

# Rough 집합 이론을 이용한 원격 탐사 다중 분광 이미지 데이터의 특징 추출

## Features Extraction of Remote Sensed Multispectral Image Data Using Rough Sets Theory

원성현 · 정환목\*

Sunghyun Won and Hwanmook Chung

지산대학 전산정보처리과  
\*대구효성가톨릭대학교 전자·정보공학부

### 요 약

본 논문에서는 초 다중 밴드 환경의 효과적인 데이터 분류를 위해서 Rough 집합 이론을 이용한 특징 추출 방법을 제안한다. 다중 분광 이미지 데이터의 특성을 분석하고, 그 분석 결과를 토대로 Rough 집합 이론의 식별 능력을 이용하여 가장 효과적인 밴드를 선택할 수 있도록 한다. 실험으로는 Landsat TM으로부터 취득한 데이터에 적용시켰으며, 이를 통해 전통적인 밴드 특성에 의한 밴드 선택 방법과 본 논문에서 제안하는 러프 집합 이론을 이용한 밴드 선택 방법이 일치됨을 보이고 이를 통해 초다중 밴드 환경에서의 특징 추출에 대한 이론적 근거를 제시한다.

### ABSTRACT

In this paper, we propose features extraction method using Rough sets theory for efficient data classifications in hyperspectral environment. First, analyze the properties of multispectral image data, then select the most efficient bands using discernibility of Rough sets theory based on analysis results. The proposed method is applied Landsat TM image data, from this, we verify the equivalence of traditional bands selection method by band features and bands selection method using Rough sets theory that proposed in this paper. Finally, we present theoretical basis to features extraction in hyperspectral environment.

### 1. 서 론

최근 인공위성으로부터 수집한 원격 탐사 이미지 데이터(remote sensed image data)를 처리하여 실생활에 활용하고자 하는 노력이 사회 각 분야에서 폭넓게 진행되고 있다. 원격 탐사 데이터에 대한 구체적인 활용은 지리 정보 시스템(Geographic Information System, 이후 GIS로 약칭) 등으로 귀결될 수 있는데, 이러한 지리 정보 시스템을 통하여 효과적인 환경 감시, 국토 개발, 자원 탐사, 기상 예측, 군사 동향 파악 및 농업 생산량 추정 등을 수행할 수 있다[10-12, 14].

원격 탐사 데이터에 대한 처리 과정에서 발생할 수 있는 주된 연구 이슈는 첫째, 원격 탐사 이미지 데이터의 효과적인 처리라고 할 수 있다[1, 4-5, 7]. 원격 탐사 데이터는 단순한 화상 이미지와는 달리 다중 분광 이미지 데이터(multispectral image data)라는 큰 특징이 있다. 즉, 파장대에 따라 나뉘어진 여러개의 밴

드로부터 수집된 데이터를 처리한다는 의미이다. 처리에서 가장 중요하게 인식되는 과정은 분류(classification)라고 할 수 있다. 분류라 함은 어떤 대상이 되는 객체(혹은 화소(pixel))가 있을 때, 그 객체가 어떤 범주(category)에 해당되는지에 대한 결론을 내리는 것이다. 또 하나의 중요한 문제는 하나의 객체가 두개 이상의 범주로 분류되었을 때, 이에 대한 정확한 성분 비율 추정 문제이다[2, 6]. 두번째 연구 이슈는 최근에 부각된 것으로 다중 밴드(multi band)에서의 유효 밴드 선택에 관한 문제이다[3]. 다중 밴드의 차원이 4차원 혹은 7차원 등이었을 때, 실제 시스템이 처리하는 밴드는 이들 전부가 아니고 그 중 의미있는 2-3개의 밴드만을 추출한다. 이때 기존에는 파장대의 물리적 특성에 따라 분류하고자 하는 대상에 맞는 밴드가 이미 고정되어 있었다. 그러나, 최근 다중 밴드의 차원이 220개 등으로 확장되면서 이러한 전통적인 방법에 의한 밴드 선택에 무리가 있음이 밝

혀졌다[3]. 따라서, 수많은 분광 밴드에서 사용 가능한 유효 밴드 선택 문제가 새로운 연구 이슈로 떠오르게 된 것이다.

본 논문에서는 이와 같이 원격 탐사 데이터 처리에서 발생할 수 있는 몇가지 연구 이슈 중에서 유효 밴드 선택 문제에 대해 집중 연구하기로 한다. 아직 7개의 밴드를 넘는 초다중(hyperspectral) 분광 데이터에 대한 일반적인 처리는 이루어지지 않고 있으나 향후, 이것이 보편화될 것으로 예측되고 나아가서는 수백 개 이상의 밴드로 확장될 가능성이 매우 높은 현 시점에서 이를 대비한 합리적인 밴드 선택 알고리즘을 개발하는 것은 매우 중요한 일이라 할 수 있다.

최근까지도 밴드 특징 추출에 대한 연구는 거의 진행되지 않았다. 그 이유는 필요성이 제기되지 않았기 때문이다. 그러므로, 지금까지 많은 연구자들의 주된 연구 이슈는 데이터 자체에 대한 새로운 분류 기술을 개발하는 것이었다. 기존의 통계적 기법들이 갖는 한계에 인공지능 기법들을 접목하여 0.01%의 분석 정확도라도 향상시키려는 노력들이 바로 그것이다. 사실, 7개의 다중 밴드 환경이 정착되어 데이터를 수집한 것도 그리 오래되지 않은 일이라는 것이 이 사실을 입증하고 있다. 통계 기법들은 오래 전부터 컴퓨터 공학을 포함한 응용 과학 전 분야에 걸쳐 가장 우수한 데이터 처리 도구로 인정되어왔고 지금도 그렇다.

본 논문은 다중 분광 밴드 환경이 머지않아 초다중 분광 밴드 환경으로 바뀌더라도 좀더 양호한 데이터 분석 정확도를 얻어내기 위해 합리적인 밴드 특징 추출 방법이 마련되어 있어야 한다는 필요성에서 출발한다.

## 2. 다중 분광 이미지 데이터의 분류에 대한 기존 연구

본 장에서는 본 연구의 효과적인 진행을 위해서 다

중 분광 이미지 데이터의 기본 사항을 기술한다. 또한, 일반적인 원격 탐사 이미지 데이터에 대한 처리 과정을 소개하고 다중 분광 밴드의 특징 추출이 왜 필요한지에 대해 문제점을 제기한다.

### 2.1 다중 분광 이미지 데이터

다중 분광 이미지 데이터는 통상 원격 탐사 이미지 데이터, 위성 데이터(satellite data)라고도 부른다. 즉, 인공위성 등에서 수집한 영상을 분석하여 자원 탐사, 농수산업, 임업, 국토 계획 및 지도 제작, 기상·해양 오염 등의 부문에 응용한다. 다음 표 1에 그 응용 분야를 요약했다[11].

다중 분광 이미지 데이터의 처리 및 응용에 대한 연구는 현재 한국과학기술원의 시스템공학센터, 동력자원연구소, 해양연구소, 기상연구소, 농촌진흥청, 해양경찰, 서울대학교 등을 주축으로 위성 자료 분석 기법 개발, 자원 탐사, 해양 연구, 토지 이용 분석, 기상예보, 농작물 수확량 예측, 해양 오염 감시 등 기본 기술 개발 및 응용 연구를 수행하고 있다. 한편, 산업계를 중심으로 디지털 지도 제작 및 응용 시스템을 국내에 토착화하기 위한 연구가 진행되고 있다[11, 14].

### 2.2 일반적인 다중 분광 이미지 데이터의 처리

원격 탐사된 다중 분광 이미지에 대한 일반적인 처리 과정은 다음과 같다[10, 12, 15, 16].

#### 2.2.1 일반적인 데이터 처리 과정

Landsat TM과 SPOT HRV 등을 비롯한 많은 위성에 탑재된 센서를 통해 수집된 위성 데이터는 위성에 따라서는 일반 사용자에게 공급하지 않는 데이터도 물론 있지만 주기적으로 지상 수신국에 송신된다.

수집된 데이터들 중에는 사용할 수 없는 데이터들이 포함되어 있다. 즉, 수집되는 과정에서 여러 가지 왜곡 현상이 발생할 수 있는데 그 한 예로는 위성이 진행되는 동안 지구의 자전을 생각할 수 있다. 물론

표 1. 다중 분광 이미지 데이터의 응용 분야

응용 부문	응 용 업 무
자원 탐사	광물 자원 분포 조사, 석유 탐사
농업	농작물 수확량 예측, 병충해 및 가뭄 피해 조사, 농작물 재배 면적 조사
수산업	어장 분포 조사
임업	산림 자원 분포 조사, 산림 병충해 및 화재 피해 조사, 목재 수확량 예측
국토 계획 및 지도 제작	토지 이용 현황 조사, 단지 및 입지 선정, 지도 제작 및 수정, 고속도로·철도 등의 노선 선정
기상	일기 예보, 재난 예보
수자원 및 토목	수자원 분포 조사, 홍수 피해 상황 조사, 적설량 조사, 토양도 작성, 산지 침식 현황 조사
해양	해안 지형 변화, 해수 및 해류 특성 분석, 부유 퇴적물 분포 및 이동, 해저 탐사
지질	지질 구조 조사
오염	대기·수질·해양 오염 실태 조사 및 피해 상황 분석

이러한 지구의 이동을 고려하여 센서가 작동하지만 간혹 이로부터 불규칙 데이터가 발생할 가능성이 매우 높다고 볼 수 있다. 또한, 대기 상태 등이 센서의 작동에 영향을 미칠 가능성도 있다. 따라서, 데이터가 수집된 후 가장 먼저 수행해야 할 일은 데이터의 보정 작업이다. 그 다음 단계에서 할 일은 이미지 분석인데 이는 이미지 강화와 같은 전통적인 이미지 처리 과정에 필요한 제반 사항을 수행하게 된다. 그 다음에는 분류 알고리즘을 통해 화소들에 대한 범주 귀속 작업이 수행된다. 이렇게 되면 일단 이미지 처리는 종료된 것이다. 그 결과를 이용해서 전자 사진을 만들 수도 있고, GIS로의 입력으로 활용할 수 있게 된다.

**2.2.2 데이터 분류**

데이터 분류는 크게 교사 분류와 비교사 분류로 나누어 생각할 수 있다[10].

교사 분류는 분류 항목에 해당되는 학습 데이터를 그 항목의 대표적 분광 특징을 갖는다고 가정 하에 분류 과정의 지표로 삼는 기법을 말한다. 세부 알고리즘으로는 학습 데이터의 각 밴드에 대한 상한값과 하한값을 구하여 그 범위 내에서 분류를 수행하는 평행육면체 분류(parallelepiped classification), 화소가 속하는 각 분류 항목의 평균 벡터와 그로부터의 거리를 계산하여 가장 가까운 항목에 귀속시키는 최소 거리 분류(minimum distance classification), 분류 항목의 분포가 정규 분포를 이룬다는 가정 하에 공분산을 이용하여 최소 거리 분류를 도입한 Mahalanobis 거리 분류, 임의의 화소가 각 범주에 속할 경우 발생하는 확률 밀도 함수를 모든 분류 항목에 대해 구하고, 그 중 확률이 가장 커지는 경우의 범주로 결정하는 최대 공산 분류(maximum likelihood classification) 등이 있는데 이 중 마지막의 최대 공산 분류가 일반적으로 많이 사용된다.

비교사 분류는 군집화라고도 하는데 분석자의 분류 항목에 대한 정의에 따라 결합 혹은 분리시켜 같은 속성을 갖는 데이터들끼리 무리지어는 방법으로 순차적 군집, ISODATA 군집화, RGB 군집화 등이 있다.

그 밖에 최근에는 이들 두 가지 기법과 인공 지능 기법을 혼합하여 분류 정확도를 향상시키는 방법들도 제안되고 있다.

**2.3 다중 분광 밴드 상의 특징 추출의 필요성 및 기존 연구**

최근까지의 원격 탐사 데이터에 대한 처리는 다중 분광 밴드 상에서의 처리가 중심이 되었다[10]. 센서의 종류에 따라 밴드의 수는 다르지만(Landsat TM의

경우는 7개 밴드, SPOT HRV의 경우는 4개의 밴드) 어느 것이든 밴드의 수가 그리 많지 않았다. 그러므로, 그 중 실제 데이터 분석에 사용할 밴드를 선택하는 것이 큰 이슈가 되지 않았다. 그런데 7개의 밴드로 데이터를 수집하던 탐사 기술이 최근 220개의 초다중 밴드 환경으로 발전하게 되면서 전통적인 밴드 선택 방식에 문제가 있음이 밝혀졌다[3]. 220개나 되는 초다중 밴드에서는 기존의 7개 혹은 4개의 파장대보다 파장 간격이 훨씬 세분되기 때문에 이 중에서 실제 사용할 3개의 밴드를 선택하기 위해서는 보다 합리적인 근거를 갖는 방법에 의해야 할 것이라는 주장이 제기되고 있다. 그럼에도 불구하고, 이에 대한 연구는 현재 거의 진행되지 않고 있고 Jimenez 등에 의해 최초로 초기 연구가 시작되고 있으나 이들의 연구는 지나치게 복잡한 수학적 처리 절차를 요구하고 있기 때문에 최적의 해를 얻는데 큰 연산 오버헤드가 발생할 것으로 생각된다. 그러므로, 보다 간단하게 이 과정을 수행할 수 있는 방법에 대한 요구가 증대되고 있다.

**3. Rough 집합 이론**

Rough 집합 이론은 1982년 폴란드의 과학자 Pawlak에 의해 제안된 방법으로 객체들의 속성값들의 특징을 분석하여 객체들을 군집화하는 도구로 많이 사용되는 이론이다[8,9]. Rough 집합 이론은 불충분하거나 일관성이 없는 정보로부터 체계적이고 자동적으로 정보를 가려냄으로써 여러 속성값들로부터 구분화 혹은 근사화를 수행할 수 있다[8,9]. 본 장에서는 이들의 기본 개념과 주요 성질에 대해 소개한다.

**3.1 기본 개념**

본 절에서는 Rough 집합 이론의 근간이 되는 몇가지 정의에 대해 소개한다.

(정의 3.1) 속성값 기술 함수  $P_X$

$$P_X : Q \rightarrow V : P_X(q) = P(X, q) \tag{1}$$

(정의 3.2) 식별 불능 관계(indiscernibility relation)  $ind(Q)$

속성 집합  $X_i$ 와  $X_j$ 가 어떤 속성  $q$ 에 관하여 식별 불능 관계라면 다음과 같이 나타낸다.

$$\begin{aligned} P(X_i, q) &= P(X_j, q) \\ (X_i, X_j) &\in ind(q) \\ R &= ind(Q) \end{aligned} \tag{2}$$

그러므로,  $X_i, X_j$ 가  $P \subset Q$ 에 의하여 식별될 수 없을 경우에 다음과 같이 나타낸다.

$$(X_i, X_j) \in \text{ind}(P)$$

$$\text{ind}(P) = \bigcap_{q \in P} \text{ind}(q) \quad (3)$$

여기서,  $P=Q$ 일 때,  $(X_i, X_j) \in \text{ind}(q)$ 라면  $X_i$ 와  $X_j$ 는 상호 식별될 수 없는 관계가 되고,  $\text{ind}(q)$ 는 동치 관계이므로 이것에 의한 분할은 상집합(quotient set)이 된다.

$$X/\text{ind}(q) = [X_i] \mid X_i \in X \quad (4)$$

(정의 3.3) 상한 근사  $R^*(F)$

부분집합족  $F = \{Z_1, Z_2, \dots, Z_n\}$ 의 상한 근사는 다음과 같이 나타낸다.

$$R^*(F) = R^*(Z_1), R^*(Z_2), \dots, R^*(Z_n) \quad (5)$$

(정의 3.4) 하한 근사  $R^*(F)$

부분집합족  $F = \{Z_1, Z_2, \dots, Z_n\}$ 의 하한 근사는 다음과 같이 나타낸다.

$$R_*(F) = R_*(Z_1), R_*(Z_2), \dots, R_*(Z_n) \quad (6)$$

### 3.2 Rough 집합 이론을 이용한 동치 클래스 구성

다음과 같은 정보 시스템 테이블이 구성되어 있다고 가정하자.  $x_i$ 는 환자이고,  $q_i$ 는 검사 항목,  $V_{q_i}$ 는 검사 결과라고 한다면 다음 표 2는 환자들의 검사 항목에 대한 검사 결과를 나타내는 표라고 할 수 있다. 검사의 결과는 0, 1, 2의 세 가지 종류로 얻어졌다.

이 정보 시스템 테이블을 Rough 집합의 동치 관계를 이용하여 동치 클래스화하면 다음과 같은 분류가 가능하다.

$$X/\text{ind}(q_1) = [x_1, x_2, x_3], [x_4, x_5] \quad (7)$$

$$X/\text{ind}(q_2) = [x_1, x_2, x_3, x_4, x_5] \quad (8)$$

$$X/\text{ind}(q_3) = [x_1, x_2], [x_3, x_4, x_5] \quad (9)$$

$$X/\text{ind}(q_4) = [x_1, x_2, x_5], [x_3], [x_4] \quad (10)$$

$$X/\text{ind}(q_1, q_2) = [x_1, x_2, x_3], [x_4, x_5] \quad (11)$$

$$X/\text{ind}(Q) = [x_1, x_2], [x_3], [x_4], [x_5] \quad (12)$$

즉, 식 (7)에 의하면  $q_1$ 이라는 속성값에 의해  $x_1, x_2, x_3$

표 2. 환자들의 임상 검사 항목에 대한 정보 시스템 테이블

검사항목(Q)				
	$q_1$	$q_2$	$q_3$	$q_4$
환자(X)				
$x_1$	1	1	0	1
$x_2$	1	1	0	1
$x_3$	1	1	1	0
$x_4$	0	1	1	2
$x_5$	0	1	1	1

은 식별 불능 관계, 다시 말하면 조합 가능 관계이지만 이들과  $x_4, x_5$ 는 식별 가능 관계 즉, 조합 불가능 관계라고 말할 수 있다.

## 4. Rough 집합 이론을 이용한 다중 분광 이미지 데이터의 특징 추출

본 장에서는 본 논문에서 주로 제안하는 Rough 집합 이론을 이용한 다중 분광 밴드 상에서의 특징 추출 방법에 대해 기술한다.

### 4.1 Rough 집합 이론을 이용한 유효 밴드 선택

원격 탐사 이미지 데이터를 다중 분광 밴드 상에서 수집하는 이유에 대해서는 앞서 기술한 바와 같이 데이터의 분석 정확도를 높이기 위함이다. 예를 들어, 단일 밴드로 데이터를 수집하는 경우 어떤 객체의 분광 강도가 44인 화소도 실제로는 숲(forest)이라는 범주로 귀속되고, 동시에 91이라는 분광 강도를 갖는 화소도 똑같이 숲으로 귀속된다. 이때 우리는 이들 분광 강도만으로 올바른 데이터 분류가 매우 어려움을 느끼게 된다. 그런데, 이것이 2차원으로 밴드로 확장되면 2개의 밴드에서의 분광 강도를 조합하면 단일 밴드보다는 상이한 분광 강도를 갖는 화소들이 동일한 범주로 분류되는 경우가 훨씬 줄어들게 된다. 같은 방법으로 3차원 밴드로 확장하면 거의 분류 오차가 발생되지 않기 때문에 실제로 원격 탐사 데이터의 분류를 수행할 때는 3차원 밴드를 조합하게 되는 것이다.

#### 4.1.1 다중 분광 이미지 데이터의 참조표

통상적으로 임의의 밴드에서 특정 범주에 속하는 화소들의 분광 강도는 거의 정해져 있다. 물론, 데이터 수집 과정에서 왜곡 현상이 발생할 수 있고, 의미 없는 값(garbage)들이 수집되는 경우도 있지만 이들은 이미 개발되어 있는 여러 보정 방법들에 의해서 보정되므로 정해져 있는 분광 강도에서의 표준편차는 그리 크지 않다. 그러나, 어느 정도의 편차는 발생할 수 있기 때문에 다음과 같은 일반화된 다중 분광 이미지 데이터의 참조표(look up table)를 만들어 분류시 활용한다. 이 참조표는 오랜 동안의 데이터 분석 과정 시 축적된 경험과 밴드별 객체들의 분광 특성에 의한 것이다.

다음 그림 1은 n개의 면(frame)에 대한 범주와 밴드간의 관계를 나타내는 분광 강도의 참조표이다. 각 면에서 C는 범주들의 벡터로  $C_1$ 부터  $C_m$ 까지의 m개로 구성되어 있다. 즉, 분류하고자 하는 대상들이 귀속될 수 있는 범주를 의미하는데 예를 들면, 숲, 수역

$B \backslash C$	$C_1$	$C_2$	...	$C_m$
$B_1$	$v_{111}$	$v_{112}$	...	$v_{11m}$
$B_2$	$v_{121}$	$v_{122}$	...	$v_{12m}$
...	...	...	...	...
$B_7$	$v_{171}$	$v_{172}$	...	$v_{17m}$

$B \backslash C$	$C_1$	$C_2$	...	$C_m$
$B_1$	$v_{211}$	$v_{212}$	...	$v_{21m}$
$B_2$	$v_{221}$	$v_{222}$	...	$v_{22m}$
...	...	...	...	...
$B_7$	$v_{271}$	$v_{272}$	...	$v_{27m}$

$B \backslash C$	$C_1$	$C_2$	...	$C_m$
$B_1$	$v_{n-111}$	$v_{n-112}$	...	$v_{n-11m}$
$B_2$	$v_{n-121}$	$v_{n-122}$	...	$v_{n-12m}$
...	...	...	...	...
$B_7$	$v_{n-171}$	$v_{n-172}$	...	$v_{n-17m}$

$B \backslash C$	$C_1$	$C_2$	...	$C_m$
$B_1$	$v_{n11}$	$v_{n12}$	...	$v_{n1m}$
$B_2$	$v_{n21}$	$v_{n22}$	...	$v_{n2m}$
...	...	...	...	...
$B_7$	$v_{n71}$	$v_{n72}$	...	$v_{n7m}$

그림 1. 다중 분광 이미지 데이터의 참조표.

등이다. 또한,  $B$ 는 밴드들의 벡터로  $B_1$ 부터  $B_7$ 까지의 7개의 밴드로 설정했다. 본 논문의 궁극적인 목표는 초다중 분광 밴드 환경에서의 특징 추출이지만 그 근거는 7개의 밴드 환경인 다중 분광 밴드에서 만들고자 한다.  $v_{111}$ ,  $v_{27m}$  등은 특정 면( $v_{111}$ 은 1면,  $v_{27m}$ 은 27면)의 특정 밴드( $v_{111}$ 은 밴드 1,  $v_{27m}$ 은 밴드 7)에서 특정 범주( $v_{111}$ 은 범주 1,  $v_{27m}$ 은 범주  $m$ )로 귀속되는 화소가 갖는 분광 강도이다.

그림 2는 이들  $n$ 개의 면에 대해 각 면의 동일한 범주와 동일한 밴드에서 발생하는 다양한 분광 강도에 대한 표준편차를 구하여 오차의 한계를 인정한 것이다.  $\sigma$ 는 표준편차를 의미하는 것으로  $\sigma_{11}$ 은 1면부터  $n$ 면까지에서 밴드 1( $B_1$ )에서 범주 1( $C_1$ )로 분류되는  $n$ 개의 분광 강도에 대한 표준편차를 의미한다.

#### 4.1.2 탐사된 화소들의 범주 귀속

그림 3은 탐사된 객체에 대한 화소들의 분광 강도 표이다. 여기서,  $P$ 는  $k$ 개의 탐사 화소  $P_1, P_2, \dots, P_k$ 의

집합을 의미한다. 또한,  $pv_{11}$ 은 화소  $P_1$ 이 밴드 1에서 나타낸 분광 강도를 말한다.

그림 2와 3을 이용하여 다음 그림 4와 같이 각 화소들을 특정 범주로 귀속시킨다. 이때, 귀속되는 원칙은 화소의 분광 강도가 그림 2의 범위 내에 포함되는 경우에는 해당 범주로 귀속시키고, 그렇지 않은 경우는 가장 인접한 범주로 귀속시킨다.

#### 4.2 Rough 집합 이론을 이용한 동치 클래스 생성 및 해석

그림 2와 3을 이용하여 구성한 그림 4로부터 Rough 집합 이론을 이용하여 동치 클래스를 생성시킨다.

이로부터 동치 클래스를 추출하면 다음과 같다.

$$B/\text{ind}(P_1) = \{[B_1, B_7], [B_2, B_4, B_5], [B_3, B_6]\} \quad (13)$$

$$B/\text{ind}(P_2) = \{[B_1, B_3, B_4], [B_2, B_5], [B_6, B_7]\} \quad (14)$$

$B \backslash C$	$C_1$	$C_2$	...	$C_m$
$B_1$	$v_{111} \pm \sigma_{11}$	$v_{112} \pm \sigma_{12}$	...	$v_{11m} \pm \sigma_{1m}$
$B_2$	$v_{121} \pm \sigma_{21}$	$v_{122} \pm \sigma_{22}$	...	$v_{12m} \pm \sigma_{2m}$
...	...	...	...	...
$B_7$	$v_{171} \pm \sigma_{71}$	$v_{172} \pm \sigma_{72}$	...	$v_{17m} \pm \sigma_{7m}$

그림 2. 다중 분광 이미지 데이터에 허용된 오차의 한계.

$B \backslash P$	$P_1$	$P_2$	...	$P_k$
$B_1$	$pv_{11}$	$pv_{12}$	...	$pv_{1k}$
$B_2$	$pv_{21}$	$pv_{22}$	...	$pv_{2k}$
...	...	...	...	...
$B_7$	$pv_{71}$	$pv_{72}$	...	$pv_{7k}$

그림 3. 화소들의 밴드별 분광 강도.

$B \backslash P$	$P_1$	$P_2$	...	$P_k$
$B_1$	$C_1$	$C_n$	...	$C_2$
$B_2$	$C_2$	$C_1$	...	$C_1$
$B_3$	$C_{n-1}$	$C_n$	...	$C_1$
$B_4$	$C_2$	$C_n$	...	$C_3$
$B_5$	$C_2$	$C_1$	...	$C_2$
$B_6$	$C_{n-1}$	$C_3$	...	$C_n$
$B_7$	$C_1$	$C_2$	...	$C_n$

그림 4. 탐사된 화소들의 범주 귀속.

$$B \text{And}(P_k) = \{[B_1, B_5], [B_2, B_3], [B_4], [B_6, B_7]\} \quad (15)$$

$$B \text{And}(P) = \{[B_1], [B_2], [B_3], [B_4], [B_5], [B_6], [B_7]\} \quad (16)$$

그러므로, 화소  $P_1$ 에 대해 밴드 1과 밴드 7은 식별 불능 관계이고, 밴드 2, 밴드 4, 밴드 5역시 상호 식별 불능 관계임을 나타낸다.

### 4.3 적용 예

#### 4.3.1 다중 분광 밴드 상의 문제 설정

다음 그림 5와 같이 밴드와 범주 간의 참조표가 있다고 하자. 물론 그 이전에  $n$ 개의 면에 대한 참조표가 있어야 하지만 여기서는 생략하고 그들에 대한 표준편차를 구하여 각 분광 강도에  $\pm$ 하여 오차의 허용 한계로 설정된 것만 나타냈다. 또한, 센서에 의해 그림 6과 같은 화소들의 분광 강도가 수집되었다고 하자.

#### 4.3.2 Rough 집합을 이용한 밴드 특징 추출과 해석

그림 5와 6를 통해 탐사된 화소들이 귀속될 가능성이 가장 높은 범주들을 구하면 다음 그림 7과 같다. 여기서, 그림 내의 수치는 범주 번호를 의미한다.

그림 7로 나타난 결과들에 대한 해석은 다음과 같이 할 수 있다. 다음 식에서 [ ] 안에 있는 수치들은 밴드 번호를 의미한다.

$B \backslash P$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
$B_1$	13.5	71.2	18	45.4	51.1	135.1	66.1
$B_2$	30.3	27	35.7	91.1	42.8	143.2	89.5
$B_3$	57.1	12	14.8	31.1	100.6	133.4	78.4
$B_4$	66.2	90.9	28.1	53.2	78.7	109.7	55.6
$B_5$	26.5	21.2	121.2	21.3	65.4	134.4	57.4
$B_6$	82.1	19.1	28.5	47.5	55.5	129.7	89.7
$B_7$	43.4	35.8	35.1	82.8	56	110.5	78

그림 5. 밴드와 범주간의 분광 강도 관계(허용 오차 고려).

$B \backslash P$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$
$B_1$	14	71	66	51	70
$B_2$	27	36	35	91	30
$B_3$	15	57	16	31	14
$B_4$	91	28	27	53	66
$B_5$	21	26	121	21	20
$B_6$	19	18	19	18	19
$B_7$	36	35	35	83	44

그림 6. 화소들의 밴드별 분광 강도.

$B \backslash P$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$
$B_1$	1	2	7	5	2
$B_2$	2	3	3	4	1
$B_3$	3	1	3	4	3
$B_4$	2	3	3	4	1
$B_5$	4	1	3	2	2
$B_6$	2	2	2	2	2
$B_7$	2	3	3	4	1

그림 7. 탐사된 화소들의 범주 귀속.

$$B \text{And}(P_1) = \{[1], [2, 4, 6, 7], [3], [5]\} \quad (17)$$

$$B \text{And}(P_2) = \{[1, 6], [2, 4, 7], [3, 5]\} \quad (18)$$

$$B \text{And}(P_3) = \{[1], [2, 3, 4, 5, 7], [6]\} \quad (19)$$

$$B \text{And}(P_4) = \{[1], [2, 3, 4, 7], [5, 6]\} \quad (20)$$

$$B \text{And}(P_5) = \{[1, 5, 6], [2, 4, 7], [3]\} \quad (21)$$

$$B \text{And}(P_2, P_5) = \{[1, 6], [2, 4, 7], [3], [5]\} \quad (22)$$

$$B \text{And}(P) = \{[1], [2, 4, 7], [3], [5], [6]\} \quad (23)$$

식 (17)의 경우는 화소 1( $P_1$ )에 대해 각 밴드별 식별 가능성을 분석한 결과 밴드 2, 4, 6, 7간에는 식별 불능 즉, 조합 가능임을 알 수 있다. 만일, 식 (22)과 같이 화소 2와 화소 5에 대해 밴드별 식별 가능성을 분석한 결과 밴드 1과 6, 밴드 2, 4, 7 등이 상호 식별 불능 즉, 조합 가능임을 알 수 있었다. 식 (23)의 경우는 즉  $P_1, P_2, P_3, P_4, P_5$  모두에 대해서 식별 가능성을 분석하면 오로지 밴드 2, 4, 7만이 식별 불능임을 알 수 있다. 결국, 이들 3개의 밴드만이 조합 가능하다는 결론을 얻을 수 있다.

## 5. 실험 및 결과의 고찰

본 논문에서 제안하는 특징 추출의 타당성을 검증하기 위해 다음과 같은 실제 데이터를 이용하여 특징 추출 작업을 수행했다.

### 5.1 실험 대상 선정

water (band 2)

분광 강도	화소의 수	비중 (%)	히스토그램
44	1	1.33	*
45	12	16.00	*****
46	24	32.00	*****
47	10	13.33	*****
38	16	21.33	*****
49	11	14.67	*****
50	1	1.33	*

(a)

water (band 4)

분광 강도	화소의 수	비중 (%)	히스토그램
29	4	5.33	***
30	61	81.33	*****
31	8	10.67	*****
32	1	1.33	*
33	1	1.33	*

(b)

그림 8. (a) 밴드 2, (b) 밴드 4에서의 물에 대한 다중 분광 데이터.

본 논문에서 실험 대상으로 선정한 것은 1992년 6월 2일의 Landsat TM 데이터이고 지역은 서울의 한강과 행주대교 및 그 인근 지역이며 보조 도구로 동일 지역의 국립지리원의 1:10000 지형도를 사용하였다. 본래 Landsat TM의 장면 당 지상 면적은 170×185 km이고 5965×6920개의 화소로 구성되어 있는데 앞서 소개한 지역에 해당되는 부분만 선택하여 136×136개의 화소로 실험을 수행했다. 또, 대상이 되는 지역에는 많은 범주가 있을 수 있지만 수역(water), 곡물(crop), 도시 지역(urban), 숲(forest)의 4가지로만 제한하여 실험을 수행했다.

5.2 실험 내용

실험으로 사용하고자 하는 다중 분광 이미지 데이터의 분포 특성은 다음과 같다. 실험에 사용된 136×136개의 화소 중 명확히 4개의 범주에 해당되지 않는 화소와 4개의 화소에 포함될 가능성이 있는 화소들을 제외하고, 명확히 4개의 범주에 해당되는 화소로만 제한했기 때문에 전체 화소의 수는 불규칙할 수 있다. 또한, 밴드에 대한 분포 특성을 모두 여기에 나타내지는 못했다.

5.3 다중 분광 이미지 데이터의 분포 특성표

학습 데이터의 패턴 분포 특성을 각 범주별로 분석하면 다음과 같다.

crop (band 2)

분광 강도	화소의 수	비중 (%)	히스토그램
47	2	1.67	**
48	20	16.67	*****
49	38	31.67	*****
50	44	36.67	*****
51	15	12.50	*****
52	1	0.83	*

(a)

crop (band 7)

분광 강도	화소의 수	비중 (%)	히스토그램
8	1	0.83	*
9	9	7.50	*****
10	13	10.83	*****
11	16	13.33	*****
12	6	5.00	*****
13	9	7.50	*****
14	17	14.17	*****
15	13	10.83	*****
16	7	5.83	*****
17	7	5.83	*****
18	4	3.33	****
19	4	3.33	****
20	3	2.50	***
21	1	0.83	*
22	1	0.83	*
23	1	0.83	*
24	3	2.50	***
25	3	2.50	***
26	0	0	
27	2	1.67	**

(b)

그림 9. (a) 밴드 2, (b) 밴드 7에서의 곡물에 대한 다중 분광 데이터.

5.4 Rough 집합을 이용한 밴드 특징 추출

이와 같은 분포 특성을 갖는 데이터를 4.1, 4.2의 밴드 특징 추출 방법에 의해 시뮬레이션 프로그램으로 수행한 결과는 다음과 같다. 실험의 특성상 실험 결과를 본 논문에서 상세히 기술하는 것은 어렵기 때문에 실험 결과를 분석하여 기술하기로 한다. 밴드 6은 거의 공통적으로 식별 가능 즉, 조합 불가능한 밴드로 판명되었다. 다른 밴드에 비해 분광 강도의 값이 전체적으로 높게 나타났기 때문이다. 이는 Landsat TM의 경우 밴드 6은 열 밴드(thermal band)이기 때문인 것으로 사료된다. 또한, 화소의 수가 과다하지 않다면 (10개 내외) 전체적으로 밴드 2, 4, 7은 식별 불가능한 밴드로 분석되어 조합 가능하다는 결론을 내릴 수 있었다. 물론 화소의 수를 약간 줄이면 다른 밴드도 조합 가능할 수 있었으나 그 수가 늘어날수록 점차 밴

urban (band 2)

분광강도	화소의 수	비중(%)	히스토그램
56	2	1.55	**
57	2	1.55	**
58	0	0	
59	2	1.55	**
60	0	0	
61	8	6.20	*****
62	15	11.63	*****
63	21	16.28	*****
64	18	13.95	*****
65	19	14.73	*****
66	16	12.40	*****
67	10	7.75	*****
68	3	2.33	***
69	4	3.10	****
70	2	1.55	**
71	2	1.55	**
72	5	1.33	*****

(a)

urban (band 4)

분광강도	화소의 수	비중(%)	히스토그램
61	2	1.55	**
65	1	0.78	*
67	1	0.78	*
71	1	0.78	*
72	1	0.78	*
73	3	2.33	***
74	3	2.33	***
75	4	3.10	****
76	6	4.65	*****
77	6	4.65	*****
78	2	1.55	**
79	6	4.65	*****
80	6	4.65	*****
81	7	5.43	*****
82	12	9.30	*****
83	8	6.20	*****
84	7	5.43	*****
85	9	6.98	*****
86	6	4.65	*****
87	5	3.88	****
88	6	4.65	*****
89	8	6.20	*****
90	5	3.88	****
91	5	3.88	****
92	2	1.55	**
93	3	2.33	***
95	4	3.10	****

(b)

그림 10. (a) 밴드 2, (b) 밴드 4에서의 도시지역에 대한 다중 분광 데이터.

드 2, 4, 7로 모이는 경향이 발견되었다. 단, 화소의 수가 과다하면 그 어느 밴드도 식별 가능하여 방대한 화소들을 동시에 처리하는 데는 약간의 문제점이 있

[water]

	밴드 2	밴드 4	밴드 7
최소 분광 강도	44	29	5
최대 분광 강도	50	33	9
평균	46.87	30.12	6.96
표준편차	1.41	0.56	0.94
중간값	47	30	7

[crop]

	밴드 2	밴드 4	밴드 7
최소 분광 강도	47	46	85
최대 분광 강도	52	60	27
평균	49.44	52.31	14.36
표준편차	0.99	2.80	4.28
중간값	49.5	52	14

[forest]

	밴드 2	밴드 4	밴드 7
최소 분광 강도	56	61	47
최대 분광 강도	72	95	122
평균	64.50	82.74	91.27
표준편차	3.06	6.57	14.44
중간값	64	83	94

[urban]

	밴드 2	밴드 4	밴드 7
최소 분광 강도	41	74	21
최대 분광 강도	53	111	53
평균	45.31	90.77	30.96
표준편차	3.09	8.66	9.51
중간값	44	90	27

그림 11. 학습 데이터의 패턴 특성.

을 수 있음이 밝혀졌다. 다음 그림 12는 데이터 분석 시 발생한 식별 가능·불가능 여부에 대한 판단 결과이다.

### 5.5 실험 결과의 평가

일반적으로 ERDAS와 같은 대부분의 원격 탐사 데이터 처리 도구들은 수역, 곡물, 숲, 도시 지역 등을 분류하는데 밴드 2, 4, 7을 사용한다[10]. 본 논문에서 Rough 집합을 이용하여 밴드의 특징 추출을 수행한 결과 역시 밴드 2, 4, 7은 상호 식별 불능 관계임을 알 수 있었고 이들은 상호 조합 가능한 관계로 해석할 수 있다. 물론 분류의 기준이 되는 화소와 그 수에 따라 군집화 경향이 약간 다를 수는 있으나 전체적으로는 밴드 2, 4, 7로 군집화된다고 보는데 큰 무리가 없을 것으로 사료된다. 따라서, Rough 집합을 이용한 결과와 기존의 물리적, 파장적 특성을 중심으로 한 밴드 선택방법과 동일함을 확인할 수 있었고 이를 확장



화소 수	2개 이하	3~5개	6~8개	8~10개	10개 초과
군집화 경향	불규칙한 군집화	$[B_1, B_3]$ $[B_2, B_4, B_7]$ $[B_5]$	$[B_1, B_5]$ $[B_2, B_3, B_4, B_7]$	$[B_1, B_3, B_5]$ $[B_2, B_4, B_7]$	불규칙한 군집화

그림 12. 실험의 종합적 결과.

하변 수백개의 밴드를 갖는 초다중 분광 환경하에서도 이와 같은 방법으로 용이하게 밴드의 특징을 추출할 수 있을 것으로 평가된다.

## 6. 결 론

앞의 실험 결과와 같이 본 논문에서 제안하는 밴드 선택 즉, 특징 추출 방법에 의하면 기존의 밴드 조합시 사용하는 밴드들을 똑같이 추출해낼 수 있음을 알 수 있다. 이미 언급했듯이 본 논문에서의 특징 추출 방법은 다중 분광 환경이 초다중 분광 환경으로 발전할 가능성을 가정하여 다중 분광 환경으로부터 합리적인 특징 추출 근거를 제시하는 것이 그 연구의 주목적이다. 따라서, 초다중 분광 환경으로 상황이 확장되었을 때도 본 논문에서 제안하는 방법을 그대로 활용함으로써 좀더 용이한 특징 추출을 통해 이미지 데이터의 분석 및 분류 효율을 높일 수 있을 것으로 사료된다.

향후로는, 특징 추출뿐 아니라 각 화소의 분류 과정에도 Rough 집합 이론을 도입하여 분류 정확도를 향상시킬 수 있는 방법이 모색될 수 있는 연구가 진행되어야 할 것이다.

## 참고문헌

- [1] P. D. Heermann and N. Khazenie, "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Trans. on Geosci & Remote Sensing*, **30**(1), Nan., 1992.
- [2] A. R. Huete, "Separation of soil-plant spectral mixtures by factor analysis," *Remote Sensing of Environmental*, **19**, pp. 237-251, 1986.
- [3] L. Jimenez and D. Landgrebe, "Supervised classification in high dimensional space :geometrical, statistical and asymptotical properties of multivariate data," *IEEE Trans. on Systems, Man, and Cybernetics*, Jan., 1998.
- [4] J. I. Kim and S. C. Kim, "Fuzzy approaches for improving a classification accuracy in remote sensing image data," *Proceedings of IEEE/SMC '94 Conference*, **1**, pp. 354-359, Oct., 1994.
- [5] J. I. Kim and S. C. Kim, "A fuzzy partitioning method of spectral space for remote sensing image classification," *Proceedings of FUZZ-IEEE/IFES '95 Joint Conference*, **3**, pp.1125-1130, 1995.
- [6] F. J. Knorr and J. H. Futrell, "Separation of mass spectra of mixtures by factor analysis," *Anal. Chem.*, **51**, pp. 1236-1241.
- [7] J. H. Lee and W. D. Philpot, "Spectral texture pattern matching: a classifier for digital imagery," *IEEE Trans. on Geosci. & Remote Sensing*, **29**(4), pp. 545-554, July, 1991.
- [8] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Science*, 1982.
- [9] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publisher, 1991.
- [10] ERDAS Field Guide, ERDAS Inc., 1991.
- [11] *Management & Computer*, 2월호, 1988.
- [12] 김진일, "위성 탐사 이미지 분류를 위한 다중 분광 패턴 공간에서의 퍼지 규칙 생성에 관한 연구," 서강대학교 대학원, 박사학위논문, 1994.
- [13] 심재홍 역, 정보위성, 군사위성이 벌이는 정보전쟁의 세계, 민컴, 1991.
- [14] 양영규 외 9인, "원격 탐사 관련 기술 실용화 연구 (I, II, III)," 한국과학기술연구원, 시스템공학센터, 1889-1991.
- [15] 원성현, "러프 집합에서의 식별 불능 관계를 이용한 다중 분광 이미지 데이터의 밴드 분류," *경영정보연구*, 제1호, pp. 401-412, 1997.
- [16] 원성현, 이병성, 정환목, "러프 집합을 이용한 다중 분광 이미지 데이터의 분류," *한국퍼지 및 지능시스템학회 학술발표논문집*, 제7권, 제 1호, pp. 205-208, 1997.



**원성현(Sunghyun Won) 정회원**

1964년 12월 19일생  
1990년: 서강대학교 전자계산학과 졸업(학사)  
1992년: 서강대학교 대학원 전자계산학과 졸업(석사)  
1998년: 대구효성가톨릭대학교 대학원 전산통계학과 졸업(박사)

1992년~현재: 지산대학 전산정보처리과 교수 재직 중  
주요관심분야: 지능시스템공학, 퍼지시스템, 신경망 컴퓨터, 지능형 교수 시스템 등

**정환목(Hwanmook Chung)**

한국퍼지 및 지능 시스템학회 논문지  
제 7권 제 2호 참조