

퍼지 알고리즘의 융합에 의한 다중분광 영상의 패턴분류

(Pattern Classification of Multi-Spectral Satellite Images based on Fusion of Fuzzy Algorithms)

전 영 준 [†] 김 진 일 ^{**}
(Young Joon Jeon) (Jin Il Kim)

요 약 본 논문에서는 다중분광 영상의 분류를 위하여 퍼지 G-K(Gustafson- Kessel) 알고리즘과 PCM 알고리즘을 융합한 분류방법을 제안하였다. 제안된 방법은 학습데이터를 이용하여 퍼지 G-K 알고리즘을 수행한 후 그 결과를 이용하여 PCM 알고리즘을 수행한다. PCM 알고리즘과 퍼지 G-K 알고리즘 분류결과를 비교하여 그 결과가 일치하면 해당 항목으로 분류항목을 결정한다. 일치하지 않는 화소는 PCM 알고리즘의 평균내부거리 안쪽에 있는 화소들을 새로운 학습데이터로 하여 베이시안 최대우도 분류를 수행하여 분류항목을 결정한다. 평균내부거리 안쪽에 있는 화소 데이터는 정규분포형태를 보여준다.

다차원 다중분광 영상인 IKONOS와 LANDSAT TM 위성영상을 이용하여 제안된 알고리즘의 효율성을 검증한 결과 퍼지 G-K 알고리즘과 PCM 알고리즘 그리고 전통적인 분류 방법인 최대우도 분류 알고리즘보다 전체 정확도가 더 높은 결과를 얻을 수 있었다.

키워드 : 퍼지 G-K 알고리즘, PCM 알고리즘, 베이시안 최대우도 분류, 원격탐사 위성영상

Abstract This paper proposes classification of multi-spectral satellite image based on fusion of fuzzy G-K (Gustafson-Kessel) algorithm and PCM algorithm. The suggested algorithm establishes the initial cluster centers by selecting training data from each category, and then executes the fuzzy G-K algorithm. PCM algorithm perform using classification result of the fuzzy G-K algorithm. The classification categories are allocated to the corresponding category when the results of classification by fuzzy G-K algorithm and PCM algorithm belong to the same category. If the classification result of two algorithms belongs to the different category, the pixels are allocated by Bayesian maximum likelihood algorithm. Bayesian maximum likelihood algorithm uses the data from the interior of the average intracluster distance. The information of the pixels within the average intracluster distance has a positive normal distribution. It improves classification result by giving a positive effect in Bayesian maximum likelihood algorithm.

The proposed method is applied to IKONOS and Landsat TM remote sensing satellite image for the test. As a result, the overall accuracy showed a better outcome than individual Fuzzy G-K algorithm and PCM algorithm or the conventional maximum likelihood classification algorithm.

Key words : fuzzy Gustafson-Kessel algorithm, PCM algorithm, Bayesian maximum likelihood algorithm, remote sensing satellite image

1. 서 론

원격탐사 기술은 사진 기술, 우주과학 기술, 정보통신

기술, 컴퓨터 기술 등 첨단과학 기술의 빠른 발전의 결과와 더불어 비약적인 성장을 거듭하고 있다. 인공위성 영상자료를 통해 지도 제작, 자원 탐사, 기상 관측, 환경 감시, 재해 방제 등으로부터 군사적인 관측에 이르기까지 다양한 작업이 수행되어지고 있다. 원격탐사 위성영상 자료에 나타나는 어떤 대상물의 밝기나 색은 그것이 반사하는 파장에 따라 독특한 특성을 나타내게 되며, 이를 수학적으로 분석하고 정리함으로써 그 대상물을 알

본 논문은 2004학년도 동의대학교 연구년 지원에 의해서 수행되었습니다.

[†] 학생회원 : 동의대학교 컴퓨터공학과
j4017@chol.com

^{**} 종신회원 : 동의대학교 컴퓨터공학과 교수
jikim@deu.ac.kr

논문접수 : 2004년 11월 10일
심사완료 : 2005년 5월 23일

아내고자 하는 것이 바로 원격탐사의 목적이라 할 수 있다. 인공위성으로부터 얻어진 화상으로부터 지표면의 특징을 분류하여 주제도를 작성하는 것은 원격탐사 분야에서 중요하게 다루어지는 분야이다.

위성영상의 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 작업자가 지정한 일정한 기준에 따라 하나의 클래스 또는 값을 갖게 된다. 한 분류항목 내의 화소는 다른 분류항목 내의 화소보다는 서로 더 큰 유사성을 지니지만, 동일 집단 내의 화소라 할지라도 어느 정도 편차는 가지고 있다[1]. 원격탐사자료를 이용하여 영상을 추출하고 분류하는 대표적인 방법으로는 평행육면체 분류기법(Parallelepiped Classification), 최단거리 분류기법(Minimum Distance Classification), 최대우도 분류기법(Maximum Likelihood Classification)과 같은 감독분류(supervised classification)와 K-means 방법, ISO-DATA(Iterative Self Organizing Analysis Technique) 방법 등의 무감독분류(unsupervised classification)를 이용한 방법이 있고[2-4], 기타 퍼지로직을 이용한 분류방법[5,6], 신경망의 응용[7,8], 여러 방법들을 통합한 혼성방법[9,10] 등이 있다.

본 연구에서는 Bezdek이 제안한 목적함수기반(objective function based) 퍼지 규칙을 사용한 FCM(Fuzzy C-Means) 알고리즘을 확장한 형태인 퍼지 G-K 알고리즘(Fuzzy Gustafson-Kessel Algorithm)과 PCM(Possibilistic C-Means) 알고리즘을 통합하여 원격탐사 위성영상을 분류하는 알고리즘을 제안하였다. 데이터를 분류하는 기존의 클러스터링 방법으로는 하드 클러스터링, FCM 등의 방법이 있다. 하드 클러스터링 기법은 주어진 데이터 상호간의 관계가 명확하다는 가정 하에 각 패턴을 분할하는 방식이다. 따라서 이 방법은 다루고자 하는 데이터의 경계가 명확하지 않을 경우 실제 데이터 상호간의 군집성을 묘사하기에 부적절할 뿐만 아니라, 주어진 데이터 분포의 성질을 손실하는 결과를 초래할 수도 있다. Bezdek이 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[11,12]. 그러나 FCM 알고리즘은 데이터의 각 클러스터에 관한 소속함수 값이 1이 되는 확률적 제약 조건을 이용하므로 소속함수 값이 소속성(belonging)이나 적합성(compatibility) 정도의 직관적인 개념과 항상 일치하지 않는다. FCM의 경우에는 잡음이 있는 상황에서 심각한 문제점을 야기시킨다.

이는 FCM의 경우 소속함수의 값이 다른 클러스터와 관계가 있는 상대적인 값이기 때문이다. 이러한 문제점을 개선하기 위해 Raghu는 PCM 알고리즘을 제안하였다[13]. PCM의 경우는 소속함수 값이 다른 클러스터와 관계가 없고, 속성 데이터와 해당 클러스터 중심간의 거리에만 의존하기 때문에 이러한 문제점을 개선할 수 있다. 그러나 PCM 방법을 적용하기 위해서는 각 영역에 해당하는 클러스터들의 크기, 평균내부거리, 초기 중심값을 사전에 알아야 하는 단점이 있다. 따라서 본 논문에서는 다차원 다중분광 위성영상 데이터의 분류를 위해서 공분산 행렬을 적용한 퍼지 G-K 알고리즘[14,15]을 이용하여 먼저 입력 데이터를 분류하고, 분류된 결과로서 평균내부거리와 초기 클러스터의 중심값을 구하여 PCM 알고리즘을 수행한다. 두 알고리즘의 수행결과를 비교하여 동일 항목으로 분류된 화소에 대하여 먼저 분류항목을 할당하고, 다른 항목으로 분류된 화소는 베이시안 최대우도 분류기법을 이용하여 분류를 수행하여 분류항목을 할당하여 최종적으로 전체 영상을 분류한다.

본 논문의 구성은 2장에서는 관련연구로서 제안된 위성영상 분류기법에 사용되는 퍼지 G-K 알고리즘과 PCM 알고리즘에 대해서 살펴보고, 3장에서는 제안된 분류기법으로서 퍼지 G-K 알고리즘과 PCM 알고리즘을 결합한 분류기법에 대해서 설명하고, 4장에서는 IKONOS와 Landsat 위성영상을 이용하여 제안된 알고리즘을 검증하고 분류결과를 분석하고, 5장에서 결론을 맺는다.

2. 관련연구

2.1 퍼지 G-K 알고리즘

Fuzzy C-Means (FCM)이 가지는 문제점에 대하여 성능을 확장한 것이 G-K 알고리즘이다. 이 알고리즘은 데이터 집합에서 서로 다른 기하학적인 형태의 클러스터를 검출하기 위하여 유클리디언 거리측정과 함께 적응적인 거리 측정치를 적용한 표준 FCM의 확장 형태이다[14,15]. 각각의 클러스터는 다음과 같은 내적 놈(norm)을 가지는 유도된 행렬 A_i 를 가진다.

$$D_{ijA_i}^2 = (x_j - v_i)^T A_i (x_j - v_i) \quad (1)$$

여기서 행렬 A_i 는 c-means 함수에서 최적화 변수로 이용되며, 데이터의 위상학적 구조에서 각 클러스터에 대하여 거리 측정에 적용적이다. A_i 를 유도된 행렬 $A_i = (A_{i1}, A_{i2}, \dots, A_{in})$ 의 c 개의 조합으로 기술한다. 여기서 G-K 알고리즘의 목적함수는 다음과 같이 정의한다.

$$J(X; U, V, A) = \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m D_{ijA_i}^2 \quad (2)$$

데이터 집합 X 와 희망하는 클러스터의 수 c 가 주어졌을 때, $U = [\mu_{ij}] \in [0, 1]^{c \times n}$ 은 데이터 $X \in R^{c \times n}$ 의 퍼지 분할행렬이다. $V = [v_1, v_2, \dots, v_c], v_i \in R^n$ 은 클러스터의 중심행렬이고, $m \in [1, \infty)$ 는 가중지수이다. 고정된 A_i 에 대하여, 식 (3)과 (4)는 직접적으로 적용된다.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{D_{ijA_i}}{D_{kjA_i}} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq c, 1 \leq j \leq N \quad (3)$$

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

거리 놈(norm) $D_{ikA_i}^2$ 은 데이터 집합에서 서로 다른 기하학적인 형태의 클러스터들에 대하여 계산할 수 있다. 여기서 A_i 는 공분산 행렬 F_i 에 대한 놈-유도(norm-inducing) 행렬로서 식 (5)와 같이 계산된다.

$$A_i = \det(F_i)^{1/n} F_i^{-1} \quad (5)$$

여기서 F_i 는 i 번째 클러스터의 퍼지 공분산 행렬로 다음 식 (6)과 같이 정의된다.

$$F_i = \frac{\sum_{j=1}^N (\mu_{ij})^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^N (\mu_{ij})^m} \quad (6)$$

여기서 공분산은 U 에서 소속정도 정도에 의해 가중 처리된다. 퍼지 G-K 알고리즘은 식 (3), 식 (4)를 반복 수행하여 목적함수의 최소치를 만족하는 분포를 구할 수가 있다.

2.2 PCM 알고리즘

Bezdek이 제안한 FCM 알고리즘은 각 패턴이 특정 클러스터의 중심에 속하는 소속정도를 줌으로써 보다 정확한 정보를 획득하도록 하였다. 하지만 FCM 알고리즘은 확률적 제약조건(probabilistic constraint)을 이용하므로 소속함수 값이 소속성(belonging)이나 적합성(compatibility) 정도의 직관적인 개념과 항상 일치하지는 않는다. FCM의 경우에는 잡음이 있는 상황에서 심각한 문제점을 야기시킨다. 이는 FCM의 경우 소속함수 값이 다른 클러스터와 관계가 있는 상대적인 값이기 때문이다. 따라서 최근에 믿음 이론(belief theory)과 가능성 이론(possibility theory)등이 이와 같은 문제점을 개선하기 위해 연구되어져 왔다. Raghu는 PCM 알고리즘을 제안하였고, 이 방법은 소속정도의 값이 다른 클러스터와 관계가 없고, 각 패턴과 해당 클러스터 중심사이의 거리에만 의존하기 때문에 위의 문제점을 개선할 수 있었다[13]. 그러나 PCM 알고리즘은 각 영역에 해당하는 클러스터들의 크기와 초기 중심 값을 사전에 알아야 하는 단점이 있다.

PCM 알고리즘과 목적함수는 다음과 같이 정의되어진다.

$$J_m(U, V) = \sum_{j=1}^N \sum_{i=1}^c (u_{ij})^m d_{ij}^2 + \sum_{i=1}^c \eta_i \sum_{j=1}^N (1 - u_{ij})^m \quad (7)$$

목적함수 J_m 의 전역 최소치를 만족하는 U 는 식 (7)을 u_{ij} 에 대해 미분하고 미분치를 0으로 놓을 때 다음 식을 얻을 수 있다.

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i} \right)^{\frac{1}{m-1}}} \quad (8)$$

식 (8)에서, u_{ij} 의 값이 $[0, 1]$ 사이에 놓여 있다는 것을 알 수 있다. η_i 값은 클러스터에서 속성 데이터의 소속정도가 0.5가 되는 거리를 결정한다. 그러므로 이것은 각 클러스터의 가능성 분포의 적절한 대역폭에 의존하여 선택되어지는 것이 바람직하다. η_i 는 목적함수에서 두 번째 항이 첫 번째 항과 비교하여 얼마나 중요한가를 나타내는 상대적인 정도를 결정한다. 여기서 η_i 는 클러스터의 전체 모양과 크기에 관계되며 Raghu는 η_i 를 식 (9)와 같이 추정하였다.

$$\eta_i = K \frac{\sum_{j=1}^N u_{ij}^m d_{ij}^2}{\sum_{j=1}^N u_{ij}^m} \quad (9)$$

식 (9)에서 N 은 속성 데이터의 개수이고, $K=1$ 이다. η_i 는 소속함수에 의해 가중치 된 각 클러스터의 평균내부거리이다. PCM 알고리즘은 식 (4), 식 (8)을 반복 수행하여 목적함수의 최소치를 만족하는 가능성 분포를 구할 수가 있다.

3. 퍼지 알고리즘의 융합에 의한 패턴분류기법

본 논문에서는 퍼지 G-K 알고리즘과 PCM 알고리즘을 결합한 다중분광 영상인 원격탐사 위성영상에 대한 패턴분류방법을 제안하였다. 위성영상 데이터는 여러 개의 밴드를 가지고 있는 다차원의 데이터로서 각 군집별 데이터의 밀도가 틀리거나 하나의 군집에 데이터가 편중되는 경우가 많다. 즉, 클러스터의 다양한 모양이나 다른 크기를 가지고 있다. 일반적인 퍼지 클러스터링 알고리즘의 경우 정상적인 데이터 분포를 갖는 경우 클러스터 파라미터 추정에 좋은 성능을 가지는 반면, 데이터가 편중되거나 각 데이터의 밀도가 다른 경우 클러스터의 파라미터가 정확하게 클러스터를 표현하지 못하는 문제점이 있다. 이러한 문제점을 해결하기 위하여 퍼지 G-K 알고리즘을 이용하였다. 퍼지 G-K 알고리즘은 데이터집합에서 서로 다른 기하학적인 형태의 클러스터를 검출하기 위하여 적용적 거리 측정을 적용한 표준 FCM의 확장 형태이다. 즉 G-K 알고리즘은 크기나 밀도가

다른 형태의 클러스터를 검출하고 클러스터 중심을 추정하기 위해서 공분산 행렬을 응용한 거리를 적용하였다. 그러나 퍼지 G-K 알고리즘은 FCM 알고리즘과 같이 데이터로부터 각 클러스터들에 대한 소속정도의 합이 1이 되는 확률적 제약조건(probabilistic constraint)을 이용하므로 잡음을 포함하고 있는 상황에서 문제점을 야기시킨다. 따라서 이러한 문제점을 해결하기 위하여 PCM 알고리즘을 결합하였다. PCM 알고리즘은 소속정도의 값이 다른 클러스터와 관계가 없고, 각 패턴과 해당 클러스터 중심간의 거리에만 의존하기 때문에 위의 문제점을 개선할 수 있었다.

본 연구에서는 퍼지 G-K 알고리즘의 결과를 PCM 알고리즘의 입력으로 하여 클러스터들의 크기와 중심값을 얻어서 PCM 알고리즘을 이용한 분류를 수행한다. 그리고 PCM 알고리즘과 퍼지 G-K 알고리즘 수행 결과를 비교하여 분류결과가 일치하는 화소들의 분류항목을 결정한다. 일치하지 않는 화소는 최대우도 분류를 수행하여 분류항목을 결정한다. 베이지안 최대우도 분류를 수행시 평균내부거리 안쪽의 데이터를 이용하여 공분산 행렬을 구한 후 분류를 수행한다. 평균내부거리 안쪽의 데이터들의 집합을 내부클러스터라 한다. 내부클러스터는 클러스터의 크기와 밀도에 비례한다. 그리고 정규분포 형태를 나타낸다. 따라서 분류의 정확도를 높일 수 있으며, 주어진 학습데이터가 작은 경우에도 분류를 효과적으로 할 수 있다. 퍼지 G-K 알고리즘과 PCM 알고리즘 수행시 일반적으로 반복적인 과정을 통해서 분류를 하지만 본 논문에서는 학습데이터를 이용하여 반복적인 과정을 거치지 않음으로서 분류 시간을 줄일 수 있다.

그림 1은 퍼지 알고리즘의 융합에 의한 다중분광영상의 패턴분류를 위한 시스템의 전체 구성을 나타낸다. 본 연구의 패턴분류 시스템은 학습데이터를 선택하는 과정, 퍼지 융합알고리즘을 이용한 분류를 수행하는 과정, 결

과를 출력하고 저장하는 과정의 3가지 모듈로 구성되어 있다.

3.1 학습데이터의 선정

위성영상의 감독 분류는 특정 지역을 분류하기 위한 사전 학습데이터가 필요하다. 학습 과정의 전반적인 목표는 특정 영상에서 구분되는 모든 토지피복 종류에 대한 분광반응패턴을 설명할 수 있는 통계집단을 수집하는 것이다. 좋은 분류결과를 산출하기 위해, 학습데이터는 대표성과 완벽성을 가져야 한다. 이 말은 영상의 각각 정보항목을 구성하고 있는 모든 분류항목들의 학습영역에 대한 통계정보를 영상분석자가 정확하게 추출하여야 한다는 의미이다. 일반적으로 학습데이터를 선정하는 방법은 첫째, 야외조사, 참조정보(항공사진, 지도, 영상육안추정)를 이용하여 다중시기 영상으로부터 실측된 데이터에서 얻을 수 있으며, 둘째로는 획득한 영상을 디스플레이 하여 시각적인 분석가가 분류항목에 따라 일부 영역을 샘플로 취득하여 사용하는 방법이 있다. 첫째 방법은 학습데이터를 얻는데 많은 시간과 비용이 걸린다. 따라서 본 연구에서는 두 번째 방법을 이용하여 학습데이터를 선정한다. 두 번째 방법에 의하여 학습데이터를 정확하게 획득하려면 입력 위성영상을 히스토그램 평활화에 의한 영상강화 과정을 거쳐서 선명한 영상으로 디스플레이 한 후 분류하고자 하는 항목을 설정하고 항목별로 몇 지역을 학습데이터로 선정한다. 그림 2는 학습데이터의 선정과정을 보여준다.

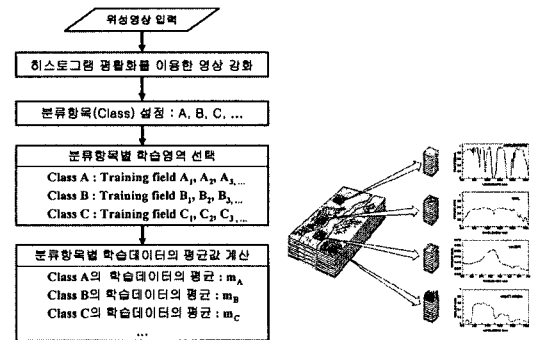


그림 2 학습데이터의 선정과정

3.2 퍼지 융합 알고리즘을 이용한 분류

본 연구에서는 다중분광 위성영상의 분류를 위하여 퍼지 G-K 알고리즘과 PCM 알고리즘을 결합한 퍼지 융합 알고리즘에 의하여 분류를 한다. 먼저 학습데이터를 이용하여 퍼지 G-K 알고리즘을 먼저 수행하고 그 분류결과를 이용해서 클러스터들의 크기와 초기 중심값을 구한 후 각 클러스터에 대한 평균내부거리를 구하여 PCM 알고리즘을 수행한다.

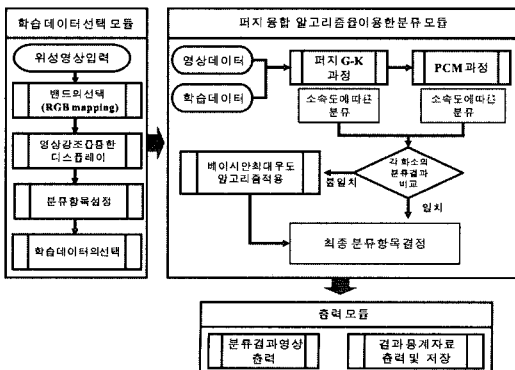


그림 1 퍼지 알고리즘의 융합에 의한 분류 시스템 구조

퍼지 G-K 알고리즘은 데이터 집합에서 서로 다른 기하학적인 형태의 클러스터를 검출하기 위하여 공분산 행렬을 이용한 적응적인 거리 측정(adaptive distance measure)을 사용하여 다차원 다중분광 위성영상의 분류에 효과적으로 적용 가능하다. PCM 알고리즘은 각 데이터와 특정클러스터와의 거리에 소속정도를 부여하는 퍼지 C-means 알고리즘과 달리 소속정도를 각 데이터와 클러스터 중심 간의 절대거리에 의존하는 방법으로 퍼지 C-means 알고리즘이 가지는 상대성 문제를 해결하였다. 학습데이터의 화소 값에 특별히 큰 값이나 작은 값의 잡음이 있는 상황에서 분류에 좋은 결과를 얻을 수 있다. PCM 알고리즘은 각 영역에 해당하는 클러스터들의 크기와 초기 중심 값을 사전에 알아야 한다. 따라서 퍼지 G-K 알고리즘을 먼저 수행하고 그 분류결과를 이용해서 PCM 알고리즘을 수행한다. 퍼지 G-K 알고리즘과 PCM 알고리즘 수행시 일반적으로 반복적인 과정을 통해서 분류를 하지만 본 연구에서는 학습데이터를 이용하여 수행함으로써 반복적인 과정을 거치지 않는다. 따라서 분류 시간을 줄일 수 있다.

퍼지 알고리즘 융합 및 적용은 다음과 같은 과정에 의하여 이루어진다.

주어진 입력영상에 대하여 분류항목의 수 $c(2 \leq c \leq n)$ 와 가중 먹지수 $m(1 \leq m \leq \infty)$ 을 선정하고, 선정된 학습데이터에서 각 분류항목별 화소의 평균값을 구한다. 일반적으로 $m=2$ 로 설정한다.

단계 1) 중심값 v_i 를 계산한다. 초기 중심값은 학습데이터의 평균값을 이용하며, 초기 소속도 μ_{ij} 는 모든 분류항목에 대하여 1로 설정한다. $1 \leq j \leq N$ 는 화소를 나타낸다.

$$v_i = \frac{\sum_{j=1}^N (\mu_{ij})^m x_j}{\sum_{j=1}^N (\mu_{ij})^m}, 1 \leq i \leq c \quad (10)$$

단계 2) 각 분류항목별 공분산 행렬들을 계산한다.

$$F_i = \frac{\sum_{j=1}^N (\mu_{ij})^m (x_{pj} - v_{pi})(x_{pj} - v_{pi})^T}{\sum_{j=1}^N (\mu_{ij})^m}$$

$$= \frac{\sum_{j=1}^N \left(\begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right) \cdot \left(\begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right)^T}{N}, 1 \leq i \leq c \quad (11)$$

단계 3) 각 분류항목별 중심값과 입력영상의 각 화소 간의 공간상의 거리를 계산한다. p 는 차원의 수(위성영상의 밴드 수)를 나타낸다.

$$d_{ij}^2(x_j, v_i) = (x_{pj} - v_{pi})^T [det(F_i)]^{1/p} F_i^{-1} (x_{pj} - v_{pi}) \quad (12)$$

$$d_{ij}^2(x_j, v_i) = \left(\begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right)^T [det(F_i)]^{1/p} F_i^{-1} \left(\begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right) \quad (13)$$

단계 4) 소속도 μ_{ij} 를 새롭게 갱신한다. 만약 $1 \leq i \leq c, 1 \leq j \leq N$ 에 대하여 $d_{ij}(x_j, v_i) > 0$ 이면

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}(x_j, v_i)}{d_{kj}(x_j, v_i)} \right)^{\frac{2}{m-1}}} \quad (14)$$

이고, 아니면 $d_{ij}(x_j, v_i) > 0$, $\sum_{i=1}^c \mu_{ij} = 1$ 이고, $\mu_{ij} \in [0, 1]$ 이면 $\mu_{ij} = 0$ 이다.

단계 5) 새롭게 갱신된 소속도를 이용하여 단계 1)부터 단계 4)의 과정을 다시 한 번 수행하여 소속도 μ_{ij} 를 계산한다.

단계 6) 갱신된 소속도 μ_{ij} 에 따라 입력 영상의 각 화소를 분류한다.

단계 7) 퍼지 G-K 알고리즘에 의한 입력영상의 각 분류항목별 분류결과를 PCM 알고리즘의 입력으로 한다.

단계 8) 퍼지 G-K 알고리즘에 의하여 분류된 각 분류항목에 대하여 각각의 중심값 v_i 를 식 (10)에 의하여 구한다.

단계 9) 각 분류항목의 중심값과 그 항목에 속하는 화소와의 거리를 식 (13)에 의하여 구한다.

단계 10) 각 분류항목별로 평균내부거리 η_i 를 식 (15)에 의하여 구한다. j 는 각 분류항목에 속하는 화소이며 K 는 1로 설정한다.

$$\eta_i = K \frac{\sum_{j=1}^N (u_{ij})^m d_{ij}^2(x_j, v_i)}{\sum_{j=1}^N (u_{ij})^m} \quad (15)$$

단계 11) 새로운 소속도 u_{ij} 를 식 (16)에 의하여 계산한다.

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2(x_j, v_i)}{\eta_i} \right)^{\frac{1}{(m-1)}}} \quad (16)$$

단계 12) 단계 11)에서 구한 소속도 u_{ij} 에 따라 각 화소를 분류한다.

단계 13) 단계 6)과 단계 12)의 각 화소에 대한 분류결과가 서로 같은 항목을 나타내면 해당 항목으로 분류를 최종 확정하고 분류결과가 서로 다른 경우에는 베이시안 최대우도 분류과정으로 진행한다.

3.3 베이시안 최대우도 분류 알고리즘 적용과정

퍼지 G-K 알고리즘의 분류결과와 PCM 알고리즘의 분류결과를 비교하여 각 화소의 분류결과가 같은 분류

항목이라면, 해당 분류항목으로 화소를 할당한다. 그러나 두 알고리즘의 분류결과가 다른 항목인 경우에는 베이시안 분류에 의하여 최종 분류항목을 결정한다. 베이시안 최대우도 분류에서는 위성영상과 같이 다차원 다중분광 데이터를 분류할 경우 다차원 정규확률밀도 함수를 이용한다[16]. 어떤 화소가 특성치 x 를 가지며 동시에 w_i 에 속할 확률 $P(X|w_i)P(w_i)$ 를 구하면 다음과 같다.

$$P(X | w_i)P(w_i) = \frac{P(w_i)}{(2\pi)^{N/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(X-U_i)^T \Sigma_i^{-1}(X-U_i)\right] \quad (17)$$

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

$$D_i(X) = \ln P(w_i) - \frac{N}{2} \ln 2\pi - \frac{1}{2} \ln \left| \Sigma_i \right| - \frac{1}{2} (X-U_i)^T \Sigma_i^{-1} (X-U_i) \quad (18)$$

일반적으로 사전확률을 가지는 베이시안 최대우도 분류기법은 다변량 정규확률밀도함수에 있어서의 판별식 $D_i(X)$ 의 값이 최대가 되는 분류항목으로 화소를 분류하는 과정이다. 이때 $P(w_i)$ 는 c 개의 분류항목 중 i 분류항목이 발생할 사전확률을 의미한다.

본 논문에서는 식 (15)에서 구한 평균내부거리를 이용하여 각 분류항목별로 평균내부거리 안쪽에 속하는 데이터들을 획득한다. 그 데이터를 이용하여 베이시안 최대우도 분류를 수행한다. 베이시안 최대우도 분류기법은 식 (19)에 의하여 구해진다.

$$D_i(X) = \ln P(w_i) - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(F_i)) - A \quad (19)$$

$$A = \frac{1}{2} \left[\left(\begin{pmatrix} x_{1ij} \\ x_{2ij} \\ \vdots \\ x_{pij} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right)^T \cdot \left[\det(F_i) \right]^{-\frac{1}{2}} F_i^{-1} \cdot \left(\begin{pmatrix} x_{1ij} \\ x_{2ij} \\ \vdots \\ x_{pij} \end{pmatrix} - \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{pi} \end{pmatrix} \right) \right]$$

$P(w_i)$ 는 사전확률로서 퍼지 G-K 알고리즘과 PCM 알고리즘의 분류 결과에 의하여 분류가 결정된 화소의 분류항목별 비율이다. 판별식 $D_i(X)$ 의 값이 최대가 되는 항목에 화소를 분류한다. 베이시안 알고리즘에 의하여 나머지 화소의 최종 분류항목을 결정함으로써 전체 이미지의 분류를 마치고 결과영상을 출력한다.

4. 실험 및 결과분석

본 논문에서는 제안한 분류 알고리즘의 효율성을 검증하기 위하여 1992년 6월 서울 한강의 일부 영역을 촬영한 LANDSAT TM 위성영상과 부산의 수영만 지역을 촬영한 IKONOS 위성영상을 이용하여 실험하였다.

각 실험의 결과를 얻기 위해 주어진 학습데이터에 의하여 각 알고리즘을 30회 정도 수행하여 그 평균값에 해당되는 실험결과를 나타낸 것이다.

그림 3은 IKONOS 위성 영상 3, 2, 1번 밴드와 Landsat TM 7, 4, 2밴드를 각각 RGB 컬러로 합성한 영상을 나타낸 것이다.

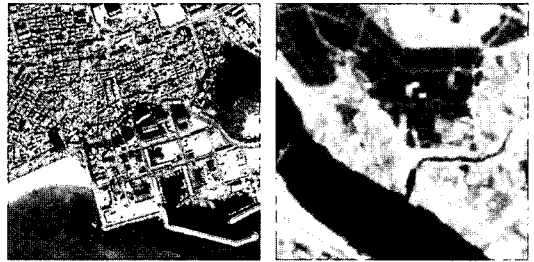


그림 3 IKONOS 위성영상과 LANDSAT TM 위성영상

IKONOS 위성영상은 고해상도 위성영상으로서 흑백 및 다중분광 영상 모두 11비트의 정보를 가지고 있으며, 1m 전정색 밴드의 흑백 영상데이터는 가시광선에서 근적외선 파장대의 정보를 가지고 있다. 4m 다중분광밴드 영상은 적색(R), 녹색(G), 청색(B)과 근 적외선(NIR)의 각 4개의 밴드로 구성되어 있다. 본 논문에서는 4m 다중분광 영상을 이용하여 실험하였다. 실험에 이용된 영상의 크기는 1000×1000 화소를 가지고 있다. 11비트 영상의 화소들의 밝기 값은 0-2047 사이의 값을 가지게 된다.

본 연구에서는 4개의 분류항목 Forest, Water, Soil, Urban을 선정한 후 디스플레이 된 영상에서 각 항목별로 3-6개 영역의 학습지역을 선정하였다. 표 1은 4개의

표 1 각 분류항목별 학습데이터에 대한 통계량

분류항목	학습데이터 화소의 갯수	최대 화소값	최소 화소값	평균값
Forest	14618	밴드 1 : 489	136	169.869
		밴드 2 : 633	98	146.995
		밴드 3 : 597	34	91.445
		밴드 4 : 598	26	156.067
Water	52860	밴드 1 : 372	170	204.212
		밴드 2 : 438	134	180.413
		밴드 3 : 389	42	71.905
		밴드 4 : 395	11	32.134
Soil	15866	밴드 1 : 636	183	333.325
		밴드 2 : 851	166	428.780
		밴드 3 : 843	69	409.543
		밴드 4 : 796	25	409.494
Urban	117257	밴드 1 : 1298	106	224.940
		밴드 2 : 1826	71	223.010
		밴드 3 : 1896	24	165.464
		밴드 4 : 1919	19	154.893

표 2 PCM 알고리즘에서 평균내부거리 안쪽에 속하는 분류항목별 화소의 수

	Forest	Water	Soil	Urban
평균내부거리	49,419	15,210	32,912	28,926
평균내부거리 중심값	204,536	209,465	346,389	213,741
	203,227	186,271	443,197	204,549
	158,094	76,985	418,941	144,719
	243,400	36,591	406,966	130,531
평균내부거리 안쪽의 화소 수	34531	95212	24638	387408

분류항목에 대하여 디스플레이 된 영상에서 선택한 학습 데이터의 화소의 개수와 그것에 대한 각 밴드별 화소의 최소값, 최대값, 평균값 등의 통계량을 나타낸 것이다.

표 2는 PCM 알고리즘에서 평균내부거리와 평균내부거리 안쪽에 속하는 분류항목별 화소의 수, 밴드별 중심값을 나타낸 것이다. 평균내부거리 안쪽의 화소들은 베이시안 최대우도 분류의 새로운 학습데이터로 이용된다. PCM 알고리즘과 퍼지 G-K 알고리즘 수행 결과를 비교하여 분류결과가 일치하지 않는 화소는 최대우도 분류를 수행하여 분류항목을 결정한다. 이때 PCM 알고리즘의 평균내부거리 안쪽에 있는 화소 값들을 새로운 학습데이터로 하여 분류를 수행한다. 평균내부거리 안쪽에 속하는 화소들의 집합은 클러스터의 크기와 밀도에 비해 하게 되며, 좋은 정규분포를 보여주므로 최대우도 분류의 정확도를 높여주었다. 그림 4는 IKONOS 위성에서 4개의 분류항목 중 Water 영역에 대한 학습데이터의 히스토그램과 평균내부거리 내부의 데이터에 대한 히스토그램을 나타낸 것이다. 평균내부거리 안쪽의 데이터가 더 정규분포에 가까워지는 것을 보여준다. 따라서 베이시안 최대우도 분류를 적용함에 있어서 좋은 결과를 얻을 수 있다. 반복적인 테스트를 수행한 결과 퍼지 G-K 알고리즘과 PCM 알고리즘에 의한 분류 결과가 일치하는 화소의 수가 약 909702개다. 일치하는 화소들은 해당 항목으로 분류항목을 결정한다. 일치하지 않는 화소들은 베이시안 최대우도 분류를 수행한다.

표 3은 각 분류 알고리즘을 이용하여 입력영상 전체를 분류하였을 때, 각 분류항목별로 선정된 학습영역들에 대한 데이터들의 분류결과가 얼마나 정확하게 분류되었는가를 나타낸 것이다. 그 결과 제안된 방법이 퍼지 G-K만 수행한 결과나 PCM만 수행한 결과보다 정확한 분류결과를 얻을 수 있음을 확인할 수 있다. 또한 일반

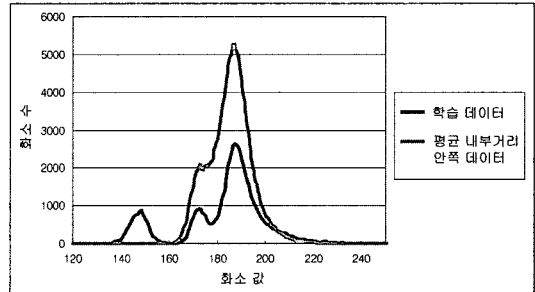


그림 4 Water 영역에 대한 히스토그램

적으로 가장 많이 사용되는 분류기법인 최대우도 분류기법보다 더 좋은 결과를 얻을 수 있음을 확인할 수 있었다. 표 3에서 잘못 분류된 오차의 경우 그림자 영역과 각 분류항목 사이의 경계지역의 일부 화소들이 오분류되는 결과가 나타났다. 하지만 제안된 방법은 기존의 분류방법보다 그 오차를 줄였다.

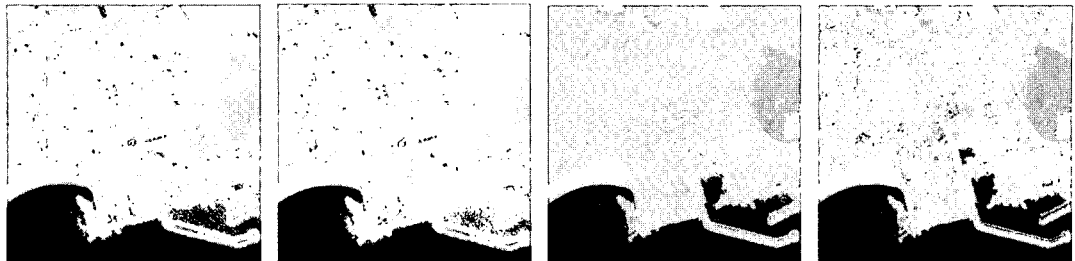
그림 5는 IKONOS 위성영상을 퍼지 G-K 알고리즘, PCM 알고리즘, 최대우도 분류알고리즘, 제안된 알고리즘으로 분류한 결과영상이다.

IKONOS 위성영상의 분류와 같은 방법으로 LANDSAT TM 위성영상에 대한 실험을 수행하였다. 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. 4개의 분류항목 Forest, Water, Crop, Urban 들을 선정된 후 각 항목별로 3~5개 영역의 학습영역을 선정하였다.

표 4는 LANDSAT TM 위성영상에 대하여 각 분류

표 3 학습데이터의 분류결과에 대한 정확도의 비교

		Forest	Water	Soil	Urban	전체정확도
학습데이터		14618	52860	15866	117257	200,601
분류방법	MLC	11800	50903	13889	112100	94.1%
	퍼지 G-K	10981	50006	13584	109999	92%
	PCM	11381	48506	12584	117257	91%
	제안된 알고리즘	12381	52009	14090	112499	95.2%



(a) 퍼지 G-K

(b) PCM

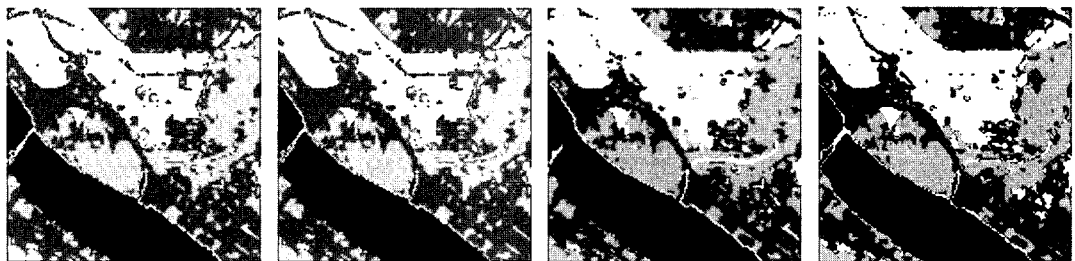
(c) MLC

(d) 제안된 알고리즘

그림 5 분류 결과 영상

표 4 학습데이터의 분류결과에 대한 정확도의 비교

		Forest	Water	Crop	Urban	전체정확도
학습데이터		1024	1232	1265	1008	4,529
분류방법	MLC	960	1228	1225	927	95.8%
	퍼지 G-K	940	1222	1201	895	94%
	PCM	935	1228	1195	896	94%
	제안된 알고리즘	988	1232	1255	960	98%



(a) 퍼지 G-K

(b) PCM

(c) MLC

(d) 제안된 알고리즘

그림 6 분류결과 영상비교

알고리즘을 이용하여 입력영상 전체를 분류하였을 때, 각 분류항목별로 선정한 학습영역들에 대한 데이터들이 각각의 알고리즘 수행 후 처음 선정된 분류항목으로 얼마나 정확하게 분류되었는가를 나타낸 것이다. 그 결과 제안된 방법이 일반적으로 많이 사용되는 최대우도 분류 알고리즘(MLC)보다 역시 좋은 결과를 얻을 수 있었으며, 또한 퍼지 G-K 알고리즘만 수행한 결과나 PCM 알고리즘만 수행한 결과보다 정확한 분류결과를 얻을 수 있음을 확인할 수 있다.

그림 6은 퍼지 G-K 알고리즘, PCM 알고리즘, 최대우도 분류알고리즘, 제안된 알고리즘을 이용하여 분류를 수행한 결과 영상이다.

고해상도 위성인 IKONOS 위성과 중·저해상도 위성인 LANDSAT TM 위성영상에 대하여 각각 제안된 방법으로 실험한 결과 모두 좋은 분류 결과를 보여 주었다.

5. 결론

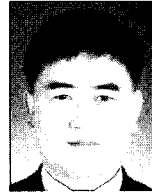
본 논문에서는 다중분광 원격탐사 위성영상을 분류하

기 위하여 퍼지 G-K 알고리즘과 PCM 알고리즘의 장점을 결합한 퍼지 알고리즘의 융합에 의한 분류방법을 제안하였다. 제안된 방법은 입력영상에서 학습데이터를 선택하여 초기 중심값을 구하고, 퍼지 G-K 알고리즘을 수행한 후 그 결과를 이용하여 PCM 알고리즘을 수행한다. 이 때 계산 시간을 감소시키기 위해 반복과정을 수행하지 않는다. 최종 분류항목의 결정은 PCM 알고리즘과 퍼지 G-K 알고리즘 수행 결과를 비교하여 분류결과가 일치하는 화소들은 해당항목으로 결정한다. 일치하지 않는 화소는 PCM 알고리즘의 평균내부거리 안쪽에 있는 화소들을 새로운 학습데이터로 하여 베이시안 최대우도 분류를 수행하여 분류항목을 결정한다. 제안된 알고리즘을 다차원 고해상도 IKONOS 위성영상과 LANDSAT TM 위성영상을 이용하여 실험한 결과 퍼지 G-K 알고리즘만 적용하거나 PCM만 적용한 결과보다 더 좋은 결과를 얻을 수 있었다. 그리고 평균내부거리 안쪽의 데이터는 좋은 정규분포를 보여주므로 베이시안 최대우도 분류의 정확도를 높여주었다.

향후 연구과제는 그림자 지역과 혼합 화소 지역에 대한 분류 정확도를 향상시키기 위한 추가적인 연구가 뒤따라야 하겠다.

참고 문헌

- [1] John A. Richards, Remote Sensing Digital Image Analysis : An Introduction, Second, Revised and Enlarged Edition, pp.229-262, Springer-Verlag, 1994.
- [2] Pierce, L., Samples, G., Dobson, M.C., Ulaby, F., "An automated unsupervised/ supervised classification methodology," Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International, Volume: 4, pp.1781-1783, 6-10 July 1998.
- [3] Hoffbeck, Joseph P. and David A. Landgrebe, "Classification of Remote Sensing Images having High Spectral Resolution," Remote Sensing of Environment, Vol. 57, No. 3, pp 119-126, September 1996.
- [4] Rego, L.F.G., Koch, B., "Automatic classification of land cover with high resolution data of the Rio de Janeiro City Brazil," Remote Sensing and Data Fusion over Urban Areas, 2003. 2nd GRSS/ISPRS Joint Workshop on, pp.172-176, 22-23 May 2003.
- [5] 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.
- [6] Nakashima, T., Nakai, G., Ishibuchi, H., "Constructing fuzzy ensembles for pattern classification problems," Systems, Man and Cybernetics, 2003. IEEE International Conference on, Volume: 4, pp.3200-3205, 5-8 Oct. 2003.
- [7] Weiyang Zhou, "Verification of the nonparametric characteristics of backpropagation neural networks for image classification," Geoscience and Remote Sensing, IEEE Transactions on, Volume: 37, Issue: 2, pp.771-779, March 1999.
- [8] 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.
- [9] 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, Volume: 4, pp.2525-2527, 24-28 June 2002.
- [10] 한종규, 이상구, "뉴로-퍼지 알고리즘을 이용한 원격탐사 화상의 지표면 패턴분류 시스템의 구현", 퍼지 및 지능시스템학회 논문지, 제9권 제5호, pp.472-479, 1999.
- [11] James C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NY, 1981.
- [12] N.R. Pal and J.C. Bezdek : On cluster validity for the fuzzy c-means model, IEEE Transactions on Fuzzy Systems, Vol. 3, No. pp.3370-379, 1995.
- [13] R. Krishnapuram and J. M. Keller : A possibilistic approach to clustering, IEEE Trans. on Fuzzy Systems, Vol. 1, No. 2, pp.98-110, 1993.
- [14] D.Gustafson and W.Kessel, "Fuzzy clustering with a fuzzy covariance matrix," In Proc. IEEE CDC, San Diego, USA, pp.761-766, 1979.
- [15] Babuka, R., van der Veen, P.J., Kaymak, U., "Improved covariance estimation for Gustafson-Kessel clustering," Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on, Volume: 2, pp.1081-1085, May 2002.
- [16] 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, 2002.



전 영 준

1992년~1998년 동의대학교 전산통계학과 학사 졸업(이학사). 1998년~2000년 동의대학교 전산통계학과 석사 졸업(이학석사). 2005년 동의대학교 컴퓨터공학과 박사 졸업(공학박사). 관심분야는 영상처리 및 패턴인식, 위성영상응용, 모바일응용, 인공지능 및 컴퓨터응용분야



김 진 일

1976년~1982년 경희대학교 전자공학과 졸업(학사, 공학석사). 1994년 서강대학교 전자계산학과 졸업(공학박사). 1982년~1984년 미국 Bon Scours 시스템, 맥도널드글라스 우주항공회사 등 연구원 1996년~1997년 미국 Purdue Univ. 전기및컴퓨터공학부 연구교수. 1998년~2001년 위성탐사 영상처리연구회(한국과학재단) 회장. 1988년~현재 동의대학교 컴퓨터공학과 교수. 관심분야는 패턴인식, 퍼지논리, 위성영상응용, GIS/GPS/RS, 인공지능 및 컴퓨터응용분야