



고해상도 위성영상의 효율적 지형분류기법 연구 A Study on Efficient Topography Classification of High Resolution Satellite Image

임혜영* · 김황수** · 최준석*** · 송승호****

Lim, Hye Young · Kim, Hwang Soo · Choi, Joon Seog · Song, Seung Ho

요 旨

위성영상에서 실제 지표면의 형태와 지상물체를 구분하여 분류하는 것은 원격탐사의 중요한 목적중의 하나이다. 다중분광영상을 이용한 분류는 일반적인 토지피복도의 제작에 이용되어지고 있으며 영상분류의 방법에는 많은 이론들이 사용되어지고 있다. 본 연구는 대구 달성군 지역의 IKONOS 영상을 MLC(Maximum Likelihood Classification), ANN(Artificial neural network), SVM(Support Vector Machine), Naive Bayes 분류기법들을 이용하여 각각의 분류정확도를 비교분석하였다. 또한 PCA/ICA 전처리 과정을 거친 분류기법들 결과와, Boosting 알고리즘 과정을 거친 후의 결과를 비교하였다. 본 연구의 목적은 적절한 전처리과정과 분류기법을 수행함으로써 가장 효율적인 지형분류 방법을 획득하는데 그 목적이 있다.

핵심용어 : IKONOS, MLC, ANN, SVM, Naive Bayes

Abstract

The aim of remotely sensed data classification is to produce the best accuracy map of the earth surface assigning each pixel to its appropriate category of the real-world. The classification of satellite multi-spectral image data has become tool for generating ground cover map. Many classification methods exist. In this study, MLC(Maximum Likelihood Classification), ANN(Artificial neural network), SVM(Support Vector Machine), Naive Bayes classifier algorithms are compared using IKONOS image of the part of Dalsung Gun, Daegu area. Two preprocessing methods are performed-PCA(Principal component analysis), ICA(Independent Component Analysis). Boosting algorithms also performed. By the combination of appropriate feature selection pre-processing and classifier, the best results were obtained.

Keywords : IKONOS, MLC, ANN, SVM, Naive Bayes

1. 서 론

IKONOS 위성영상은 1999년 발사된, 1m 공간 해상도를 제공하는 상업화된 영상이다. 이 영상은 Landsat, KOMPSAT 등 기존의 다른 위성영상보다 더 세밀하고 다양한 정보를 제공한다. 고해상도 영상을 이용한 분류 기법으로는 Maximum Likelihood 분류기법이 많이 사용되어지고 있으나 자료가 정규분포(Gaussian분포)를 이루지 않으면 분류가 잘되지 않는다. 이러한 단점을 피하기 위하여 ANN을 이용한 분류, SVM을 이용한 분류방법이 사용되어지고 있다.

IKONOS 위성영상을 이용하여 Neuro-Fuzzy, Neural

network, MLC기법으로 토지피복을 분류한 연구에서 Neuro-Fuzzy기법이 혼합된 항목(나대지, 마른초지, 침엽수)을 분류하는데 더 정확한 것으로 나타났다⁽¹⁾.

본 논문은 MLC, ANN, SVM, Naive Bayes 기법을 사용하여 분류정확도를 비교하였으며 또한, PCA/ICA 전처리 과정을 거친 후의 분류기법 결과들을 비교하였고 분류가 잘 되지 않은 항목들의 분류정확도를 높이기 위해 Boosting 기법을 적용하여 비교하였다. 연구의 흐름도는 그림 1과 같다. 본 연구에서 사용된 영상분류기법들의 프로그램 코딩화면은 그림 2와 같고 수행화면은 그림 3과 같다.

2005년 8월 24일 접수, 2005년 9월 26일 채택

* 정회원, 주성대학 토목과 시간강사 (mminha@hotmail.com)

** 경북대학교 컴퓨터학과 교수 (hsk@knu.ac.kr)

*** 정회원, 주성대학 토목과 겸임교수 (hivary@empal.com)

**** 정회원, 주성대학 토목과 부교수 (shsong@jsc.ac.kr)

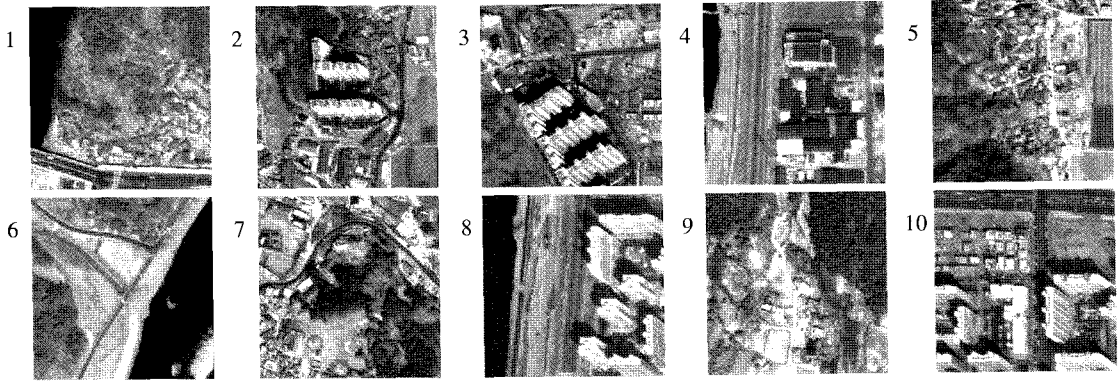


그림 4. 실험영상

SVM은 과다학습의 위험을 최소화하는 예측을 구현할 수 있는 통계적 학습이론을 장점으로 가지고 있다. 비선형 분류는 kernel함수의 내적으로 이루어진다. 일반적인 kernel 함수에는 linear, polynomial, Gaussian이 있지만 분류결과 Polynomial kernel을 사용한 경우 차수가 낮을 때 좋은 결과를 나타냈으며 RBF kernel인 경우 $C = 32, \sigma = 0.2$ 를 사용하였을 때 가장 좋은 결과를 얻을 수 있었다.

2.1.4 Naive Bayes 분류과정 및 분석

Naive Bayes 분류를 하기 위해서는 각 밴드별 클래스의 확률을 구해야 한다. 우선 주어진 훈련자료의 각 밴드값을 오름차순으로 정렬하여 각 밴드에 해당되는 클래스의 빈도수를 계산하여 확률을 구하였다. 영상 분류시 분류 화소의 각 밴드값을 중심으로 8개의 훈련자료확률을 가지고 LSM을 이용하여 각 밴드별 클래스의 확률을 구하였다. 이렇게 구해진 확률을 Naive Bayes분류식에 적용하여 분류하였다.

2.2 MLC, ANN, SVM, Naive Bayes 분류결과 비교

실험 영상에 대한 분류결과는 그림 5와 같다. 이 결과를 비교하여 볼 때 건축구조물과 아파트의 분광적 특성이 비슷하여 오분류가 많이 발생하는 것을 알 수 있다. 전체적인 영상들의 분류정확도를 비교하였을 때 표 1과 같다.

ANN, SVM 분류가 비슷한 분류 정확도를 보였으며 MLC 순으로 나타나는 것을 볼 수 있다.

Naive Bayes 분류시 정확도가 가장 낮은 것을 볼 수 있다.

2.3 PCA 후 영상 분류

분류하고자 하는 영상에서 각 밴드별 평균을 제하고 zero-mean 자료를 만들었다. 4개의 zero-mean 입력 밴드, 4개의 출력 주성분, $\eta = 0.0001$ 로 설정하여 가중치의 변화가 없을 때까지 GHA 알고리즘을 반복 수행하였다.

구해진 가중치를 zero-mean 자료에 곱하고 제하였던 평균을 더하여 새로운 영상 자료를 구성하였다. 이 영상자료를 사용하여 각 클래스 별 훈련자료를 만들어 MLC,

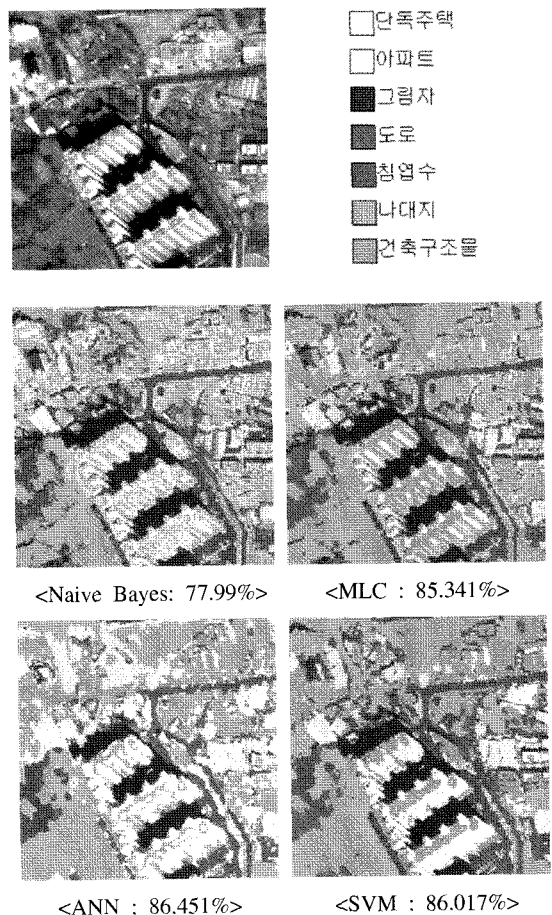


그림 5. MLC, ANN, SVM, Naive Bayes 분류결과

표 1. MLC, ANN, SVM, Naive Bayes 분류 결과

	MLC	ANN	SVM	NB
image 1	84.21	85.46	85.15	79.34
image 2	82.15	84.49	83.30	74.87
image 3	85.34	86.45	86.02	77.99
image 4	86.28	87.24	88.43	79.28
image 5	83.99	84.77	86.39	78.54
image 6	84.04	85.67	85.70	76.47
image 7	83.34	85.28	85.48	75.81
image 8	86.43	89.46	88.84	80.12
image 9	83.88	84.52	84.63	76.49
image10	80.06	82.94	81.27	73.51
평균	83.97	85.63	85.52	77.24

ANN, SVM, Naive Bayes 분류기법을 적용하였다. 여러 개의 확장된 밴드들 사이의 상관성은 다중 분광 영상을 분석하는 경우 자주 접하게 되는 문제점이다. 즉 다양한 파장대의 밴드에서 얻어진 디지털 영상 자료들은 각기 유사하게 보이며, 거의 동일한 정보를 제공하는 경우가 있다. 다중 분광 영상이 가진 여분의 밴드 수를 줄여서 정보를 집약하고 불필요한 분광대를 제거하기 위한 처리로서 PCA를 사용하였다.

PCA 전처리후 영상의 분류는 그림 6과 같다. PCA 전처리 한 영상에 Naive Bayes 분류를 했을 때 분류 정확도가 많이 향상되었으나 나머지 분류 결과에는 별 차이를 보이지 않았다. PCA 전처리 한 경우에도 표 2와 같이 ANN, SVM분류가 비슷한 정확도를 보였으며 MLC가 그 다음 정확도를 보였다.

2.4. ICA 후 영상분류

ICA는 복합된 신호에서 구성되는 신호들을 찾아내는 신호 처리방법이다. 다중분광 영상도 파장에 따라서 서로 다른 분광적 특성을 나타내므로 고차원 통계를 이용하여 분광적으로 서로 다른 신호들을 독립 요소 (Independent Component)로 분리할 수 있다.

영상을 작은 조각으로 나누어 각 조각을 임의의 벡터로 나타내고, 이 벡터에 ICA를 적용하여 분광적으로 서로 다른 요소들을 추출해 낸다. 자료에서 먼저 평균을 빼고 구성 요소간의 상관관계가 없고 편차가 단일한 whitening 단계를 거쳐 FastICA 알고리즘을 수행하였다.

구해진 기중치를 영상자료 밴드 값에 곱하여 새로운 영상 자료를 만들었다. 이것을 이용하여 각 클래스 별 훈련자료를 만들어 MLC, ANN, SVM과 Naive Bayes 분류기법을 적용했다. 모든 분류 기법에서 ICA 전처리를 한 경우 분류 정확도가 향상된 것을 표 3에서 알 수 있다. 특히 Naive Bayes, SVM 분류 정확도가 많이 향상 되었다.

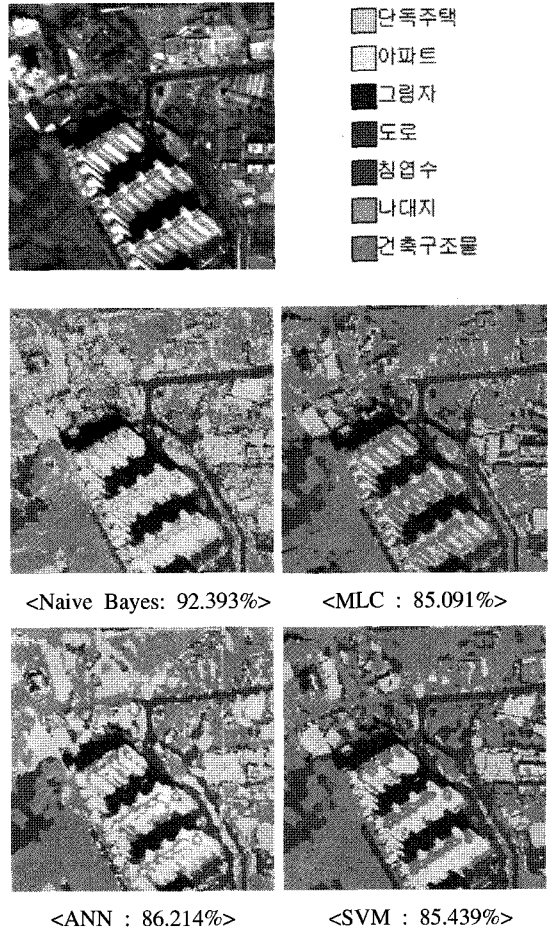


그림 6. PCA 후 영상의 분류결과

표 2. PCA 후 영상의 분류결과

	P/MLC	P/ANN	P/SVM	P/NB
image 1	84.08	86.12	84.72	87.52
image 2	82.11	84.53	83.01	88.86
image 3	85.09	86.21	85.44	92.39
image 4	87.43	88.97	87.75	91.82
image 5	85.22	85.17	85.92	88.35
image 6	83.68	85.32	85.45	89.77
image 7	84.40	85.10	84.84	86.14
image 8	87.25	88.43	87.31	90.44
image 9	83.70	84.94	84.24	87.51
image10	80.27	82.75	81.06	85.29
평균	84.32	85.75	84.97	88.81

다. ANN 분류에서는 전처리를 하기 전과 비슷한 결과를 보였으며 MLC에서는 PCA 전처리를 한 경우와 비슷한 결과를 보였다. Naive Bayes 분류를 하기 위해 ICA 전처

리를 한 훈련자료에서 각 밴드별 클래스의 확률을 구하였다. 영상을 분류 할 때 구해진 훈련자료의 확률을 기초로 각 화소에 대한 확률을 구하여 Naive Bayes 분류식을 적용하였다.

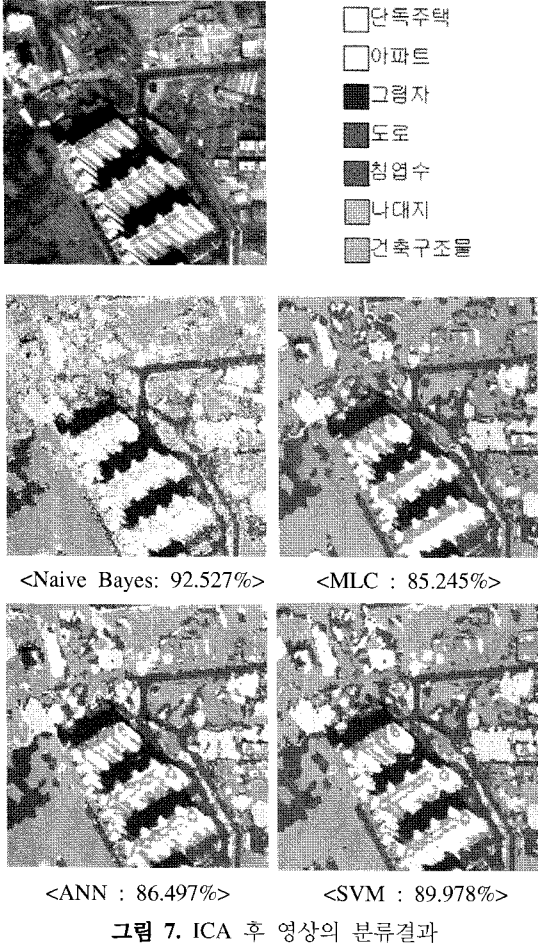


그림 7. ICA 후 영상의 분류결과

표 3. ICA 후 영상의 분류결과

	(단위: %)			
	I/MLC	I/ANN	I/SVM	I/NB
image 1	85.06	85.06	86.23	89.56
image 2	82.04	85.04	86.08	90.94
image 3	85.25	86.50	89.98	92.53
image 4	87.31	88.82	89.10	92.39
image 5	85.20	84.47	86.81	89.78
image 6	84.83	85.73	86.37	90.49
image 7	84.49	85.16	85.94	88.12
image 8	88.55	88.93	89.97	91.88
image 9	84.51	84.92	85.31	89.43
image 10	80.81	82.64	83.88	87.25
평균	84.80	85.73	86.97	90.24

각 화소에 대한 확률을 구하는 방법으로 LSM(Least Square Method)과 RBF(Radial Basis Function)를 사용하였다. LSM을 사용한 경우 각 화소에 대하여 확률을 매번 구해야 함으로 처리시간이 많이 걸렸지만 RBF를 사용한 경우 각 밴드별 클래스의 확률을 미리 훈련하여 실제 영상 분류시 처리속도를 향상시키고 LSM 보다 나은 분류 정확도를 얻을 수 있었다. 그림 8은 10개의 클래스중 단독주택, 아파트, 그림자, 도로 등 4개 클래스의 band 1에 대한 확률의 LSM, RBF fitting을 나타낸다.

2.5 Boosting을 이용한 분류

2.5.1 Ensemble Classifier

여러 개의 분류기를 사용하는 개념은 많은 분류기를 조합하여 하나의 분류기를 만들고 구성된 각각의 분류기 보다 더 좋은 성능을 갖는 것이다. 여러 가지 방법으로 분류기를 조합할 수 있으나 가장 일반적인 방법은 Ensemble이다.

Ensemble은 각 분류기가 모든 자료에 대하여 접근하

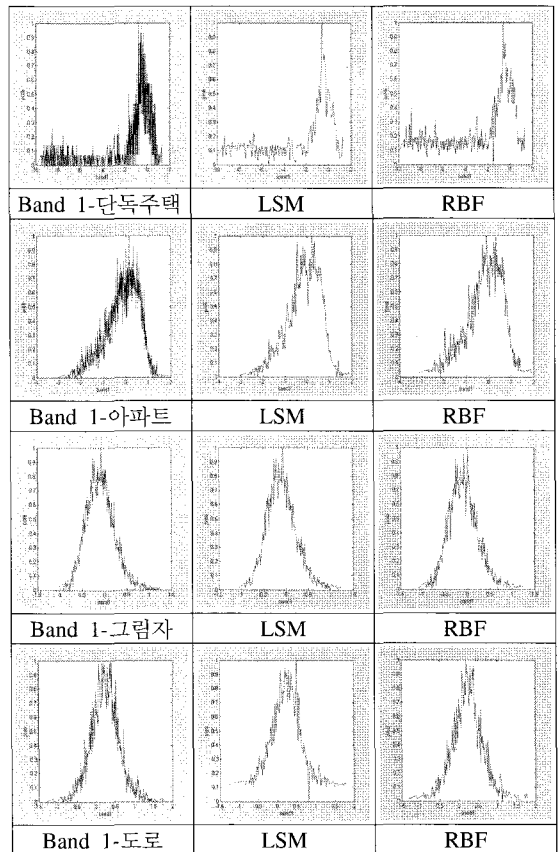


그림 8. Band 1에 대한 확률의 LSM, RBF fitting

고 모든 분류기의 결과가 전체 시스템의 분류를 결정하기 위하여 조합된다. 이러한 조합은 일반적으로 구성원인 분류기에 의해서 자료가 가장 많이 분류된 클래스로 전체 시스템의 분류를 결정하는 다수 투표 방식을 택한다. 각 분류기는 상호간에 완전히 독립적이어서 고차원의 병렬 시스템을 구성함으로써 매우 효율적인 다중처리 시스템을 구현할 수 있다. 하나의 분류기보다 여러 개로 구성된 분류기가 더 좋은 결과를 얻기 위해서는 구성되는 분류기가 달라야만 한다. 만약 다르지 않으면 모든 분류기가 같은 결과를 낼 것이며 이것은 처리 시간만 가중시키는 것이다. 각각의 분류기를 유일하게 하기 위해서는 ANN의 경우 층의 수를 다르게 하거나 뉴런의 수를 다르게 할 수 있고 서로 다른 분류기를 사용하거나 서로 다른 훈련 자료를 사용할 수 있다.

최근 몇 년 동안 원격탐사에서 여러 개의 분류기를 조합하여 하나의 분류기를 만드는 연구들이 진행되어왔다. Ensemble classifier를 사용함으로써 개별의 분류기보다 분류 정확도가 높게 나타나고 있다. Ensemble classifier는 그림 9와 같이 개별 분류기들의 결정에 대해 투표함으로써 최종 분류결정을 한다. 이렇게 함으로써 개별의 분류기보다 더 정확한 분류를 할 수 있다.

본 연구에서는 이러한 Ensemble 방법의 하나인 Boosting 기법을 사용하였다.

2.5.2 Boosting

Boosting은 많은 훈련자료가 제한된 훈련 자료를 부분 대치함으로써 만들어진다. 각 자료의 분류결과에 따라 가중치를 부가하는데, 가중치가 높을수록 분류기에 영향을 많이 준다. 분류가 시행될 때 마다 해당 분류기의 성능에 따라서 가중치가 조정되고 잘못 분류된 자료의 가중치가 커진다. 이렇게 함으로써 전 단계의 분류기에서 잘못 분류된 자료가 더 많이 다음 분류기의 훈련 자료에 포함된다. 최종 분류기는 각각의 분류기 결과들을 투표하여 분류를 결정한다.

본 연구에서는 다음의 AdaBoost.M1 알고리즘을 사용하였다.

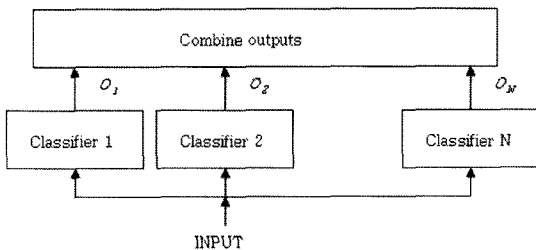


그림 9. Ensemble Classifier

입력:

$$L = \{(x_1, g_1), \dots, (x_N, g_N); x_i \in X, g_i \in G\}$$

$$G = \{1, \dots, |G|\}, \text{ weak classifier } h: X \rightarrow G$$

T : boosting round의 수

초기화 :

$$D_1(i) = \frac{1}{N}$$

For $t = 1, \dots, T$:

- 분포 D_t 를 가지고 weak classifier를 훈련한다. h_t 는 다음의 ϵ_t 를 줄여간다.

$$\epsilon_t = \sum_i D_t(i) [|h_t(x_i) \neq g_i|]$$

- $\epsilon_t \geq \frac{1}{2}$ 이면 $T = t-1$ 상태에서 출력을 결정한다.

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

- 분포 D_t 를 수정한다.

$$D_{t+1}(i) = \frac{D_t(i) e^{-2\alpha_t [|h_t(x_i) = g_i|]}}{Z_t}$$

Z_t : normalization factor

출력 :

최종분류

$$H(x) = \operatorname{argmax}_{g \in G} \left(\sum_{t=1}^T \alpha_t [|h_t(x) = g|] \right)$$

알고리즘을 반복할 때마다 이전 분류기에서 잘못 분류된 훈련자료에 집중하여 분류기를 훈련한다. 각 분류기의 훈련자료는 주어진 전체 자료를 대표하지 않고 이전 분류기에서 분류하기 힘들었던 부분의 집합으로 구성된다.

2.5.3 Boosting 분류과정 및 분석

처음 분포를 균등하게 초기화한 후 각 분류기에 따라 훈련하고 그에 대한 분류결과에 따라서 alpha의 분포를 구하였다. 새로 구해진 분포에 따라서 훈련 자료를 다시 구성하는데, 잘못 분류된 자료의 확률이 높아지므로 잘못 분류된 자료의 수가 증가한 새로운 훈련 자료를 구성하여 훈련, 분류 과정을 반복한다.

Boosting을 수행했을 때 처음에는 오류가 증가하다가 반복횟수가 증가할수록 점차 오류가 감소하고 alpha가 증가한다. Boosting을 이용한 최종 분류 결과 오류 발생 부분이 많이 해결되어 표 4 와 같이 분류 정확도가 향상되었다.

표 4. Boosting 분류결과

(단위: %)

	MLC	MLC/Boosting	ANN	ANN/Boosting	SVM	SVM/Boosting
image 1	84.21	85.97	85.46	85.97	85.15	87.23
image 2	82.15	85.34	84.49	86.45	83.30	86.43
image 3	85.34	86.43	86.45	87.53	86.02	87.55
image 4	86.28	86.75	87.24	88.45	88.43	89.11
image 5	83.99	84.10	84.77	85.12	86.39	88.97
image 6	84.04	86.45	85.67	86.49	85.70	88.49
image 7	83.34	85.36	85.28	86.27	85.48	88.25
image 8	86.43	88.42	89.46	90.35	88.84	89.31
image 9	83.88	85.99	84.52	87.83	84.63	86.18
image10	80.06	84.22	82.94	85.43	81.27	85.42
평균	83.97	85.90	85.63	86.99	85.52	87.69

3. 실험결과 및 분석

본 연구에서는 여러 가지 분류기법을 적용하여 IKONOS 위성영상을 분류하였으며 그 결과를 비교·분석하였다. 10개의 영상으로 나누어 다음과 같은 분류기법을 적용하여 분석하였다.

- (1) MLC(Maximum Likelihood Classification)
- (2) ANN
- (3) SVM(Support Vector Machine)
- (4) Naive Bayes
- (5) PCA(Principal Component Analysis) 전처리 후 영상분류
- (6) ICA(Independent Component Analysis) 전처리 후 영상분류
- (7) Boosting기법을 적용한 분류

분류항목으로는 수역, 침엽수, 활엽수, 아파트, 단독주택, 그림자, 나대지, 아스팔트, 습지, 비닐하우스, 농경지, 도로로 구분하였다.

전처리를 하지 않은 경우 ANN, SVM 분류 정확도가 비슷했으며 Naive Bayes 분류가 가장 낮은 정확도를 보였다. PCA 전처리 과정을 거친 경우 Naive Bayes 분류에서 분류 정확도가 많이 향상되었지만 MLC, ANN과 SVM 분류는 전처리를 하지 않은 경우와 비슷한 결과를 보였다. ICA 전처리 과정을 거친 경우 Naive Bayes 분류 결과가 가장 좋았으며 SVM, ANN 분류, MLC 순으로

나타 났다. MLC 결과는 ICA/PCA 전처리를 했을 때 분류결과가 더 좋았으며 ANN 분류 결과는 모든 경우에서 비슷했다. SVM 분류에서는 ICA 전처리한 분류결과가 가장 좋았고 Naive Bayes 분류에서는 PCA / ICA 전처리를 한 경우 분류 정확도가 많이 향상 되었다. Boosting 했을 때는 하기 전 보다 더 좋은 분류결과를 보였다. 이러한 영상을 사용하여 여러 가지 분류기법을 적용하였을 때 분류정확도의 순서는 다음과 같이 나타났다.

- (1) ICA/Naive Bayes
- (2) PCA/Naive Bayes
- (3) ICA/SVM
- (4) ANN, PCA/ANN, ICA/ANN, PCA/SVM
- (5) ICA/MLC, PCA/MLC
- (6) MLC
- (7) Naive Bayes

그림 10은 훈련자료 클래스 중 아파트와 단독주택의 각 밴드분포를 나타낸다.

각 분류기법의 수행시간의 비교는 ICA 전처리 후 영상을 분류할 경우 걸리는 시간을 비교하였다. ANN 분류시 속도가 가장 빨랐으며, Naive Bayes 분류시 LSM을 사용하였을 때 가장 오래 걸렸다. LSM은 각각의 화소에 대해 확률계산을 하기 때문에 처리시간이 많이 걸리지만 RBF를 이용하여 확률을 구하는 경우 한번의 훈련과정을 거쳐 분류를 시행함으로써 시간이 단축된다. 그림 11은 분류기법의 수행시간의 비교를 나타낸다.

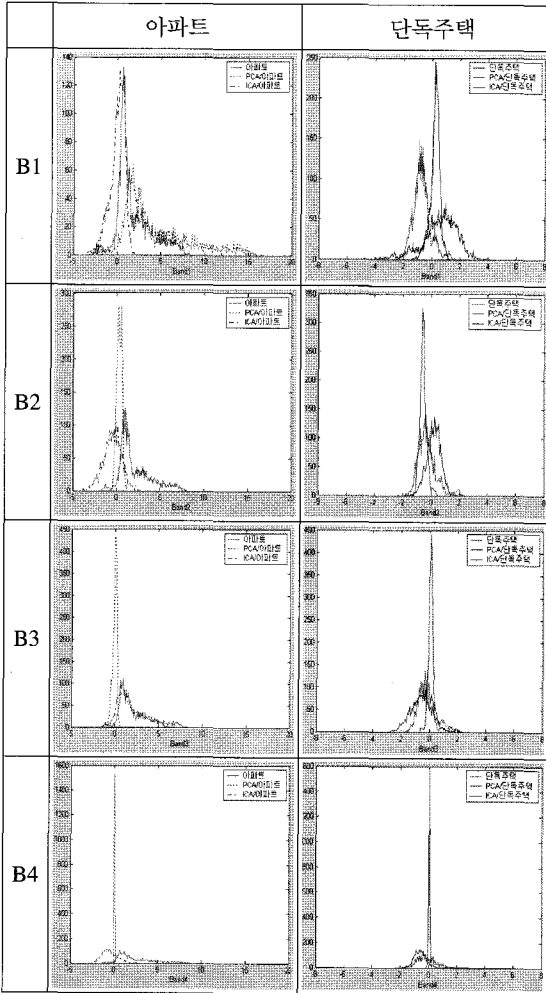


그림 10. 클래스별 밴드 분포

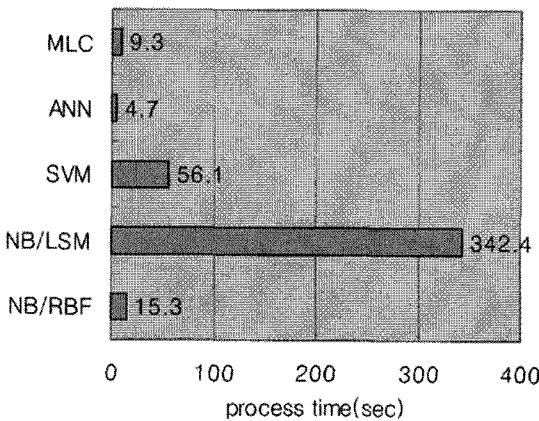


그림 11. 분류기법의 수행시간 비교

4. 결 론

본 연구에서는 IKONOS 고해상도 위성영상을 효율적으로 분류하기 위한 기법들을 고찰하고 각 기법들을 적용하여 실험하고 분류결과를 비교·분석함으로써 다음과 같은 결과를 얻을 수 있었다.

1. MLC는 영상 분류에서 가장 많이 쓰이는 기법중의 하나로 훈련자료의 확률과 분포를 이용하여 분류할 수 있으나 훈련자료 밴드간의 상관관계가 높거나 정규분포(Gaussian분포)가 아닌 경우에는 분류가 제대로 되지 않는다. 또한 MLC에서 ICA/PCA 전처리를 수행한 후의 분류 정확도는 향상되는 것을 알 수 있었다.
2. ANN 분류에서는 주어진 훈련자료에 따라서 가중치를 조정해가는 과정을 거침으로 확률이나 분포에 상관없이 분류 정확도가 높은 것을 볼 수 있다. PCA/ICA 전처리를 한 경우에도 비슷한 분류 정확도를 보이는 것을 알 수 있었다.
3. SVM 분류에서는 ICA 전처리 했을 때 분류 정확도가 향상되었다.
4. Naive Bayes 분류는 PCA/ICA 전처리를 했을 때 분류결과가 향상됐으며 분류 영상의 확률을 fitting할 때 RBF를 사용하면 처리 속도도 빠르고 분류 정확도도 향상됨을 볼 수 있었다. Boosting은 잘못 분류 되는 항목을 집중적으로 훈련함으로 일반적인 분류기법보다 분류 정확도가 향상되었다.

본 연구를 통하여 분류 정확도는 분류기법에 의존하는 것을 알 수가 있었으며, 특징 추출 방법과 분류기의 조합으로 지형의 분류정확도를 향상시킬 수 있었다.

참고문헌

1. J.Han, S. Lee, K.Chi, K.Ryu, "Comparison of Neuro-Fuzzy, Neural Network, and Maximum Likelihood Classifiers for Land Cover Classification using IKONOS Multispectral Data", *IEEE Geoscience and Remote Sensing Symposium*, 2002.
2. Ramanathan Sugumaran, Manoj Kumar Pavuluri, *Student Member, IEEE*, and Daniel Zerr, "The Use of High-Resolution Imagery for Identification of Urban Climax Forest Species Using Traditional and Rule-Based Classification Approach", *IEEE*, 2003.
3. V.N. Vapnick, *Statistical Learning Theory*, John Wiley and Sons Inc., 1998.
4. C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification", *Int. J. Remote sensing*, Vol.23, No. 4, pp. 725-749, 2002