였고, 그 분류결과를 베이지안 최대우도 분류기의 사전확률로 이용하여 분류를 수행함으로써 전체적인 분류 성능을 향상시킬수 있었다.

본 논문의 구성은 2장에서는 관련연구로 몇 가지 감독 분류와 무감독 분류 알고리즘에 대하여 설명하고, 3장에서는 본 연구에서 제안하는 분류방법으로서 ISODATA 클러스터링과 퍼지 C-Means 알고리즘, 베이지안 최대우도 분류기법을 통합한 하이브리드 분류방법에 대해서 설명한다. 4장에서는 Landsat TM 위성영상을 입력으로 하여 실험 및 결과를 분석하고, 5장에서는 향후 연구 및 결론을 맺는다.

2. 관련 연구

위성영상 분류의 방법에는 크게 감독 분류와 무감독 분류로 나누어진다. 감독 분류기법은 영상에서 분류하고자 하는 정보 집단의 위치나 분광특성을 사용자가 정확히 알고 있을 때 사용하는 방법이다. 일단 영상자료로부터 집단의 각 밴드별 화소값을 추출해야 한다. 이렇게 추출된 영상의 일부분을 학습 데이터라고 하며, 분류작업시 학습 데이터의 통계적 수치를 계산하여 입력자료로 사용한다. 분류하고자

하는 자료의 학습 데이터가 완성되면, 이들과 비슷한 통계 값을 갖는 화소들을 계속 수집하여 집단의 크기를 확장해감으로써 전체 영상에 대한 분류작업을 완성하게 된다.

감독 분류기법은 사용자의 판단이 전체 결과에 매우 중요한 요소로 대두하기 때문에 많은 경험과 정확한 사전지식이 요구된다. 평행육면체(parallelepiped) 분류기법은 데이터 파일의 값들을 상한과 하한의 테두리 속에서 분류하는 것이며, 최소거리(minimum distance) 분류기법은 화소자료와 분류항목 특징과의 유사도를 특징공간에 있어서의 거리로 나타내고, 거리가 가장 짧은(유사도가 가장 큰) 분류항목에 화소자료를 분류하는 방법이다. 즉, 각 분류항목의 평균벡터와 분류하고자 하는 화소의 특징벡터간의 분광적 거리를 계산하여 가장 가까운 값을 가지는 항목으로 화소가 분류되는 기법이다. 마하라노비스 거리(Mahalanobis distance) 분류기법은 밴드의 히스토그램이 정규분포를 이룬다는 가정하에 실행이 되며, 정규분포가 확실히 아니라면 다른 분류기법을 사용하는 것이 좋다. 마하라노비스 거리 분류기법은 최소거리 분류와 유사하지만, 거리를 구할 때 공분산 행렬이 포함되어 있으므로 분산이 아주 높은 군집이 유사한 성질의 항목으로 분류가 되도록 유도하는 장점을 지닌다. 최대우도(maximum likelihood) 분류기법은 하나의 분류항목에 대한 확률함수가 정규밀도 함수에 근사한다고 가정하고서, 학습 데이터로부터 취득된 통계 값들을 이용하여 영상의 화소들을 특정한 항목들로 분류하는 알고리즘이다. 즉, 각 분류항목에 대한 화소자료의 우도를 구하고 최대우도 분류항목에 그 화소를 분류하는 방법이다.

무감독 분류기법은 사전정보 없이 영상을 구성하는 화소값의 공간적, 분광적 특성만을 이용하여 분류작업을 수행하는 기법이다. 이 방법은 순수한 통계처리의 기법에 의해 수행된다는 점에서 군집화라고 불리며, 분류할 군집의 개수와 각 분광 군집 사이의 한계 거리만을 지정해 줌으로써 작업이 가능하다. 이와 같은 무감독 분류기법에는 순차 군집분류기법, K-Means 군집분류기법, ISODATA 군집분류기법, 퍼지 C-Means 군집분류기법 등이 있다[3, 4]. 순차 군집분류기법은 화소를 한 개씩 순차적으로 검사하는 방법이다. 각 분석된 화소와 이미 계산된 군집중심 사이의 분광 거리를 계산하여 각 화소를 새로운 군집중심의 형성에 사용되거나, 이미 존재하는 군집에 분배한다. 사용자의 요구 군집보다 많은 수의 군집이 생성되었을 경우 군집들의 병합이 일어나며 모든 화소를 단일검색으로 군집기법이 종료하기 때문에 아주 빠른 수행을 한다. K-Means 군집분류기법은 영상을 통해 얻고자 하는 군집의 개수를 사용자가 지정하도록 되어있고, 사용자가 지정한 개수만큼의 군집 중점을 다차원 측정공간상에 위치시킨다. 영상의 각 화소들은 군집들의 임시 평균벡터들 중에서 가장 가까운 군집에 할당되

며 모든 화소들이 이러한 방법으로 분류된 후, 군집은 새로운 군집중심이 계산된다. 이렇게 수정된 군집중심은 영상 재분류의 근거로 사용되며 군집중심의 위치가 더 이상 변하지 않을 때까지 알고리즘의 반복이 지속된다. ISODATA 군집분류기법은 매 반복 단계마다 표본의 평균을 군집의 중심으로 정한다는 점에서 K-Means 군집분류기법과 원리상으로 유사하지만, 부가적인 군집의 분리와 병합을 한다는 점에서 K-Means 군집분류기법과 다르다. 퍼지 C-Means 군집분류기법은 하드 C-Means(HCM) 군집분류기법의 퍼지모델로 알고리즘의 결과가 군집중심과 퍼지 C-Means 분할행렬을 동시에 구할 수 있는 자기 조직화, 무감독 분류기법이다.

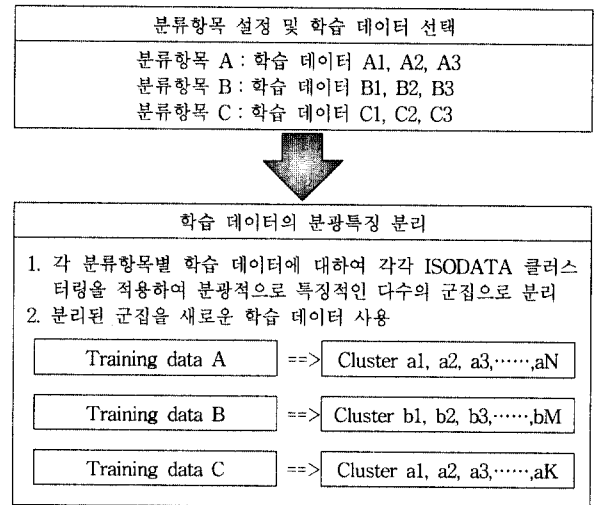
3. 하이브리드 분류 기법

본 논문에서는 ISODATA 클러스터링, 퍼지 C-Means 알고리즘, 베이시안 최대우도 분류기법을 통합한 하이브리드 분류기법을 제안하였다. 분석자에 의하여 분류항목별 학습 데이터를 선정할 후 이를 ISODATA 클러스터링을 이용하여 각각의 분류항목별로 분광특징에 따라 학습 데이터를 세분화하여 새로운 학습 데이터를 선정하였다. 새롭게 선정된 학습 데이터를 이용하여 퍼지 C-Means 알고리즘을 이용하여 분류를 수행하고 그 결과를 베이시안 최대우도 분류기의 사전확률로 적용하여 분류를 수행하였다.

3.1 학습 데이터의 선정

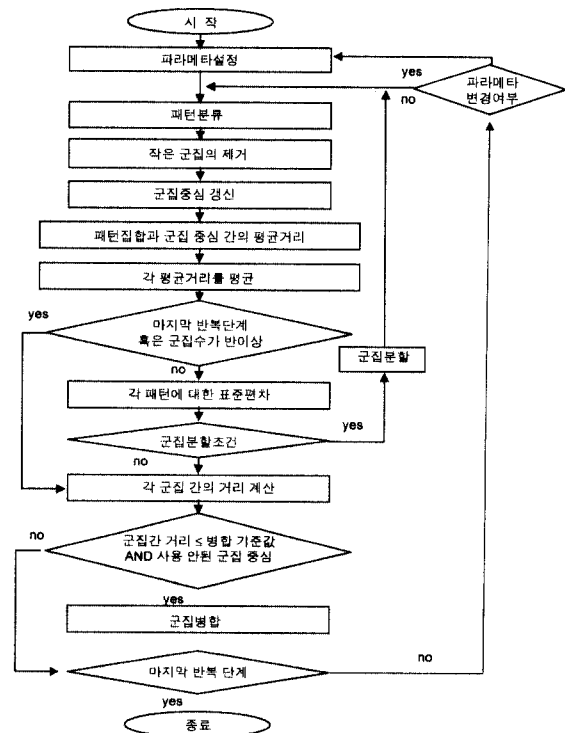
위성영상의 감독 분류는 특정 지역을 분류하기 위한 사전 학습 데이터가 필요하다. 학습 과정의 전반적인 목표는 특정 영상에서 구분되는 모든 토지피복 종류에 대한 분광 반응패턴을 설명할 수 있는 통계집단을 수집하는 것이다. 좋은 분류결과를 산출하기 위해, 학습 데이터는 대표성과 완벽성을 가져야 한다. 이 말은 영상의 각각 정보항목을 구성하고 있는 모든 분류항목들의 학습영역에 대한 통계정보를 영상분석자가 정확하게 추출하여야 한다는 의미이다.

본 연구에서는 분류 성능의 향상을 위한 학습 데이터의 선정에 위하여 분석자에 의하여 선정된 분류항목별 학습 데이터를 분광특징별로 다수의 데이터로 분리하고, 각각의 군집을 새로운 분류항목별 새로운 훈련데이터로 설정하여 분류를 수행하였다. 학습 데이터의 선정 과정은 분석자가 먼저 분류항목별로 입력영상에 대하여 학습 데이터로서 여러 영역을 선정한다. 다음으로 분류항목별로 각각 선정된 학습 데이터에 대하여 ISODATA 클러스터링을 이용하여 분광특징에 따라 학습 데이터를 세분화하여 새로운 학습 데이터를 선정하였다. (그림 1)은 본 연구의 분류항목별 학습 데이터 선정과정을 나타낸다.



(그림 1) 학습 데이터의 선정과정

ISODATA 클러스터링은 통계치를 계산하고 분류하는 과정을 반복적으로 수행한다. 매 반복 단계마다 표본의 평균을 군집의 중심으로 정하며, 군집의 삭제, 분리, 병합을 통해 자기 조직화가 가능하며, 고정된 수의 군집들을 처리하는 것이 아니라, 최종 군집의 개수는 사용자의 요구 군집 개수에 의존하지 않고 적당한 군집의 개수를 생성한다는 특징을 가지고 있다[14, 15]. ISODATA 클러스터링 알고리즘은 (그림 2)와 같다.



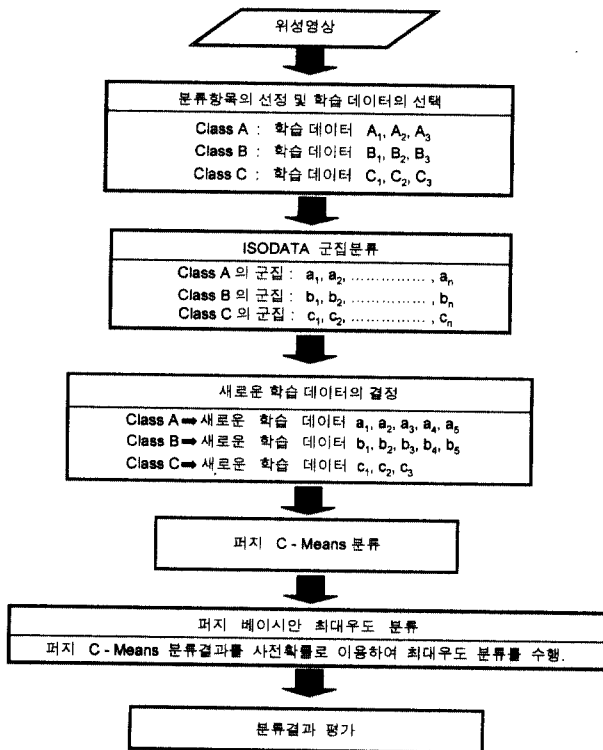
(그림 2) ISODATA 클러스터링 알고리즘

ISODATA 클러스터링은 각 군집마다 허용되는 샘플들

의 최소 개수, 병합이 일어나지 않는 군집중심 사이에 허용되는 최소 거리, 군집의 분리를 조절하는 파라미터, 각 반복에서의 군집 병합의 최대 수, 알고리즘 반복의 최대 수를 입력 값으로 한다. 그리고 본 연구에서는 ISODATA 클러스터링 알고리즘에 있어서의 픽셀과 군집 중심과의 거리는 마하라노비스 거리를 사용하였다. 마하라노비스 거리는 분산의 차이에 대해 각 축간 모집단의 분포상관을 고려한 보정을 행하고 있는 거리이다. 본 연구에서는 ISODATA 클러스터링을 적용하여 새롭게 선정된 학습 데이터를 이용하여 분류를 수행함으로써 분석자가 선정한 학습 데이터의 분광적인 분포형태에 관계없이 분류를 수행할 수 있다.

3.2 퍼지 베이시안 분류기법

퍼지 C-Means 알고리즘을 이용한 베이시안 최대우도 분류기법의 과정은 먼저 ISODATA 클러스터링에 의해 새롭게 선정된 학습 데이터로부터 각 군집의 평균값을 구하여 퍼지 C-Means의 중심값으로 설정하고 각 중심값에 대한 소속도를 구하여 분류를 수행하였다. 퍼지 C-Means의 분류결과를 베이시안 최대우도 분류기의 사전확률로 이용하여 분류를 수행하였다. (그림 3)은 본 연구에서 제안한 분류과정을 나타낸 것이다.



(그림 3) 퍼지 C-Means를 이용한 최대우도 분류

3.2.1 퍼지 C-Means 알고리즘

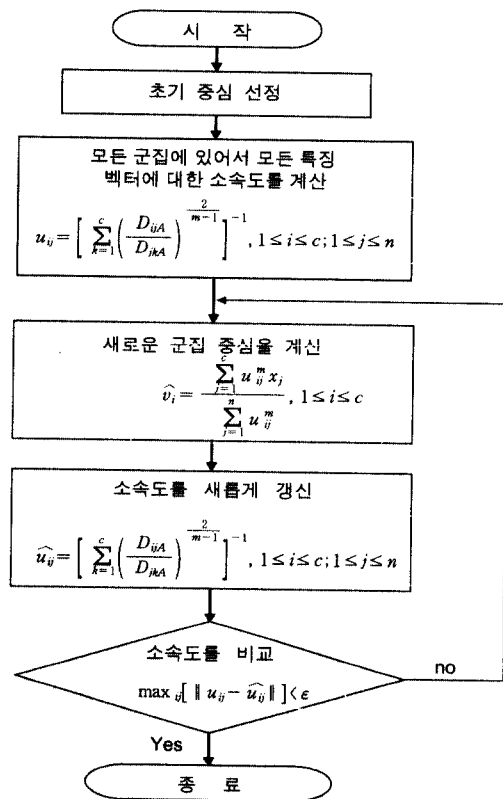
Bezdek에 의해 제안된 퍼지 C-Means 알고리즘은 목적

함수기반(objective function based) 퍼지 규칙을 사용한다 [16-17]. 퍼지 C-Means 알고리즘의 목적은 주어진 데이터 집합을 주어진 수의 퍼지 클러스터로 분할하는 것이다. 주어진 데이터 집합 $X = x_1, \dots, x_n$ 와 분할하고자 하는 군집 수 c 에 대해서 아래의 목적함수 J_m 을 최소화함으로써 퍼지 분할 $F = F_1, F_2, \dots, F_c$ 를 구하는 것이다.

$$J_m(U, V; X) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2 \quad (1)$$

여기서 u_{ij} 는 퍼지 클러스터 F_i 에 대한 x_j 의 소속정도를 나타내며, $(c \times n)$ 패턴행렬 $U = [u_{ij}]$ 의 원소가 된다. $V = v_1, \dots, v_n$ 는 퍼지 클러스터들의 중심 벡터의 집합이다. 퍼지 C-Means 군집분류 알고리즘은 각 화소들이 군집에 소속될 소속도와 군집중심을 생성한다. 따라서 위의 목적함수 J_m 의 값을 최소화시키는 u_{ij} 와 v_i 를 구하는 것이 목표이다. 여기에서 m 은 퍼지 식별자라고 하고 만약 $m=1$ 인 경우는 HCM이며, 퍼지 C-Means 알고리즘에서는 $1 < m < \infty$ 의 값을 사용하며 일반적으로 $m=2$ 값을 사용한다.

$$M_{fcm} = \{ U \in R^{cn} | u_{ij} \in [0, 1] \forall i, j; \langle \sum_{j=1}^n u_{ij} < n \forall i, \text{ and } \sum_{i=1}^c u_{ij} = 1 \forall j \} \quad (2)$$



(그림 4) 퍼지 C-Means 알고리즘

퍼지 C-Means 알고리즘에서는 목적함수 J_m 의 값을 최소화시키는 u_{ij} 와 v_i 를 다음과 같이 계산한다.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{D_{iIA}}{D_{jKA}} \right)^{\frac{2}{m-1}} \right]^{-1}, 1 \leq i \leq c; 1 \leq j \leq n, \quad (3)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, 1 \leq i \leq c \quad (4)$$

$$D_{ij} = \|x_j - v_i\|_A > 0 \forall i, j \quad (5)$$

본 논문에서는 군집의 중심 v_i 와 특징 화소와의 거리 D_{ij} 값을 마하라노비스 거리값을 이용하였다. (그림 4)는 퍼지 C-Means 알고리즘의 수행과정을 나타낸다.

퍼지 C-Means는 반복과정에 의한 중심값의 변화에 대한 소속도를 구하는 것이 일반적이는데, 본 연구에서는 중심값 v_i 를 새롭게 선정된 분류항목에 대한 학습 데이터의 각 평균값으로 설정하여, 반복횟수를 여러 번 수행하여 분류를 하는 것이 아니라 한번으로 소속도 u_{ij} 를 구하여 분류를 수행함으로써 계산 과정을 빠르게 하였다. 이렇게 구해진 퍼지 C-Means 알고리즘의 분류결과에서 각 분류항목별 분포 확률을 베이지안 최대우도 분류기의 사전확률로 이용되어 진다.

3.2.2 베이지안 최대우도 분류기법

베이지안 최대우도 분류기법은 각 분류항목에 대한 화소 자료의 우도를 구하고 최대우도 분류항목에 화소를 분류하는 방법으로서 베이즈의 이론에 기초를 두고 있다. 베이즈의 정리(Bayes' theorem)는 Thomas Bayes가 발표한 이론으로서 어떤 실험결과에서 나온 정보를 이용하여 어떤 사건의 처음 확률을 개선시킬 수 있는데, 여기서 처음 확률은 사전확률이라 하고, 개선된 확률을 사후확률이라고 한다. 베이즈의 정리를 이용한 의사 결정은 특징값이 주어진 상태에서 가장 가능성이 있는 분류항목을 선택하는 것을 말한다. 어떤 화소에 대해 분류 결정을 하기 위해서는 화소가 특성치 x 를 가질 때 w_i 에 속할 확률인 사후확률을 알아야 하며 이 확률 $P(w_i|x)$ 을 우도라고 하고 베이즈 정리에서 유도된다[5, 6, 18].

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)} \quad (6)$$

$P(w_i)$: 영상에서 분류항목 w_i 가 발생할 사전확률

$P(x|w_i)$: 학습 데이터로부터 계산될 확률(분류항목 w_i 로부터 x 가 관측되는 조건부 확률)

$P(x)$: 위치 x 에서 어떤 분류항목으로부터 픽셀을 찾을 확률

$$\Rightarrow P(x) = \sum_i P(x|w_i)P(w_i)$$

$P(x|w_i)P(w_i)$: 어떤 화소가 특성치 x 를 가지며 동시에 w_i 에 속할 확률

분류의 결정규칙은 사후확률에서 비롯된다. 즉 화소가 특성치 x 를 가질 때, 화소의 분류항목에 대한 결정방법은 다음과 같이 이루어진다.

$$x \in w_i, \text{ if } P(w_i|x) > P(w_j|x) \text{ for all } j \neq i \quad (7)$$

계산식의 간편화를 위해 양변에 로그를 취하면 다음과 같은 판별함수를 얻을 수 있다.

$$D_i(x) = \ln\{P(w_i|x)P(w_i)\} = \ln p(x|w_i) + \ln P(w_i) \quad (8)$$

따라서 다음과 같이 판별함수에 의해 화소가 속하는 분류항목을 결정할 수 있다.

$$x \in w_i, \text{ if } D_i(x) > D_j(x) \text{ for all } j \neq i \quad (9)$$

지금까지 각 분류항목에 대한 화소자료의 우도를 구하고 최대우도 분류항목에 그 화소를 분류하는 방법에 대해서 설명하였다. 이를 확장하여 하나의 분류항목에 대한 확률함수가 정규밀도 함수에 근사 한다고 가정하고서, 학습 데이터로부터 취득된 통계 값들을 이용하여 영상의 화소들을 특정한 항목들로 분류할 수 있다. 확률밀도함수로서 다차원 정규분포를 가정하고, 분류항목의 평균 벡터나 공분산 행렬은 미지수로서 학습 데이터로부터 추정한다. 학습 데이터로부터 추정되는 N밴드에 대한 확률인 우도 $P(x|w_i)$ 는 다음과 같다.

$$P(X|w_i) = \frac{1}{(2\pi)^{N/2} |\sum i|^{1/2}} \times \exp\left[-\frac{1}{2}(X-U_i)^T \sum i^{-1}(X-U_i)\right] \quad (10)$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad U_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{iN} \end{bmatrix} \quad \sum i = \begin{bmatrix} \sigma_{i11} & \sigma_{i12} & \cdots & \sigma_{i1N} \\ \sigma_{i21} & \sigma_{i22} & \cdots & \sigma_{i2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{iN1} & \sigma_{iN2} & \cdots & \sigma_{iNN} \end{bmatrix}$$

N : 밴드 수

X : 데이터 벡터

U_i : 분류항목 i 에 대한 N 개 밴드의 평균벡터

$\sum i$: 분류항목 i 의 공분산 행렬

C_{kl} : 두 개의 밴드들 사이의 공분산(k 와 l 밴드)

$$C_{kl} = \frac{\left(\sum_j (x_{kj} - u_k)(x_{lj} - u_l) \right)}{n-1}$$

밴드 수 : $k = 1, 2, \dots, N$

밴드 수 : $l = 1, 2, \dots, N$

픽셀 값 : $i = 1, 2, \dots, N$

$|\sum_i i|$: 공분산 행렬 \sum_i 의 determinant

$\sum_i i^{-1}$: \sum_i 의 역행렬

$(X - U_i)^T$: 벡터 $(X - U_i)$ 의 전치행렬 $l = 1, 2, \dots, N$

어떤 화소가 특성치 x 를 가지며 동시에 w_i 에 속할 확률 $P(X|w_i)P(w_i)$ 를 구하면 다음과 같다.

$$P(X|w_i)P(w_i) = \frac{P(w_i)}{(2\pi)^{N/2} |\sum_i i|^{1/2}} \times \exp\left[-\frac{1}{2}(X - U_i)^T \sum_i i^{-1}(X - U_i)\right] \quad (11)$$

계산식의 간편화를 위해 양변에 로그를 취하여 다음과 같은 판별함수를 얻을 수 있다.

$$D_i(X) = \ln P(w_i) - \frac{N}{2} \ln 2\pi - \frac{1}{2} \ln |\sum_i i| - \frac{1}{2}(X - U_i)^T \sum_i i^{-1}(X - U_i) \quad (12)$$

위 식은 다음과 같이 나타낼 수 있다.

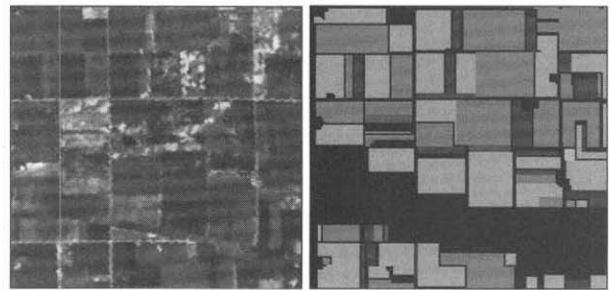
$$D_i(X) = \ln P(w_i) - \frac{N}{2} \ln 2\pi - \frac{1}{2} \begin{bmatrix} c_{i11} & c_{i12} & \dots & c_{i1N} \\ c_{i21} & c_{i22} & \dots & c_{i2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{iM1} & c_{iM2} & \dots & c_{iMN} \end{bmatrix} A \quad (13)$$

$$A = \frac{1}{2} \left[\begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{bmatrix} - \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{Ni} \end{bmatrix} \right]^T \times \begin{bmatrix} c_{i11} & c_{i12} & \dots & c_{i1N} \\ c_{i21} & c_{i22} & \dots & c_{i2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{iM1} & c_{iM2} & \dots & c_{iMN} \end{bmatrix}^{-1} \times \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{bmatrix} - \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{Ni} \end{bmatrix}$$

일반적으로 사전확률을 가지는 베이시안 최대우도 분류기법은 다변량 정규확률밀도함수에 있어서의 판별식 $D_i(X)$ 의 값의 값이 최대가 되는 분류항목으로 화소를 분류하는 과정이다. 이때 $P(w_i)$ 는 c 개의 분류항목 중 i 분류항목이 발생할 사전확률을 의미한다. 본 연구에서는 퍼지 C-Means 분류결과를 사전확률로 이용한다. 최대우도 분류방법은 가장 많은 변수를 고려하기 때문에 정확도가 높은 분류방법으로 알려져 있다. 그리고 공분산 행렬로서 분류항목의 가변성을 충분히 고려하였다. 이러한 최대우도 분류기법에 영상의 사전 분포상황을 고려한 사전확률로서 퍼지 C-Means 분류결과를 이용함으로써 분류의 평균확률 오차를 최소화함으로써 분류의 정확도를 향상시킬 수 있다.

4. 실험 및 결과분석

본 논문에서는 $30m \times 30m$ 의 공간 해상도를 가지는 Landsat TM 위성영상을 입력으로 하여 제안한 하이브리드 분류기법의 정확도를 실험하였다. 제안된 알고리즘의 정확성을 테스트하기 위해서 사용된 이미지는 미국의 Purdue 대학에서 수집한 데이터로서 1986년에 관측된 미국의 Indiana의 Tippecanoe County 지역에 대한 Landsat TM영상이다 [19]. 본 영상을 이용하게 된 이유는 알고리즘의 정확성의 평가를 위해서 실측된 데이터(ground truth data)가 필요하기 때문에 실측 데이터(ground truth data)가 제시되어 있는 본 영상을 사용하였다. 영상의 크기는 가로와 세로의 크기가 169×169 화소이고, 밴드 수는 7개이며, 파일형식은 BIL 포맷이다. 하나의 화소는 8bit로 구성되어 있다. (그림 5)는 7개의 밴드 중 4, 3, 2밴드를 선택하여 RGB로 표현한 위성영상과 그 지역에 대한 실측데이터이다.



(그림 5) 4, 3, 2밴드를 RGB로 표현한 영상과 실측데이터

Purdue 대학에서 수집한 (그림 5)의 위성영상에 나타난 지역에 대한 실측데이터를 바탕으로 corn, soybean, wheat, alfalfa/oats, pasture의 5개의 분류항목을 정의하고, 각 분류항목별로 학습 데이터를 선정하였다. 각 분류항목별로 선정된 학습 데이터에서 평균값, 평균+표준편차, 평균-표준편차 값을 초기 군집의 중심값으로 설정하고, 분리계수는 10, 병합계수는 5로 설정하여 ISODATA 클러스터링을 수행하여 분광특징별로 다수의 군집으로 분리하여 새롭게 학습 데이터를 생성하였다. 이때 전체 분류항목에 대하여 총 8개의 학습 데이터를 새롭게 생성하였다.

<표 1>은 분석자에 의해 선정된 학습 데이터의 밴드별 통계치를 나타낸다. <표 2>는 ISODATA 클러스터링을 수행한 후의 밴드별 통계치를 나타낸 것이다. ISODATA 클러스터링에 의하여 분광특징별로 다수의 군집으로 분리하여 선정한 학습 데이터는 비교적 작은 값의 밴드별 분산값을 가지며, 정규분포형태의 분포를 보였다.

<표 3>은 ISODATA 클러스터링을 수행한 후 그 결과를 퍼지 C-Means 알고리즘에 의하여 각 분류항목별로 분류된 결과를 표로 나타낸 것이다. 전체 정확도가 68.8%를 나타내고 있다.

<표 1> 분류항목별 학습데이터에 대한 밴드별 통계치

밴드	분류항목									
	corn		soybean		wheat		alfalfa/oats		pasture	
	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
1	85.3	3.1	90.4	7.5	96.7	5.4	93.0	7.0	88.7	2.4
2	32.4	2.2	37.5	4.6	43.2	3.8	40.2	3.8	38.2	1.8
3	28.9	3.6	35.6	9.4	46.0	7.5	39.9	9.3	35.2	2.0
4	121.3	13.4	114.8	25.9	102.9	12.1	103.6	25.5	116.6	9.1
5	80.4	13.3	122.4	18.3	127.3	17.3	112.5	21.9	101.4	5.5
6	24.3	9.0	48.0	16.2	49.3	10.9	45.7	17.4	32.0	4.1
7	147.9	1.9	150.2	3.2	152.1	2.1	152.1	3.1	151.3	0.9

<표 2> ISODATA 클러스터링을 수행한 후의 각 학습 데이터에 대한 밴드별 통계치

class	corn		soybean				wheat		alfalfa/oats				pasture			
	A		B		A		B		A		B		A			
	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차		
1	85.005	2.287	100.496	7.712	87.310	3.502	99.639	7.300	97.747	4.480	92.904	5.547	97.146	11.376	88.544	1.676
2	32.154	1.621	42.147	4.251	35.610	2.097	43.029	4.635	43571	2.691	39.819	3.595	42.286	5.388	38.561	1.475
3	28.522	2.334	46.112	7.601	31.636	4.045	46.849	9.390	47.140	5.263	40.523	8.240	41.896	9.908	35.333	1.668
4	120.686	11.827	99.895	16.972	119.921	24.351	102.399	17.977	102.601	10.663	99.515	18.873	117.475	26.120	117.211	6.902
5	78.603	7.669	122.818	16.950	114.766	9.706	147.124	14.617	128.182	11.798	109.122	21.157	122.921	13.306	102.105	5.540
6	23.090	4.961	56.772	14.851	40.762	6.832	70.962	13.577	50.163	7.798	44.815	15.300	49.647	11.946	31.456	2.616
7	147.888	1.624	151.249	2.049	149.257	2.215	154.022	3.129	152.075	2.045	152.807	2.298	153.406	3.578	150.930	0.896

<표 3> 퍼지 C-Means 분류결과에 대한 confusion matrix

분류항목	Fuzzy C-Means					Total
	corn	soybean	wheat	alfalfa/oats	pasture	
corn	6604	847	25	336	15	7,827
soybean	515	5242	55	646	2	6,460
wheat	65	228	134	75	0	502
alfalfa/oats	453	1080	23	450	3	2,009
pasture	738	282	14	249	22	1,305
total	8375	7679	251	1756	42	18,103

<표 4>는 퍼지 C-Means를 수행한 결과에 의한 분류항목별로 주어진 사전확률값을 나타낸다. corn 지역이 가장 넓게 분포함을 알 수 있다.

<표 4> 퍼지 C-Means 결과에 의한 분류항목별 사전확률값

분류항목	corn	soybean	wheat	alfalfa/oats	pasture
사전확률	0.432	0.357	0.028	0.111	0.072

<표 5>는 일반적인 최대우도 방법에 의한 분류를 수행한 결과이다. <표 6>은 본 연구에서 제안한 방법에 의하여

퍼지 베이시안 최대우도분류를 수행한 결과이다.

<표 5> 일반적인 최대우도분류 방법에 대한 confusion matrix

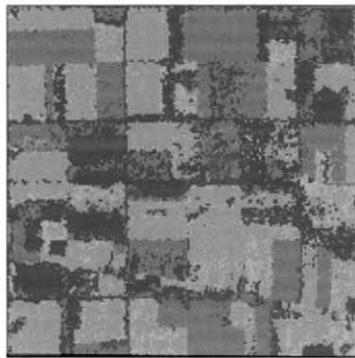
분류항목	Maximum likelihood					Total
	corn	soybean	wheat	alfalfa/oats	pasture	
corn	6420	670	12	154	1	7,257
soybean	355	4718	2	121	0	5,196
wheat	155	611	205	136	1	1,108
alfalfa/oats	645	1594	22	1244	0	3,505
pasture	800	86	10	101	40	1,037
total	8375	7679	251	1756	42	18,103

<표 6> 퍼지 베이시안 분류에 대한 confusion matrix

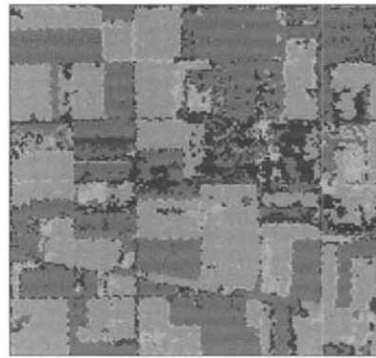
분류항목	Fuzzy Bayesian maximum likelihood					Total
	corn	soybean	wheat	alfalfa/oats	pasture	
corn	7211	696	21	251	14	8,193
soybean	854	6465	34	499	4	7,856
wheat	23	129	163	35	1	351
alfalfa/oats	285	386	31	966	1	1,669
pasture	2	3	2	5	22	34
total	8,375	7,679	251	1,756	42	18,103

〈표 7〉 분류 정확도 평가

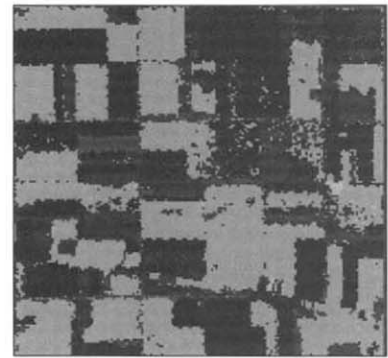
종 류 분류항목	Maximum likelihood				Fuzzy Bayesian maximum likelihood			
	Producer's accuracy	User's accuracy	Overall accuracy	Kappa 계수	Producer's accuracy	User's accuracy	Overall accuracy	Kappa 계수
corn	76.7%	88.4%	70%	55%	86%	88%	82%	70%
soybean	61.4%	91%			84%	82%		
wheat	81.7%	18.5%			65%	46.4%		
alfalfa/oats	71%	35.5%			55%	57.9%		
pasture	95%	3.8%			52%	64.7%		



(a) 일반적인 MLC 분류결과



(b) 퍼지 C-Means 분류결과



(c) 퍼지-베이시안 최대우도 분류결과

(그림 6) 알고리즘에 따른 분류결과

〈표 7〉의 분류 정확도 평가를 살펴보면 기존의 일반적인 최대우도분류를 수행한 결과는 분류의 전체 정확도가 70% 이었으며, 본 연구에서의 퍼지 베이시안 최대우도분류는 〈표 1〉에 의하여 전체 정확도가 82%로 성능이 향상되었음을 보여 주었다. Kappa 계수는 옳게 분류된 화소에서 우연히 옳게 분류될 수 있는 화소를 제거한, 순수하게 옳게 분류된 화소만을 고려하여 정확도 평가를 하려는 의도에서 도입된 통계량으로서 최대우도분류가 55%, 퍼지 베이시안 최대우도분류가 70%로서 더 좋은 효과를 나타내고 있다. (그림 6)은 각 방법에 의하여 분류가 수행된 결과를 보여주고 있다.

실험 결과 분광특징에 따른 학습 데이터의 분할에 의한 분류를 수행한 후에 그 결과를 다시 분석자가 초기에 설정한 분류항목으로 재설정함으로써 분석자가 처음 의도했던 분류 결과를 얻을 수 있었다. 그리고 각 클래스의 초기 학습데이터를 선정할 때 분광특징이 다양하게 나타나는 지역을 선정했을 경우에 세부항목으로 분리되는 클러스터의 수가 많아져서 분류에 시간이 좀 더 걸리지만, 분광특징이 다양하고 분산값이 크게 나타나는 분포가 넓은 분류항목의

분류에 좋은 효과를 나타내었다.

5. 결 론

본 논문에서는 위성영상의 분류에 대한 성능 개선을 위하여 ISODATA 클러스터링, 퍼지 C-Means 알고리즘, 베이시안 최대우도 분류기법을 이용하여 감독 분류를 기반으로 부분적으로 무감독 분류기법을 적용한 하이브리드 분류기법을 제안하였다. 본 연구에서는 위성영상 이미지의 분류에 있어, 분석자에 의하여 분류항목별 학습 데이터를 선정한 후 이를 ISODATA 클러스터링을 이용하여 각각의 분류항목별로 분광특징에 따라 학습 데이터를 세분화하여 새로운 학습 데이터를 선정하였다. 이렇게 함으로서 새롭게 생성된 군집은 비교적 적은 값의 밴드별 분산값을 가지며, 정규분포 형태의 분포를 보여주어 분류의 적용에 적절한 자료가 되며, 분석자가 선정한 분류항목별 학습 데이터의 분광적인 특징에 관계없이 분류를 수행할 수 있다. 분류항목별로 세분화된 학습 데이터로부터 평균값을 구하여 퍼지 C-Means 알고리즘의 중심값으로 설정하고 각 중심값에 대

한 소속도를 구하여 분류를 수행하였다. 퍼지 C-Means의 분류 결과를 베이지안 최대우도 분류기의 사전확률로 이용하여 분류를 수행하였다. 제안된 기법은 Landsat TM 위성 영상을 이용하여 그 적용성을 시험하였으며, 그 결과 기존의 분류방법보다 위성영상의 분류의 성능을 개선할 수 있었다.

향후 연구과제는 클러스터링 알고리즘에서 몇 가지 변수들의 입력이 요구되는데, 최적의 변수를 효과적으로 선정하는 방법과 퍼지 알고리즘의 최적 클러스터 수를 결정하는 방법에 대한 연구가 뒤따라야 하겠다.

참 고 문 헌

- [1] John A. Richards, 'Remote Sensing Digital Image Analysis : An Introduction,' Second, Revised and Enlarged Edition, Springer-Verlag, pp.229-262, 1994.
- [2] The Canada Centre for Remote Sensing, <http://www.ccrs.nrcan.gc.ca>
- [3] R. Schowengerdt, 'Techniques of Image Processing and Classification in Remote Sensing,' 1st Ed., Academic Press, pp.1-58, 1983.
- [4] David Landgrebe, "Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data," Chapter 1 of Information Processing for Remote Sensing, edited by C. H. Chen, published by the World Scientific Publishing Co., Inc., pp.1-30, Spring, 1999.
- [5] B.Gorte and A. Stein. "Bayesian classification and class area estimation of satellite images using stratification," IEEE Trans, On Geoscience and Remote Sensing, 36(3), p.303, 1998.
- [6] Amal S. Perera, Masum H. Serazi, William Perrizo, "Performance Improvement for Bayesian Classification on Spatial Data with P-Trees," 15th International Conference on Computer Applications in Industry and Engineering(CAINE 2002), Nov., 2002.
- [7] Melgani, F., Hashemy B. A. R. and Taha S. M. R. "An explicit fuzzy supervised classification method for multispectral remote sensing images," Geoscience and Remote Sensing, IEEE Transactions on, Vol.38, Issue 1 Part 1, pp.287-295, 2000.
- [8] Nakashiim T., Nakai G. and Ishibuchi, H., "Improving the performance of fuzzy classification systems by membership function learning and feature selection," Fuzzy Systems, FUZZ-IEEE '02. Proceedings of the 2002 IEEE International Conference on, Vol.1, pp. 488-493, 2002.
- [9] Mehmet I Saglam, Bingul Yazgan, Okan K Ersoy, "Classification of Satellite Images by using Self-organizing map and Linear Support Vector Machine Decision tree," GISdevelopment Conference Proceedings of Map Asia 2003.
- [10] Zhacong Wu, "Research on remote sensing image classification using neural network based on rough sets," Info-tech and Info-net, 2001, Proceedings. ICII 2001-Beijing. 2001 International Conferences on, Vol.1, pp.279-28429, Nov., 2001.
- [11] Heermann, P. D. and Khazenie, N., "Classification of multispectral remote sensing data using a back-propagation neural network," IEEE Trans. on Geosci. and Remote Sensing, Vol.30, No.1, pp.81-88, Jan., 1992.
- [12] Nishii, R., "A Markov random field-based approach to decision-level fusion for remote sensing image classification," Geoscience and Remote Sensing, IEEE Transactions on, Vol.41, Issue 10, pp.2316-2319, Oct., 2003.
- [13] Yuyu Zhou, Hong Chen, Qijiang Zhu, "The research of classification algorithm based on fuzzy clustering and neural network," Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, Vol.4, pp.2525-2527, June, 2002.
- [14] F. R. D. Velasco, "Thresholding using the isodata clustering algorithm," IEEE Trans. Syst., Man, Cybern., Vol.SMC-10, pp.771-774, 1981.
- [15] ISOCLUS, Isodata Clustering Program, http://www.pcigeomatics.com/cgi-bin/pcihlp/m_isoclus/.
- [16] James C. Bezdek, 'Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NY, 1981.
- [17] N. K. Pal and J. C. Bezdek, "On Cluster Validity for the Fuzzy C-Means Model," IEEE Trans. Fuzzy Syst., Vol.3, No.3, pp.370-379, 1995.
- [18] Sun, W., "A New Information Fusion Method for Land-Use Classification Using High Resolution Satellite Imagery," Ph.D. Dissertation, <http://archimed.uni-mainz.de/pub/2000/0004>, 2000.
- [19] Purdue/Laboratory for Applications of Remote Sensing <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/>.



전 영 준

e-mail : j4017@chol.com

1992년~1998년 동의대학교 전산통계학과
학사(이학사)

1998년~2000년 동의대학교 전산통계학과
석사(이학석사)

2000년~2003년 동의대학교 컴퓨터공학과
박사과정 수료

관심분야 : 패턴인식, 위성영상응용, 인공지능 및 컴퓨터응용분야



김 진 일

e-mail : jikim@dongeui.ac.kr

1976년~1982년 경희대학교 전자공학과
(학사, 공학석사)

1994년 서강대학교 전자계산학과(공학박사)

1982년~1984년 미국 Bon Scours 시스템,
맥도널다글라스 우주항공회사 등
연구원

1996년~1997년 미국 Purdue Univ. 전기 및 컴퓨터공학부
연구교수

1998년~2001년 위성탐사 영상처리 연구회(한국과학재단) 회장

1988년~현재 동의대학교 컴퓨터공학과 교수

관심분야 : 패턴인식, 퍼지논리, 위성영상응용, GIS/GPS/RS, 인
공지능 및 컴퓨터응용분야