

원격탐사 데이터의 이차계획법에 의한 토지피복분류에 관한 연구 A Study for the Land-cover Classification of Remote Sensed Data Using Quadratic Programming

전형섭* · 조기성**

Jeon, Hyeong-Seob · Cho, Gi-Sung

요 旨

본 연구에서는 원격탐사 데이터의 분류방법으로서 이차계획법을 토지피복 추출에 적용하였으며, 주제도 추출에 일반적으로 사용되는 최대우도법과 신경망의 분류결과와 정확도를 비교하여 그 적용성을 검토하였다. 그 결과, 이차계획법에 의한 분류방법이 최대우도법에 비하여 평균 6%정도의 향상된 분류결과를 도출할 수 있어서 원격탐사 데이터의 분류에 이차계획법을 적용할 수 있으리라 판단되었다. 또한 이차계획법에 의한 분류에서는 클래스 구성비가 클래스 결정에 직접적인 영향을 주어 기존의 이진적인 분류방법에서 무시되었던 결과 값들을 명확하게 나타낼 수 있었다.

ABSTRACT

This study present the quadratic programming as the classification method of remote sensed data applying to the extraction of landcover and examine it's applicable capability by comparing the classification accuracy of quadratic programming with that of neural network and maximum likelihood method which are used in the extraction of thematic layer. As the results, as drawing the more improved classification results by 6% than maximum likelihood method, we could discern that the method of quadratic programming is applicable to classifying the remote sensed data. Also, in the classification of quadratic programming method, we could definitely indicate the results which was ignored in the previous extreme(binary) classification method by affecting the class decision with the class composition proportion.

1. 서 론

일반적으로 원격탐사는 자연현상을 매핑, 모니터링, 예측의 3가지 응용분야로 대별할 수 있으며, 원격탐사 데이터로부터 주제도를 추출하는 매핑분야는 다른 연구분야의 기초자료로서 사용되기 때문에 우선적으로 요구되는 연구분야이다.¹⁾ 그러나 원격탐사 데이터로부터 추출된 주제도는 실제 사용하기에 여전히 그 정확도가 낮다고 판단되며, 그 중 가장 흔히 응용되는 토지피복도 추출 역시 정확하게 얻을 수 있다고 판단하기에는 아직 이르다고 할 수 있다.²⁾ 이에 대한 원인으로는 원격탐사 데이터의 낮은 공간해상도, 분광해상도, 분류항목에 있어

상세한 항목화의 결여, 그리고 매핑에 사용되는 방법의 낮은 분류정확도 등의 문제점으로 대별 할 수 있다. 최근 과학기술의 발달과 연구노력으로 위의 문제점들은 어느 정도 보완되었지만, 분류방법의 낮은 정확도 문제는 원격탐사 데이터의 폭넓은 사용을 계속적으로 제한하고 있다.

일반적으로 위성영상은 통계학적인 이론에 근거한 최대우도법(maximum likelihood method)과 80년대 후반부터 원격탐사 분야에 등장한 인공 신경망(artificial neural network)을 이용한 방법 등으로 분류되고 있다. 그러나 최대우도법은 각 항목이 다차원 정규분포를 따르지 않을 때, 분류정확도는 현저하게 저하된다는 문제점이 있으며, 인공 신경망을 이용한 경우에는 망의 구조와 클래스의 특성에 관계하는 크기나 구성과 같은 훈련 집합의 특성에 따라 분류결과가 일정치 않다는 점과 망을 훈련하는데 시간을 요한다는 점, 그리고 처리가 블랙

*전북대학교 공업기술연구소 연구원

**전북대학교 공과대학 토목환경공학부 부교수,
전북대학교 공업기술연구소 연구원

박스 형태이어서 결과를 설명하기가 어렵다는 문제점을 가지고 있다.^{3),4)}

따라서 본 연구에서는 원격탐사 데이터로부터 주제도 추출의 가장 일반적인 응용인 토지피복분류에 있어서 최적화기법의 하나인 이차계획법을 원격탐사 데이터의 분류에 적용하여 기존분류방법에서 탈피한 새로운 분류방법을 모색하였다. 또한 분류결과를 시험하기 위하여 기존의 분류방법중 대표성을 띤 최대우도법과 신경망을 이용한 분류방법과 비교하여 그 적용성을 시험하였다.

2. 연구방법 및 기본이론

2.1 연구방법

본 연구는 크게 영상분류를 위한 데이터 준비과정인 전처리, 가정될 순수한 픽셀 값을 추출하는 과정, 각 분류법에 의한 영상분류 그리고 분류정확도 평가로 나눌 수 있으며, 이에 따른 연구개요도는 그림 1과 같다.

먼저 전처리 과정에는 사용될 데이터와 분류에 사용될 분류알고리즘의 특성 파악, 그리고 원격탐사의 일반적인 전처리 단계인 기하보정과 방사보정이 포함되나 본 연구에서는 정보손실의 문제와 대상지역의 협소함을 감안하여 기하보정만을 실시하였다.

다음으로 제안된 분류방법의 적용성을 고찰하고, 현재로서 정의하기 어려운 순수한 분광값을 가정하기 위하여 일반적인 분류과정에서 가장 많이 사용되고 있는 최대우도법과 신경망을 적용하였다. 즉, 두 분류방법의 결과에서 동일한 위치이며, 각 클래스에 대한 결과 값이 상위 일정부분에 해당하는 픽셀들은 토지피복을 구성하는데

단일의 피복클래스로 구성되어 순수한 분광값을 갖는다고 가정하였으며, 이렇게 얻어진 순수한 분광 값을 최적화 방법인 이차계획법(quadratic programming method)에 적용하여 토지피복클래스 구성비(각 클래스의 소속도)를 구하였다. 분류의 최종분류방법으로 픽셀에 대한 최대 결과값을 갖는 클래스로 그 픽셀의 출력값을 결정하는 방법에 의하여 동일한 훈련데이터와 시험데이터를 분류하였다. 그리고 정확도 평가로서 훈련데이터와 시험데이터에 대한 분류결과와 정확도 분석을 하여 이차계획에 의한 분류방법의 유용성을 검증하였다.

2.2 이차계획법에 의한 분류방법

본 연구에서 제안된 이차계획법은 공학설계나 경제학의 분야에서 사용되는 최적화 이론중 하나로, 기존의 최적화방법에 비하여 제약조건에 존재하는 비선형의 목적함수의 최적 해에 수렴하는 속도가 굉장히 빠른 방법이다.⁵⁾

원격탐사 영상의 경우 각 픽셀에 기록되는 분광 값은 기본적으로 해당 픽셀의 면적에 해당되는 지표요소의 특성에 의해 좌우된다. 따라서 각 클래스에 대한 순수한 분광값이 주어지고, 각 밴드에 나타나는 픽셀의 실제 분광값이 각 클래스의 토지피복비(면적비)에 선형으로 변화한다고 가정하면, 순수한 분광값과 실제 위성영상 분광값을 사용하여 각 픽셀의 토지피복구성비를 추정할 수 있으며, 추정된 토지피복 클래스 구성비로부터 분류를 수행할 수 있다.

따라서 본 연구에서는 각 픽셀에 있어서 토지피복구성비에 따른 순수한 분광값의 결과와 실제 위성영상 분광값과의 오차를 최소화시킬 수 있는 최적의 토지피복 구성비를 추정하고자 하였으며, 이러한 문제를 이차식의 목적함수와 제약조건이 일차식인 이차계획의 문제로 변형하여 적용하였다. 즉, 지상 피복의 c 개 타입과 n 개의 분광밴드가 있다고 하면, 일반적으로 판별조건을 만족하기 위해 $n \geq c$ 를 사용해야 한다.⁶⁾ 한 픽셀의 n 개 분광밴드를 표기하기 위하여 $x = [x_1, \dots, x_n]^T$ 를 사용하고, 지상피복 c 개 타입 각각에 의하여 점유되는 픽셀안에서 면적의 비를 표기하기 위하여 $f = [f_1, \dots, f_c]^T$ 를 사용한다면, 선형 혼합모델에 따라 식 (2-1)과 같이 표현될 수 있다.

$$x = Mf + e \quad (2-1)$$

여기서 M 은 순수한 분광 값이며, x 는 클래스의 해당 분광 값이고, e 는 센서로부터 영상데이터를 전송하는 동안

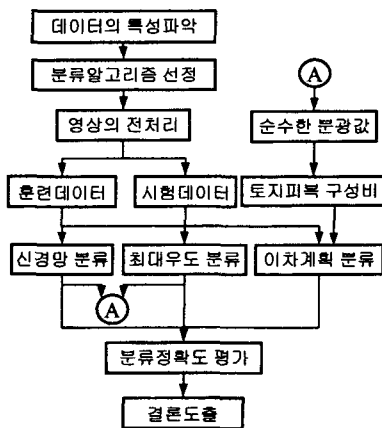


그림 1. 연구개요도

발생한 오차와 기기 노이즈에 따른 이상적인 분광값으로부터의 편차를 나타내는 관측오차이다.

문제는 순수한 분광값 M 과 미지의 픽셀 분광값 x 가 주어졌을 때, 식 (2-3), (2-4)의 제약조건 I, II를 만족시키며 목적함수를 최소화시키는 토지피복 클래스 구성비 (cover class proportion)를 찾는 것이다.

$$\text{Minimize} \|Mf - x\|^2 \quad (2-2)$$

제약조건

$$\text{I) } 1^T f = [1 \ 1 \ \dots \ 1] f = 1 : \text{sum to one} \quad (2-3)$$

$$\text{II) } f_j \geq 0, j = 1, \dots, c \text{ or } f \geq 0 : \text{positivity} \quad (2-4)$$

기존하는 많은 선형 비혼합 응용들은 토지피복 클래스 구성비를 추정하기 위해 최소제곱 근사식을 이용하였으나^{7),8)} 최소제곱법에 의한 해는 전역적 최적해가 아닌 국지적 최적 해를 구할 염려가 있어 목적함수를 만족시키지 못할 경우가 있다. 또한 라그랑지 승수 (lagrange multiplier)를 사용하여 제약조건 I을 목적함수에 결합시키는 방법도 제안되었으나⁹⁾ 이 방법 역시 각 클래스의 값이 양이어야 한다는 제한조건을 만족시키지 못했다.

따라서 본 논문에서는 제약조건 I과 II를 동시에 만족시킬 뿐만 아니라 목적함수의 전역적 최적 해를 신속하게 구할 수 있는 식 (2-5)와 같은 이차계획법을 사용하였다.

$$\text{Minimize} \left(\frac{1}{2} f^T H f - e^T f \right) \\ \text{Subject to } 1^T f = 1, f_i \geq 0, i = 1, \dots, c. \quad (2-5)$$

여기서 $H = M^T M$ 이고, $c = M^T x$ 이다.

이와 같은 방법으로 추출되는 최적의 토지피복 클래스 구성비를 사용하여 위성영상의 토지피복 분류를 수행하였다.

3. 적용 및 고찰

3.1 연구대상지역 및 사용데이터

본 연구의 대상지역으로는 전라북도과 충청남도를 연결하는 금강하구언을 중심으로 한 65.61 km²(270×270 pixel)지역이다. 이 지역은 그림 2에 나타난 바와 같이 금강을 사이에 두고, 좌하단으로 군산 시가지의 일부가

포함되어 있으며, 중앙 상단부에는 비교적 넓은 농경지가 형성되어있다. 또한 금강이라는 뚜렷한 수계와 인접한 넓은 습지가 존재하고 있으며, 서해안 고속도로 건설로 비교적 넓은 나대지가 분포하고 있다. 즉, 도시, 농경지, 삼림, 수역, 습지 나대지 등의 토지피복분류에 좋은 표본조건을 갖추고 있어 연구대상지로 선정하였다.

연구에 사용된 데이터는 Landsat-5 TM센서에 의해 1994년 6월 1일과 1999년 3월 11일에 취득된 영상으로 계절변화가 뚜렷한 우리나라의 특성을 고려하여 3월 영상은 동절기의 대표 영상으로, 6월 영상은 하절기의 대표 영상으로 사용하였다.

3.2 영상의 전처리

3.2.1 기하보정

연구에 사용된 영상은 Bulk Correction의 전처리 수준을 가진 것이어서 방사보정은 필요가 없었으며, 축척 1/25,000의 수치지도를 사용하여 Image to Map방식으로 기하보정 하였다. Landsat TM 1/4 Subscene의 경우 허용오차가 0.5픽셀이어서 일반적으로 Affine 변환을 사용하는 것으로 충분하다고 하나, 본 연구의 경우 1차 다항식으로는 이러한 조건을 만족시킬 수 없어 전체적으로 3차 다항식을 사용하였으며, 그 결과 평균제곱근오차가 1994년 영상의 경우 0.4497픽셀, 1999년 영상의 경우 0.4308픽셀이었다. 픽셀의 재배열은 원 데이터를 왜곡시

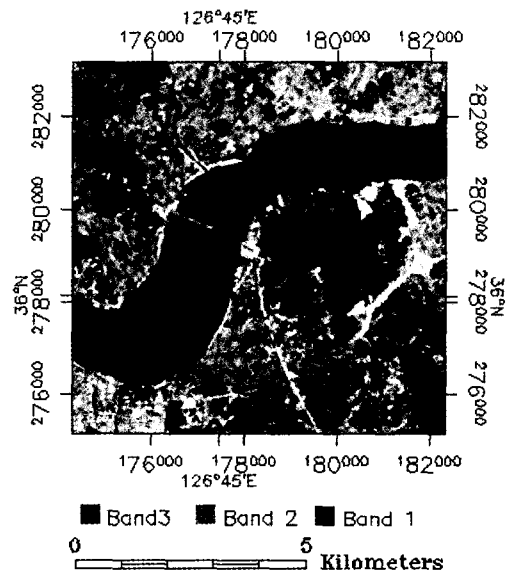


그림 2. 연구대상지역 영상(1999/03/11)

켜 데이터 가치를 떨어뜨릴 수 있기 때문에 데이터를 보존하기 위하여 최근린 내삽법을 사용하였다.

3.2.2 분류항목의 선정

실제적인 분류를 위해서는 먼저 대상지역에 대하여 적용할 분류항목을 선정하는 단계가 선행되어야 하며, 이를 위하여 토지이용/피복분류체계에 있어서 일반적으로 통용되고 있는 USGS의 분류체계중 Landsat TM의 경우에 해당되는 레벨 1의 항목을 참조하였고, 대상지역에서 명확히 나타나는 것을 분류항목선정의 기본적인 조건으로 하여,¹⁰⁾ USGS의 9개 클래스중 도시, 농경지, 삼림, 수역, 습지, 나대지의 6개 클래스를 선정하였다.

3.2.3 훈련데이터와 시험데이터의 추출

3.2.3.1 훈련데이터의 추출

훈련데이터의 추출은 수치지도와 현지답사를 통하여 화면상에서 이루어졌으며 분류기법 상호간의 비교를 위하여 최대우도법과 신경망, 이차계획법에 의한 분류에서 동일한 훈련데이터를 사용하는 것을 원칙으로 하였다. 또한 다른 시기에 취득된 영상에 대하여 토지피복형태가 변형되지 않은 지역을 선별하여 동일한 지리적 위치의 훈련데이터를 추출하였다.

최대우도분류기의 경우 훈련데이터를 각 클래스에 대하여 10 n(n:밴드수)화소 이상의 훈련 데이터가 수집되어야 한다는 것이 일반적이거나,¹⁰⁾ 신경망이나 이차계획법의 경우에는 이러한 제한 조건이 존재하지 않는다. 본 연구에서는 각 분류방법의 비교를 위하여 각 클래스에 대하여 150픽셀씩 동일하게 추출하여 그림 3(a),(b)와 같은 900픽셀의 훈련데이터를 생성하였다.

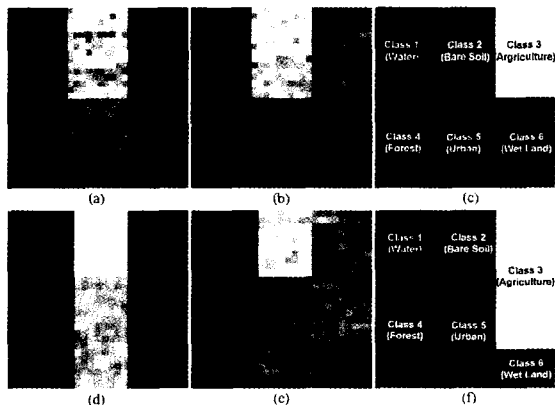


그림 3. 연구지역의 훈련데이터(a:1994, b:1999, c:index)와 시험데이터(d:1994, e:1999, f:index)

3.2.3.2 시험데이터의 추출

연구결과의 평가를 위한 시험데이터의 선정은 연구의 성과를 좌우하는 것으로 본 연구에서는 시험데이터 즉, 검증데이터는 연구 여건상 훈련데이터와 같이 수치지도와 야외조사를 통하여 수행되었으나, 다음과 같은 두 가지 경우, 즉 원격탐사 전문가가 수치지도와 야외조사를 통하여 수집한 시험데이터(case 1)와 연구에 관련이 없는 일반인이 모니터상의 수치지도로부터 수집하게 한 시험데이터(case 2)를 구분하여 추출하였다.

시험데이터의 크기는 오차행렬에서 각 토지피복 클래스에 대하여 최소 50개의 샘플을 수집되어야 한다고 제시한 Congalton의 제안을 참고하여¹⁰⁾ 최소 54점에서 최대 189점까지 지역의 클래스 분포를 고려하여 선택하였다. 이렇게 추출된 시험데이터는 각 시기와 지역에 따라 그림 3(d),(e)와 같이 27×27의 시험데이터 집합으로 생성하여 각 분류에 적용하였다.

3.3 영상의 분류

3.3.1 최대우도법에 의한 분류

최대우도법은 각 밴드의 각 클래스에 대한 훈련 데이터 통계가 정규분포라 가정하고, 사상백터가 나타날 가능성이 가장 큰 클래스에 픽셀을 할당하는 분류방법이다.

본 연구에서는 최대우도의 이론에 따른 분류방법의 분류정확도를 평가해 보기 위하여 최대우도분류를 C++로 프로그래밍하였다. 훈련데이터는 그림 3(a),(b)와 같은 900개의 훈련데이터를 사용하였으며, 이들로부터 표본의 평균과 공분산 등, 통계학적 매개변수들을 추출하였다. 또한 사전확률의 항은 사전에 분류를 한 결과라든지 분류지역에 대한 분포비 등의 사전정보를 얻기 어렵기 때문에 각 클래스에 대하여 동일하다고 가정하였다.

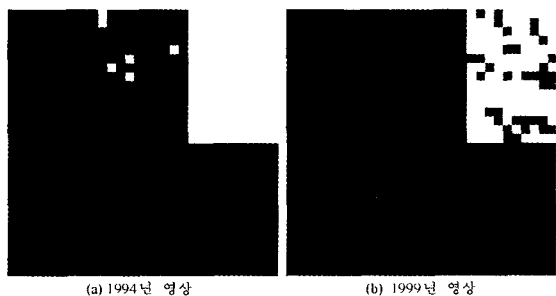


그림 4. 훈련데이터의 최대우도법에 의한 분류결과

먼저 훈련지역에 대한 분류결과(그림 4)를 살펴보면 1994년 영상에서는 수역, 농경지, 산림, 습지가 매우 정확한 분류경향을 보이고 있으나, 나대지는 대부분이 도시지역으로 분류되거나 일부 산림지역으로 오분류되었다. 이러한 오분류는 그림 5의 각 밴드별 히스토그램의 예와 같이 나대지와 도시지역의 평균값과 데이터분포가 유사하고 나대지에 비하여 도시지역이 분산이 적고 정밀도가 좋다고 판단된다. 따라서 그림 6과 같은 전체적인 연구대상지역의 토지피복분류결과에 있어서도 전반적으로 양호하게 분류된 것으로 보이나 나대지의 클래스는

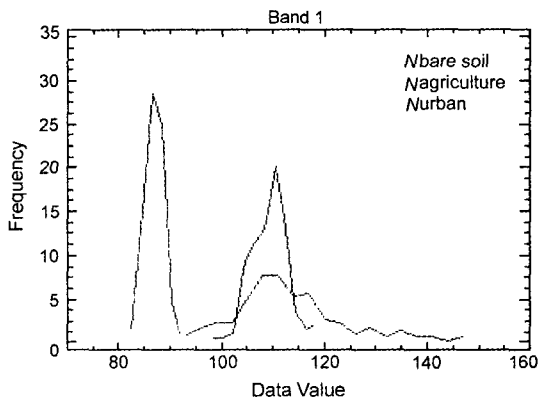


그림 5. 훈련데이터의 히스토그램 예(1994)

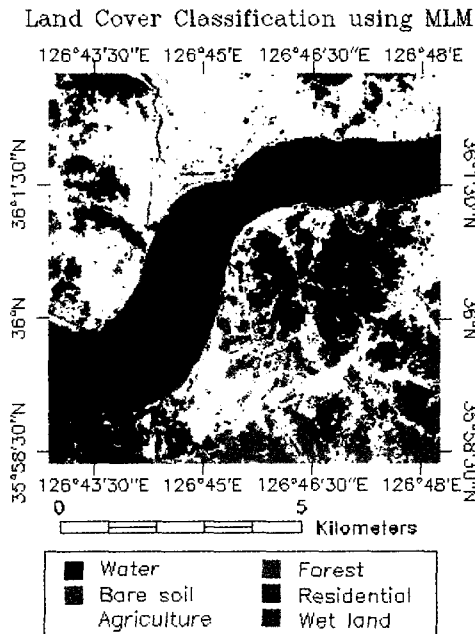


그림 6. 최대우도 분류결과(1994)

거의 보이지 않는다.

1999년 영상에서는 수역과 산림, 나대지 그리고 습지의 경우 정확한 분류가 가능하였으나 농경지가 나대지로, 도시지역이 나대지로 오분류되는 결과가 일부 보이고 있다(그림 4(b)). 이러한 원인은 1994년에 비하여 농경지나 도시지역의 정밀도가 낮음으로 인해 나대지의 분산영향을 많이 받은 것으로 사료된다. 이러한 점은 도시지역이나 특히 농경지에 있어서 평균값과 떨어져 가능성이 미약한 픽셀의 경우, 이 픽셀에 대한 나대지의 분산이 크기 때문에 나대지로 분류된 것으로 판단된다. 훈련 데이터에 나타난 이러한 경향은 연구지역 전체의 분류에서도 뚜렷하게 나타난다. 이것은 영상의 취득시기가 농사철이 아닌 휴지기이기 때문에 초봄의 건조한 기후로 인한 농경지의 함수율의 감소로 나타난 결과로 볼 수도 있다.

3.3.2 신경망에 의한 분류

연구에서 사용된 신경망은 최급하강법을 사용한 역전파 학습알고리즘이다. 신경망 구현을 위하여 C++언어를 사용하였다. 신경망의 성능을 결정하는 요소로는 은닉층의 수나 노드수, 연결강도 그리고 학습에 사용되는 파라미터인 학습률과 모멘텀상수의 값일 것이다. 이들의 값을 결정하는데는 정규화된 방법이 없고 시행착오적으로 이들 값을 구하는 방법이 일반적이다.⁴⁾

따라서 본 연구에서도 시행착오법으로 은닉층의 노드수와 학습률, 모멘텀상수의 값을 결정하였다. 그 결과 1994년도 훈련데이터에서는 학습률이 0.1이고, 모멘텀 상수가 0.5일 때 최고의 분류정확도를 나타냈고, 99년도의 훈련데이터에서는 학습률 0.1과 모멘텀 상수 0.2일 때 최고의 분류정확도를 산출하였다. 또한 전자에서 생성된 학습률과 모멘텀항으로 고정시켜놓고 은닉노드의 수를 변화시켜 13개의 최적의 은닉층 수를 추출하였다.

이러한 과정을 통하여 추출된 학습률과 모멘텀 상수의 매개변수와 13개의 은닉층 노드수를 가진 6×13×6의 삼층 구조 신경망을 사용하여 신경망을 훈련시켰으며, 훈련 알고리즘으로는 일반적인 역전파 알고리즘을 사용하였다. 훈련과정에서 추출된 각 층간의 연결강도는 다음 단계인 분류과정에 사용되었다.

- 먼저 각 시기에 따른 훈련데이터의 분류결과를 살펴보면 1994년 영상의 경우 수역, 산림 그리고 농경지와 습지 클래스가 매우 정확하게 분류되었다(그림 7(a)). 그러나 나대지 클래스가 다른 클래스에 비하여 상대적으로

정확도가 약간 떨어지나, 최대우도법에 비하여 상당히 향상된 결과를 가져왔다. 이것은 신경망이 클래스의 패턴에 대하여 그 유사도에 따라 분류하는 블랙박스 형태이기는 하지만, 영상데이터 값을 정규화하는 과정에서 최대값과 최소값을 사용하는데 서로가 비슷한 분포형태를 가짐으로서 정규화되는 값이 상대적으로 비슷하기 때문에 유사한 패턴으로 분류된 것이라고 판단된다. 1999년 영상의 경우 전반적으로 양호한 분류를 보였는데 특히 최대우도에 비하여 농경지가 나대지로 분류되는 오분류가 현저히 감소되었다는 것을 알 수 있다(그림 7(b)).

연구지역의 분류결과를 살펴보면 1994년 영상의 경우 최대우도에 있어서의 나대지가 도시지역, 산림지역, 농경

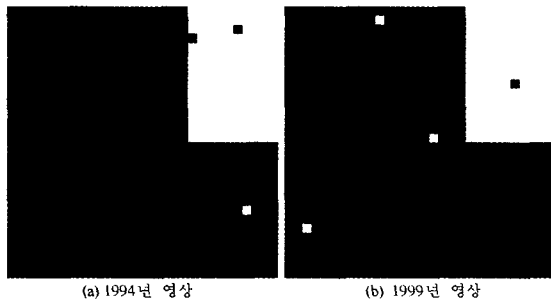


그림 7. 훈련데이터의 신경망에 의한 분류 결과

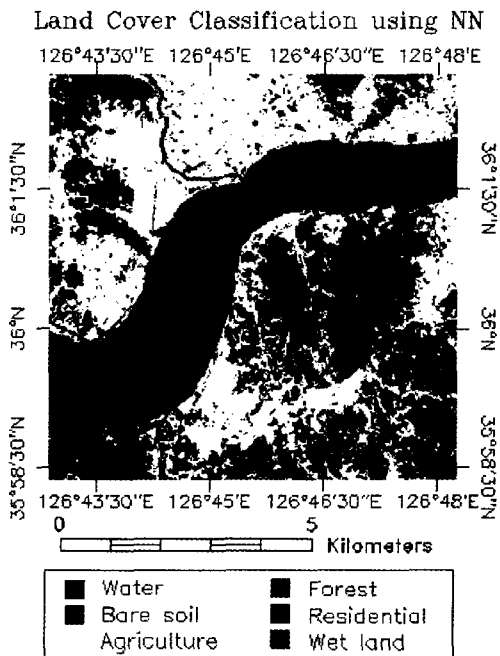


그림 8. 신경망에 의한 분류결과(1994)

지 등으로 분류되는 오분류나, 수정된 최대우도분류에서 나대지가 수역으로 오분류되는 결과를 나타내지 않고, 나대지로 정확히 분류되었다(그림 8). 1999년 영상에서는 최대우도법이나 수정된 최대우도법의 경우에 많은 농경지가 나대지로 분류되었으나, 신경망의 경우는 이러한 오분류 없이 양호한 분류결과를 나타내었다.

3.3.3 이차계획법에 의한 분류

3.3.3.1 순수한 분광값의 추정

원격탐사영상의 분광값 중에서 순수한 분광값을 추출하기는 매우 어렵다. 여기에는 지역에 따른 태양 고도각의 차에 의한 영향, 지형에 따른 영향, 대기의 영향, 그리고 클래스의 혼재 등에 의한 오차를 포함하고 있기 때문이다. 그러나 지형에 따른 영향을 제외한 고도각이나 대기의 영향은 지역이 협소하기 때문에 각 밴드 영상의 전체적인 범위에 미치는 영향으로 해석할 수 있어서 전체적인 분광값의 증감패턴으로 해석될 수 있다.

일반적으로 순수한 분광값의 추정은 원격탐사영상의 어떤 작은 지역에서 수동적으로 추출하는 방법과 각 영상 픽셀에 대한 토지피복 클래스 비를 사용하여 최소제곱법을 사용하는 방법이 있다. 본 연구에서는 신경망과 최대우도의 분류결과를 사용하였으며 전자에 가깝다고 할 수 있다. 신경망과 최대우도법은 하나의 픽셀에 대하여 각 클래스에 대한 가능성값과 출력유니트의 활성화 수준을 제공한다. 이러한 각 클래스에 대한 결과값의 통계파일로부터 상위 0.05%에서 0.5%까지 두 분류결과에서 중복되는 픽셀을 10~20개 추출하였으며 이들의 표준편차를 그림 9에 나타내었다.

대부분의 클래스에서는 표준편차의 값이 0.0에서 4.0

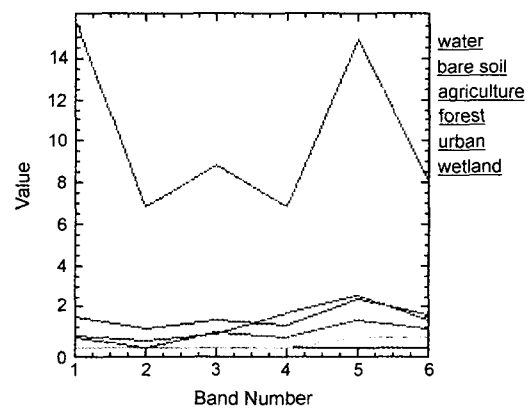


그림 9. 추출된 픽셀의 표준편차의 예(1994)

표 1. 각 클래스에 대하여 추출된 순수한 분광값

Classification		Class1	Class2	Class3	Class4	Class5	Class6
1994 Scene	Band1	91.000	134.636	87.000	78.750	108.857	92.545
	Band2	41.000	74.000	38.000	32.875	50.952	42.000
	Band3	48.000	105.591	47.000	37.625	65.762	50.000
	Band4	30.000	92.000	43.000	53.938	55.288	41.545
	Band5	15.000	164.864	77.364	67.625	92.762	51.909
	Band7	8.000	83.591	42.545	30.375	60.238	20.455
1999 Scene	Band1	67.000	116.765	88.529	67.000	94.611	80.778
	Band2	25.000	59.235	39.412	24.750	40.000	33.944
	Band3	25.000	103.176	64.824	32.875	61.944	49.944
	Band4	9.000	74.118	54.647	44.125	44.167	34.333
	Band5	6.000	150.706	123.412	58.313	81.500	57.333
	Band7	2.000	76.941	63.882	21.313	46.167	21.779

정도로 비슷한 양상을 보이고 있으나 내대지의 경우 두 시기의 영상 모두에서 높게 나타났다. 특히 1994년 영상에서 높게 나타난 원인은 대상지역의 영상내에 다른 클

래스에 비하여 상대적으로 적은 픽셀수가 존재하는데 두 분류결과 영상내에서 중복되는 픽셀을 최소 10개이상 추출하였기 때문에 나타난 결과라 판단되며, 이에 대한 개체를 줄임으로서 감소시킬 수 있다고 사료된다. 이와 같은 방법으로 추출된 픽셀에 대한 평균값이 각 클래스에 대한 순수한 분광값을 갖는다 가정하였으며 각 시기에 따른 결과는 표 1과 같다.

3.3.3.2 토지피복 클래스 구성비 추출

이차계획문제의 최적화를 위해서 2.2절에서 언급한 이론을 바탕으로 하여 토지피복 클래스 구성비의 최적해를 구하였다. 이러한 과정은 MATLAB을 사용하여 작성되었으며, 추출된 토지피복구성비의 예가 그림 10과 같다.

적색에 가까울수록 각 클래스에 대한 픽셀의 확률값이 1에 가까우며, 검정색에 가까울수록 확률값이 0에 가깝다. 검정색과 적색사이의 색깔로 표현된 픽셀은 서로 다른 클래스가 임의의 토지피복 구성비로 혼합된 픽셀(mixel)이라 할 수 있다. 즉, 토지피복 클래스 구성비는 하나의 픽셀에 있어서 각 클래스가 차지하고 있는 면적비이다. 추출된 토지피복구성비는 각 픽셀에서 클래스가 차지하고 있는 면적에 따라 0에서 1의 분포를 하고 있으며 그 분포 예가 표 2와 같다.

따라서 이와 같은 각 픽셀에 대한 토지피복 클래스 구

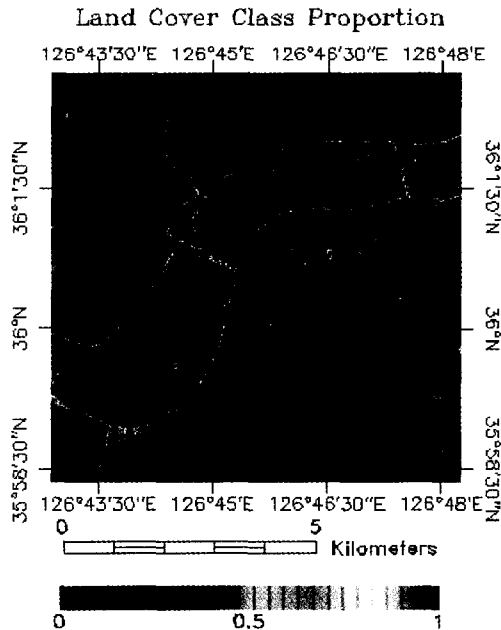


그림 10. 토지피복클래스 구성비 추정 결과의 예(1999, class1)

표 2. 토지피복 구성비 분포의 예 (1999, class1)

CCP	0.0~0.1	0.1~0.2	0.2~0.3	0.3~0.4	0.4~0.5	0.5~0.6	0.6~0.7	0.7~0.8	0.8~0.9	0.9~1.0	Sum
NOP	43356	6283	3951	1966	964	415	391	285	458	14831	72900
PFSA	59.47	8.62	5.42	2.70	1.32	0.57	0.54	0.39	0.63	20.34	100.00

NOP : Number Of Pixel, PFSA : Proportion For Study Area(%)

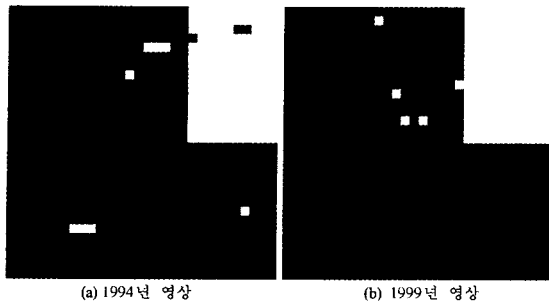


그림 11. 혼련데이터의 이차계획법에 의한 분류결과

성비를 통하여 임의의 두시기에 있어서 지역개발이나 도시개발 등에 따른 토지피복 구성비의 변화나 계절에 따른 미세한 변화를 체계적으로 감지할 수 있으리라 사료된다.

3.3.3.3 이차계획법에 의한 분류결과

이차계획을 사용하여 토지피복 구성비를 추출한 후 다른 분류방법과 동일하게 분류를 실시하였다. 이차계획법에 의한 분류에서는 신경망 출력결과와 최대우도 출력결과와 공통적인 상위 클래스의 분광정보를 순수한 클래스라 가정하였기 때문에 추출된 픽셀에 대한 분류결과는 무의미한 것으로 판단되어 신경망과 최대우도 분류에서 사용된 동일한 혼련데이터에 대하여 분류를 실시하였다. 이는 시험데이터를 사용한 검증과 같은 의미를 갖는다고 할 수 있으며 그 결과를 그림 11에 나타내었다.

1994년의 분류결과(그림 11(a))에서 나대지를 제외한 모든 클래스의 분류결과가 비교적 양호하였다. 나대지의 경우 전자의 순수한 분광값 추정과정에 언급하였듯이 분산이 크기 때문이며 이것은 소수의 픽셀을 추출함에 의해 해결될 수 있는 것으로 심각한 문제로는 사료되지 않는다. 1999년의 분류결과(그림 11(b)) 역시 상당히 양호한 결과를 보여주고 있으나 삼림과 농경지 그리고 습지에 대하여 수역 클래스의 분포가 보이고 있다. 이것의 원인으로는 산림의 경우 지형에 의한 영향으로 산배면에 속하는 픽셀이 수역으로 분류되었다고 사료되며, 도시지역의 경우도 산림과 동일한 영향이다. 즉, 빌딩 그림자에 위치한 픽셀이 이러한 분류결과를 초래하였다고 판단된다. 따라서 DEM과 같은 지형정보와 영상취득시 위성정보 등에 의한 보정과정을 통하여 보정될 수 있으리라 사료된다.

전체적인 분류영상을 살펴보면 두 시기 모두 양호한

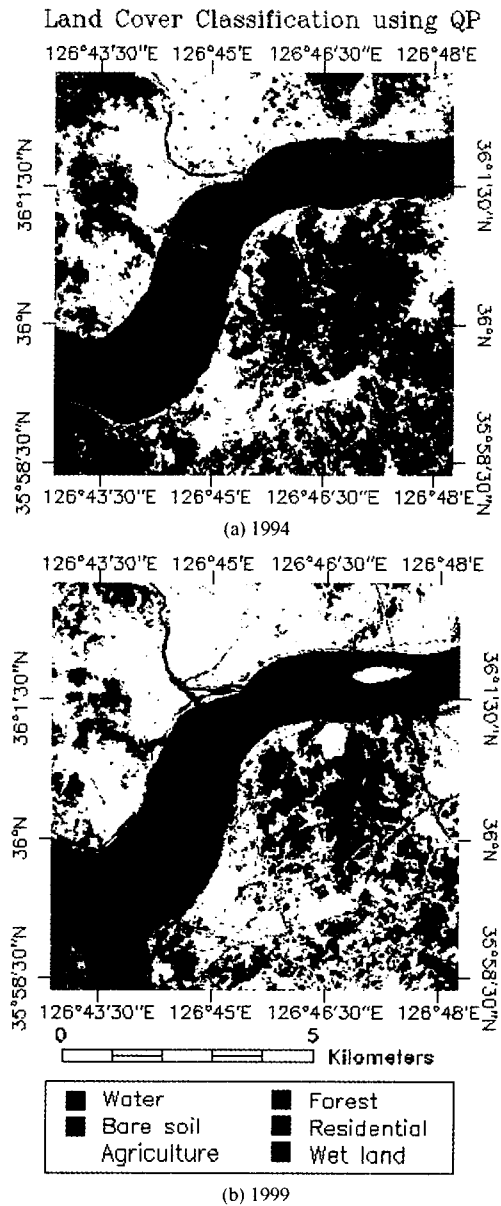


그림 12. 이차계획법에 의한 분류결과

결과를 보여주고 있으나, 1999년도 영상의 경우 혼련데이터에서 보였던 오분류가 산배면과 도시지역의 건물 그림자가 발생하는 후면부분에서 보이고 있다. 그러나 이차계획법에 의한 분류는 기존의 분류방법에서 분류할 수 없었고 무시되었던 결과 값들이 명확하게 나타나고 있음을 보여주고 있다. 즉, 신경망 분류방법에서는 양극으로 이진 분리하려는 성향으로 중간 값이 존재하지 않으며, 최대우도분류방법에서 log 함수의 사용은 미세한 차이를

극대화하였기 때문에 중간값이 존재하기 어려우나 이차 계획법에 의한 분류에서는 구성비가 클래스 결정에 직접적인 영향을 주어 세천이나 작은 소의 분류결과가 명확하게 나타나고 있음을 알 수 있다(그림 12).

3.4 분류정확도 평가 및 고찰

각 분류기에 의해 분류된 훈련데이터 집합과 시험데이터 집합에 대하여 분류정확도를 평가하였다. 특히 시험데이터는 앞에서 언급한 바와 같이 두가지 경우로 평가하였다.

3.4.1 훈련데이터의 분류정확도

먼저 훈련데이터에 대한 분류결과(표 3)에서 알 수 있듯이 1994년 영상의 최대우도에 의한 분류결과를 제외하고 94%에서 98%까지 매우 정확한 분류를 나타내었다. 그러나 Kappa 통계값이 1994년 영상의 최대우도법에 의한 분류에서는 79.87%로서 1999년 영상에 대한 동일한 분류방법 사용시, 94%에 비하면 동일한 위치의 분광값을 사용하였음에도 불구하고 상당히 낮은 분류정확도이다. 이러한 결과는 나대지의 클래스가 절대적으로 영향을 미치고 있음을 알 수 있다. 1994년의 영상이 하절기에 해당되는 영상으로서 나대지의 특성을 띄기 어렵고, 나대지와 도시지역의 통계적인 특성치가 상당히 유사함에 의한 결과라 할 수 있다. 그 반면에 비통계적인 방법이라 할 수 있는 두가지의 분류결과에서는 이와 같은 현상이 나타나지 않고 정확한 분류결과를 얻을 수 있었다. 따라서 통계적인 분류방법은 분광값의 통계적인 특

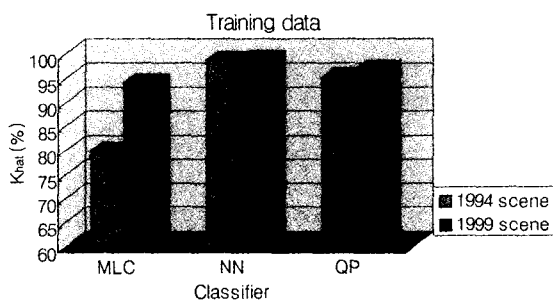


그림 13. 훈련데이터의 분류정확도 비교

표 3. 훈련데이터의 분류정확도(K_{hat})

Training data	MLC	NN	QP
1994 scene	79.87	98.67	95.07
1999 scene	94.00	98.93	96.80

성치가 비슷한 대상체의 해석에는 적용하기가 힘들며 이것이 분류방법의 낮은 정확도의 원인이라 할 수 있다. 제안된 이차계획법에 의한 분류정도는 신경망과 거의 비슷한 결과를 보이고 있다.

3.4.2 시험데이터의 분류정확도

시험데이터는 3.2.3절에서 언급한 바와 같이 729개의 픽셀을 선택하여 시험하였으며, 최대우도법, 신경망, 그리고 제안된 이차계획법으로 시험데이터를 분류하였다.

그 결과, 두 시기 영상이 비슷한 결과를 보였는데 최대우도에 의한 분류에서는 정확도가 90%정도였고, 신경망에 의한 분류와 이차계획법에 의한 방법에서는 96%의 정확도가 얻어졌다. 그러나 신경망의 경우 훈련데이터의 정확도에 비하여 시험데이터의 정확도가 떨어진 반면, 제안된 이차계획법에 의한 분류에서는 정확도가 유지되는 것을 볼 수 있어 제안된 이차계획법과 신경망은 거의 비슷한 분류정확도를 갖는다고 할 수 있다. 원격탐사에 대한 개념이 없는 일반인이 시험데이터를 구성한 경우 (Case 2)에 있어서도 Case 1의 경우와 비슷한 결과를 나타내었다(그림 14, 표 4). 따라서, 서로 상이한 분류방법의 정확도를 비교를 하는데 사용되는 Kappa통계값을 비교하여 볼 때 제안된 이차계획법에 의한 분류가 일반적으로 사용되는 최대우도법에 비하여 평균 약 6%정도 향상된 분류결과를 가져왔으며, 신경망과 비슷한 정도로 원격탐사 영상을 정확하게 분류할 수 있을 것으로 판단된다.

표 4. 각 분류기에 따른 영상별 시험데이터의 분류정확도

Testing data	MLC	NN	QP
Case 1			
1994 scene	88.87	96.31	96.45
1999 scene	89.61	96.18	96.86
Case 2			
1994 scene	89.01	96.86	96.04
1999 scene	94.04	97.54	97.68
Mean(K_{hat})	90.38	96.72	96.76

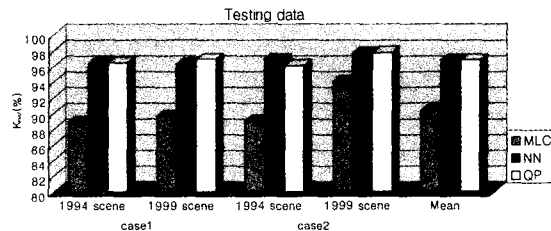


그림 14. 시험데이터의 분류정확도 비교

4. 결 론

본 연구에서는 이차계획법을 원격탐사 데이터의 분류 방법으로서 제시하고자 토지피복 추출에 적용하였으며, 주제도 추출에 일반적으로 사용되는 최대우도와 신경망의 분류정확도와 비교하여 그 적용성을 검토한 결과, 다음과 같은 결론을 얻을 수 있었다.

첫째, 분광값의 통계적인 특성치가 비슷한 대상체의 해석에 있어서는 최대우도법과 같은 통계적인 분류방법보다는 비통계적인 분류방법 즉, 이차계획법이나 신경망을 이용한 분류방법이 우수함을 알 수 있었다.

둘째, 시험데이터를 이용하여 각 분류방법의 정확도를 분석한 결과, 최대우도분류법에 비하여 이차계획법에 의한 분류방법이 평균 6%정도의 향상된 분류결과를 도출할 수 있다

셋째, 이차계획법에 의한 분류에서는 클래스 구성비가 클래스 결정에 직접적인 영향을 주어 기존의 단적인 분류방법에서 무시되었던 결과 값들을 명확하게 나타낼 수 있었다.

넷째, 이차계획을 원격탐사데이터의 분류방법으로서 사용하기 위해서는 우리 나라 특성에 맞는 각 클래스의 순수한 분광 값의 개발과 각 클래스의 분광 데이터베이스의 구축이 요망된다.

참고문헌

1. Giles M. Foody and Manoj K. Arora, "Incorporating mixed pixels in the training, allocation and testing stages

of supervised classifications", Pattern Recognition Letters 17, 1996, pp. 1389-1398.

2. P. M. Atkinson, N. J. Tate, "Advances in Remote Sensing and GIS Analysis", John Willey & Sons, 1999, pp. 17-37

3. Justin D. Paola, Robert A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum likelihood classifier for urban land use classification", IEEE TG&RS, Vol. 33, No. 4, 1995, pp. 981-996

4. Justin D. Paola, Robert A. Schowengerdt, "The effect of neural network structure on a multispectral landuse/landcover classification", PE&RS, Vol. 63, No. 5, 1997, pp. 535-544.

5. Jasbir S. Arora, "최적설계입문", 반도출판사, 1994.

6. Kent, J. T., K. V. Mardia, "Spatial classification using fuzzy membership models", IEEE Trans. Pattern Recognition Machine Intell., PAMI-10. 1988.

7. Drake, N., K. White, "Linear mixture modeling of LANDSAT Thematic Mapper data for mapping the distribution and abundance of gypsum in the tunisian southern", Spatial Data 2000 (Nottingham : Remote Sensing Society), 1991.

8. Quarmby N. A., "Linear mixture modeling applied to AVHRR data for crop area estimation", IJRS, Vol. 13, No. 3, 1992.

9. Settle J. J., and N. A. Drake, Linear "Mixing and the Estimation of Ground Cover Proportions", IJRS, Vol. 14, No. 6, 1993.

10. Jensen, J. R., "Introductory digital image processing : A Remote sensing perspective", 2nd ed. Prentice-Hall, 1996, pp. 197-256.