

Fuzzy C-Mean 알고리즘을 이용한 토지피복분류기법 연구

A study of Land-Cover Classification technique Using Fuzzy C-Mean Algorithm

신석효¹⁾ · 안기원²⁾ · 이주원³⁾ · 김상철⁴⁾

Shin, Sok Hyo · Ahn, Ki Won · Lee Ju Won · Kim Sang Chol

¹⁾ 경상대학교 공학연구원 건설연구센터 연구원(E-mail:s_shshin@gshp.gsnu.ac.kr)

²⁾ 경상대학교 공과대학 토목공학과 교수(E-mail:kwahn@nongae.gsnu.ac.kr)

³⁾ 경상대학교 공학연구원 자동화·컴퓨터 센터 연구원 (E-mail:elecwave@bioee.gsnu.ac.kr)

⁴⁾ 경상대학교 공과대학 토목공학과 박사과정(E-mail:sangcholyi@hanmail.net)

Abstract

The advantage of the remote sensing is extraction the information of wide area rapidly. Such advantage is the resource and environment are quick and efficient method to grasps accurately method through the land cover classification of wide area. Accordingly this study is used to the high-resolution (6.6 m) Electro-Optical Camera (EOC) panchromatic image of the first Korea Multi-Purpose Satellite 1 (KOMPSAT-1) and the multi-spectral Moderate Resolution Imaging Spectroradiometer (MODIS) image data(36 bands).We accomplished FCM classification technique with MLC technique to be general land cover classification method in the content of research. And evaluated the accuracy assessment of two classification method.

1. 서 론

최근의 위성영상들은 1m 이하 급의 공간해상도를 지니기 때문에 고해상도의 위성영상자료는 기존의 Landsat 위성영상이나 SPOT 위성영상에 비해 보다 정밀한 자료획득이 가능한 특성을 가지고 있다. 인공위성 영상 디지털데이터는 지표면에 있는 식물, 토양, 물, 인공구조물 등의 물체로부터 각각 다른 전자파 에너지를 수집, 대상물에 따른 전자파 특성을 이용하여 대상물의 위치, 상태, 크기, 형상 등에 관한 정보를 얻을 수 있다. 위성 영상분류는 지형정보자료 구축을 위한 중요한 자료형태의 하나로서 많이 이용되어 오고 있으며, 객관적인 정보 특성을 갖는 위성 영상을 이용한 지형정보시스템구축에 대한 수요 및 이와 관련된 연구는 최근 들어서 많은 연구가 진행되어 오고 있다. 이러한 영상분류는 원격탐사에서 가장 중요한 분석기법으로 영상을 구성하는 각각의 화소 중 비슷한 분광특성을 갖는 것 끼리 집단화 시켜주는 방법을 말한다. 영상분류에 사용될 인공위성 영상은 되도록이면 고해상도이면서 토지피복상태의 분류가 가능한 다파장대의 영상이어야 한다. 하지만 이러한 고해상도의 영상 중 대표적인 미국의 IKONOS 영상은 고가이면서 한 scene의 피복면적이 좁고, 구입에 소요되는 기간이 많이 걸리기 때문에 시간과 비용의 활용측면에서는 비경제적일 수 있다. 1999년 12월 KOMPSAT-1(Korea Multipurpose Satellite - 1) EOC 위성의 발사에 성공하면서 국내 인공위성자료의 활용수요가 증가하고 있는 추세이다. 또한 다행히도 한국항공우주연구원에서는 해상력(250 m, 500 m 및 1 km)은 다소 떨어지지만 다파장대(36개 bands)인 MODIS(Moderate Resolution Imaging Spectroradiometer) 영상과 해상력(약 6.6 m)이 비교적 높고 영상 취득이 용이한 단 밴드의 panchromatic 영상인 KOMPSAT-1 EOC (Electro-Optical Camera) 영상을 매일 실시간으로 수신하고 있기 때문에 비용과 시간측면에서 경제적인 장점이 있다. 따라서 본 연구에서는 현재 관측 수집되고 있는 KOMPSAT-1 EOC 영상과 MODIS영상을 이용하여 정형화된 분류방법인 MLC분류기법과 Fuzzy C-Mean 알고리즘을 통한 분류기법을 수행하여 두

방법의 분류정확도를 평가 하고자 한다. KOMPSAT-1 EOC 영상과 MODIS영상을 이용하여 분류정확도를 높이기 위한 본 연구의 흐름도는 그림 1과 같다.

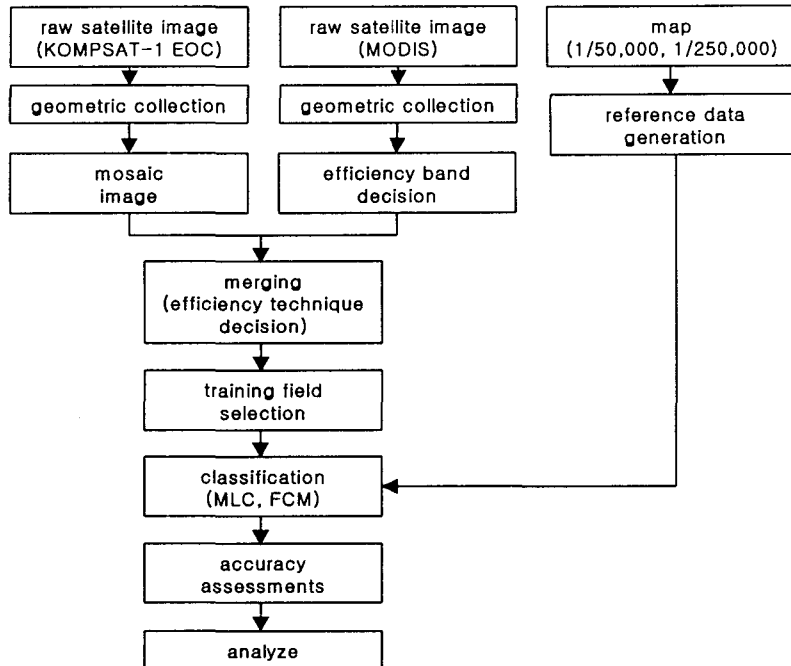


그림 1. 연구흐름도.

2. 연구대상지역 및 사용데이터

본 연구에서 사용된 인공위성 영상데이터는 표 1에서 보는바와 같이 1999년 12월에 우리나라에서 발사하여 현재 운용중인 KOMPSAT-1 EOC 영상과 NASA의 지구관측시스템의 일환으로 개발된 Terra위성과 Aqua위성에 탑재되어 있는 MODIS 영상데이터를 이용하였고, KOMPSAT-1 EOC 영상은 2002년 6월에 수집된 데이터, MODIS 영상은 2002년에 수집된 영상을 이용하여 최대한 경년변화를 줄일 수 있도록 데이터를 수집하였다. 연구대상지역으로 절출된 지역은 북한 황해도 대동강유역 일대 간척사업지역으로서 그 크기는 중횡방향으로 43.8 km × 34.6 km로서 면적은 1515.48 km²이다.

표 1. 사용된 위성영상데이터

Satellite	KOMPSAT-1		TERRA
	Left Orbit Image	Right Orbit Image	
Sensor	EOC	EOC	MODIS
Spatial Resolution	6.6	6.6	250m~1000m
Date of Acquisition	June 29,2002	June 3,2002	July 26, 2002
Number of Bands	1(Pan)	1(Pan)	36
Clouds	0%	0%	0%
Wavelet Length (μm) of Used Bands	0.51~0.73	0.51~0.73	Band1: 0.620~ 0.670 Band2: 0.841~ 0.876 Band3: 0.459~ 0.479 Band4: 0.545~ 0.565

사용된 영상 중 MODIS 영상은 36개 band로서 영상분류에 사용될 band를 결정하여야 하는데 본 연구에서는 해상력이 250 m 또는 500 m 인 band 1 ~ band 7 까지의 7개 영상 중에서 선택하였다. 7개 영상 중에서 KOMPSAT -1 EOC 영상의 분광특성(0.51 ~0.70 μm)을 고려하여 밴드 1 및 밴드 2 영상과 해상도 500 m인 밴드 3 및 밴드 4인 4개의 영상만을 유효밴드로 선정하였다. 여기서 밴드 5, 6, 7영상들은 분광특성이 1 μm이상인 관계로 KOMPSAT -1 EOC 영상과의 분광특성이 많은 차이를 보여 배제를

시켰다. 결정된 MODIS 영상의 band들은 해상력이 최고 250m인 관계로 KOMPSAT -1 EOC 영상과 중합영상을 생성시켜 이용하였다. 중합방법으로는 대표적인 중합방법인 IHS기법과 PCA기법을 통하여 평가해본결과 IHS기법을 통한 중합영상이 PCA 기법을 통한 중합영상보다는 영상분류를 위한 적절한 방법으로 선정되었다. 따라서 본 연구에서도 PCA 중합영상을 이용하여 토지이용분류를 수행하였다.

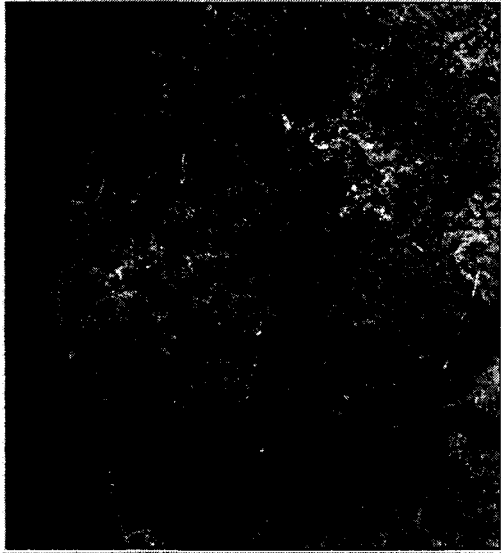


그림 2. KOMPSAT-1 EOC image data.

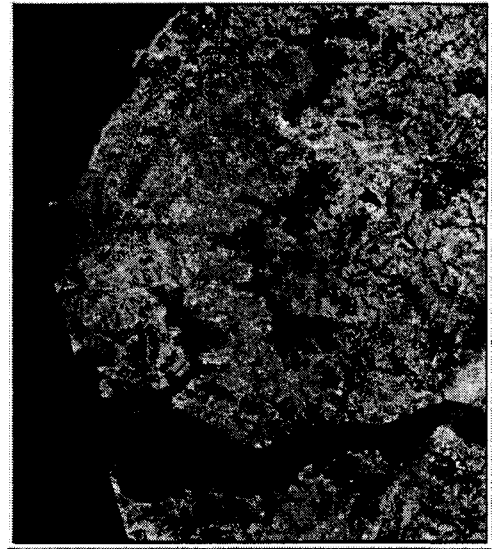


그림 3. MODIS image data.

3. 영상분류

위성 영상분류는 지형정보자료 구축을 위한 중요한 자료형태의 하나로서 많이 이용되어 오고 있으며, 객관적인 정보 특성을 갖는 위성 영상을 이용한 지형정보시스템구축에 대한 수요 및 이와 관련된 연구는 최근 들어서 많은 연구가 진행되어 오고 있다. 이러한 영상분류는 원격탐사에서 가장 중요한 분석기법으로 영상을 구성하는 각각의 화소 중 비슷한 분광특성을 갖는 것 끼리 집단화 시켜주는 방법을 말한다. 분류기법은 감독분류(Supervised Classification)나 무감독 분류(Unsupervised Classification)중 하나를 수행하게 된다. 무감독분류는 군집화 과정이라 하며, 영상자료의 분포특성을 이용하여 분광특성이 다른 여러 개의 군집을 형성하는 순차적 군집화, 통계적 군집화, ISODATA, RGB 군집화 등이 있다. 이 분류기법은 분석자가 몇 개의 간단한 변수만을 지정하여 분류가 실시되며 군집에는 군집의 중심, 밴드간의 공분산등으로써 분광적 분포특성이 표현된다. 무감독분류의 알고리즘은 군집을 형성하는 과정에서 효율성의 기준에 따라 그 접근방법이 다르다. 감독분류는 분석자가 영상 내에서 알고 있는 화소들을 지정하면 각 군집의 중심 밴드간 공분산등을 이용하여 자료내의 모든 화소를 분석하여 가장 유사한 분포특성을 가지는 군집에 할당시키는 과정이다. 감독분류는 각 군집의 유사성을 결정하는 기준에 따라 최소거리법(minimum distance), 평행사변형법(parallel-piped), 최대우도법(maximum likelihood)등이 있다. 최근에는 고전적인 영상분류기법이 가지고 있는 여러 가지 단점을 극복하고 상대적인 분류결과에 영향을 줄 수 있는 여러 가지 환경인자들의 영향을 최소화 시켜 분류 정확도가 뛰어나면서도 단순한 분류법들이 개발되어 오고 있는데 이러한 방식들 중의 하나인 퍼지분류법이 있다. 본 연구에서는 토지피복분류를 위한 선행분류 방법은 일반적으로 많이 사용되는 무감독분류를 먼저 수행하고, 다음으로 감독분류 방법 중 최대우도법을 사용하여 재분류를 수행하였다. 분류항목의 선정에 있어서는 USGS(U.S. geological survey)의 토지분류체계를 기본으로 하여, KOMPSAT-1 EOC 영상과 MODIS 영상의 IHS중합영상, 1:250,000 지세도 및 1:50,000 군용지도를 참고하여 분류항목을 선정하였다. 분류항목은 6개의 항목으로 구분하였다. 또 다른 분류 방법으로는 Fuzzy C mean 방법을 이용하였다.

3.1 FCM(Fuzzy C-Mean)

퍼지 C-mean 알고리즘은 최적 클러스터를 분할하는 반복 분할 방법이다. 이 방법은 클러스터의 중심을 생성하고 클래스의 소속 함수 행렬을 발생시킨다. 최적의 퍼지 클래스 분할은 아래의 식과 같이 최소자승 오차 함수를 최소화시키는 것이다.

$$J_m(U, v) = \sum_{k=1}^N \sum_{i=1}^c (u_{ki})^m \|y_k - v_i\|^2 A \quad (1)$$

여기서, $Y = \{y_1, y_2, y_3, \dots, y_n\} \in R^2$ 인 데이터 집합이고, c 는 $Y: 2 \leq c < n$ 에 포함된 클러스터의 수, m 은 $1 \leq m < \infty$ 상의 지수 가중치이며, $U = \{u_{ki}\}$ 는 Y 의 퍼지 C-분할, $\|y_k - v_i\| A$ 는 R^n 상에 1-norm, A 는 정의된 양의 $n \times n$ 가중치 행렬이다. 가중치 지수 m 은 클러스터에서 관찰 소속에 존재하는 것의 양에 의해 함수 자승 거리 오차를 감소시키는 영향을 가진다. 이때 m 은 최소자승오차함수 J_m 이 최소가 되는 분할이 되도록 점차적으로 단단하게 조정된다. 바꾸어 말하면 m 의 높은 값은 한 샘플의 클러스터 소속을 부드럽게 하는 경향이 있어 이 분할은 점차적으로 부드럽게 된다. 일반적으로 m 은 실험적으로 얻어야 한다. 퍼지 c-mean(FCM)알고리즘의 처리단계는 초기 소속행렬을 랜덤하게 발생시키고, 각 클러스터에 속하는 몇몇의 샘플들의 가중치와 같은 이 랜덤한 소속행렬을 사용한다. 그리고 각 클러스터의 중심을 연산한다. 이 새로운 클러스터의 중심은 소속 행렬의 갱신할 때 사용된다. 수정된 소속행렬은 이전의 값들과 비교하며, 만약 그 차이가 설정된 역치보다 매우 크면, 성능개선을 위해 반복 연산을 지속하고, 아닐 경우에는 알고리즘은 종료한다.

FCM 알고리즘은 다음의 절차로 이루어져 있다.

1. c, A, m, e , 반복횟수 $t-1$ 을 설정한다.
2. $N \times C$ 개의 소속행렬 U 를 랜덤 하게 만든다.
3. 아래 식(2.4.7)으로 각 클러스터의 중심을 연산한다.

$$v_i^{(t)} = \frac{\sum_{k=1}^N (u_{ki}^{(t)})^m y_k}{\sum_{k=1}^N (u_{ki}^{(t)})^m} \quad (2)$$

4. 다음의 식(2.4.8)에 의해 소속 행렬을 수정한다. 여기서 $d_{ki} = \|y_k - v_i\| A$ 이다.

$$u_{ki}^{(t+1)} = \left(\sum_{i=1}^c \left(\frac{d_{ki}}{d_{ki}'} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (3)$$

5. 만약 $\max\{|U_{ki}(t) - U_{ki}(t-1)|\} > e$ 이면 학습 횟수 t 를 증가시키며, 단계 3으로 재실행한다.
6. 학습을 종료한다.

본 연구에서는 이 FCM알고리즘은 이용하여 중합영상의 토지이용상태를 50개의 클러스터로 분류하였다. 이용분류를 하기 위해 FCM의 입력 벡터는 다중밴드의 R(Red image), G(Green image), B(Blue image)가 결합된 $Y = \frac{1}{2}R, G, \frac{3}{4}B$ 을 사용하였다.

3.2 분류결과

본 연구에서는 토지피복분류를 위한 선행분류 방법은 일반적으로 많이 사용되는 먼저 무감독분류를 시행하였다. IHS중합영상 1, 2, 3밴드를 사용하였으며, 50개의 항목으로 분류하였다. 이 때 분류 계산의 반복횟수는 24회, 수렴되는 백분율은 95%가 되었다. 다음으로 감독분류 방법 중 최대우도법을 사용하여 재분류를 수행하였다. 이 방법은 training field가 정규분포를 따른다는 가정 하에 데이터의 평균벡터와 공분산행렬을 계산한 후 확률밀도함수를 이용하여 각 픽셀을 확률이 가장 높은 항목으로 분류하는 방법이다. 분류항목의 선정에 있어서는 USGS(U.S. geological survey)의 토지분류체계를 기본으로 하여,

KOMPSAT-1 EOC 영상과 MODIS 영상의 IHS중합영상, 1:250,000 지세도 및 1:50,000 군용지도를 참고하여 분류항목을 선정하였다. 표 2와 같이 6개의 항목으로 구분하였다.

표 2. Land cover classes for study area

	Land cover		
Land	Vegetation	Forest	Forest bodies
		Non Forest	reclaimed
			vegetation
	farmland		
Non Vegetation	Urban		
Water	Water bodies		

본 연구의 분류 정확도 평가는 분류 오차표(error matrix)의 계산을 통한 총 분류정확도(overall accuracy)와 kappa 계수값(KHAT)으로 평가되었다. 총 분류정확도의 값은 계산과 이해가 용이하여 주로 사용되는 것으로, 분류 오차표에서 대각요소의 합에 대한 전체 참조자료와의 백분율로 표현된다. 이때 대각요소의 값들은 각각의 지피항목에 대해 정확히 분류된 참조자료의 수이다. 반면 kappa 계수값은 대각요소 뿐만 아니라 모든 행렬요소를 대상으로 하는 것이기 때문에, 총 분류정확도보다 분류정확도에 대한 신뢰도가 높다. 이 계수 값은 다음의 식(4)로 표현된다.

$$KHAT = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} \times X_{+i})} \quad (4)$$

이때 r은 분류 오차표에서 행의 수이며, X_{ii}는 행 i와 열 i에서의 값이다. X_{i+}와 X_{+i}는 각각 행 i와 열 i에서의 한계합(marginal totals)이고, N은 참조자료의 전체 개수이다 Kappa 계수값이 0.8~1.0일 경우 분류가 최고로 잘 되었음을 의미하며, 0.6~0.8일 경우 매우 잘된 것임을, 0.4~0.6이면 잘된 것으로 평가된다. 또한 0.2~0.4일 경우 합당하고, 0.0~0.2는 불충분하며 0.0이하일 경우에는 아주 나쁜 것을 의미한다(표 3).

표 3. The grade of KHAT value

Kappa value	Classified grade
0.8 - 1.0	최고의 분류단계
0.6 - 0.8	매우 잘된 단계
0.4 - 0.6	잘된 단계
0.2 - 0.4	보통 단계
0.0 - 0.2	불충분 단계
0.0이하	아주 나쁜 단계

그림 4는 KOMPSAT-1 EOC 중합영상의 최대우도법(MLC) 분류 결과이다. 무감독 분류를 수행한 뒤, 지세도와 군용지도를 통하여 수집한 자료를 사용하여 감독분류 하였다. 그림 5는 FCM(Fuzzy C-Mean) 분류 결과를 나내고 있으며, 다음으로 KOMPSAT-1 EOC 중합영상의 분류결과와 지세도, KOMPSAT-1 EOC 중합영상의 육안판독결과 비교를 통해 표 4~5와 같은 분류 오차표와 총 분류정확도를 계산하였다. 또한 표에서 보는 바와 같이 분류결과의 평가를 위한 정확한 참조자료의 선정은 연구의 성과와 분류정확도 선정을 위해서 중요한 역할을 하게 된다. 따라서 본 연구에서는 분류 정확도를 평가하기 위하여 참조자료를 이용하여 random point 100점을 이용하였다. 이와 같은 평가방법은 참조자료의 크기가 제한적인 관계로 평가된 결과가 신뢰성에서는 미흡한 면이 있지만 대상지역이 비접근 지역으로서 일정한 수준의 정확도가 보장되는 참조자료를 취득할 수 없는 상황에서 선택한 방법이다.

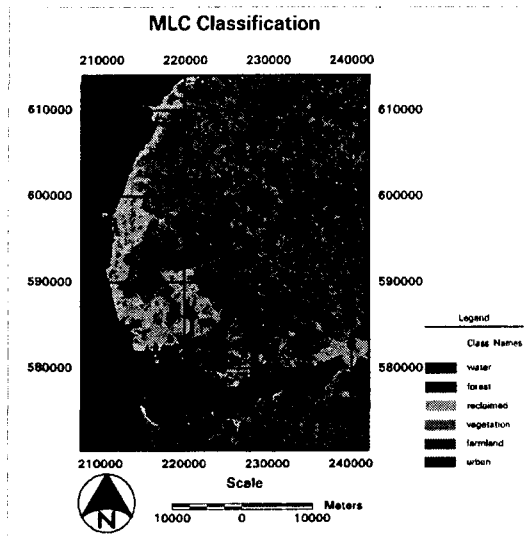


그림 4. MLC Classified result image

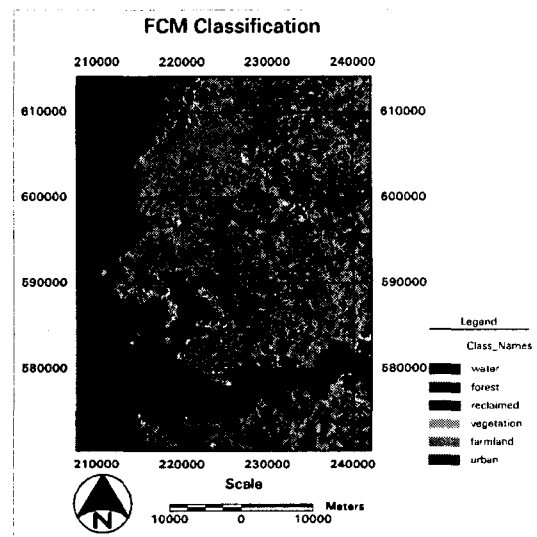


그림 5. FCM Classified result image

표 4. The error matrix of MLC classification

Classified Data	water	forest	reclaimed	vegetation	farmland	urban	Row Total	Accuracy (%)	
water	37	0	0	0	0	0	37	100.00	
forest	0	21	0	2	3	1	27	77.78	
reclaimed	0	0	8	0	0	0	8	100.00	
vegetation	0	1	0	11	1	1	14	78.57	
farmland	0	0	0	0	7	1	8	87.50	
urban	0	0	0	1	0	5	6	83.33	
Column Total	37	22	8	14	11	8	100	-	
Accuracy(%)	100.00	95.45	100.00	78.57	63.64	62.50			
overall accuracy	89(%)								
KHAT	0.8791								

표 5. The error matrix of FCM classification

Classified Data	water	forest	reclaimed	vegetation	farmland	urban	Row Total	Accuracy (%)	
water	31	0	0	0	0	0	31	100.00	
forest	0	21	0	1	0	0	22	95.45	
reclaimed	0	0	5	1	0	0	6	83.33	
vegetation	0	0	0	11	1	0	12	91.67	
farmland	0	0	0	2	16	0	18	88.89	
urban	0	0	0	0	1	10	11	90.91	
Column Total	31	21	5	15	18	10	100	-	
Accuracy(%)	100.00	100.00	100.00	73.33	88.89	100.00	-		
overall accuracy	91(%)								
KHAT	0.9034								

분류 오차표에서 세로방향은 지세도와 KOMPSAT-1 EOC 중합영상의 육안판독결과 즉, 참조자료에 대한 분류정확도(생산자 정확도)를 의미하는 것이고, 가로방향은 토지이용분류 영상의 분류정확도(사용자 정확도)를 의미한다. 이 분류오차표를 사용하여 위성영상자료의 분류결과를 평가한 결과는 다음과 같다. 평가 결과 일반적으로 분류정확도는 전체적으로 높은 정확도를 보였다. 또한, 분류의 특성상 지피 군집의 크기가 상대적으로 크고 동질적이며 연속적인 것에 대해서는 분류정확도가 높게 나타났다. 이들은 주로 수역으로서 하천이나 저수지등의 수역인 water, 그리고 간척지인 reclaimed 의 경우만 생산자 정확도가 100%이고 다른 항목들은 아니었다. 여기서 다른 항목들(임야, 농경지, 도심지, 식생)의 분류정확도가 높지 않은 이유는 구성요소가 다양하고 복잡하며, 불연속적인 관계로 나타남으로서 분류 정확도가 상대적으로 낮게 나타나는 것을 볼 수 있다. 또한 저해상도에서 자주 발생했던 픽셀의 혼합효과에 따른 특성도 나타남으로서 두 요소간의 분류의 오류가 발생되었다고 할 수 있겠다. 결과적으로 표 4와 5를 비교하면 FCM 기법을 적용한 분류영상이 MLC기법을 이용한 분류영상보다 전체 정확도측면에서는 2%, KAPPA 계수값은 0.024정도 증가함을 알 수 있었다.

4. 결 론

본 연구에서는 KOMPSAT-1 EOC 영상과 MODIS영상을 이용하여 정형화된 분류방법인 MLC분류기법과 Fuzzy C-Mean 알고리즘을 통한 분류기법을 수행하여 분류정확도에 대한 평가에서 다음과 같은 결론을 얻었다.

1. FCM 기법을 적용한 분류영상이 MLC기법을 이용한 분류영상보다 전체 정확도측면에서는 2%, KAPPA 계수값은 0.024정도 증가함을 알 수 있었다.
2. 임야, 농경지, 도심지, 식생의 항목 등에서는 구성요소가 다양하고 복잡하며, 불연속적인 관계로 나타남으로서 분류 정확도가 상대적으로 낮게 나타나는 것을 볼 수 있다.
3. MODIS영상과 KOMPSAT-1 EOC이 중합영상을 이용함으로써 저해상도의 단점인 분류항목의 세분화와 판독의 정확성과 신뢰도를 향상 시킬 수 있었다. 또한, 우리나라에서 매일 실시간으로 수신하고 있기 때문에 비용과 시간측면에서 경제적인 장점을 이룰 수 있었다.

참고문헌

안기원, 이효성, 서두천, 신석효 (1998), IRS-1C PAN데이터와 Landsat TM 데이터의 IHS중합화상을 이용한 토지이용분류 정확도 분석, 한국측지학회지, 한국측지학회, 제16권 제2호, pp. 187-194.

정재준, 김용일, 유기윤, (2002), Evidence Theory를 적용한 다중영상데이터의 토지피복분류, 대한토목학회 논문집, 대한토목학회, 제 22권, 제 1-D호, pp.141-150.

최현, 강인준, 홍순현 (2001), 신경망을 이용한 합성영상에서의 분류정확도에 대한 평가, 대한토목학회 논문집, 대한토목학회, 제 21권, 제 1-D호, pp.115-123.

사공호상,황승미(2002), 원격탐사와 GIS 연계 활용방안 연구, 국토연구원, pp. 61-75.

Amini, J. et al (2002), Fuzzy logic system for road identification using IKONOS images, Photogrammetric Record, 99 : pp. 493-504.

Chavez, P.S., Jr. and Howell J.A. (1988). "Comparison of the spectral information content of Landsat Thematic Mapper and SPOT for three different sites in the Phoenix, Arizona region." PE & RS. Vol. 54. No. 12. pp. 1699-1708.

Carper, W.J., Lillesand, T.M. and Kiefer, R.W. (1990). "The use of intensity- hue-saturation transformations for merging SPOT panchromatic and multispectral image data." PE & RS. Vol. 56. No. 4. pp. 459-467.