

# 표준화 주성분 분석(Standardized PCA)을 이용한 LANDSAT 위성자료 분류 (Classification)의 정확도 향상 LANDSAT remotely sensed data's Classification accuracy improvement Using Standardized Principal Components Analysis

<sup>1</sup>장 훈(Chang, Hoon) · <sup>2</sup>윤완석(Yoon, Wan Seok)

<sup>1</sup>성균관대 토목환경공학과 전임강사(hchang@skku.edu, 031-290-7524)

<sup>2</sup>성균관대 토목환경공학과 석사과정(bettynet@skku.edu, 031-290-7525)

**요약** - 본 연구에서는 2000년 LANDSAT ETM+ 수도권 영상을 이용하여 도시지역 10개소, 식생지역 10개소를 선정해서 각각에 대해 표준화 주성분 분석을 적용하여 두 지역간의 고유벡터 매트릭스를 비교·분석해보았다. 도시 지역과 식생 지역 각각에 대해 총 6개의 주성분이 생성되었으며 PC-2와 고유벡터 부호가 변한 밴드 (band2, band7)를 RGB로 조합하여 수원 지역을 대상으로 분류(Classification)한 결과의 정확도를 분광서명 분별 분석 (Signature Separability Analysis) 통해 얻은 밴드조합 (band1, band3, band5) 영상의 분류결과와 비교해 보았다. 수원지역 2000년 IKONOS 영상의 다중분광 밴드 (4×4m)와 전정색 밴드 (1×1m)를 융합한 영상이 분류 정확도를 판단하는 기준으로 사용되었다. 비교결과 분류 전체 정확도는 각각 87.7%, 77.29% Khat 지수는 0.83, 0.68로 나타나 PC-2, 밴드2, 밴드7을 이용했을 때 분류 정확도를 높일 수 있다는 결과를 얻었다.

## 1. 서론

주성분 분석 (Principal Components Analysis, PCA)은 다중분광영상자료를 향상시키고 판독하는데 쓰이는 통계학적 방

법으로 인공위성영상을 효과적으로 압축하고 토지 피복 분류 및 식생의 변화탐지 등에 유용하게 사용되는 이미지 처리 (image processing) 기법이다.

LANDSAT-5 TM 영상의 열 밴드를 제외한 6개 밴드에 대해 주성분 분석을 적용하면 데이터의 97%를 처음 3개의 주성분 (PC-1, PC-2, PC-3)이 포함하고 있고, 마지막 3개의 주성분 (PC-4, PC-5, PC-6)이 2.6%를 포함한다. (Sabins, 1987)

PCA를 이용한 인공위성영상은 원래의 영상보다 해석하기가 수월하다. 분광특성이 다른 개개의 물체들이 다른 밴드에서보다 PC이미지에서 더 명확하게 구별된다.

표준화 주성분 분석 (Standardized PCA)은 공분산 매트릭스 (Covariance Matrix) 대신에 상관 매트릭스 (Correlation Matrix)를 이용하여 고유 값 (eigenvalue) 및 고유 벡터 (eigenvector)를 구해서 영상에 적용하는 기법이다. (Conese et al., 1988; Estman and Falk, 1993; Rencher 1995). 표준화 주성분 분석은 영상 해석 능력을 향상시키고, 통계적으로 다루기 용이하며, 더 정확한 분류 정확도를 나타낸다. 특히 다중시대 영상 (multi-temporal image)의 고유벡터 매트릭스의 밴드간 부

호변화를 이용하여 변화탐지(change detection)에 유용하게 사용된다.(Byrne et al., 1980; Fung and LeDrew, 1988; Muchoney and Haack, 1994; Coppin and Bauer, 1996; Macleod and Congalton, 1998). 두 시대 영상의 고유벡터 매트릭스간에 각 주성분에 대해 비교했을 때, 밴드의 부호가 변한 주성분이 나타내는 정보는 두 영상의 변화이다. 본 연구에서는 표준화 주성분 분석을 다중 시대 영상에 동시에 적용하여 시간의 변화에 따른 변화탐지(Change detection)를 목적으로 하는 기존의 연구방법과 달리 2000년 영상인 LANDSAT ETM+에 도시지역을 나타내는 10개 곳, 식생지역을 나타내는 10개 곳을 샘플링하여 적용한 후 두 지역의 고유벡터 매트릭스를 비교하여 변화를 나타내는 주성분을 이용하여 분류(Classification)의 정확도 향상을 꾀하였다.

## 2. 연구 방법

본 연구에서는 2000년 LANDSAT ETM+ 수도권 영상을 이용하여 도시지역 10개소, 식생지역 10개소를 선정해서 각각에 대해 표준화 주성분 분석을 적용하여 두 지역간의 고유벡터 매트릭스를 비교·분석해보았다. 도시 지역과 식생 지역 각각에 대해 총 6개의 주성분이 생성되었으며 두 지역간에 밴드의 부호 변화를 나타내는 주성분과 부호가 변한 밴드를 RGB로 조합하여 수원지역을 대상으로 분류해보았다. 분류 결과의 정확도를 분광서명 분별 분석(Signature Separability Analysis) 통해 얻은 밴드조합 영상과 비교해 보았다. 분류 정확도 평가를 위해 수원지역 2000년 IKONOS 영상(해상도 1×1m)을 이용하였다.

### 1) 도시지역과 식생지역에 표준화 주성분 분석 적용

LANDSAT ETM+영상을 이용하여 도시지역과 식생지역을 각각 10개소를 샘플링하였다. 샘플링한 지역은 그림 1.과 같다.

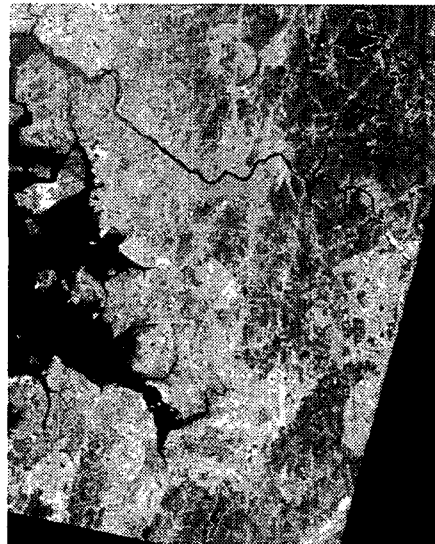


그림1. 도시지역 및 식생지역 선정

각 지역에 대해서 표준화 주성분 분석을 적용해 고유벡터 매트릭스(eigenvector matrix)를 산출하였다. 표준화 주성분 분석은 열 밴드와 전정색 밴드를 제외한 6개 밴드를 이용하였으며, 도시와 식생 지역 각각에 대해 총 6개의 주성분(PC-1, PC-2, PC-3, PC-4, PC-5, PC-6)이 생성되었다. 고유벡터 표는 표1과 같다.

표1.은 다른 도시 및 식생 지역의 고유벡터에 대한 대표적인 값을 나타낸 것으로 각 샘플링 지역은 표1.과 거의 비슷한 값을 갖는다. 도시 지역과 식생지역의 PC-1을 비교해보면 부호의 변화가 없으며 특이할 만한 정도 없다. 그러나 PC-2를 비교해보면 밴드2와 밴드7의 고유벡터 부호가 변한 것을 확인할 수 있다. PC-2뿐만 아니라 PC-3, 4, 5, 6에서도 같은 밴드간

주성분 band	주성분											
	도시 지역						식생 지역					
	1	2	3	4	5	6	1	2	3	4	5	6
band1	0.3828	-0.6256	0.2356	0.5117	-0.2985	-0.2357	0.4106	-0.1845	-0.6973	0.5497	0.0937	0.0007
band2	0.4246	-0.3799	-0.1770	-0.1373	0.7346	0.2925	0.4356	0.2374	-0.2778	-0.6767	0.4606	-0.0816
band3	0.4205	-0.1532	-0.1832	-0.7362	-0.4610	-0.1080	0.4367	-0.3109	-0.0276	-0.3340	-0.7720	-0.0661
band4	0.3948	0.3536	-0.6676	0.4208	-0.2162	0.2226	0.2750	0.8398	0.0634	0.2073	-0.2992	0.2874
band5	0.4197	0.4472	0.1579	0.0159	0.3118	-0.7081	0.4421	0.0351	0.4930	0.2813	0.1545	-0.6762
band7	0.4055	0.3402	0.6394	-0.0021	-0.1215	0.5443	0.4239	-0.3263	0.4345	0.0789	0.2641	0.6701

표1. 도시 지역과 식생 지역의 고유벡터 표

의 고유벡터의 부호 변화를 확인할 수 있다. 각 주성분의 픽셀 값은 다음과 같은 식에 의해 결정된다.

$$newBV_{i,j,p} = \sum_{k=1}^n a_{kp} BV_{i,j,k} \quad --(1)$$

BV<sub>i,j,p</sub>=주성분 p의 i번째 행, j번째열의 픽셀값,  
a<sub>kp</sub>=주성분 p의 k밴드 고유벡터  
BV<sub>i,j,k</sub>=k밴드의 i번째 행, j번째열의 픽셀값  
n=총 밴드 수

즉, 각 주성분의 픽셀 값은 각 밴드의 픽셀 값에 그 밴드에 해당하는 고유벡터 값을 곱하여 전부 더한 것이다. 그러므로 임의 주성분에서 도시 지역과 식생 지역의 픽셀 값의 차이는 두 지점의 각 밴드별 픽셀 값의 차이에 그 밴드의 고유벡터를 곱한 값을 더한 것이다. 식으로 표현하면 다음과 같다.

$$newBV_{i,j,p} - newBV_{i,m,p} = \sum_{k=1}^n a_{kp} (BV_{i,j,k} - BV_{i,m,k}) \quad ---- (2)$$

i,j: 도시지역을 나타내는 픽셀,  
i,m: 식생지역을 나타내는 픽셀

표1. PC-2에서 고유벡터 부호의 변화가 있는 밴드2와 밴드7은 두 지역의 변화를 잘 나타낸다고 볼 수 있다. 그러므로 PC-2와 밴드2, 밴드7을 RGB로 조합하여 분류를 하게되면 분류정확도를 높일 수 있다고 예상할 수 있다. 고유벡터 부호의 변화가 있는 다른 주성분을 이용할 수도 있겠으나 표1.에서 보는 바와 같이 식생 지역을 잘 나타내는 근적외선 밴드(밴드 4)와 그 외 밴드(밴드 1,2,3)의 차이가 가장 극명하게 차이를 보여 도시와 식생을 구분하는데 결정적인 영향을 미치는 성분이 PC-2이며, 연구대상지역인 수원에서 PC-2가 포함하는 데이터가 전체의 약 14%로 가장 많이 포함하고 있으므로 PC-2와 밴드2, 밴드7을 이용하여 분류 정확도를 높일 수 있는지를 연구해 보았다.

주성분 비율	1	2	3	4	5	6
고유값 (eigenvalue)	4.8951	0.8234	0.2268	0.0395	0.0096	0.0057
Percentage	81.56%	13.72%	3.78%	0.66%	0.16%	0.1%

표2. 수원지역 표준화 주성분 분석 결과 고유값

2) 분류(Classification)

수원지역을 대상으로, 조합된 영상(PC-2, BAND2, BAND7)과 분광서명 분석(Signature Separability Analysis) 통해 얻은 밴드조합(BAND 1, BAND 3, BAND 5) 영상을 분류하였다. 분류 방법은 RGB로 조합한 2개의 영상 가운데 대상 지역의 클래스간 분광특성을 잘 구별할 수 있는 영상을 판별하기 위해 무감독 분류를 실시하였다. 초기 클래스 설정은 각 영상에 대해서 20개로 하였으며 분류 후 5개의 클래스(하천 및 호수, 산림, 경작

지, 도시, 나지)로 다시 재분류하였다. 분류한 결과는 그림2와 같다.

왼쪽 영상(PC-2, band 2, band 7)이 오른쪽(band 1, band 3, band 5)보다 경작지 비율이 더 많은 대신 도시와 나지의 면적이 작게 나타난 것을 육안으로도 확인할 수 있다.

3)분류 정확도 평가

수원지역 2000년 IKONOS 영상의 다중분광 밴드(4×4m)와 전정색 밴드(1×1m)를 융합한 영상이 분류 정확도를 판단하

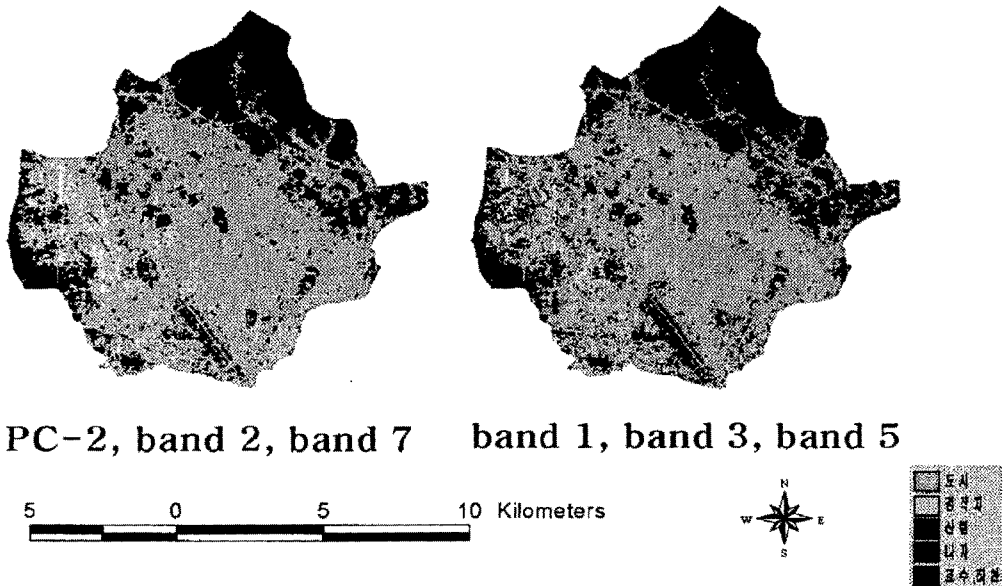


그림2. 분류 결과

는 기준으로 사용되었다. 정확도 분석을 위해 샘플 포인트(sample point)갯수는 다음 공식을 이용하였다.(van Genderen and Lock, 1977; Fitzpatrick-Lins, 1981, Marsh *et al.*, 1994).

$$N = \frac{Z^2 pq}{E^2} \quad \text{---(3)}$$

N=샘플의 갯수, p=기대 정확도,  
q=100-p, E=허용 오차(%)  
Z=표준정규분포 95% 값(1.96)

기대 정확도를 60%, 허용오차를 5%로 설정했을 때 최소 샘플 포인트의 개수 370개를 클래스 면적비율로 분배하였으며 신뢰있는 정확도 평가를 위해 클래스 최소 샘플포인트는 20개, 도시 지역을 제외한 나머지 클래스에는 원래의 샘플포인트보다 50%많은 수를 사용하였다. 그래서 총 480개의 샘플포인트를 이용하여 분류 정확도를 평가하였으며 그 결과는 다음 표3.

과 같다.

### 3. 결론

본 연구에서는 2000년 LANDSAT ETM+ 수도권 지역의 영상을 이용하여 도시지역 10개소, 식생 지역 10개소를 선정하여 각각에 대해 표준화 주성분 분석을 적용한 결과를 이용하여 산출한 고유벡터 표로부터 도시지역과 식생지역의 고유벡터를 분석하였으며 그 결과 PC-2에서 도시와 식생지역 간에 BAND2와 BAND7의 고유벡터 부호가 변화됨을 알 수 있었다. PC-2, BAND2, BAND7 각각에 RGB를 할당하여 조합한 영상으로 분류한 결과는 분광서명 분별 분석(Signature Separability Analysis)을 통해 얻은 밴드 조합을 이용한 것보다 높은 분류 정확도를 얻을 수 있었다.

클래스	PC-2, band 2, band 7		band 1, band 3, band 5	
	생산자 정확도(%)	사용자 정확도(%)	생산자 정확도(%)	사용자 정확도(%)
1.하천 및 호수	100	100	100	83.33
2.산림	93.02	97.56	79.71	91.67
3.경작지	67.23	86.96	44.19	57.78
4.도시	100	83.44	99.47	83.48
5.나지	66.67	70.97	53.13	56.67
전체 정확도	87.70%	-	77.29%	-
KHAT	0.83	-	0.68	-

표3. 분류 정확도 비교

참고문헌

1. Byrne, G.F., Crapper, P.F., and Mayo, K.K.(1980),Monitoring land-cover change by principal components analysis of multitemporal Landsat data. Remote Sens. Environ. 10:175-184
2. Cansas, A., Barnett.M, 1985. The generation and interpretation of false-colour composite principal component image. International Journal of Remote Sensing 6, 867-881
3. Congalton, R.G., R. Oderwald, and R. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogrammetric Engineering and Remote Sensing 49(12): 1671-1678.
4. Congalton, R.G. 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing 54(5): 593-600.
5. Congalton R.G. and K. Green. 1998. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, New York.
6. Eklundh, L. and Singh, A.(1993), A comparative analysis of standardized and unstandardized principal components analysis in remote sensing. Int J. Remote Sens. 14(7): 1359-1370.
7. Estman, R.J.(1992), Time series map analysis using standardized principal components, in ASPRS/ACSM/RT 92 Technical Papers, Volume 1: Global Change and Education, 3-8 August, Washington, DC, pp. 195- 204
8. Estman, R.J., and Fulk, M(1993), Long sequence time series evaluation using standardized principal components. Photogramm. Eng. Remote Sens. 59 (8) : 1307-1312
9. Fung, T., and LeDrew, E.(1987). Application of principal components analysis to change detection. Photogramm. Eng. Remote Sens. 53(12):1649-1658
10. Hay, A.M. 1979. Sampling designs to test land use map accuracy. Photogrammetric Engineering and Remote Sensing 45(40): 529-533.
11. Jensen, J. R. 1996. Introductory Digital Image Processing. A Remote Sensing Perspective. Upper Saddle River, New Jersey: Prentice Hall.
12. Sabins, F., 1987. Remote Sensing: Principles and Interpretation
13. Singh, A. (1989), Digital Change detection techniques using remotely sensed data: review article. Int J. Remote Sens. 10:989-1003