

# IRS-1C PAN 데이터와 Landsat TM 데이터의 PCA 중합화상을 이용한 토지이용 분류 정확도 분석

An Analysis of the Landuse Classification Accuracy Using PCA Merged  
Images from IRS-1C PAN Data and Landsat TM Data

안 기 원 \*      이 호 성 \*\*      서 두 천 \*\*      신 석 호 \*\*  
Ahn, Ki-Won      Lee, Hyo-Sung      Seo, Doo-Chun      Shin, Sok-Hyo

## 要 旨

본 연구에서는 고 해상력의 IRS-1C PAN 데이터와 다중 관측과장대를 갖고 있는 Landsat TM 데이터를 PCA(principal component analysis)기법으로 중합하고, 중합화상을 이용하여 토지이용분류를 수행, TM 원화상과 분류정확도를 비교, 분석하고자 하였다. 분류결과를 평가하기 위해서 10개의 분류항목으로 구성된 sample data를 생성시켰으며, 생성된 sample data의 전체정확도(overall accuracy)로서 분류결과를 평가하였다. 그 결과 여러 밴드를 사용할 수 있다는 이점을 가진 PCA기법으로 Landsat TM 데이터와 IRS-1C PAN 데이터를 중합하여 토지이용분류를 수행할 경우, IRS-1C PAN데이터와 TM13457의 중합화상에서 95.1 %로 분류정확도가 가장 높았으며, 특히 TM123457 또는 TM 13457의 중합화상의 경우 TM원화상보다 분류정확도가 가장 크게 향상(6.2 %)되었다. 또한 토지이용분류에 있어서 중합칼라합성화상이 TM 원화상 보다 전체적으로 약 5.16 %의 정확도 향상을 가져왔다.

## ABSTRACT

The main object of this study was to prove the effectiveness of PCA(principal component analysis) merged images produced by PCA method using high resolution IRS-1C PAN data and multispectral Landsat TM data. A sample data which has ten classes was generated for evaluation of the overall classification accuracy. In result, merged sample image which TM13457 bands with IRS-1C PAN data by PCA method showed best result (95.1 %). Especially, the largest improve (6.2 %) in classification accuracy was resulted when IRS-1C PAN data was merged with TM123457 or TM13457 images. In addition, landuse classification accuracy of the PCA merged images was improved (5.16 %) than original color composite images of Landsat TM data.

\* 경상대학교 공과대학 토목공학과 교수  
(경상대학교 부속 생산기술연구소 연구원)

\*\* 경상대학교 대학원 토목공학과 박사과정

## 1. 서 론

지구자원탐사용 인공위성에 탑재된 센서의 종류가 다양해지고 그 해상력이 크게 향상되고 있기 때문에 이들 인공위성에서 관측 수집된 화상데이터를 각종 주제도 작성에 활용할 필요성이 많아졌다.

특히 주제도 작성 중에서도 토지이용도를 작성하여 도시계획, 수문 및 유출특성 분석 등에 많이 활용되고 있다. 각기 다른 센서에서 관측 수집된 화상데이터를 이용하여 토지이용분류를 수행할 경우, 원화상을 직접 사용하기보다는 보다 유효한 중합화상을 작성하여 사용하는 것이 토지이용 분류 정확도를 향상시키고 아울러 주제도 작성에 필수적인 분광반사특성과 높은 공간해상력이 보존된 화상을 얻을 수 있다.

국내외적으로도 화상데이터의 중합 및 그 활용성에 관하여 많은 관심을 갖고 활발한 연구가 이루어지고 있다.

Carper et al(1990)은 IHS(intensity-hue-saturation) 방법이 식생지역의 판독에는 적절하지 않음을 제시한 바 있으며, Chavez et al.(1991)은 PCA방법이 IHS 및

HPF(high-pass filter)방법보다 분광반사특성의 왜곡이 제일 작다는 결과를 발표하였다.

국내 연구로는 손덕재와 이종훈(1991)은 IHS변환방법으로 SPOT PAN 데이터와 XS 데이터를 중합하여 시각적 분석가능성의 향상을 제시하였다.

또한 IRS-1C PAN데이터와 SPOT XS 데이터 또는 Landsat TM 데이터의 중합방법을 비교 분석하여 PCA방법이 특히 산림 및 식생지역에서 공간해상력이 잘 보존되고 있음을 제시한 바 있다.(안기원 등, 1998; 안기원 등, 1999).

본 연구에서는 고 해상력의 IRS-1C PAN 데이터와 다양한 분광반사특성을 갖고 있는 Landsat TM 데이터를 사용하여 PCA기법으로 중합화상을 작성하고 토지이용분류를 수행, TM 데이터의 원화상과 분류정확도를 비교, 분석하고자 하였다.

## 2. 사용데이터 및 기하보정

사용된 인공위성 화상데이터는 1995년 12월 인도 ISA(Indian Space Agency)에서 발사된 IRS-1C 위성

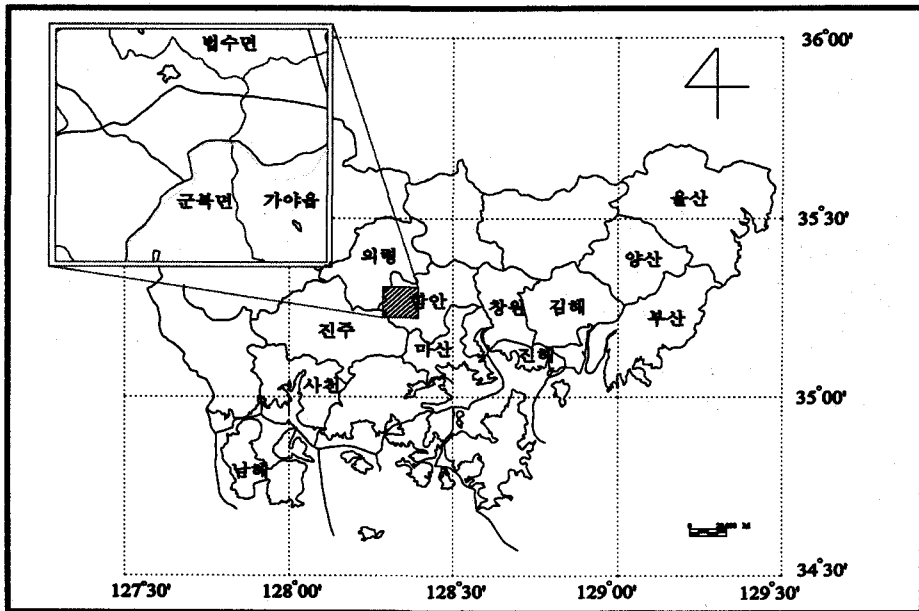


그림 2.1 Location of the study area

의 PAN 데이터와 Landsat TM 데이터로서 그 주요 재원은 표 2.1에서 보는 바와 같다.

구입된 화상데이터는 공급기관에서 표준적으로 채택하고 있는 지구타원체 및 평면직각좌표 투영법이 우리 나라 국토기본도의 경우와 다르기 때문에 기하보정을 실시하여 위성데이터를 지도좌표체계로 맞추어야 한다. 본 연구에서는 지상기준점의 지도좌표를 사용하는 기하보정방법을 수행하였으며, 화상좌표와 지도좌표간의 좌표변환식은 1차다항식을, 화소값의 재배열(resampling)방법으로는 공1차내삽법(bilinear interpolation)을 사용하였다. 재배열에 있어서 한 화소에 해당하는 지상크기는 IRS-1C PAN 데이터 및 Landsat TM 데이터 모두 5 m × 5 m가 되도록 하였다. 기하보정시의 평균제곱근오차(RMSE)는 각각 ± 0.5 화소 이내가 되도록 하였다.

기하보정 화상에서 IRS-1C PAN 데이터 및 Landsat TM 데이터 모두 1800 × 1860 (row × column) 화소크기로 질출하여 연구대상지역으로 선정하였다. 이 지역은 그림 2.1에서 보는 바와 같이, 위도 35° 14' ~ 35° 19' , 경도 128° 18' ~ 128° 24' 사이의 경상남도 함안군 일대의 가야읍, 군북면 및 범수면

을 포함하는 지역으로서, 도시역, 산림지역, 식생지역 및 수역이 적절히 포함되도록 하였으며, 지상크기는 종횡방향으로 9 km × 9.3 km 이고 면적은 83.7 km<sup>2</sup>이다.

### 3. 분류항목 및 Training field 선정

토지이용 분류결과를 검증하기 위하여는 비교할 자료나 화상이 필요한데 실측을 행하거나 자치단체발행의 토지이용현황도를 활용하여야 하지만 실측의 경우는 현지조사의 작업량이 방대하고 토지이용현황도는 작성시기가 위성데이터의 관측시기와 상이하거나, 내용이 부족하기 때문에 위의 방법들 모두 유효한 비교자료가 되기 어렵다.

따라서 본 연구에서는 1차 선행분류를 행하여 분류결과가 비교적 양호한 항목별 화소들을 추출하여 sample data를 만들고 비교화상으로 활용하였다.

#### 3.1 Sample data의 구축을 위한 1차선행분류

해석대상영역의 분류항목 및 sample data 선정을 위해서 비교사분류인 ISODATA (iterative self-organizing data analysis technique)에 의해 1차적으로 토지이용분류를 수행하였다. 이 방법은 군집화를 실행하기 이전에 분류하고자 하는 군집의 수, 수렴 임계치 (반복계산 시 변화하여 분류되는 화소수의 최대 백분율) 및 반복계산의 최대 수행 회수, 한 군집 내에 허용되는 화소의 최소수를 지정하여, 반복계산마다 각 군집에는 새로운 평균이 계산되어 그 평균의 변화가 수렴 임계치 이내일 경우 작업을 끝내는 방법이다.

본 연구에서는 화상처리용 소프트웨어인 ERDAS IMAGINE 8.3.1을 이용하여 초기 항목 수를 30개로 선정하였고 전체 화소의 군집에 대한 수렴치를 98%로 부여하여 수행하였다. 그러나 처리 결과, 한 군집에 포함되는 화소수가 너무 적거나 다른 항목과의 유사성이 나타나지 않는 군집이 생겨나 최종적으로 10개의 군집을 얻게 되었다.

얻어진 10개의 군집을 검토하여 최종적인 분류항목

표 2.1 Image data used

Parameter	IRS-1C PAN	Landsat TM
Spatial resolution	5.8 m(at nadir)	30 m
Scene size	70 km × 70 km	185 km × 170 km
Date of acquisition	April 19, 1997	March 30, 1997
Path-Row	143-045	114-036
Number of bands	1	7
Cloud	0%	0%
Spectral bands	0.5 ~ 0.75 $\mu$ m	0.45~2.35 $\mu$ m (without band 6)
Sun azimuth/elevation	145.2° /62°	132° /48°
Data correction	Level 2	Bulk

명을 부여하였다. 이 작업에 있어서는 TM1, TM2, TM3 및 IRS-1C PAN 데이터를 사용하여 PCA기법으로 중합된 화상을 고해상도 컬러 모니터(21", 1600×1200 resolution)에 출력하여 1:25,000 지형도와 1:50,000 식생도 등과 비교 검토하였다.

그림 3.1은 TM123의 true color화상이고 그림 3.2는 본 검토 작업에서 사용된 PCA중합화상이다. 검토결과, 산림지역에서는 침엽수와 혼효림의 구분이 가능한 것으로 판단되었다. 따라서 최종적으로 주거지(residential), 침엽수(evergreen), 혼효림(mixed forest),

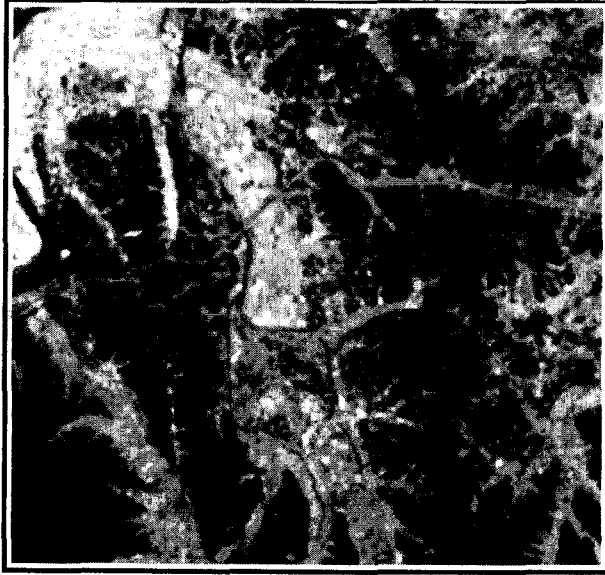


그림 3.1 Original Landsat TM image of study area (true color ,TM123)

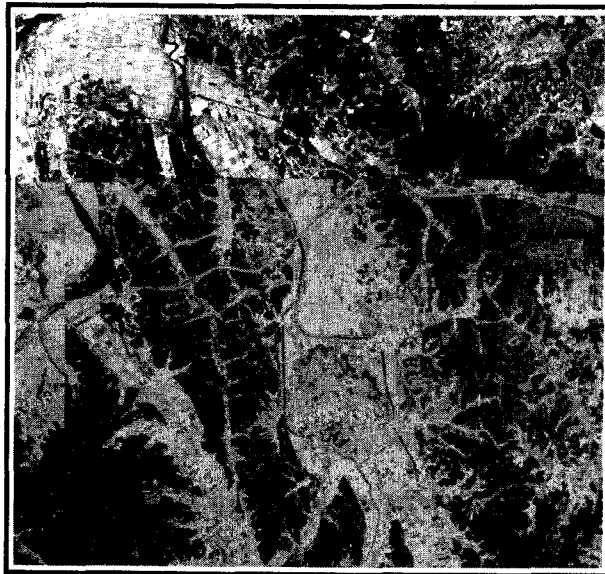


그림 3.2 Color displayed merged image with Land-sat TM123 and IRS-1C PAN data using PCA method

논(cropland), 식생(vegetation), 초지(brush), 습지(wetland), 나지(barren land), 비닐하우스(vinyl house) 및 물(water)의 총 10개의 분류항목을 확정하였으며 항목명은 USGS(U.S. Geological Survey)의 토지분류 체계를 참조하였다.

### 3.2 Sample data의 구축 및 평가

1차 선행분류결과화상에서 sample data를 구축할 때에는 확률적으로 표본을 추출하는 것이 적합하다. 하지만 경우에 따라서는 완전한 확률표본보다는 약간의 제약을 주는 추출계획을 실현함으로써 보다 더 유효한 정보를 수집할 수도 있다. 일반적으로 표본추출의 방법으로는 단순확률추출법(simple random sampling), 층화확률추출법(stratified random sampling), 집락확률추출법(cluster random sampling) 및 계통추출법(systematic sampling)이 있다. 이 4가지 표본추출기법중 본 연구에서는 층화확률추출법을 사용하여 sample data를 구축하고자 하였다. 이 기법은 모집단의 특성에 따라 여러 개의 층으로 구분될 수 있는 경우에 표본을 층별로 할당하여 추출하는 기법이다. 그리고 각 층은 서로 이질적(heterogeneous)이어야 하고, 층내의 데이터들이 동질적(homogeneous)인 경우에 적합하다(김우철 등, 1998; 김정년, 1998).

표본의 크기는 조사목적과 추론결과의 요구되는 정

밀도에 따라 달라질 수 있다. 일반적으로 인공위성에서 수집되는 데이터는 관측시기, 센서의 종류, 지역의 특성 등 많은 요소에 따라 화상의 특성이 결정된다.

이들 모두를 고려하여 표본의 크기를 결정하는 것은 현실적으로 불가능하므로 임의의 모집단에 대하여 근사적으로 적용될 수밖에 없다.

따라서 본 연구에서는 사용되는 화상데이터의 기본 통계값을 이용하여 임의적으로 표본의 크기를 결정하고자 하였다. 필요한 표본분산이 기지인 경우  $100(1-\alpha)\%$  오차한계를  $d$ 이하로 하는데 필요한 표본크기는 식 (3.1)에 의해서 구해진다(김우철 등, 1998).

이때 화상의 평균에 대한 추정을 90%, 오차한계를 0.3 이내, 표준편차 45로 결정하였으며, 필요한 표본의 크기는

$$n \geq (1.645 \times 45 / 0.3)^2 = 60885.6 \text{ 로서}$$

약 60,886개이다.

위의 표본 크기와 화상데이터 구축시의 편리성을 고려하여 sample data의 크기를  $250 \times 250$  화소(column $\times$ row, 62500 화소)로 결정하였다.

$$n \geq (z_{\alpha/2} \sigma / d)^2 \tag{3.1}$$

여기서  $n$  : 표본크기

$z_{\alpha/2}$  :  $\alpha/2$  % 상방 백분위수

$\sigma$  : 표준편차

$d$  : 오차한계

sample data의 각 항목의 화소수 결정에 있어서는

표 3.1 Number of pixels of generated sample data using stratified random sampling

Class	Pixels (first classified image)	Standard deviation (mean of 6 TM bands)	Pixels (sample data)
Unclassified	2214594	-	-
Residential	106798	4.91	6875
Evergreen	183055	3.09	6875
Mixed forest	122223	3.80	6250
Cropland	255849	4.16	14375
Vegetation	137467	4.61	8750
Brush	163679	3.88	8125
Wetland	11437	6.38	625
Barren land	64416	6.61	5000
Vinyl house	77033	4.84	5000
Water	11449	4.37	625
Total	3348000		62500

모집단의 표준편차를 고려하여 아래와 같은 식에 의해서 구해진다.

$$n_i = n \frac{N_i \sigma_i}{\sum N_i \sigma_i} \quad (3.2)$$

여기서

$n_i$  : sample data의  $i$  번째 분류항목의 화소수

$n$  : sample data의 총화소수 (62500화소)

$N_i$  : 1차 분류된 화상에서  $i$  번째 분류항목의 화소수

$\sigma_i$  : 1차 분류된 화상에서  $i$  번째 분류항목의 표준편차

$i$  : sample data의 항목수 ( $i = 1 \sim 10$ )

표 3.1은 1차분류화상의 항목별 화소수, 표준편차 및 결정된 sample data의 항목별 화소수이다. 표 3.1의 표준편차는 각 항목의 TM밴드별 평균을 취한 것이다. sample data 구축에 있어서는, random sampling이 될 수 있도록 난수 발생 프로그램을 작성하여 사용하였다. 결정된 10개의 분류항목과 총 62,500 화소 크기의 sample data를 구축하기 위하여 62,500 화소 각각 난수를 발생시켜 random한 row 및 column 값이 얻어지면 그 위치의 Landsat TM, IRS-1C PAN 화상데이터 값 및 1차선행 분류결과의 분류항목 값을 sample data의 해당 항목 화상으로 그림 3.5의 참조화상과 같은 설계안대로의 sample data를 구축하였다.

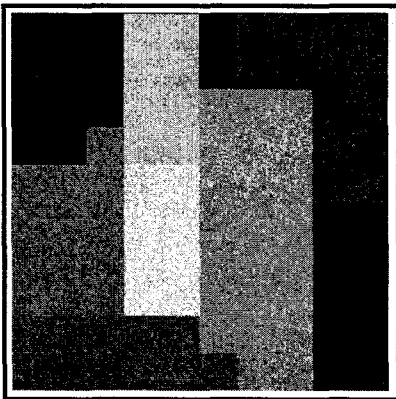


그림 3.3 Generated sample data of Landsat TM234 using stratified random sampling

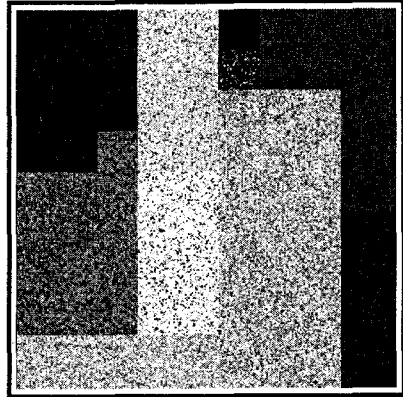


그림 3.4 Generated sample data of IRS-1C PAN data using stratified random sampling

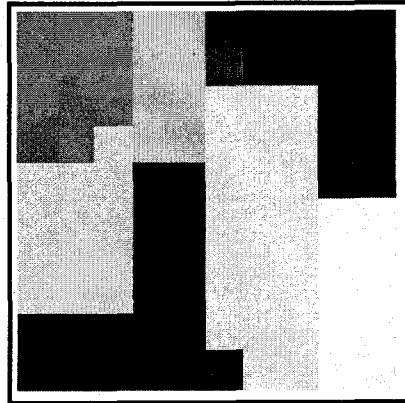


그림 3.5 Location of each classes of the sample data as reference image.(R:residential, E:evergreen, MF:mixed forest, C:cropland, V:vegetation, B:brush, WL: wetland, BL:barren land, VH:vinyl house, W:water)

그림 3.3, 그림 3.4 및 그림 3.5는 sample data중 TM 234 false color화상, IRS-1C PAN 화상 및 참조화상이다.

#### 4. PCA 중합화상의 작성과 토지이용분류

##### 4.1 PCA(Principal Component Analysis) 중합

분광반사특성의 왜곡이 적은 PCA기법을 통하여 토

지이용분류에 사용될 Landsat TM과 IRS-1C PAN 데이터의 중합화상을 작성하였다. PCA는 그림 4.1에서 보는 바와 같이 다수의 변량변동을 소수의 지표를 이용하여 표현하는 방법으로서, 필요한 정보를 거의 상실하지 않으면서도 특징공간의 차원 즉, 자료량이나 패턴의 분류처리시간을 줄이기 위하여 많이 사용된다.

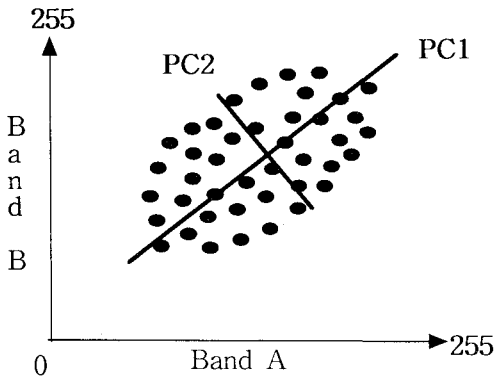


그림 4.1 First and second principal component

다중과장대 화상에서는 각 밴드간 상관관계가 높은 경우가 많기 때문에 PCA를 이용하여 현재의 화상들에 포함되어 있는 대부분의 정보를 가상적인 소수 밴드로 나타낼 수 있다.

PCA방법에 의한 IRS-1C PAN 데이터와 Landsat TM 데이터의 중합방법과 절차는 일반적으로 많이 사용되는 IHS방법과 유사하다. 그러나 IHS방법에서는 단지 3개 과장대의 화상만을 이용하여 I, H 및 S 성분화상으로 변환할 수 있는 반면에, PCA방법은 3개 이상의 화상들을 사용하여 중합할 수 있다는 장점이 있으며 TM6을 제외한 총 42가지( $6C3+6C4+6C5+6C6=20+15+6+1=42$ )의 중합화상을 생성시킬 수 있다.

그러나 이들 각각을 화상중합하여 토지이용분류를 할 경우 컴퓨터 처리시간 및 저장공간의 확보 등과 같은 문제를 수반하고 있다.

Toll(1985)은 TM 데이터를 이용하여 토지이용분류를 수행한 결과, 3개 밴드로 조합할 경우 TM247이 72.9%, 4개 밴드 TM3457이 75.0%, 5개 밴드 TM13457이 78.1%이며, 6개 밴드인 경우 TM123457이

78.7%로서 모든 밴드조합의 경우보다 가장 정확도가 높았다는 것을 제시한 바 있다. 따라서 본 연구에서는 위의 4가지 밴드조합방법을 PCA방법의 유효갈라합성 밴드로 선정하였으며, TM 247, TM3457, TM 13457 및 TM 123457 경우마다 각각 3개, 4개, 5개 및 6개의 주성분화상을 얻었다. 그리고 4가지 조합 각각의 경우에 마지막 주성분을 IRS-1C PAN 데이터로 대체한 다음 PCA 역변환을 수행하여 고해상도의 IRS-1C PAN 데이터 공간해상력이 포함된 3개, 4개, 5개 및 6개의 중합 sample data를 다시 작성하여 토지이용분류에 사용하였다.

#### 4.2 Sample data의 토지이용분류

4가지 조합의 중합화상을 사용하여 토지이용분류를 행하였는데, 토지이용 분류기법으로는 최대우도법을 사용하였다. 이 방법은 다른 방법에 비하여 알고리즘이 복잡하고 처리 시간이 많이 소요되지만 분류정확도의 향상이 두드러지며, 확률 통계적 이론을 기본으로 하기 때문에 토지이용분류에 많이 사용되는 방법이다.

sample data내의 training field는 컬러모니터에 출력된 sample data화상을 보면서 각 항목을 대표할 수 있는 위치를 중심으로 정사각형크기로 택하였다. 각 항목별 training field의 화소수는 통계량 산출이 충분하도록 항목별 화소수의 약 7%내외가 되도록 하였다.

이와 같은 요령으로 결정된 training field를 이용하여 모두 4가지 밴드조합으로 중합된 sample data를 대상으로 각각 토지이용분류를 수행하였다.

#### 5. 분류정확도 분석 및 평가

sample data 중합화상의 분류결과를 고찰하기 위하여 error matrix를 작성하여 그 결과를 비교 평가하였다. 표 5.1과 그림 5.1은 3, 4, 5 및 6개의 밴드를 조합하여 Landsat TM 원화상만을 사용한 경우와 IRS-1C PAN 화상과 TM 원화상을 사용하여 PCA중합된 화상을 사용한 경우에 대한 토지이용분류 정확도를 나타낸다. 그 결과 IRS-1C PAN 데이터와

TM247의 중합화상에서 87.96%, TM3457의 중합화상에서 94.65%, TM13457의 중합화상에서 95.10%, TM123457의 중합화상에서 94.99%이고, 이 중 TM13457의 중합화상의 분류정확도가 가장 높았다.

또한 밴드수를 추가하여 중합한 화상일수록 토지이용분류정확도가 다소 향상됨을 알 수 있었다. 그 이유는 보다 많은 개수의 화상을 사용할수록 보다 다양한 분광반사특성이 확보되고, 많은 정보를 포함하고 있기 때문인 것으로 판단되었다.

Landsat TM 원화상만을 사용한 경우와 IRS-1C PAN

화상과 TM 원화상을 사용한 PCA 중합화상의 경우에 6개 밴드조합의 TM 123457 또는 5개 밴드조합의 TM13457중합화상의 경우에 분류 정확도가 가장크게 (6.2 %) 향상되었다.

4가지 경우의 PCA 중합화상이 TM 원화상과 비교하여 분류정확도는 평균적으로 5.16%정도 향상되었으며, 특히, TM 4개 밴드이상만 IRS-1C PAN 화상과 PCA 중합하여 사용한다면 높은 분류정확도를 얻을 수 있음을 알 수 있었다. 그 이유 중 하나로 PCA기법이 산림, 논, 밭 등과 같은 식생지역에 있어서 다른

표 5.1 Classification accuracy of each band combination

(unit:%)

Band combination	Landsat TM original	PCA merging images (Landsat TM and IRS-1C PAN)	Improved
Six-bands (TM123457)	88.80	94.99	6.19
Five-bands (TM13457)	88.93	95.10	6.17
Four-bands (TM3457)	88.94	94.65	5.71
Three-bands (TM247)	85.41	87.96	2.55
Average	88.02	93.18	5.16

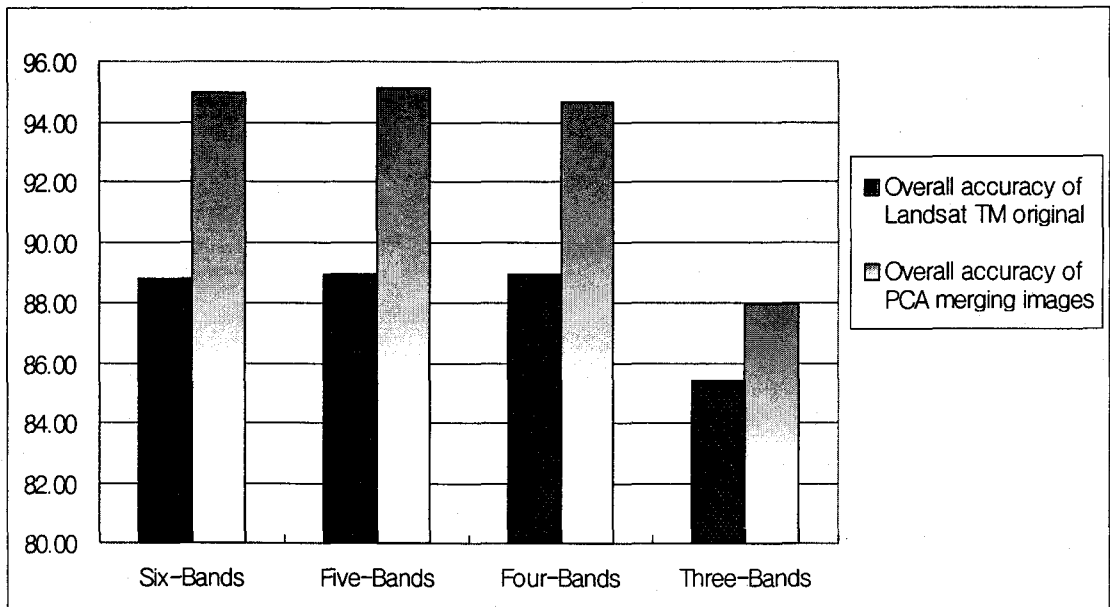


그림 5.1 Classification accuracy of each band combination



중합방법보다 분광반사특성의 보존성이 우수한 중합 방법이기 때문인 것으로 판단된다.

### 참 고 문 헌

### 6. 결 론

본 연구에서는 높은 해상력의 IRS-1C PAN 데이터와 다양한 관측과장대와 분광반사특성을 갖고 있는 Landsat TM 데이터를 사용하여, 대표적 화상중합방법인 PCA방법으로 중합화상을 작성한 후, 이 중합화상의 토지이용 분류정확도를 평가해본 결과, 다음과 같은 결론을 얻을 수 있었다.

1. Landsat TM 데이터와 IRS-1C PAN 데이터를 사용하여 PCA 중합화상을 작성하고 토지이용 분류를 수행할 경우, IRS-1C PAN 데이터와 TM123457 또는 IRS-1C PAN 데이터와 TM13457의 중합화상에 있어서 그 분류정확도가 Landsat TM 원화상보다 가장 크게 향상(약 6.2 %)되었다.
2. Landsat TM 데이터와 IRS-1C PAN 데이터를 사용한 PCA 중합화상이 TM 원화상보다 평균적으로 약 5.16 %의 토지이용 분류정확도 향상을 가져왔다.
3. IRS-1C PAN 데이터를 TM의 4개 밴드 이상의 화상과 PCA 중합할 경우 TM의 3개 밴드 중합화상보다 높은 분류 정확도를 얻을 수 있는 것으로 판단되었다.

1. Carper, W.J., T.M Lillesand, and R.W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data", PE & RS, Vol. 56, No.4, 1990, pp.459-467.
2. Chavez, P.S. Jr., S.C. Sides, and J.A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data : Landsat TM and SPOT Panchromatic", PE & RS, Vol 57, No.3, 1991, pp.295-303.
3. 손덕재, 이종훈, "色變換과 映像改善技法을 이용한 SPOT P-mode와 XS-mode 映像合成", 한국측지학회지, 제 9 권, 제 2호, 1991, pp.103-113.
4. 안기원, 서두천, "TRS-1C PAN 데이터와 Landsat TM 데이터의 중합방법 비교분석", 대한원격탐사학회지, 1998, 제14권 제2호, pp.149-164.
5. 안기원, 이효성, 신석효 "IRS-1C PAN 데이터와 SPOT XS 데이터의 중합방법 비교분석" 대한토목학회논문집, 1999, 제19권, 제3-1호, pp.69-79.
6. 김우철 외, 현대통계학, 영지문화사, 1998, pp.117-137.
7. 김정년, 통계학, 경문사, 1998, pp.131-174.
8. Toll, D.L., "Landsat-4 Thematic Mapper Scene Characteristics of a Suburban and Rural Area", PE & RS, Vol 51, No.9, 1985, pp.1471-1482.