

## 분광 상호정보를 이용한 하이퍼스펙트럴 영상분류

# Classification of Hyperspectral Images Using Spectral Mutual Information

변영기\* · 어영담\*\* · 유기윤\*\*\*

Byun, Young Gi · Eo, Yang Dam · Yu, Ki Yun

### 要 旨

하이퍼스펙트럴 영상자료는 객체에 대한 많은 정보를 함유하고 있어 객체의 보다 정확한 분류가 가능하다. 본 논문에서는 하이퍼스펙트럴 영상분류를 위하여 SMI(Spectral Mutual Informaiton)이라는 새로운 스펙트럼 유사도 측정기법을 제안하였다. 본 방법은 정보이론 분야에서 대두된 상호정보량의 개념을 차용하여 고안되었으며 스펙트럼간의 통계적 의존성을 측정할 수 있다. SMI는 영상의 각 화소스펙트럼을 확률변수로 간주하고 두 스펙트럼간의 유사 상호정보량을 통하여 유사도를 측정함으로써 영상을 분류한다. 제안된 기법의 효율성을 평가하기 위해 기존에 개발된 SAM, SSV 분류기법을 이용하여 동일지역에 대해 분류를 수행하고 분류 정확도를 비교 평가하였다. 실험결과 제안한 SMI 기법은 하이퍼스펙트럴 영상분류에 유용하게 적용될 수 있으리라 판단된다.

**핵심용어** : 상호정보, 하이퍼스펙트럴, 엔트로피, 토지피복분류

### Abstract

Hyperspectral remote sensing data contain plenty of information about objects, which makes object classification more precise. In this paper, we proposed a new spectral similarity measure, called Spectral Mutual Information (SMI) for hyperspectral image classification problem. It is derived from the concept of mutual information arising in information theory and can be used to measure the statistical dependency between spectra. SMI views each pixel spectrum as a random variable and classifies image by measuring the similarity between two spectra form analogy mutual information. The proposed SMI was tested to evaluate its effectiveness. The evaluation was done by comparing the results of preexisting classification method (SAM, SSV). The evaluation results showed the proposed approach has a good potential in the classification of hyperspectral images.

**Keywords** : Mutual Information, Hyperspectral, Entropy, Land cover classification

## 1. 서 론

하이퍼스펙트럴 영상자료는 지표물의 분광반사특성곡선(spectral reflectance curve)정보를 제공하고, 다중분광 영상자료에 비해 높은 분광해상도를 가지므로 지표물에 대한 세밀한 분석이 가능하다. 그러나 하이퍼스펙트럴 영상의 많은 밴드수로 인하여 신뢰도가 낮은 트레이닝 데이터를 취득할 경우 영상을 분류하는데 있어서 오히려 낮은 분류정확도를 나타낼 수 있다. 따라서 기존의 다중분광 영상에 적용되었던 일반적인 분류기법들을 그대로 적용하는데 어려움을 겪을 수 있다.

이러한 문제를 해결하기 위해 하이퍼스펙트럴 데이터

에 적합한 분류 알고리즘 개발에 대한 연구가 활발히 진행되어왔다.

특히 물체의 분광특성을 가장 잘 표현하는 하이퍼스펙트럴 영상의 장점을 최대한 활용하여 보다 정확한 지표물의 종류를 측정하기 위한 여러 가지 스펙트럼 분석기법들이 제시되고 있다(James, 2003; Chein-I Chang, 1999; Shon et al., 2002). 스펙트럼 분석기법은 기존에 알려져있는 대상물체의 기준 분광반사값을 이용하여 하이퍼스펙트럴 영상에서 얻어지는 반사값과의 분광특성 유사성을 분석하여 각 화소에 대한 대상물체의 종류를 식별하는 방법으로써, 유사도를 측정하는 방법과 기준에 따라 다양한 분류 알고리즘이 연구되어 왔다. 기존의 하이퍼스펙

2007년 7월 16일 접수, 2007년 8월 21일 채택

\* 서울대학교 공과대학 지구환경시스템공학부 공간정보연구실 박사과정 (kko071@snu.ac.kr)

\*\* 국방과학연구소 기술연구본부 선임연구원 (eoandrew@yahoo.co.kr)

\*\*\* 정희원·서울대학교 공과대학 지구환경시스템공학부 부교수 (kiyun@snu.ac.kr)

트럴 영상에 적용된 유사도 측정방법에는 크게 거리계수, 상관계수, 거리계수와 상관계수를 통합한 방법 등으로 나눌 수 있다.

거리계수(distance coefficient)는 다양하게 정의된 스펙트럼 공간상에서의 거리를 가지고 스펙트럼간의 비유사성을 측정하는 방법이다. 대상간의 거리는 유사성의 반대 개념으로, 유사도가 높은 두 스펙트럼간의 거리는 짧다. 하이퍼스펙트럴 영상분류에 사용된 대표적인 거리계수로는 유클리드거리(euclidean distance)와 SAM(Spectral Angle Mapper)등이 있다(Shon et al., 2002).

상관계수(correlation coefficient)는 비교하고자 하는 두 스펙트럼을 표현하고 있는 화소값들의 벡터쌍에 대한 상호관계도를 측정하는 방법이다. 하이퍼스펙트럴 영상분류에 대표적인 상관계수는 피어슨 상관계수에 기반한 SCS(Spectral Correlation Similarity)기법이 있다(Carvalho et al., 2000). 또한 최근에는 유클리드거리와 상관계수를 통합한 방법인 SSV(Spectral Similarity Value)기법이 제안된 바 있다(James, 2003).

하지만 이러한 기존의 유사도의 측정방법들은 두 변수간에 선형관계를 갖는지의 여부와 그 강도를 측정하는 결정론적인 방법으로 변수 간의 상호 의존성을 반영하지 못한다.

이에 본 연구에서는 두 변수간의 유사도 측정문제에 대하여 정보 이론적(information-theoretic)개념인 상호정보를 차용한 통계적 접근방법인 SMI(Spectral Mutual Information)기법을 제안하고 이를 하이퍼스펙트럴 영상에 적용하여 제안된 방법의 성능을 평가하는데 그 목적이 있다. 본 방법은 두 스펙트럼이 완전 일치할 경우 SMI값이 최대화된다는 가정 하에 기준스펙트럼과 대상스펙트럼간의 유사도를 측정함으로써 영상을 분류한다.

## 2. 유사 상호정보량 최대화를 통한 영상분류

일반적으로 하이퍼스펙트럴 영상을 이용한 감독분류(Supervised classification)의 과정은 사용자가 각 분류항목의 특성을 지정하는 트레이닝(Training)단계와 트레이닝 단계에서 추출된 특징을 기반으로 하는 분석단계로 이루어진다. 트레이닝 단계에서는 영상에서 각 분류항목별로 확실한 영역(Training Field)을 지도나 영상을 통하여 설정하고, 이 트레이닝 데이터로부터 분류항목별 특징을 추출하게 된다. 분석단계에서는 선정된 특징값을 이용하여 일정한 기준에 의하여 모든 화소들을 순차적으로 분류하게 된다. 이때 분류 기준설정 방법에 따라 SAM, SSV등 다양한 기법이 적용된다.

본 연구에서는 분류항목별 특징으로서 평균벡터를 이

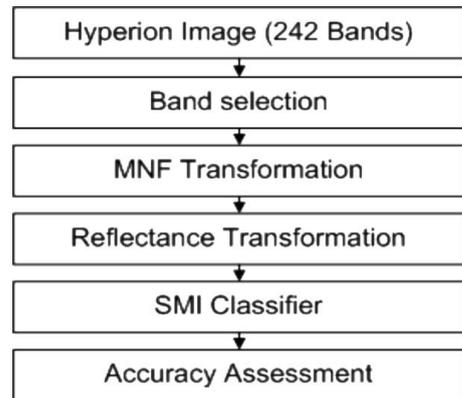


그림 1. 연구 흐름도

용하였고, 이 항목별 평균스펙트럼을 참조스펙트럼으로, 영상의 개별 화소 스펙트럼을 대상 스펙트럼으로 정의하고, 두 스펙트럼이 완전 일치할 경우 유사상호정보로부터 유도된 SMI 값이 최대화된다는 가정 하에 스펙트럼간의 유사도를 측정함으로써 영상분류를 수행하였다. 그림 1은 본 논문에서 제안한 SMI 최대화를 통한 영상분류 단계의 연구 흐름도를 나타낸 것이다.

### 2.1 엔트로피와 상호정보량

엔트로피는 원래 물질계의 열적상태를 나타내는 물리량으로 계의 무질서 정도를 나타내는 지표로 사용될 수 있다. Shannon은 이러한 엔트로피 개념을 정보이론 영역에 도입하여 정보량을 정의하였다. 엔트로피 개념을 확장하여 상호정보량(mutual information)을 정의하며, 이는 한 변수가 다른 변수에 대해 담고 있는 정보의 양을 측정하는 척도로 두 변수의 유사성 측정에 사용될 수 있다. Shannon의 엔트로피는 식 (1)과 같이 정의된다.

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (1)$$

이때  $X$ 는  $x$ 를 값으로 갖는 이산 확률 변수를 의미하고  $H(X)$ 는  $X$ 에 대한 엔트로피, 그리고  $p(x)$ 는  $x$ 가 발생할 확률값을 의미한다. 엔트로피는 확률변수에 대해 확률이 균등하게 분배되어 있을수록 큰 값을 가지게 된다.

상호정보량은 한 확률 변수가 다른 확률 변수에 대해 담고 있는 정보량에 대한 척도이다. 또는 다른 확률 변수에 대한 지식으로 인해 한 확률 변수에 대한 불확실성이 줄어드는 정도이다. 두 확률 변수  $X$ 와  $Y$  및 이들의 결합확률 질량함수  $p(x, y)$ , 주변(marginal)확률 질량함수  $p(x)$ 와  $p(y)$ 가 주어져 있을 경우, 상호정보량  $I(X; Y)$ 는 식 (2)와 같이 정의된다(Cover et al., 1991).

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (2)$$

상호정보량  $I(X; Y)$ 의 정의를 다시 쓰면 다음과 같이 엔트로피와 조건부엔트로피, 결합엔트로피로 표현될 수 있고 이를 통해 각종 엔트로피와 상호정보량의 관계를 확인 할 수 있다.

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned} \quad (3)$$

식 (3)을 통해 상호정보량  $I(X; Y)$ 는  $Y$ 에 관련된 지식으로 인해  $X$ 의 불확실성이 줄어드는 정도이며, 한 확률 변수 자신에 대한 상호정보량은 확률변수의 엔트로피임을 알 수 있다. 이러한 이유로 엔트로피는 때때로 자기 정보(self-information)라고 불리기도 한다. 식 (3)에서 알 수 있듯이, 각 확률변수의 엔트로피와 조건부엔트로피 또는 결합엔트로피를 알고 있을 경우 두 확률변수의 상호 의존성을 나타내는 상호정보량을 구할 수 있다.

### 2.2 Spectral Mutual Information

본 연구에서 하이퍼스펙트럴 영상분류 시, 트레이닝 데이터로부터 추출된 클래스 항목별 참조 스펙트럼과 영상화소 스펙트럼과의 유사도 측정을 위해 유사상호정보량에 근간을 둔 SMI 유사계수를 고안하였다. 주어진 참조 스펙트럼 벡터  $X = (x_1, \dots, x_N)^T$ 와 영상화소 스펙트럼 벡터  $Y = (y_1, \dots, y_N)^T$ 는 적절한 확률분포를 정의함으로써 확률변수로 취급할 수 있다. 스펙트럼 벡터는 파장과 분광 반사값을 축으로 하는 히스토그램을 볼 수 있기 때문에 각각의 벡터 원소를 다음과 같이  $p_i = x_i / \sum_{i=1}^N x_i$ ,  $q_i = y_i / \sum_{i=1}^N y_i$  정규화 함으로서 확률변수  $X$ ,  $Y$ 에 대한 확률 벡터  $p$ ,  $q$ 를 구할 수 있게 된다. 본 연구에서 제안한 SMI 계수는 식 (3)의 상호정보량에서 결합엔트로피  $H(X, Y)$ 를 두 확률분포 합인 엔트로피  $H(X+Y)$ 로 대체하여 다음과 같이 정의한다.

$$\begin{aligned} SMI(X; Y) &= H(X) + H(Y) - H(X+Y) \quad (4) \\ H(X) &= \sum_{i=1}^N p_i \log_2 p_i, \quad H(Y) = \sum_{i=1}^N q_i \log_2 q_i \\ H(X+Y) &= \sum_{i=1}^N (p_i + q_i) \log_2 (p_i + q_i) \end{aligned}$$

고안된 SMI 측도는 상호정보량을 구하는 식과 유사하지만, 확률분포에 대해서만 정의될 수 있는 상호정보량의

제한조건을 만족하지 않기 때문에 상호정보량의 성질을 차용한 유사상호정보량에 근간을 둔다 할 수 있다.

제안한 SMI 측도의 물리적 의미는 그림 2에서와 같이 두 개의 정규분포  $X$ ,  $Y$ 와 이 두 분포를 더해 만든  $X+Y$  분포를 통해 확인 할 수 있다.

$X+Y$ 의 분포의 경우 전체값을 더하면 2가 되기 때문에 엄밀한 의미에서  $X+Y$ 는 확률분포라 할 수 없다. 따라서  $X+Y$  분포에 대해서는 엔트로피를 구할 수 없지만 식 (1)의 엔트로피 계산식을 그대로 적용하여  $X+Y$ 에 대한 유사 엔트로피를 구할 수는 있다.

SMI의 값은  $X$ 와  $Y$ 가 겹쳐있는 정도가 클수록, 즉 다시 말해 두 분포간의 유사성이 클수록 더욱 큰 값을 갖게 되는데 이를 살펴보기 위해 실제로 분포  $X$ 와  $Y$ 의 상대적인 위치를 변화 시키면서 구한 차이 값을 그려보면 그림 3 과 같이 정규곡선과 유사한 모양을 갖게 된다. 여기에서는 편의상 곡선의 최소값을 0보다 크게 만들기 위해 곡선을  $y$ 축에 대해 적당히 평행 이동하여 도시하였다.

그림 3에서와 같이 SMI은 두 분포간의 거리에 따라 정규분포와 유사한 확률분포를 갖기 때문에 두 분포간의 유사도에 대한 확률적으로 유의미한 지표로 사용될 수 있다.

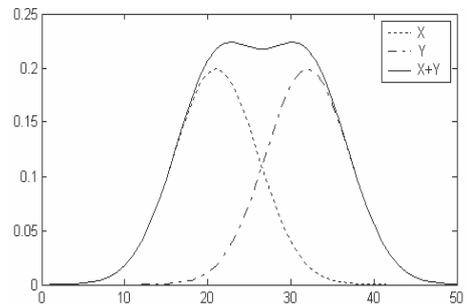


그림 2. 두 개의 정규분포  $X$ ,  $Y$ 와 이 두 분포를 더한 새로운 분포  $X+Y$

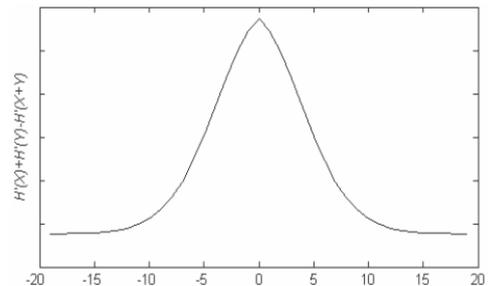


그림 3. 두 정규분포간의 거리에 따른  $H(X) + H(Y) - H(X+Y)$  값의 변화

### 3. 연구 대상지역 및 자료처리

#### 3.1 연구 대상지역

Hyperion 센서는 가시/근적외(VNIR) 분광계와 단파적외(SWIR) 분광계로 구성되어 있다. 다중분광센서와 달리 연속적인 분광영역에 대해 242개의 고유한 분광채널을 갖는다. USGS에서 제공하는 레벨-1B 영상의 경우에 두 분광기 자료의 정렬 및 불량 픽셀 보간, 포화된 검출소자 보고서 등을 통하여 보정이 이루어졌으나, 여전히 검출소자의 민감도 차이로 인하여 많은 밴드에서 수직 방향의 보정되지 않은 줄무늬를 볼 수 있다. 본 연구에서 실험에 사용한 자료는 2002년 4월 3일에 취득한 EO-1 Hyperion 센서 영상으로 안양과 군포 일대의 250pixels × 500pixels 영상을 대상지역으로 사용하여 알고리즘을 평가하였다. 자세한 제원은 표 1과 같다.

#### 3.2 전처리 과정

하이퍼스펙트럴 센서의 특성상 매우 많은 영상촬영소자를 이용하게 되므로 측정기기 내의 열로 인한 잡음 현상과 검출기의 불량이나 민감도 등의 이유로 영상 내에 노이즈가 발생한다. 그러므로 사용자의 Band selection이나 노이즈제거와 같은 전처리 과정이 필요하다. 본 연구에서는 초기 Hyperion 영상에 수 많은 수직선 형태의 노이즈들이 많이 존재하므로 이를 고려하여 Band selection 과정과 원영상에 MNF 변환을 실시하여 신호대잡음비(SNR)가 매우 우수한 MNF 성분들만을 이용하는 방법을 적용하여 영상 전처리를 수행하였다(최재완 등, 2006).

##### 3.2.1 MNF(Minimum Noise Fraction) 변환

MNF 변환은 주성분분석(PCA) 변환과 유사한 변환이다. 하지만 자료들의 산포도를 최대화하는 주성분분석과는 달리, MNF변환은 신호대잡음비를 최대화하는 성분을 선정하는 선형변환이다. MNF변환 후에 생성되는 낮은

성의 성분들에는 증가된 잡음이 존재하기 때문에 필터링 등의 후처리를 실시하고 이를 MNF 역변환을 실시하여 노이즈를 감소시킨다.

##### 3.2.2 Band Selection

MNF를 사용한 후에도 영상 전체가 의미 없는 정보를 담고 있는 밴드를 처리를 할 수 없다. 특히, Hyperion 영상에는 이러한 밴드들이 많이 존재하는데, 정보가 전혀 없는 밴드와 노이즈로만 채워져 있는 밴드로 구분할 수 있다. 이를 제거하기 위해서는 다양한 알고리즘을 적용할 수 있지만, 본 연구에서는 USGS에서 공식적으로 보정되지 않은 밴드로 명명한 밴드 44개와 수작업을 통하여 노이즈라고 판단되는 밴드 63개를 제거하여 242개의 밴드로부터 135개의 밴드만을 추출하였다.

##### 3.2.3 Reflectance Transformation

하이퍼스펙트럴 영상에서 얻을 수 있는 가장 중요한 정보는 각 화소에 해당하는 지표물의 연속적인 분광반사값(spectral reflectance)이라 할 수 있다. 현재 하이퍼스펙트럴 영상은 대부분 센서에서 감지된 복사량(radiance)으로 표현된 자료로 제공되고 있다. 추출된 135개의 밴드 자료는 모두 16bit의 복사량이다. 영상의 복사량은 각 파장에 대하여 지표에서 반사된 에너지의 양이기 때문에 밴드별로 스케일이 모두 다르다. 따라서 이들 값들은 지표의 분광반사 특성 값들로 변화시켜주어야 분류의 성능을 향상시킬 수 있다. 또한, 변환된 분광 반사값들은 0~1의 값을 가지기 때문에 스케일 문제 또한 해결할 수 있다. 분광반사로의 변환식은 식 (5)와 같다.

$$\rho_p = \frac{\pi \times L_\lambda \times d^2}{ESUN_\lambda \times \cos \theta_s} \quad (5)$$

여기서,

- $\rho_p$  : 영상의 반사율
- $L_\lambda$  : 영상의 분광 복사휘도
- $d$  : 지구와 태양사이의 거리
- $ESUN_\lambda$  : 센서의 태양 방사도
- $\theta_s$  : 태양 천정 각이다.

## 4. 분석 및 평가

#### 4.1 데이터 집합

본 연구에서는 영상분류에 적용된 클래스 선정에 대한 객관성을 높이기 위해 환경부에서 제시하는 토지피복분류체계를 사용하였다. 하이퍼스펙트럴 영상의 공간해상도 30m를 감안하여 환경부 토지피복 대분류체계를 기본

표 1. 사용 데이터 제원

항목	특성
위성	EO-1
센서	Hyperion
공간해상도	30m
영상크기	256×974pixels
밴드 수	242
데이터 정량화	16-bit
촬영지역	서울, 경기
데이터 취득일	2002년 4월

으로 하였고 이중에서 실험 대상 영상 내에서 트레이닝 가능하고 충분히 식별 가능한 클래스를 검토한 결과 시가지 지역, 산림, 나지, 수역 등 4개의 클래스를 일차적으로 선정하였다. 하이퍼스펙트럴 영상의 높은 방사해상도와 분광해상도를 고려하여 일차적으로 선정된 시가지 지역 및 산림 지역의 클래스를 밝고 어두운 지역으로 세분하여 총 6개의 클래스를 선정하였다.

본 연구에서는 Visual C++을 이용하여 완전 자동화된 하이퍼스펙트럴 영상분류시스템을 구현하였고 제안된 알고리즘의 분류정확도 측정을 위해서 영상에서 그림 1과 같이 최대한 검증지역이 고르게 분포하도록 각 클래스별 검증데이터를 획득하였다. 표 1은 실험에 쓰인 클래스별 트레이닝데이터와 검증데이터의 개수를 보여주고 있다.

표 1에서 볼 수 있듯이 전체 6개의 클래스가 실험에 사용되었으며, 각 클래스별로 트레이닝데이터로 총 1061개를 구성하였고, 검증데이터로 총 2094개를 정확도 평가에 사용하였다.

표 2. 각 클래스별 데이터 집합

	나지	산림1	산림2	도시1	도시2	수계	합계
트레이닝 데이터	56	429	176	173	38	189	1061
검증 데이터	101	534	468	415	307	269	2094

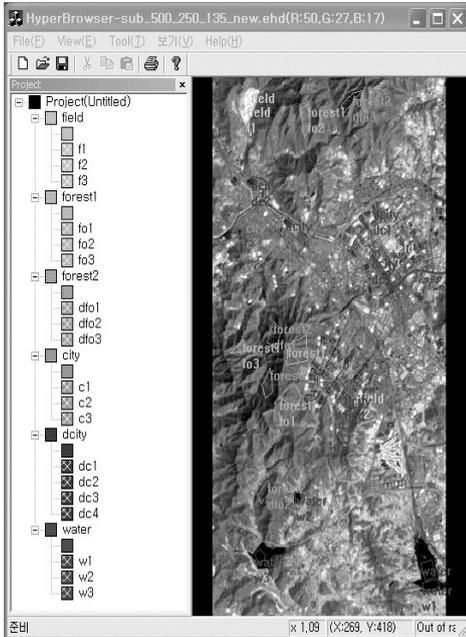


그림 4. 각 클래스별 트레이닝 및 검증데이터

#### 4.2 분류 정확도 평가 측정치

제안한 기법의 정량적인 성능 평가를 위해서 본 연구에서는 오차행렬(Confusion matrix)을 이용한 전체정확도 측정 및 F1-measure법을 적용하여 분류정확도를 측정하였다. 오차행렬을 이용한 분류정확도 평가방법은 위성영상을 이용한 영상분류결과 측정에 있어 어느 정도 표준화된 방법이다. F1-measure는 데이터마이닝 분야에서 많이 사용되는 평가도구로서 정확률(precision)과 재현률(recall)을 함께 고려하여 분류성능을 측정하는 도구이다(Slonim et al., 2000).

$$Precision(i, j) = n_{ij}/n_j \quad Recall(i, j) = n_{ij}/n_i \quad (6)$$

- $n_{ij}$ : 실제로 클래스 i에 속한 화소인데 클래스 j에 속한다고 분류된 화소 수
- $n_i$ : 실제 클래스 i에 포함된 화소 수
- $n_j$ : 분류결과 클래스 j에 포함된 화소 수

정확률은 어떤 클래스에 속한다고 분류된 화소 중 제대로 분류된 화소의 비율을 말하고, 재현률은 실제로 어떤 클래스에 속하는 화소 중 제대로 분류된 화소의 비율을 말한다.  $n_{ij}$ ,  $n_i$ ,  $n_j$ 는 오차행렬에서 바로 구할 수 있다.

본 연구에서는 정확률과 재현률을 동등하게 고려하여 F-measure의  $\alpha$ 의 값을 1로 설정하였고 실제 클래스 i와 분류결과 클래스 j에 대한 F-measure는 식 (7)과 같다.  $n_{ij}=0$ 인 경우는  $F_1(i, j)=0$ 으로 처리한다. 전체 클러스터링 결과는 식 (8)과 같이 측정한다.

$$F(i, j) = \frac{(\alpha + 1) \times Recall(i, j) \times Precision(i, j)}{\alpha \times Precision(i, j) + Recall(i, j)} \quad (7)$$

위 식에서 설명한 F-measure는 각 클래스별의 분류성능을 개별적으로 평가하는 것이다. 모든 클래스에 대한 평균적인 성능을 평가하기 위해 식 (8)을 이용하여 전체 분류성능을 평가한다.

$$F = \sum_i (Max F(i, j)) \frac{n_i}{n} \quad n = \sum_i n_i \quad (8)$$

#### 4.3 실험결과 및 분석

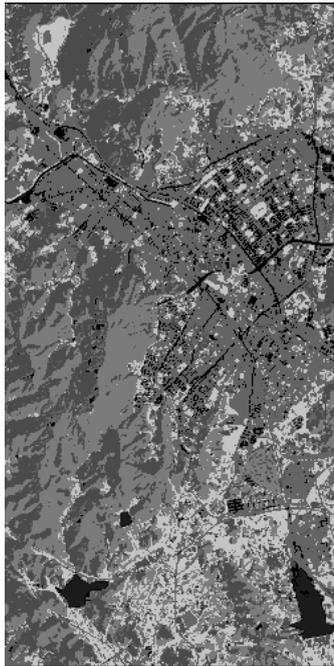
개발된 방법의 성능을 검증하기 위하여 동일데이터에 대한 기존의 기법들 중 가장 정형화된 SAM와 거리계수와 상관계수를 통합한 통합계수 기반의 SSV 기법과의 시각적 비교와 정량적 정확도에 대한 비교 실험을 수행하였다.

그림 5의 결과 영상에서 볼 수 있듯이 SAM 분류기법에 비해 SSV와 SMI 기법을 이용한 분류결과영상이 도심지의 고층 밀집 건물지대와 도로의 선형특성을 잘 나타내

고 있음을 알 수 있다. 하지만 SSV분류결과와 경우 지나치게 나대지 지역을 산림지역으로 오분류 하였고 어두운 도심지를 과대분류 하였음을 알 수 있다. 이는 SSV 분류기법의 특성상 거리계수와 상관계수의 독립성을 가정하므로 인한 편이 때문이라 사료된다. (c)영상은 본 연구에서 제안한 SMI 방법으로 분류한 결과로서, 시각적으로 볼 수 있듯이 전체적으로 토지피복상태를 잘 표현하고 있으나 일부 나대지 지역과 도심지역이 약간 혼합되어진 것을 볼

수 있다.

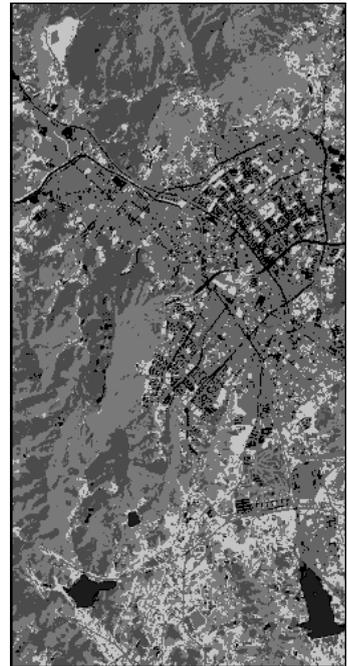
각 기법별 분류결과와 검증 데이터의 클래스 일치여부를 비교한 표 4의 분류정확도 비교표에서 볼 수 있듯이 제안한 SMI 분류결과가 전체정확도, Kappa, F1 measure에서 가장 높은 분류 정확도를 보였다. 제안한 SMI 분류기법의 경우 시각적 평가 및 정확도 평가 결과에서 알 수 있듯이 전체적으로 만족할 만한 결과를 보임을 알 수 있다.



(a) SAM 분류결과



(b) SSV 분류결과



(c) SMI 분류결과

그림 5. 기법에 따른 분류결과 영상

표 3. SMI 분류결과 오차행렬

		분류 결과						
		나지	산림1	산림2	도심1	도심2	수계	계
기준 데이터	나지	92	0	0	9	0	0	101
	산림1	0	485	49	0	0	0	534
	산림2	0	28	363	33	44	0	468
	도심1	14	6	8	378	9	0	415
	도심2	2	0	1	127	177	0	307
	수계	0	0	0	0	1	268	269
	계	108	519	421	547	231	268	2094
overall accuracy(1763/2094)=84.2%								
KAPPA index=0.803 F1-measure=0.84								

표 4. 각 기법별 분류정확도

	SAM	SSV	SMI
Overall accuracy	82.4	76.3	84.2
Kappa	0.78	0.70	0.80
F-measure	0.82	0.75	0.84

## 5. 결 론

본 연구에서는 두 스펙트럼이 완전 일치할 경우 유사상호정보량이 최대화된다는 가정 하에 기준스펙트럼과 대상스펙트럼간의 통계적 의존관계를 계산하는 SMI 유사계수를 제안하였고 이를 하이퍼스펙트럴 영상 분류에 적용하였다.

또한, 기존의 정형화된 SAM 분류기법, 통합계수 기반의 SSV분류기법과 제안된 분류기법을 수행하여 분류정확도 평가한 결과 다음과 같은 결론을 얻었다.

제안한 SMI 기반의 분류기법은 단순히 스펙트럼간의 크기의 차이만을 이용하는 기존의 방법에 비해서, 보다 좋은 분류 결과를 보여줌을 확인 할 수 있었다. 정확도 평가 결과 제안된 분류기법을 통하여 생성된 분류영상이 기존의 방법에 비해 전체 정확도 측면에서 1.8~7.9%, F1 measure에서 0.02~0.09정도 증가함으로써 전반적으로 정확도가 향상되었다. 향후 더 다양한 항목의 분류 및 지형도를 활용한 더욱 정밀한 정확도 평가를 수행해야 할 것이다.

## 감사의 글

본 연구에 대한 서울대학교 공학연구소의 지원에 감사드립니다.

## 참고문헌

1. Chein-I Chang, 1999, Spectral information divergence for hyperspectral image analysis, Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings, IEEE 1999 International, Vol. 1, pp. 509-511
2. N. James, 2003, The Spectral Similarity Scale and its Application to the Classification of Hyperspectral Remote Sensing Data, IEEE Work Shop, pp. 92-99.
3. T.M. Cover, and Thomas, 1991, J.A., Elements of Information Theory, John Wiley & Sons, pp. 1-23.
4. Y. Shon, and N. Sanjay Rebello, 2002, Supervised and Unsupervised spectral Angle Classification, PE &RS, ASPRS, Vol. 68, No. 12, pp. 1271-1280.
5. O. Carvalho., & P. Meneses., 2000, Spectral correlation mapper(SCM): an improvement on the Spectral Angle Mapper (SAM), Airborne Geoscience Workshop Proceedings, JPL, Pasadena, CA,
6. V. Haertel., & D. Landgrebe., 1999, On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 37, No. 5, Part 2, Sept.
7. N., Slonim, and N.,Tishby, 2000, "Document clustering using word clusters via the information bottleneck method", in Proceeding of SIGIR-2000, pp. 208-215.
8. 최재완, 변영기, 김용일, 유기운, 2006, "분광유사도 커널을 이용한 하이퍼스펙트럴 영상의 Support Vector Machine(SVM)분류", 한국지형공간정보학회지, 제14권, 제4호 pp. 71-77.