극단화소 기반의 Hyperion 데이터 밴드선택
Extrema-based Band Selection for Hyperion Data

한동엽[1] · 김대성[1] · 김용일[2]
Han, Dong Yeop · Kim, Dae Sung · Kim, Yong Il

[1] 서울대학교 대학원 지구환경시스템공학부 박사과정(E-mail: hkyon2@snu.ac.kr, mutul94@empal.com)
[2] 서울대학교 공과대학 지구환경시스템공학부 부교수(E-mail: yik@plaza.snu.ac.kr)

Abstract

Among 242 Hyperion bands, there are 46 bands that contain completely no information and some other bands with various kinds of noise. It is mainly due to the atmospheric absorption and the low signal-to-noise ratio. The visual inspection for selecting clean and stable bands is a simple practice, but is a manual, inefficient, and subjective process. Though uncalibrated, overlapping, and all deep water absorption bands are removed, there still exist noisy bands. In this paper, we propose that the extrema ratio be measured for noise estimation and the unsupervised band selection be performed using the Expectation-Maximization algorithm. The Hyperion data were classified into 5 categories according to the image quality by visual inspection, and used as the reference data. The accuracy of the proposed method was compared with signal-to-noise ranking and entropy ranking. As a result, the proposed method was effective as preprocessing step for band selection.

1. 서 론

하이퍼스펙트럴 원격탐사(hyperspectral remote sensing)는 지상의 정보를 10~20nm의 파장폭의 연속적인 수백 개의 밴드 데이터로 취득하고 분석하는 분야이다(Freek, 2001). 취득된 하이퍼스펙트럴 데이터는 기존의 다중분광 데이터(multispectral data)보다 높은 분광해상도(spectral resolution)와 방사해상도(radiometric resolution)로 인해 지상의 많은 정보를 포함하고 있어 최근 목표물 탐지 (target detection), 영상 분류 (classification), 화소이하 분석 (spectral mixture analysis)과 같은 분야에서 연구가 많이 이루어지고 있다. 하지만, 높은 분광 및 방사해상도가 데이터 분석에 항상 효과적으로 작용하지는 않는다. 계산량이 기하급수적으로 늘어나며, 다중분광 데이터에 가장 효과적인 분류 알고리즘으로 알려진 MLC(Maximum Likelihood Classifier)와 같은 다변량 분석이 적절하게 적용되지 않을 수도 있다. 또한 인접 밴드간 높은 상관관계로 인해 잉여 정보가 발생하고 노이즈 밴드가 존재하여 불필요한 정보를 제공한다(Groves 등, 2004). 따라서 취득되는 밴드를 모두 사용하는 것은 분석의 신뢰도를 떨어트리고 계산량을 증가시키므로 실제 데이터 분석에 필요한 밴드들만을 선별하는 전처리 과정이 무엇보다 중요하다. 이런 전처리 과정은 애니메이션 기능을 통해 시각적으로 판단하거나 상관계수, 엔트로피와 같은 간단한 통계량을 이용하는 것이 일반적이다(ENVI tutorial, 1999). 통계량을 이용하는 방법은 데이터 특성에 따라 효과적인 밴드 선별을 수행하지 못하며, 수백 개의 밴드를 수작업을 통해 추출하는 것은 시간적으로 많은 노력을 요하는 단점이 발생한다. 이러한 문제를 해결하기 위해 극단화소값의 비율을 이용하여 노이즈 밴드를 제거하는 기법을 제안하고자 한다. 또한 기대최대화(Expectation-Maximization) 알고리즘을 적용하여 데이터의 자동화된 유효밴드를 선택하였다. 이와 같은 밴드 선택은 밴드간 상관관계를 고려하지 않으며 일련의 데이터 처리 과정에서 가장 기본적인 양질의 데이터를 확보하는 차원에서 이루어지는 밴드 순위(band ranking) 기법이다(Groves 등, 2004).

2. 밴드 선택 알고리즘

2.1 밴드 순위 기법 (Band Ranking Method)

영상에 존재하는 노이즈를 제거하는 방법은 크게 세 가지로 나눌 수 있다(Gretchen, 2005). 첫째, 영상 데이터의 변환에 의해 계산된 가중값이 가장 큰 순서대로 일정한 수의 밴드를 선택하는 방법으로 PCA(Principal Component Analysis), MNF(Minimum Noise Fraction), ICA(Independent Component Analysis) 등이 있다. 둘째, 필터링 기법을 적용할 수 있는데, 템플릿을 사용한 중앙값 필터(median filter)나 평균값 필터(mean filter)를 사용한다. 셋째, 노이즈가 존재하는 밴드를 제거한다. 이는 직관적이고 쉬운 방법이나 노이즈의 존재 여부를 판단하는 기준에 있어 객관성을 확보할 수 없다는 단점이 있다. 이를 보완하기 위해 데이터의 통계량 분석을 통해 정보를 많이 포함하고 있는 순으로 데이터를 재배열하고 일정한 한계값(threshold)을 적용하여 노이즈 밴드를 제거하는 밴드 순위 기법이 주로 사용되고 있다.

SNR(signal-to-noise ratio)는 영상을 취득할 때 발생하는 기계적인 노이즈의 양을 측정하는 것이 원칙이다. Hyperion의 SNR은 파장에 따라 190~40dB 정도로 알려져 있지만(http://eo1.gsfc.nasa.gov), 취득된 영상에 대해서 SNR을 추정하기는 쉽지 않다. 취득한 영상에 대한 SNR 계산식은 여러 가지가 제시되어 있으나 영상의 평균값과 분산을 이용한 방법이 일반적이다(Emmanuel 등, 2005).

$$MSE = \frac{1}{n_x n_y n_\lambda} \sum_{x,y,\lambda} \left(I(x,y,\lambda) - \tilde{I}(x,y,\lambda)\right)^2$$

$$SNR_{dB} = 10 \cdot \log_{10} \frac{\sigma_I^2}{MSE} \qquad \text{(식 1)}$$

여기서, I는 원영상, $\tilde{I}$는 변화된 영상, x,y,λ는 영상의 행, 열, 분광밴드를 의미한다.

엔트로피(information entropy)는 어떤 현상에 대한 무질서의 정도를 값으로 표현한 것으로 영상에 적용하게 되면 식 2와 같이 표현될 수 있다(Groves 등, 2004). 엔트로피가 높은 밴드는 많은 정보를 포함하는 밴드를 의미한다.

$$H(\lambda_i) = -\sum_{k=1}^{m} p(\phi_k^i) \ln p(\phi_k^i) \qquad \text{(식 2)}$$

여기서, λ_i는 밴드, $p(\phi_k^i)$는 화소값이 가지는 확률, m은 화소 개수 또는 Bin 수이다.

이외에 밴드간 유사도를 측정하는 상관도(correlation measure)와 주성분 분석 순위(principal component analysis ranking) 기법 등이 사용되고 있다(Groves 등, 2004).

2.2 극단화소비

Hyperion 영상에 나타나는 노이즈는 스트립(strip) 노이즈, 가우시안 노이즈, 'salt & pepper' 노이즈, 스페클(speckle) 노이즈가 있다. Hyperion 영상에서 스트립 노이즈는 destriping 알고리즘으로 우선 제거되어야 하고, 그 외 노이즈는 필터링이나 PCA 또는 MNF 변환 기법에 의해 제거되고 있다. 가우시안 노이즈에 의해 영향을 받은 화소는 노이즈 크기에 따라 극단화소로 나타날 수 있고, 'salt & pepper' 노이즈는 극단화소로 나타나며, 스페클 노이즈도 노이즈 크기에 따라 극단화소로 나타날 수 있다. 따라서 극단화소는 노이즈를 추정할 수 있는 수단으로 사용가능하다. 극단점은 모폴로지 연산을 이용하여 영상에서 빠르게 찾을 수 있다.

2.3 기대 최대화 알고리즘 (Expectation-Maximization Algorithm)

기대최대화 기법은 불완전한 자료로부터 반복적인 연산을 통해 확률 밀도 함수의 인자(평균, 분산, 사전확률)를 결정하는 방법이다 (Dempster 등, 1977). 이 알고리즘은 인자의 초기화, 인자들의 기대값을 계산하는 기대화 과정(E-step)과 최대우도를 가지는 새로운 추정값을 계산하는 최대화 과정(M-step)으로 나누어진다. 전 단계의 추정값과 새로운 추정값의 차이가 거의 없을 때 반복 계산은 멈추게 되고, 각 밀도 함수가 가지는 확률이 동일한 지점이 한계값이 된다. 추출된 한계값은 유효밴드와 노이즈밴드를 자동으로 구분할 수 있다. 한계값은 식 3의 2차 방정식을 통해 계산된다.

$$(\sigma_n^2-\sigma_c^2)\widehat{T_o^2}+2(\mu_n\sigma_c^2-\mu_c\sigma_n^2)\widehat{T_o}+\mu_c^2\sigma_n^2-\mu_n^2\sigma_c^2+2\sigma_n^2\sigma_c^2\ln[\frac{\sigma_c P(w_n)}{\sigma_n P(w_c)}]=0 \qquad \text{(식 3)}$$

여기서, $\widehat{T_0}$는 한계값, σ_n,σ_c는 표준편차, μ_n,μ_c는 평균, $P(w_n),P(w_c)$는 사전확률을 나타낸다.

2.4 연구 방법

알고리즘을 실제 적용하기 위해 사용된 데이터는 Hyperion 영상이다. 서울 남부와 안산 지역을 촬영한 2002년 4월 3일과 4월 9일의 영상에서 영상의 밝기값이 0인 밴드를 우선 제거하여 196개의 밴드를 추출하였고, 영상 대 영상 기하보정을 실시하여 350×200 크기의 3개 데이터를 구성하였다. 이를 이용하여 두 가지 경우, 같은 지역의 다른 시간에 촬영한 데이터와 다른 지역의 같은 시간에 촬영한 데이터를 비교하여, 시간이나 지역에 대해 엔트로피와 SNR 순위, 극단화소비를 산출하였다. 산출값에 기대최대화 알고리즘을 적용하여 자동화된 밴드 선택을 수행함으로써 무감독 유효 밴드를 선택하였다. 결과의 평가를 위한 참조 자료는 시각적으로 판단한 노이즈의 정도에 따라 등급을 구분하였다.

3. 알고리즘 적용 및 결과

3.1 사용데이터

Hyperion은 pushbroom 방식으로 촬영하며 30m 공간해상도, 242개의 밴드로 구성되어 있다. VNIR 파장과 SWIR 파장에 민감한 두 개의 센서로 구성되어 있어 약 850nm부터 1050nm 사이의 파장에서 겹침 밴드가 존재한다. 242개의 밴드 중 보정되지 않은 밴드(1-7, 58-76, 225-242번 밴드, 44개 밴드)와 보정은 되었으나 파장이 중복되는 밴드(56-57, 77-78번 밴드 중 하나씩)를 1차적으로 제거하여 196개의 밴드를 생성하였다. 그림 1에 나타난 실험 데이터 중에서, Data I(안산시)과 II(서울 남부)는 2002년 4월 3일 영상이며, Data III(서울 남부)은 4월 19일 영상이다.

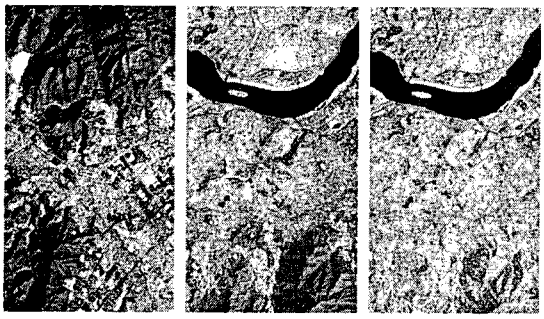

그림 1. 사용 데이터(좌로부터 Data I, Data II, Data III)

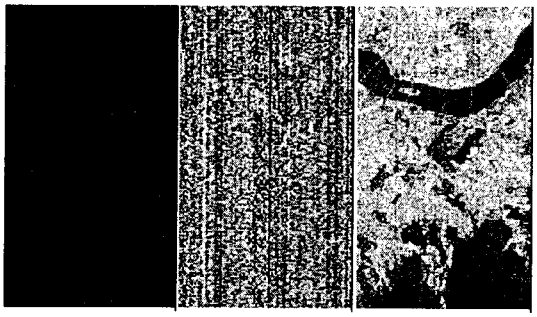

그림 2. 노이즈의 형태

Hyperion 데이터의 노이즈는 화소값이 0인 밴드(Zero Band), 랜덤 노이즈 형태(random noise), 수직 노이즈(streaking noise) 등으로 나타난다. 랜덤 노이즈는 흡수 파장(absorption band) 및 그 경계에서

나타나는 현상으로 지상의 형체를 알아볼 수 있는 것과 그렇지 못한 밴드가 존재한다. 수직 노이즈는 특정 파장을 담당하는 감지기가 역할을 하지 못할 때 발생하는데 pushbroom 센서에서 주로 발생한다. 그림 2는 Hyperion 데이터에 존재하는 노이즈의 형태를 보여준다. 수직노이즈를 제거하기 위하여 지역적 destriping 기법이 적용되었다(Datt, 2003). 일반적으로 175개의 밴드(8-56, 78-120, 129-166, 179-223번 밴드)를 최대한 사용할 수 있는 데이터(maximally useable data)로 규정하고, 155개의 밴드(8-57, 81-97, 101-119, 134-164, 182-221번 밴드)를 안정적인 데이터(stable data)로 취급하기도 한다(US Report, 2002). 그러나 이 범위에 속하는 밴드를 시각적으로 판단해보면, 일부 밴드에 대해 노이즈가 존재한다. 동일 시간에 촬영한 영상의 경우 지역이 다르더라도 노이즈의 형태가 유사하지만, 촬영시간이 다른 데이터의 경우에는 동일지역이라 하더라도 노이즈 형태가 일부 밴드에 대해 다른 양상으로 나타난다.

3.2 알고리즘 적용 및 결과

세 개의 데이터에 대해 엔트로피와 SNR 순위 기법, 극단화소비를 각각 적용하여 9가지 결과를 도출하였다(그림 3 참고). SNR 순위 기법, 극단화소비는 결과값이 높을수록, 엔트로피 값은 낮을수록 상대적으로 노이즈가 많은 밴드임을 의미한다.

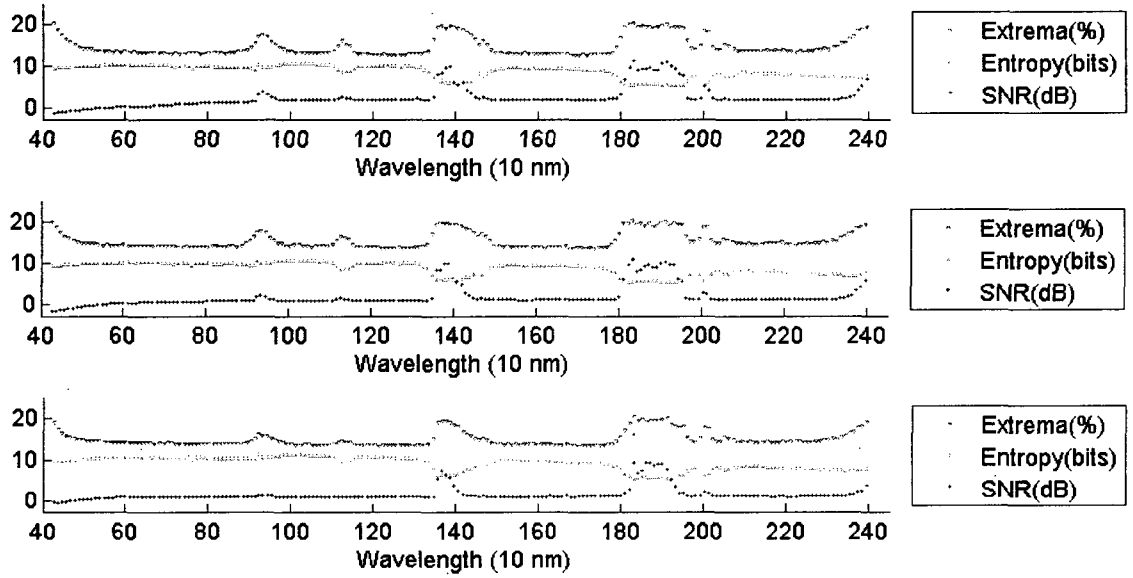

그림 3. 밴드별 SNR, 엔트로피, 극단화소비; (상) Data I, (중) Data II, (하) Data III

산출된 통계량에 기대최대화 알고리즘을 적용하여 노이즈 밴드를 자동으로 구분하는 한계값을 계산하였다. 그림 4는 극단화소비에 기대최대화 알고리즘을 적용한 결과를 보여주고 있다. 그림에서 보이는 히스토그램은 산출값을 40개의 구간(bin)으로 생성한 것이며, 노이즈 밴드의 확률 밀도 곡선은 빨간색(엔트로피는 하늘색)으로 표시되어 있다. 알고리즘을 통해 추출된 한계값과 노이즈 밴드가 제거된 유효 밴드는 표 1과 2에 정리하였다.

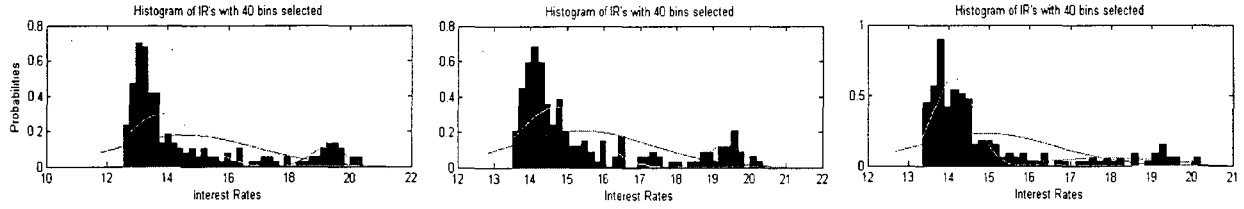

그림 4. 극단화소비의 기대최대화 기법 적용 예; 좌로부터 Data I, II, III

표 1 EM 알고리즘으로 추출된 한계값 (색 부분은 선택된 한계값을 의미함)

	SNR			Entropy			극단화소비		
	Data I	Data II	Data III	Data I	Data II	Data III	Data I	Data I	Data III
최대	4.633	3.222	1.429	12.061	12.670	13.286	24.742	24.210	15.543
최소	-6.587	-2.677	0.798	8.903	8.781	9.220	17.555	17.845	11.536

표 2 기대최대화 기법을 통해 선택된 유효 밴드

		VNIR (1-48번 밴드)	SWIR (48-196번 밴드)	총 개수
Data I	SNR	1-48	49-92, 100-137, 153-156, 158-194	171
	Entropy	1-48	49-69, 73-90, 107-132	113
	극단화소비	3-48	49-92, 101-137, 153-156, 159-193	166
	양호밴드	5-48	57-69, 73-91, 106-136, 159-161, 163-189	137
Data II	SNR	1-48	49-92, 100-137, 153-193	171
	Entropy	1-48	49-69, 71-91, 107-133	117
	극단화소비	3-48	49-92, 102-137, 153-156, 159-193	165
	양호밴드	5-48	57-69, 73-91, 106-136, 159-161, 163-189	137
Data III	SNR	1-48	49-92, 100-137, 153-156, 159-191	167
	Entropy	1-48	49-91, 106-132	118
	극단화소비	5-48	49, 53-92, 102-137, 153-156, 159-189	156
	양호밴드	5-48	57-92, 102-136, 155, 156, 159-161, 163-189	147
공통		5-48	57-69, 73-91, 107-136, 155-156, 159-161, 163-189	133

결과의 정확도 평가를 위한 참고자료로 시각적으로 양호한 밴드를 사용하였다. 5단계의 영상 패턴으로 구분하여 판별하였다; 매우 양호, 양호, 모호함, 노이즈 적음, 노이즈 많음. 이 중 모호한 밴드를 포함한 매우 양호와 양호한 밴드만을 추출하여 실험 결과와 비교하였다. 정량적인 정확도 평가는 오차 행렬(error matrix)을 이용하였다. 시각적 평가를 통해 추출한 결과를 열에 정리하고 실험 결과를 행에 정리하여 전체 정확도를 평가하였다. 극단화소비의 전체 정확도가 SNR이나 엔트로피에 비해 높게 나타났으며, 시각적인 판단에 의해 추출한 밴드와 유사한 결과를 보였다(표 3 참고). 따라서 기존의 기본적인 통계량을 통해 판단하던 노이즈의 양을 본 연구에서 제안하고 있는 극단화소비로 대체할 수 있을 것으로 판단된다. 극단화소비를 최대기대화 알고리즘과 연동함으로써 자동화된 밴드 선택이 이루어질 수 있음을 확인할 수 있었다.

표 3 정확도 평가 (색 부분은 각 데이터 적용 결과 중 가장 높은 정확도를 나타냄)

		SNR			Entropy			극단화소비		
		Data I	Data II	Data III	Data I	Data II	Data III	Data I	Data I	Data III
전체정확도		82.563	82.563	89.796	75.510	75.510	72.959	85.204	85.714	95.408
사용자정	양호밴드	100	100	100	73.723	75.182	72.109	100	100	100
확도	노이즈밴드	42.373	42.373	59.184	79.661	76.271	75.510	50.847	52.542	81.633
생산자정	양호밴드	80.117	80.117	88.024	89.381	88.034	89.831	82.530	83.030	94.231
확도	노이즈밴드	100	100	100	56.627	56.962	47.436	100	100	100

4. 결론

Hyperion 데이터의 유효 밴드 선택을 위하여 시각적인 판단에 의존하거나, 흡수밴드를 제외한 밴드를 일괄적으로 사용하는 것이 일반적이었다. 본 연구에서는 시각적 판단을 배제하고, 최대한 노이즈가 적은

밴드를 합리적으로 선택할 수 있는 자동화된 기법을 찾고자 하였다. 이를 위하여 극단화소비를 제안하였으며, 제안된 통계량으로부터 기대최대화 기법을 적용하여 자동으로 한계값을 결정하였다. 그 결과 극단화소비의 전체 정확도가 기존의 SNR 순위, 엔트로피보다 높게 나타났다. 따라서, 1차적인 밴드의 선별 방법으로 연구 결과가 사용될 수 있을 것으로 판단된다. 향후 각 밴드별 유효 정보량을 산출하기 위하여 노이즈 화소와 유효 화소의 구분에 대한 연구가 이루어져야 하고, Hyperion 데이터 분석에 사용할 수 있는 정형화된 밴드목록을 설정할 필요가 있다.

참고문헌

Boyell, R.L., Ruston, H. (1963), Hybrid Technique for Real-time Radar Simulation, IEEE Proceedings Fall Joint Computer Conference 63, Las Vegas, USA, pp. 445-458

Carr, H., Snoeyink, J., Axen, U. (2003), Computing Contour Trees in All Dimensions, *Computational Geometry*, Vol. 24, No. 2, pp. 75-94.

Datt, B. McVicar, T.R. Van Niel, T.G. Jupp, D.L.B. Pearlman, J.S. (2003), Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 6, pp. 1246-1259.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society Series(B)*, Vol. 39, No. 1, pp. 1-38.

Emmanuel Christophe, Dominique Leger and Corinne Mailhes (2005), Comparison and Evaluation of Quality Criteria for Hyperspectral Imagery, *Proceedings of SPIE*, Vol. 5668, pp. 204-213.

ENVI (1999), Tutorials, Research Systems, Boulder, Colorado.

EO-1 User Guide v. 2.3, http://eo1.gsfc.nasa.gov

Freek D. Van Der Meer and Steven M. Dejong (2001), *Imaging Spectrometry - Basic Principles and Prospective Applications*, Kluwer Academic Publishers, Dordrecht, pp. XXI-XXIII

Gretchen, M.S. (2005), *Application of Phenology to Assist in Hyperspectral Species Classification of a Northern Hardwood Forest*, Master thesis, Rochester Institute of Technology.

Groves P. and P. Bajcsy (2004), Methodology for Hyperspectral Band and Classification Model Selection, *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Washington DC, October 27, pp. 120-128.

Landgrebe D.A. (2003), *Signal Theory Methods in Multispectral Remote Sensing*, Wiley-Interscience, United States of America, pp. 237-271.

US report (2002), http://www.eoc.csiro.au/hswww.oz_oz_pi/dlbj_us_0402.htm