

훈련지역의 취득방법 및 규모에 따른 JERS-1 위성영상의 토지피복분류 정확도 평가

하성룡¹ · 경천구² · 박상영¹ · 박대희¹

Estimation of Classification Accuracy of JERS-1 Satellite Imagery according to the Acquisition Method and Size of Training Reference Data

Sung-Ryong HA¹ · Chon-Ku KYOUNG² · Sang-Young PARK¹ · Dae-Hee PARK¹

요 약

정량적인 토지피복도의 확보는 유역에 분포하는 비점오염원의 규명에 있어서 매우 중요한 과제로 인식되고 있다. 본 연구는 위성영상을 이용한 토지피복분류 과정에 있어서, 훈련지역의 취득방법 및 규모가 분류정확도에 미치는 영향을 JERS-1 OPS 위성영상을 기반으로 평가하였다. 전체 연구대상지역 중에서 0.3%, 0.5%, 1.0%를 훈련지역으로 추출함에 있어서 두 가지 기법을 제안하였다. 첫번째 기법은 해당지역에 대한 사전 지식을 갖춘 연구자가 훈련지역을 추출하였으며, 두번째 기법은 기하학적 보정을 행한 항공사진과 수치지도를 이용하여 훈련지역을 추출하였다. 영상의 토지피복 분류는 최대우도분류법을 이용하였다. 연구결과 사용자에 의한 훈련지역 취득기법보다 항공사진과 수치지도를 이용하여 훈련지역을 추출하여 최대우도분류법을 적용할 경우 전체정확도가 최대 18% 정도 향상하였다. 우리나라와 같이 복잡하고 다양한 토지이용을 가진 지형에서 JERS-1 영상을 이용하여 95%의 신뢰도를 얻기 위해서는 적어도 훈련지역을 전체지역의 약 1% 이상 추출하여야 만족할 만한 토지피복분류를 수행할 수 있었다.

주요어: 위성영상, 토지피복분류, 훈련지역 취득기법, JERS-1, 최대우도분류법

ABSTRACT

The classification accuracy of land cover has been considered as one of the major issues to estimate pollution loads generated from diffuse landuse patterns in a watershed. This research aimed to assess the effects of the acquisition methods and sampling size of training reference data on the classification accuracy of land cover using an imagery acquired by optical sensor(OPS) on JERS-1. Two kinds of data acquisition methods were considered to prepare training data. The first was to assign a certain

2002년 3월 4일 접수 Received on March 4, 2002

¹ 충북대학교 도시공학과 (simplet@cbucc.chungbuk.ac.kr, sypark119@hanmail.net, idoggy@orgio.net)

Dept. of Urban Eng, Chungbuk Nat'l University

² 한국생산성본부 (ckkyoung@kpc.or.kr) Korea Productivity Center

land cover type to a specific pixel based on the researchers subjective discriminating capacity about current land use and the second was attributed to an aerial photograph incorporated with digital maps with GIS. Three different sizes of samples, 0.3%, 0.5%, and 1.0% of all pixels, were applied to examine the consistency of the classified land cover with the training data of corresponding pixels. Maximum likelihood scheme was applied to classify the land use patterns of JERS-1 imagery. Classification run applying an aerial photograph achieved 18 % higher consistency with the training data than the run applying the researchers subjective discriminating capacity. Regarding the sample size, it was proposed that the size of training area should be selected at least over 1% of all of the pixels in the study area in order to obtain the accuracy with 95% for JERS-1 satellite imagery on a typical small-to-medium-size urbanized area.

KEYWORDS: *Satellite Imagery, Land Cover Classification, Data Acquisition Method, JERS-1, Maximum Likelihood Classifier*

서 론

최근 과학기술의 급속한 발전은 환경파괴와 자원고갈 등의 세계적인 문제를 발생시키고 있으며, 유한한 지구자원의 효율적인 활용과 보존을 위한 방법의 하나로서 원격탐사 기술의 필요성과 중요성이 널리 인식되고 있다. 원격탐사는 지구의 물리적 성질과 기하학적 특성에 관한 제반 정보를 제공할 뿐만 아니라 지구상에 존재하는 모든 대상물의 측정과 해석 및 감시를 가능하게 해준다. 전반적으로 원격탐사기법은 지형도 작성뿐만 아니라 도시의 공간정보의 분석과 활용에 있어서 중요성이 증대되고 있는 지리정보시스템(GIS: Geographic Information System)과의 통합 및 지역적 토지이용 패턴변화의 모니터링 등 여러 분야에서 활용되고 있다. 국내에서도 다목적 실용위성(Korea Multi-purpose Satellite)의 발사로 인하여 보다 저렴하고 신속하게 원격탐사자료를 활용할 수 있을 것이다. 미국의 경우, 환경정보를 GIS나 원격탐사와 연계하여 정책에 활용하고 있다. 미국 지질조사국(USGS: U. S. Geological Survey)은 1970년도 초반에 Landsat 자료를 각종 재해의 예측 및 경보에 사용한 결과 매우 유용한 결과를 얻었으며 홍수지역의 예측에 이들 자료들이 매우 효과적이라는

사실을 발표하였다(김영섭, 1998).

지금까지 많은 연구는 원격탐사 자료를 이용한 토지피복분류 결과의 분석 및 분류 알고리즘의 평가에 초점을 맞추어 왔으며(Fuller 등, 1994; Zhuang 등, 1995; Justin 등, 1997; Steven 등, 1997), 분류알고리즘을 적용하기 위한 자료의 취득 방법 및 규모에 따른 오차의 평가에 대한 연구는 미비하다. 한편, 토지피복 분류에 있어서 추출된 지역의 불확실성을 평가하기 위한 새로운 오차평가 방법들에 대한 연구가 진행되고 있으며(Frank, 1997), 보다 정교한 정확도 검증에 대한 연구가 수행되고 있다(Hardin 등, 1997; Marsh 등, 1993). 따라서, 본 연구에서는 위성영상을 기반으로 토지피복도를 생성하기 위한 분류알고리즘을 적용하는데 있어서, 그 훈련지역의 취득방법 및 규모가 분류정확도에 미치는 영향을 평가해 보고자 한다.

자료 및 연구방법

1. 자료의 취득

본 연구에서 사용한 자료는 일본 JERS-1 (Japanese Earth Resources Satellite)위성의 광학센서(OPS: optical sensor)에서 촬영한 영

상이다. JERS는 이름 그대로 지구 주위 관측과 자원 이용에 초점을 맞추어 일본 전체에 걸친 육지 측량과 농업, 임업, 수산업, 환경보호, 재난 방지, 근해의 감시등을 하는 지구 관측 위성으로 1992년 일본의 다네가시마 우주센터에서 발사되었다. JERS-1위성은 당초 계획 수명이었던 2년보다 훨씬 긴 6년 간 임무를 수행하다가 지난 1998년 10월 19일로 수명을 다해 더 이상 작동할 수는 없게 되었으나, 상대적으로 저렴한 가격과 Landsat TM(30×30m)에 비하여 고해상도의 다중밴드 영상을 제공한다는 장점이 있다. JERS-1 태양동기궤도의 고도는 약 568km이고, 경사각은 약 98도, 재방문주기는 44일, 관측폭은 75km이다. OPS는 18.3×24.2m의 지상해상도를 갖는 8개의 감지밴드를 탑재하고 있다. 즉, 3개의 가시광 및 근적외선밴드와 4개의 중적외선밴드, 1개의 입체시(Stereoscopic) 밴드로 구성되어 있다.

연구지역은 도시, 농경지, 삼림, 수역 등의 지상사상이 골고루 분포되어 토지피복 분류를 위해 좋은 표본 조건을 가지고 있는 충북 청주시와 청원군에 걸쳐 분포하는 미호천 수계의 일부지역 약 263km²를 선정하였다(그림 1). 분석에 사용된 자료는 JERS-1의 OPS에서 1993년 3월 28일에 취득된 path-row 번호 88-238인 디지털 영상이다. 이 영상의 해상도는 원래 18.3m×24.2m이지만 인공위성센터에서 일련의 처리과정을 거쳐 얻어진 데이터는 1화소(pixel)가 18m×18m이다. 연구에 사용된 영상의 TM좌표는 좌상단 (233772.431, 358143.235), (248784.431, 355713.235), 우하단 (231342.431, 343023.235), (246354.431, 340593.235)이다. 또한, 영상데이터의 기하보정에 필요한 지상기준점의 선정, 훈련지역추출, 보정 그리고 검증을 위하여 1995년 5월에 촬영된 항공사진과 이로부터 도화된 1/5,000 수치지도를 사용하였다. 영상의 처리와 분석은 ER Mapper 6.0과 ENVI 3.4 소프트웨어를 사용하였다.

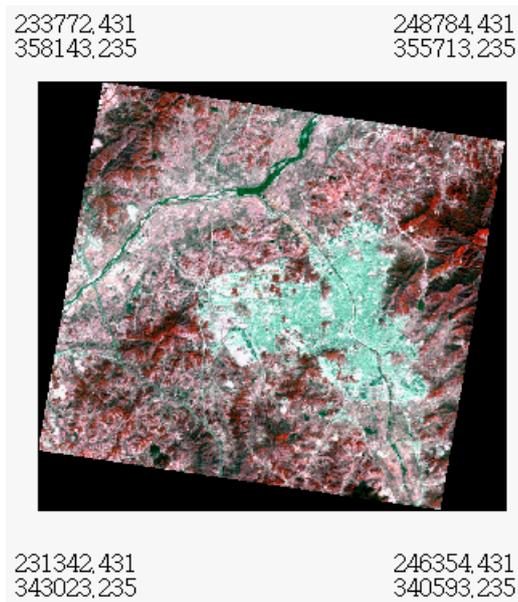


FIGURE 1. The JERS-1 OPS imagery for the study area (3. 28. 1993)

2. 연구 방법

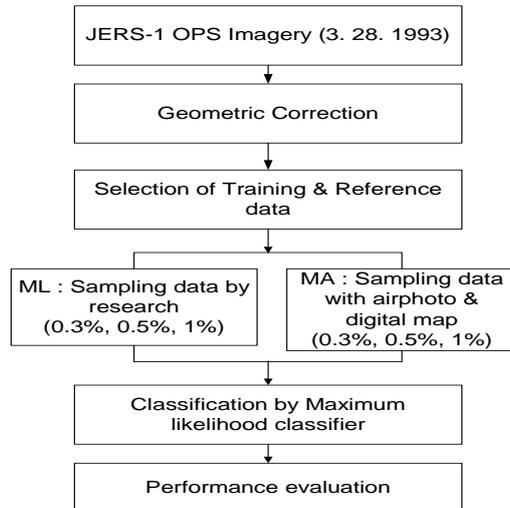


FIGURE 2. The work flow of image processing and analysis

본 연구의 영상처리 및 분석 과정을 그림 2에 도시하였다. 전처리 과정으로써 JERS-1 위

성의 지상 촬영시 발생한 기하학적 왜곡을 보정(geometric correction)하여 기본 영상자료로 이용하였다. 토지피복은 산림, 논, 초지 및 밭, 도시, 수역, 나대지의 6가지 클래스로 나누어 표본추출을 실시하였다. 이는 USGS의 토지이용 및 토지피복분류체계에 의해 JERS-1 데이터의 공간해상도 한계 내에서 분류될 수 있는 분류코드체계를 따른 것이다.

영상의 토지피복 분류에 앞서, 분류 모형의 정확도 평가를 위한 지상검정자료(reference data)의 취득은 연구자에 의한 성장기법 및 항공사진을 이용하였다. 성장기법(growing method)이란 영상에서 확실한 지표피복에 해당하는 화소를 하나 선택하고 임계값을 미리 지정하여 그 주위 화소로 확장하면서 임계값 안에 해당하는 화소를 훈련지역으로 선택하는 기법을 말한다. 이때 자료의 취득방법에 따라서 2개의 분석모형으로 나누어 연구를 진행하였다. 첫번째 모형은 해당지역에 대한 사전 지식을 갖춘 연구자가 전체 연구대상지역 중에서 0.3%, 0.5%, 1.0%를 지상검정자료로써 추출한 모형이다. 두번째 모형은 기하학적 보정을 행한 항공사진과 수치지도를 이용하여 0.3%, 0.5%, 1.0%를 지상검정자료로써 추출한 모형이다. 이것은 토지피복 특성을 대표하는 지상검정자료의 취득 방법 및 규모에 따른 분류 정확도를 평가하기 위한 것이다. 본 연구에서는 첫째 모형을 ML, 두번째 모형을 MA라 명명하였다. 제안된 모형에 대하여 최대우도분류법을 적용한 후, 그 분류정확도를 평가하였다.

3. 기하보정

우리나라에서는 횡메르카토르(Transverse Mercator) 투영법에 의해 지도가 제작되며, 경위도 좌표와 평면직각 좌표체계를 사용하고 있다. 평면직각 좌표체계는 원점으로부터의 거리로 표현하는 좌표로서 한반도를 경도 2도의 폭으로 하여 세 개의 가상좌표구역으로 나누어 사용한다. 본 연구에서는 중부좌표구역에

해당되는 지역으로서 중부원점을 중심으로 하여 원점으로부터의 거리를 미터로 표시한 직각좌표체계를 기본 좌표로 사용하였다. 보통 토지피복과 같은 연구에서는 영상을 영상 대 지도(image to map)로 기하보정한 후 다른 시기의 영상을 영상 대 영상(image to image)으로 기하보정하는 것이 일반적이거나, 본 연구에서는 위치정확도가 매우 중요하므로 1/5,000의 수치지도를 가지고 영상 대 지도(image to map)를 통하여 절대 좌표로 보정을 실시함으로써 실좌표체계로 변환, 수행하였다.

지상기준점은 영상과 수치지도에서 함께 확인할 수 있는 30지점을 선정하였다. 30지점의 지상기준점에 해당하는 영상좌표는 수치지도를 통해 평면직각 좌표를 읽은 후 영상에 표시하였다. 지도좌표에 해당하는 영상의 각 화소가 결정되어 지도좌표에 맞게 화소들의 재배열이 이루어지게 되는데, 원래 영상이 가지고 있는 분광특성을 그대로 유지하기 위해 1차 다항식을 이용한 최근린 내삽법(nearest-neighbor interpolation)을 사용하였다. 기하보정 결과에 대한 위치정확도의 평균제곱근오차(RMSE: root mean square error)는 0.8257로 JERS-1 영상의 경우 약 15m이다. 이는 공간해상도 18m에 못 미치는 수치로 본 연구에 사용하는 데 무리가 없을 것으로 사료된다.

4. 훈련지역 선정

훈련지역(training area)이란 감독분류에 사용할 통계값을 추출하기 위해서 선택한 토지피복 특성을 대표하는 영역을 말한다. 훈련지역을 선택하는 방법으로는 현장답사, 위성영상에서 모니터를 통한 직접 취득, 항공사진이나 지도를 이용한 간접 취득, 성장기법(growing method)에 의한 취득 등이 있다. GPS(global positioning system)장비를 이용하여 실제 현장에서 위치를 정확히 측정하고 토지피복 상태를 파악하는 것이 이상적이거나 비용과 시간이 많이 소모되는 단점을 가진다. 반면 위성영

상에서 모니터를 통해 직접 취득할 경우 비용과 시간이 절약되지만 해상도가 낮은 위성인 경우 토지피복 식별이 어려우므로 훈련지역 선택에 오류를 범할 수 있다. 그리고 흑백 항공사진과 지도를 이용할 경우 비용과 시간이 적게 들고 위성영상에서 선택하는 경우보다 정확하나 흑백항공사진을 오독할 수 있고 지도 또한 오차가 있으므로 현장답사를 수행하는 경우보다 정확성은 떨어진다.

본 연구에서는 해상도가 낮은 위성영상에서 직접 훈련지역을 선택할 경우 오류가 클 것으로 판단하여 토지피복 특성이 잘 나타나 있는 항공사진과 수치지도를 동시에 사용하여 훈련지역을 취득하였다. 우선 연구대상 지역에서 TM좌표를 무작위로 추출한 후 항공사진과 수치지도를 통해 토지피복을 살펴본 결과, 영상의 해상도에 해당하는 영역 안에서 훈련지역 토지피복이 균일하지 않은 경우가 많았다. 또한 나대지, 초지 클래스는 훈련지역의 수가 적었다. 따라서, 본 연구에서는 무작위로 추출한 후 그 지역이 토지피복이 균일하고 특성을 잘 나타내고 있는 경우는 그 좌표를 그대로 사용하고 그렇지 않은 경우에는 36m×36m 범위 안에서 균일한 지역을 찾아 그 좌표를 훈련지역으로 취득하여 표본 수를 증가시켰다. 그러나, 여전히 나대지, 수역 클래스에 대한 훈련지역 수가 적었는데, 이 경우 그 클래스에 대해서는 임의로 훈련지역의 좌표를 취득하여 개수를 증가시켰다. 또한, 항공사진에서 훈련지역을 추출할 때는 분석가의 주관을 배제하기 위해 위성영상을 일체 참고하지 않았다. 항공사진과 위성영상의 취득시간 불일치에 따른 훈련지역 선정오류를 방지하기 위해 도시 및 나대지 클래스는 토지피복의 변화가 2년 동안에 발생할 수 있으므로 항공사진에서 훈련지역을 추출한 후 1993년 3월 영상을 참고하여 변화된 훈련지역을 제거하였다. 항공사진을 이용하여 각각 화소를 0.3%일 때 2,645개, 0.5%일 때 4,500개, 1.0%일 때 8,750개로 추출하였

고 항공사진에 포함되어 있지 않은 수역부분에 대해서는 1/5,000 수치지도를 이용하여 추출하였다. 여기서 나타내는 퍼센트(%)는 전체 연구대상지역 중에서 훈련지역으로 취득하는 화소개수를 말한다. 즉, 0.3%는 전체지역(850 화소 × 850 화소)의 0.3%(약 2,167 화소)을 나타낸다.

5. 최대우도분류

최대우도분류법은 분류대상 영상의 화소값의 분포가 정규분포를 따른다는 가정 하에 각 분류클래스의 평균 화소값과 공분산행렬을 계산한 후 확률밀도함수를 이용하여 각각의 화소값을 확률이 가장 높은 클래스로 할당하는 통계적인 기법이다. 각 분류클래스(c)에 대한 평균 화소값 Mc 와 각 밴드에 대한 공분산행렬(Vc)을 알고, 분류하고자 하는 화소의 값을 X 라 할 때 다음과 같은 식에 대입하여 가장 높은 확률(Pc)을 갖는 클래스를 찾아낸다. $det(Vc)$ 는 공분산행렬(Vc)에 의하여 결정되며, 미지 화소에 대한 벡터 X 를 얻기 위하여 각 분류클래스에 대한 확률값(Pc)를 계산한 후 제일 큰 값을 갖는 클래스에 화소를 할당하게 된다. 최대우도분류법은 분류하고자 하는 화소가 특정 분류클래스에 해당할 확률(Pc)을 계산한 후 가장 큰 확률을 갖는 클래스로 분류하는 방법이다.

$$P_c = -0.5 \log_e [det(V_c)] - 0.5(X - M_c)^T \cdot V_c^{-1}(X - M_c) \quad (1)$$

여기서, c : 분류클래스, Vc : 공분산행렬,
 Mc : 평균화소값
 X : 분류하고자 하는 화소값
 T : 전치행렬연산자

본 연구에서 통계적 접근의 대상으로 최대우도분류를 선정한 이유는 최대우도분류가 타 기법에 비해 알고리즘의 복잡성 및 처리시간 면에서 불리한 입장이지만 본 연구는 토지이

용분류를 실행하여 분류정확도 향상을 피하고자 하는 것이므로 알고리즘의 단순성이나 소요시간의 절약 등은 큰 의미를 가지지 않는다. 더욱이 최대우도분류는 정확도가 높은 기법으로 이미 널리 알려져 있으므로 분리도 측정 등의 절차를 통해 정확도를 조금이나마 향상시킬 수 있다면 연구결과의 효용성을 곧바로 입증하는 것이 된다. 확률 통계적 이론을 근거로 하여 오분류의 확률을 최소화하고자 하는 분명한 결정원칙이 있으므로 논리적으로 유도 분석하기에 상당한 타당성을 지닌다.

토지피복 분류의 적용 결과

1. 훈련지역의 통계분석

최대우도분류기법을 적용하기 앞서 훈련지역에 대한 분석이 필요하다. 훈련지역의 분포 특성을 평균, 공분산행렬, 상관행렬 등의 지표를 이용하여 각 훈련 화소 수치별, 분류클래스

별로 계산을 하였다. 훈련지역의 모든 화소들은 가장 높은 유사도(likelihood)를 갖는 분류클래스로 할당된다. 표 1은 최대우도분류를 하기 위한 훈련지역의 분류클래스별 기본 통계치이다. 표 1에서 나타난 바와 같이 band 1에서는 논, 초지, 도시, 수역 클래스의 화소값이 약 30에서 39까지 중복되어 있으며, band 2에서도 산림과 나대지를 제외한 모든 클래스들이 35에서 44까지 중복되어 있는 것을 알 수 있다. band 3에서는 수역과 나대지 클래스만이 적게 중복되어 있다.

최대우도분류는 자료가 정규분포를 따른다는 것을 가정한 것이기 때문에 훈련지역 화소의 분포특성을 분석하였다. 그림 3은 훈련지역으로 1.0%의 화소를 추출하였을 때 밴드 1의 화소값(pixel value)이 정규분포를 따르고 있는지를 나타낸 정규확률그림이다. 자료가 정규분포를 따를 경우 직선형태에 가깝게 되는데 본 연구에서 사용된 훈련지역을 살펴보면 거의

TABLE 1. The basic statistics of each landuse category for selected training data

	Band	0.3 %			0.5 %			1.0 %		
		band 1	band 2	band 3	band 1	band 2	band 3	band 1	band 2	band 3
Paddy	range	30~44	34~56	31~49	30~44	34~56	31~49	29~44	34~56	31~49
	mean	35.27	42.71	39.68	35.31	42.73	39.87	35.13	42.61	40.22
	Stdev*	2.57	3.71	2.61	2.61	3.83	2.97	2.36	3.62	3.03
	range	30~39	34~44	31~40	30~39	34~44	31~40	29~39	34~44	31~45
Grass	mean	32.33	38.21	35.67	32.40	38.12	35.32	32.15	38.05	35.11
	Stdev	2.03	1.65	2.50	2.09	1.58	2.35	2.12	1.57	2.48
	range	29~49	30~56	23~45	29~49	30~64	23~45	29~49	30~64	23~45
	mean	38.30	44.75	35.88	37.53	43.78	34.99	37.30	43.31	34.41
Urban	Stdev	3.53	4.85	4.08	3.53	4.94	4.05	3.25	4.39	3.67
	range	24~42	21~47	10~24	24~42	20~47	10~24	24~42	20~47	10~24
	mean	27.17	24.98	12.23	27.00	24.80	12.24	26.96	24.77	13.29
	Stdev	2.09	2.75	3.10	2.02	2.73	3.19	2.03	3.02	3.21
Stream	range	20~35	17~47	19~48	20~39	17~47	19~61	20~39	17~47	19~61
	mean	26.70	29.78	32.84	26.91	30.00	32.73	27.26	30.66	32.77
	Stdev	3.17	5.01	3.90	3.09	4.75	4.12	3.01	4.64	3.93
	range	33~52	39~64	40~53	33~60	39~68	40~61	33~60	39~68	39~61
Barren	mean	42.96	53.38	46.67	43.52	53.72	47.01	42.63	52.75	46.14
	Stdev	3.50	4.71	3.44	4.40	5.25	4.15	4.68	5.35	4.04

* Stdev : Standard deviation

모든 클래스가 직선형태에 가깝다. 따라서, 훈련 지역이 정규분포를 따르고 있음을 알 수 있다.

그림 3에서 화소값이 30일 때 도시, 논 클래스가 될 확률은 5% 미만이고, 초지 클래스가 될 확률은 15%, 산림이 될 확률은 82%, 수역 클래스가 될 확률은 93%임을 알 수 있다. 다른 측면에서, 각 클래스가 95%의 확률을 가질 분포를 살펴보면 수역 클래스는 화소값이 30~33, 산림 클래스는 33~35, 초지 클래스는 36~38, 논 클래스는 39~41, 도시 클래스는 43~45, 나머지 클래스는 50이상이다.

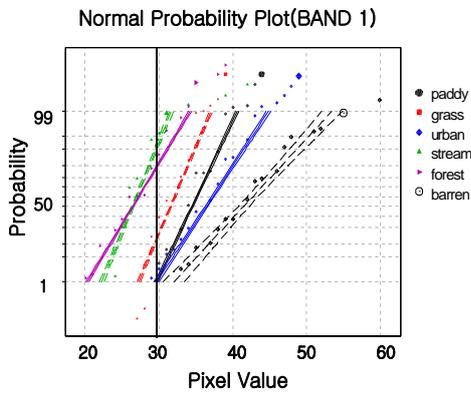


FIGURE 3. Normal probability plot of digital number for each landuse category on band 1

2. 훈련지역 검증

위성영상을 이용한 토지피복의 분류 정확도는 분류오차표(error matrix)를 이용하여 평가한다. 분류오차표란 위성영상 지표피복도와 검증자료(reference data)를 바탕으로 각 분류오차표의 해당란에 수치를 입력한 도표를 말한다. 아래 표 2는 분류오차표의 한 예이다. 보통 세로는 검증자료를, 가로는 위성영상 지표피복도의 분류클래스를 나타낸다. 정확도 검증을 하기 위해서는 반드시 분류된 영상을 비교하기 위한 검증자료(reference data)가 필요하다. 예를 들어 항공사진, 현지답사, 토지이용도, 수치지도 등 위성영상보다 더 높은 정확도

를 가진 데이터가 필요하다. 분류오차표는 다음과 같이 3가지 정확도를 평가할 수 있다.

TABLE 2. The error matrix

Classification	Reference data						Row Total
	paddy	grass	urban	stream	forest	barren	
paddy	X11	X12	X13	X14	X15	X16	ΣX1j
grass	X21	X22	X23	X24	X25	X26	ΣX2j
urban	X31	X32	X33	X34	X35	X36	ΣX3j
stream	X41	X42	X43	X44	X45	X46	ΣX4j
forest	X51	X52	X53	X54	X55	X56	ΣX5j
barren	X61	X62	X63	X64	X65	X66	ΣX6j
Column Total	ΣXi1	ΣXi2	ΣXi3	ΣXi4	ΣXi5	ΣXi6	N

분류오차표의 주대각 요소(Xii)는 각 클래스별로 정확히 분류된 화소의 총합을 의미한다. 주대각 요소의 총합을 전체 검증자료로 쓰인 화소수(N)로 나눈 값을 전체 정확도(OA: overall accuracy)라 한다.

$$OA = \frac{\sum_{i=1}^n x_{ii}}{N} \tag{2}$$

여기서, n 은 분류클래스 수이다.

위성영상 지표피복도를 만든 분석가의 관점에서는 특정지역을 분류 알고리즘이 얼마나 정확히 분류해 내는가가 관심인 경우가 있다. 이를 생산자 정확도(PA: producer's accuracy)라 하며 분류오차표의 주대각 요소(Xii)를 그 클래스의 검증자료의 화소수(column total)로 나눈 것을 말한다. 이를 누락오차(omission error)라고도 한다.

$$PA = \frac{x_{ii}}{\sum_{j=1}^n x_{ij}} \times 100 \tag{3}$$

지표피복도를 사용하는 입장에서 관심대상은 생성된 지표피복도가 과연 전체적으로 실

제 지표피복과 얼마나 일치하는 가이다. 이를 사용자 정확도(UA: user's accuracy)라하며 분류오차표의 주대각 요소(X_{ii})를 분류된 위성영상의 각 클래스 화소수(row total)로 나눈 것을 말한다. 이를 위임오차(commission error)라고도 한다.

$$UA = \frac{x_{ii}}{\sum_{i=1}^n x_{ij}} \times 100 \quad (4)$$

위의 식들에서 전체 정확도는 표본 추출시 무작위로 추출하였을 때 우연히 맞을 수 있는 것을 배제한 정확도이다. 즉 우연히 맞을 수 있는 것을 고려하여 전체 정확도를 계산하는 방법이 카파계수(kappa coefficient)를 이용한 방법이며, 이에 대한 통계값을 K_{hat} 라고 한다. K_{hat} 는 우연히 옳게 분류된 경우가 없을 때 1이며, 모두 우연히 옳게 분류된 경우에 0이 된다. K_{hat} 의 장점은 서로 다른 분석가에 의한 분류오차표(error matrix)간의 비교를 가능하게 한다는 점이다. 예를 들어 취득일, 분류기 업별, 분류과정이 다른 분류 결과물에서 얻은 분류오차표를 서로 비교 가능하게 한다.

$$K_{hat} = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} \cdot X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} \cdot X_{+i})} \quad (5)$$

여기서, r = 분류오차표의 행수, X_{ii} = i 행과 i 열의 값(대각선에 있는 값), X_{i+} = i 행의 총합, X_{+i} = i 열의 총합, N = 분류오차표의 표본의 총합

표 3은 훈련지역이 올바르게 선택되었는가를 판단하기 위한 자료이다. 최대우도분류법으로 분류한 후, 분류에 사용했던 훈련지역과 분류된 결과를 사용하여 분류오차표를 생성하였고, 이에 대한 사용자 정확도(UA)와 생산자 정확도(PA), 전체 정확도(OA), 카파계수(K_{hat})를 나타내었다. 훈련지역을 항공사진을 이용하여 1.0% 추출한 MA모형의 경우 전체정확도가 97.62%로 매우 정확함을 알 수 있다. 이는 훈련지역으로 선택한 지역이 대체로 올바르게 선택되었음을 의미한다. 그러나, 생산자 정확도와 사용자 정확도를 보면 논 클래스와 초지 클래스에 대한 정확도가 다른 클래스에 비해 낮음을 알 수 있다. 이는 논 클래스와 초지 클래스는 작황상태, 토지이용상태, 수목의 조밀

TABLE 3. Classification accuracy of classification map derived from training data

Class	ML						MA					
	0.3%		0.5%		1.0%		0.3%		0.5%		1.0%	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Paddy	71.48	83.18	78.46	83.28	77.50	79.37	82.03	73.68	96.55	91.72	94.22	99.72
Grass	81.47	33.23	78.08	48.64	88.26	45.46	80.31	35.08	97.73	96.63	98.39	98.26
Urban	90.24	91.36	91.34	83.00	79.19	86.98	64.15	91.32	92.20	95.37	92.75	96.73
Stream	97.75	99.24	98.26	98.78	89.86	100.00	98.75	100.00	99.60	97.66	99.66	97.75
Forest	71.04	99.45	70.17	96.30	79.50	98.89	70.31	98.41	100.00	98.25	99.82	99.04
Barren	82.14	76.67	84.46	50.20	89.36	59.57	92.86	38.81	95.11	97.06	100.0	61.93
Overall accuracy	79.47		77.97		81.60		76.48		96.71		97.62	
K_{hat}	0.73		0.71		0.75		0.70		0.95		0.97	

PA: producer's accuracy(measure of omission error), UA: user's accuracy(measure of commission error)

도 등에 의해 분광특성이 다를 수밖에 없는데, 이를 하나의 클래스로 묶는 것은 무리가 있으며, 나대지 클래스는 영상에서 차지하는 면적이 적음으로 인해 훈련지역 수가 적어서 분광특성을 추출하는데 어려움이 있는 것으로 사료된다. 도시 클래스인 경우 거의 계절에 의한 변화가 없을 것이라 예상할 수 있지만 논, 초지, 산림, 나대지 클래스는 계절이 봄임을 고려한다면 분광패턴의 차이가 거의 없어 다른 클래스와 비슷한 분광특성을 갖게 된다. 이를 고려하지 않고 훈련지역을 선정할 경우 분류 결과에 많은 영향을 끼치게 된다. 훈련지역을 검증한 결과 훈련지역의 개수가 많을수록 정확도가 높아졌으며, 사용자에게 의한 훈련지역의 취득보다 항공사진과 수치지도를 이용한 방법이 정확도가 높다는 것을 알 수 있었다.

3. 검증자료의 추출

검증자료(test reference data)는 최대우도 분류 후 정확도를 평가하기 위하여 추출한다. 본 연구에서는 수치지도를 이용하여 항공사진의 좌표를 추출한 후 2차 affine 보정을 수행하였고, 직접 항공사진을 육안으로 살펴보면 지표피복을 확인하였다. 그러나, 항공사진만을 이용할 경우 육안으로 지표피복을 확인할 수 없는 경우가 있는데, 이러한 경우 1/5,000 수치지도의 속성을 확인한 후 검증자료를 획득하였다.

검증자료 획득과정에서 단순 무작위 표본추출을 할 경우 특정 분류클래스의 숫자가 적어질 우려가 있으므로 무작위 표본추출(stratified random sampling)을 수행하였다. 즉, 분류된 위성영상의 지표피복도를 분류클래스 별로 나눈 후 무작위 표본추출을 수행하고, 항공사진과 수치지도를 사용하여 그 지점의 지표피복을 파악하였다. 이러한 검증지역 획득 과정은 훈련지역 선정과정과 전혀 무관하게 독립적으로 수행하였다. 수행결과 초지 클래스는 논과 산림 클래스와의 구별이 불가능한 경우가 많아서 분류오차표에 영향을 미칠 것으

로 판단된다. 그러나 통계적 객관성을 유지하기 위해 훈련지역 선택의 경우와 달리 위치이동을 수행하지 않았다. 표 4은 각 토지피복클래스별 추출된 검증자료의 수이다.

TABLE 4. The number of test reference data selected

Paddy	Grass	Urban	Stream	Forest	Barren	Total
262	452	256	286	697	47	2000

4. 분류정확도 검증

표 5는 ML 분류와 MA 분류 전체를 생산자 정확도 및 사용자 정확도로 나타낸 것이다. 대부분의 경우 추출된 훈련지역의 픽셀수가 많을수록 높은 분류정확도를 보여주고 있다. 분류의 정확도가 상대적으로 크게 떨어지는 경우 훈련지역 추출 시 분광특성에 의한 공분산이 넓게 분포되어서 나타난 결과로 볼 수 있다. ML 분류의 생산자정확도는 논, 초지, 산림이 75%이하로 매우 낮다. 이는 분류하고자 하는 토지피복 클래스들이 계절적인 요인에 의해 서로 비슷한 특성을 갖기 때문에 생기는 현상이라고 생각된다. 또한 생산자정확도와 사용자정확도가 서로 같지 않은 값을 보이고 있는데, 생산자정확도에서 사용자정확도를 뺀 값은 실제지역이 아닐 확률로 볼 수 있으며, 100%에서 생산자정확도를 뺀 값은 다른 클래스로 분류될 확률로 볼 수 있다.

그림 4와 그림 5는 ML 분류와 MA 분류 전체를 생산자 정확도 및 사용자 정확도로 나타낸 것이다. 생산자 정확도에서는 나대지만이 높은 정확도를 가지고 있으나, 0.5% MA 분류에서는 크게 낮아 있는데, 그것은 나대지 클래스가 논으로 오분류 되면서 낮아진 것이다. 또한 계절에 따라 토지피복상태가 변하는 논, 초지 클래스가 전체적으로 정확도가 넓게 분포되어 있는 것으로 보인다.

TABLE 5. Classification accuracy of classification map derived from test reference data

Class	ML						MA					
	0.3%		0.5%		1.0%		0.3%		0.5%		1.0%	
	PA	UA										
Paddy	52.67	76.24	69.85	74.69	75.19	75.48	75.95	78.04	93.13	92.08	91.60	100
Grass	73.23	48.82	69.69	60.58	90.27	61.82	80.53	54.41	97.79	92.86	98.01	92.87
Urban	85.94	70.29	90.23	61.27	74.61	66.55	71.48	76.25	83.98	88.48	83.20	93.01
Stream	97.90	98.59	92.31	96.70	75.18	100	99.30	98.96	98.95	98.26	98.95	98.26
Forest	64.71	92.42	67.58	90.58	71.88	96.91	66.14	98.09	98.85	96.91	98.85	98.43
Barren	91.49	76.79	97.87	70.77	93.62	73.33	95.75	56.96	34.04	94.12	100	71.21
Overall accuracy	73.15		75.5		77.8		76.80		94.45		95.75	
K_{nat}	0.66		0.69		0.72		0.71		0.93		0.95	

PA: producer's accuracy(measure of omission error), UA: User's accuracy(measure of commission error)

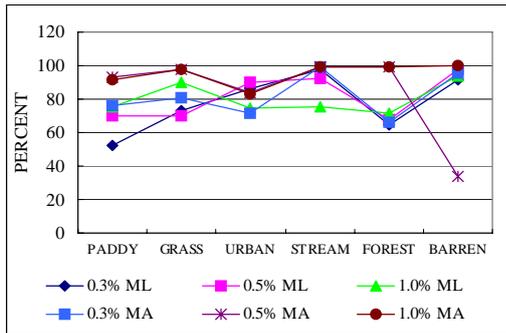


FIGURE 4. Producer's accuracy of ML and MA methods

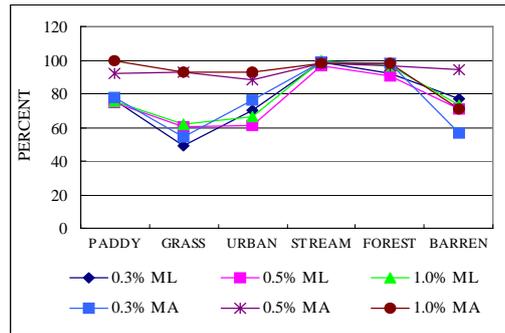


FIGURE 5. User's accuracy of ML and MA methods

결론

본 연구에서는 인공위성 데이터를 분류함에 있어서 사용자에게 의한 최대우도분류와 항공사진과 수치지도를 이용한 최대우도분류를 통해 토지피복도를 생성하였으며 분류정확도를 정량적으로 검증하였다. 본 연구의 결과는 다음과 같다.

1. 토지피복도의 생성 및 검증은 사용자에게 의한 훈련지역의 추출보다 항공사진과 수치지

도를 동시에 이용하여 추출하는 것이 분류 후 효율적이고 정량적인 분류정확도 검증을 가능하게 했다.

2. 사용자에게 의한 최대우도분류를 이용한 것보다 항공사진과 수치지도를 이용한 최대우도분류를 이용할 경우 전체정확도를 최대 18%정도 향상시켰다.

3. 우리나라와 같이 복잡하고 다양한 토지이용

을 가진 지형에 JERS-1 영상을 이용하여 95%의 신뢰도를 얻기 위해서는 적어도 훈련지역을 전체지역의 약 1% 추출하여야 만족할 만한 토지피복분류를 할 수 있었다.

4. 인공위성데이터를 사용하여 토지이용 및 토지피복 패턴을 추출하는 것은 경제적이고 합리적이며, 도시계획의 방향설정 및 환경정보시스템의 공간데이터베이스 구축을 위한 속성자료로서 중요한 기초자료가 될 것으로 판단된다.

감사의 글

본 연구는 21세기 프론티어 연구개발사업인 수자원의 지속적 확보기술개발사업단의 시공간자료 활용기술개발(과제번호 : 1-2-1) 및 우주개발사업 중 다목적 실용위성사업(과제번호 : MI-01-DA-01-0001-01-D01-01-001-1-3)의 연구비 지원에 의해 수행되었습니다. **KAGIS**

참고문헌

김영섭. 1998. 원격탐사개론. 동화기술.
 Canters, F. 1997. Evaluation the uncertainty of area estimates derived from fuzzy land-cover classification. PE & RS 63(4): 403-414.

Fuller, R. M., G. B. Groom and A. R., Jones. 1994. The land-cover map of gre at britain: An automated classification of Landsat Thematic Mapper data. PE & RS 1(12):553-562.
 Hardin, P. J. and J. M. Shumway. 1997. Statistical significance and normalized confusion matrices. PE & RS 63(6):735-740.
 Knick, S. T., J. T. Rotenberry and T. J. Zarriello. 1997. Supervised classification of Landsat Thematic Mapper imagery in a semi_arid rangeland by nonparametric discriminant analysis. PE & RS 63(1):79-86.
 Marsh, S. E., J. L. Walsh and C. Sobrevila. 1993. Evaluation of airborne video data for land-cover classification accuracy assessment in an isolated Brazilian forest. PE & RS 48:61-69.
 Paola, J. D. and R. A. Schowengerdt. 1997. The effect of neural-network structure on multispectral land-use/land-cover classification. PE & RS 63(5):535-544.
 Xin, Z., B. A. Engel, X. Xiong and C. J. Johannsen. 1995. Analysis of classification result of remotely sensed data and evaluation of classification algorithms. PE & RS 1(12):427-434. **KAGIS**