

# Hyperion 영상의 분류를 위한 밴드 추출 Feature Selection for Image Classification of Hyperion Data

한 동엽\*, 김혜진\*, 김대성\*, 조영욱\*, 김용일\*\*

Dong-Yeob Han, Hye-Jin Kim, Dae-Sung Kim, Young-Wook Cho, Yong-Il Kim

\* 서울대학교 지구환경시스템공학부, 02-880-7371, hkyon2@snu.ac.kr

\*\* 서울대학교 지구환경시스템공학부 교수, 02-880-7364, yik@snu.ac.kr

## 요약

다중분광 영상의 정확한 지형지물 분류를 수행하기 위하여 분류 클래스의 훈련지역 선정과 선정된 클래스의 분리도 분포가 중요하다. 최근에 이용되고 있는 위성탑재 초다중분광 영상은 많은 밴드를 포함하고 있기 때문에 데이터 처리가 어렵고, 노이즈로 인하여 다중분광 영상보다 분류 결과가 나쁜 경우도 나타난다. 특히 대상지역의 클래스에 따른 훈련지역의 선정시 밴드수에 비해 상대적으로 제한된 훈련화소 크기로 인하여 공분산 행렬의 계산에 어려움이 따른다. 따라서 본 연구에서는 Hyperion 데이터를 이용한 분류를 수행하기 위하여 필요한 유효 밴드 추출 방식을 알아보고, 분류영상의 정확도 평가를 통하여 추출된 밴드와 분류 클래스의 적합성 관계를 확인하고자 한다. 이 과정에서 클래스 분리도를 이용하여 정확도 평가 이전에 밴드와 클래스 선정의 타당성을 확인할 수 있다.

## 1. 서론

1972년 최초의 지구관측 위성인 Landsat-1호가 발사된 이후, SEASAT, NOAA, SPOT, IRS, Radarsat과 같은 광학 또는 SAR 센서를 탑재한 위성이 발사되어 많은 지구관측자료를 획득하고 있으며, 최근에는 IKONOS, QUICKBIRD와 같은 고해상도 위성과 EO-1 프로그램과 같은 자원 및 과학자료 관측위성이 이용되고 있다. 지표면에 대한 다양한 정보가 요구되면서 원격탐사는 지도학, 글로벌 변화 탐지, 지질학, 식생, 광물 탐지, 농업 생산량 추정 등에서 폭넓게 사용되고 있다. 고해상도 영상을 이용하여 지표면 형상의 인식 및 측정 기술이 발전되고 있는 반면

에, 고분광 영상을 이용한 지표면 특성의 정밀한 모델링에 대한 관심도 높아지고 있다.

고분광 영상은 물체의 고유한 반사특성을 구분할 수 있는 10nm의 밴드폭으로 대상을 촬영하기 때문에 광물 탐지, 수종 및 농작물 구분, 식생 관리 등에 높은 응용가능성을 보여주고 있다. 초기 항공기용 고분광 영상을 토대로 이루어진 연구는 위성 고분광 영상을 이용한 분석으로 전환되고 있다. 하지만 고분광 센서의 상이한 사양으로 인하여 항공기탑재 고분광 영상의 실험점 연구성과를 모든 고분광 영상에 적용시킬 수 없으며, 아직도 고분광 영상의 보정, 분석기법, 응용 분야의 개발은

연구가 진행되고 있다. 특히, 고분광 영상의 많은 밴드수에 포함된 유효정보와 노이즈 정보를 구분하는 것은 영상분석의 질을 향상시키고, 분석 속도 및 효율성을 높일 수 있다는 점에서 연구가 필요하다. 본 논문에서는 지표면의 토지이용 분류에 필요한 고분광 영상의 유효 정보 추출 알고리즘을 알아보고, 영상의 분류를 통하여 해당 알고리즘의 정확도를 평가하고자 한다.

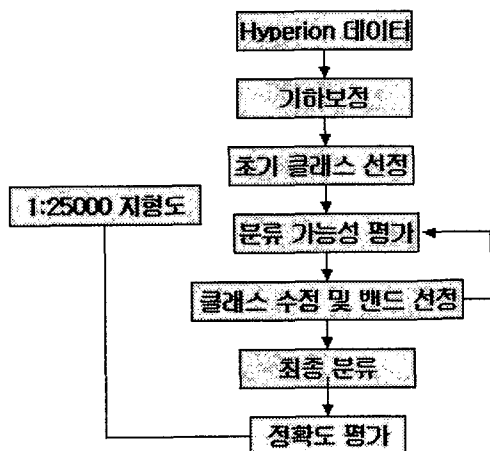


그림 1. 연구흐름도

## 2. 하이퍼분광 영상의 특징

### 2.1. 개요

지구관측용 센서는 전자기파를 이용하며, 사용하는 전자기파의 파장영역에 따라 관측되는 물체에 대한 다양한 속성 정보를 얻게 된다. 전자기파는 파장이 짧은 것부터 순서대로  $\gamma$ 선, X선, 자외선, 가시광선, 적외선, 전파로 분류되며, 원격탐사에 이용되고 있는 전자기파의 파장은 자외선의 일부( $0.3 \sim 0.4\mu\text{m}$ ), 가시광선( $0.4 \sim 0.7\mu\text{m}$ ), 적외선의 일부( $0.7 \sim 14\mu\text{m}$ ) 및 마이크로웨이브( $1\text{mm} \sim 1\text{m}$ )이다.

초기 광학 센서는  $100\text{nm}$ 의 파장영역을 하나의 밴드로 하여 가시광선에서 적외선

에 이르는 파장영역에 대한 정보를 수개의 밴드 형태로 취득하였으므로 다중분광(multispectral) 영상이라고 지칭하였지만, 고분광(hyperspectral) 센서는  $10 \sim 20\text{nm}$  폭을 가지는 40개 이상의 밴드를 연속적으로 취득하여 필드(field)나 연구실내에서 취득된 분광정보와 직접 비교할 수 있으므로, "hyperspectral remote sensing", "imaging spectrometry", "imaging spectroscopy" 등과 같은 용어로 사용되고 있다(그림 2참고).

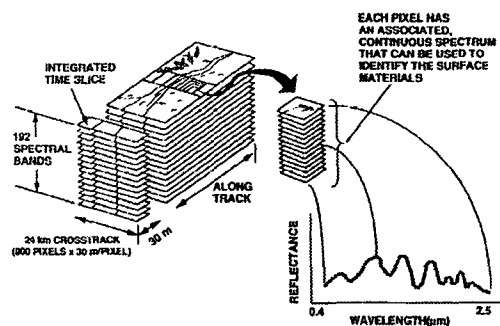


그림 2. Imaging Spectrometry 개념도

1976년에 NASA가 스펙트럴 프로파일링 기기에 대한 제안을 수락하면서 시작된 지구관측용 분광측정기의 개발은 1980년대 초에 최초의 지구관측용 고분광 센서인 AIS(Airborne Imaging Spectrometer)가 개발되고 이후 1984년에 개발된 AVIRIS는 하이퍼스펙트럴 영상의 연구를 위한 주요한 데이터원이 되었다. 이후 많은 민간회사들이 항공기 하이퍼스펙트럴 센서를 제작하였으며, 그 외에 HYDICE, NEMO, Hyperion 등이 주요한 센서로 운용되거나 계획 중에 있다.

### 2.2. 하이퍼이온 영상

Landsat과 같은 궤도를 1분 이내로 돌며 지구를 관측하는 EO-1에 탑재된 센서(ALI, Hyperion, LAC) 중 하나로 TRW에 의해 개발되었으며 220개 밴드를 가지고 있는 고분광 영상을 제공한다. Pushbroom

방식으로 스캔을 하고, Landsat과 같은 30m의 해상도를 갖지만, 185km의 관측폭을 가지는 Landsat과는 달리 5mile(약 7.7km)의 좁은 관측폭을 갖는다. 0.4~2.5  $\mu\text{m}$  파장 범위를 10nm 폭의 연속적으로 분할된 밴드로 구성되어 있다.

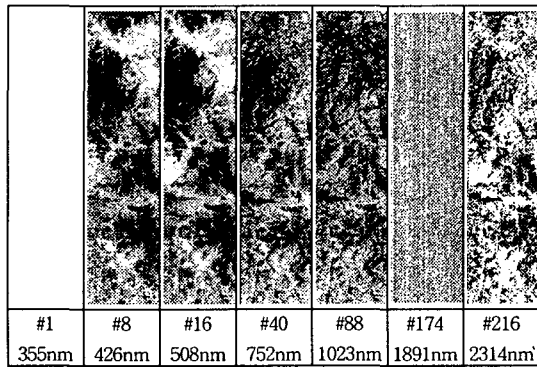


그림 3. Hyperion 밴드별 영상

### 3. 영상 밴드 추출

다중분광 영상을 분석하는 과정은 일반적으로 5단계로 나눌 수 있지만 단계별 결과를 참조하여 반복 분석해야 한다.

- ① 데이터 검토 : 데이터, 데이터 질, 일반적인 특성에 대한 검토(familiarity) 단계로 영상을 시각적으로 살펴보는 단계
- ② 클래스 정의 : 확인될 수 있는 또는 구분되어야 하는 클래스를 정의하는 것으로 training sample을 선정
- ③ 밴드 추출 : 분석에 사용될 특정 밴드가 확인(시각적)되거나 계산(정량적)되어야 한다. 즉 단순히 최적 밴드를 선택하는 과정이거나 유용한 어떤 방법으로 밴드 조합을 계산하는 단계이다.
- ④ 분석 : 구분을 위한 특정 분석 알고리즘 적용
- ⑤ 결과 평가 : 결과의 질과 특성을 결정하기 위한 정성적·정량적인 수단의 사용

위와 같은 분석과정에서 초기에 데이터를

파악하는 것이 중요하며, 5단계중에서 고분광 영상을 분류하기 위하여 필요한 ③번 단계까지의 주요 분석과정은 다음과 같다.

#### 3.1. 시각적인 판단

영상은 다양한 조건에 따라 다른 화소값을 갖게 되며 아직까지 조건을 체계화한 규칙이 마련되어 있지 않기 때문에 영상에 대한 시각적인 판단이 가장 쉽게, 때로는 컴퓨터로 처리하기 모호한 상태를 정확하게 판단할 수 있는 수단이다.

#### 3.2. 신호대잡음비(Signal-to-Noise Ratio)

SNR은 특정 밴드의 유용성을 판단하기 위하여 일반적으로 사용되는 것으로 SNR이 큰 레이더영상과는 달리 광학영상에서 SNR의 밴드별 차이를 비교함으로써 유효 밴드를 선정한다. SNR은 임의의 윈도우내에서 화소값의 평균과 표준편차의 비로 나타나며 SNR 영상을 시각적으로 확인하여, 밴드의 적합성을 판단한다.

#### 3.3. MNF 변환

MNF 변환(Minimum Noise Fraction Transform)은 영상의 고유 차원을 결정하고 노이즈를 분리하여 처리시간을 줄이기 위해 사용되는 것으로 두번의 주성분변환이 단계적으로 수행된다. 첫 번째 변환은 추정된 노이즈의 공분산 행렬에 근거하여 데이터에서의 노이즈를 재해석하는 단계이며, 이 단계를 거치면 밴드간의 상관성이 제거되고 유닛이 다양하게 나타나게 된다. 두 번째 단계는 MNF를 수행하여 노이즈가 하얗게 처리된 데이터에 표준주성분변환을 수행한다. 의미있는 데이터 차원은 최종 고유값(eigenvalue)과 이에 관련된 영상의 평가에 의해 결정되며, 선택된 고유치 임계값에 근거하여 새로운 영

상을 생성한다.

3.4. 판별 분석법(Discriminant Analysis)  
클래스내의 분산과 클래스간 분산값의 비를 최대화하는 밴드의 선형조합을 생성시키는 것으로 트레이닝 샘플의 통계값을 이용한다. 클래스간 평균값의 차이가 크지 않은 경우 비효과적이며, 신뢰할만한 추출 밴드수는 항상 클래스 수보다 하나 작다는 단점이 있다.

3.5. 결정 경계법(Decision Boundary)  
트레이닝 샘플의 통계량보다는 샘플을 직접 이용하여 유효 클래스 경계를 결정하는 것으로 원밴드를 선형결합하여 새로운 밴드 변환을 생성하고 고유값을 통한 새로운 밴드의 유효성을 평가할 수 있다. 하지만 샘플수가 작으면 비효과적이며, 너무 크면 계산시간이 길어지는 단점이 있다.

3.6. 클래스 분리도 측정  
분류 확률을 결정하는 트레이닝 단계의 최적화를 위하여 클래스간 분리도 분석을 수행하는 것으로, 밴드조합별 클래스 분리도를 통하여 분류에 이용할 적절한 밴드를 선택할 수 있다. 클래스에 따라 밴드 조합이 달라질 수 있으므로 트레이닝 샘플의 정규화도 고려해야 한다. 분리도 측정방법은 Divergency, Mahalanobis 거리, Jeffrey-Matusita 거리, Bhattacharyya 거리, 다변량 조건부 확률계산을 통한 유추 등이 있다.

4. 실험 및 결과  
사용영상은 2002년 9월 3일 EO-1 위성의 Hyperion 센서로 촬영된 것으로 영상의 기본제원은 표 1과 같다. 대상지역은 원영상의 아랫부분에 해당하는 경기도 남양, 발안, 군포, 안양을 포괄하는 곳이며, 긴

스트립 형태의 영상에서 도심지와 산, 농경지, 수계가 포함되도록 선정되었다. 실험영상의 크기는 974(행)×256(열)×242(밴드)이다.

표 1. Hyperion 데이터 헤더정보

구성항목	제원
처리단계	L1 Data Product
열	256
행	3129
밴드	242
취득일자	2002.09.03
데이터 형	2바이트
포맷	HDF Scientific Data
저장 방식	BIL(Band Interleaved)
Site 경도	126.989400
Site 위도	37.551670

4.1. 전처리  
전처리 과정에서는 영상의 노이즈 제거와 기하보정이 수행되었다. 원데이터에 나타나는 세로방향의 줄무늬(Line drop)는 영상분류에 큰 오차요인이 되기때문에 주위 화소값으로 대체하거나 평균값을 이용하여 제거하였다.  
영상의 기하보정은 1:25000 지형도(2002년 6월 수정)에서 기준점을 취득하여 1차 다항식으로 변환하였다(그림 4, 표 2 참고).

표 2. 기하보정에 사용된 기준점과 오차

#	Img X	Img Y	X Ref.	Y Ref.	RMSE
2	180.4	-91.9	195964.4	432323.3	0.19
3	140.0	-216.3	193895.1	428898.1	0.38
4	194.4	-176.2	195759.7	429716.0	0.65
5	213.8	-325.4	195285.9	425096.3	1.26
6	146.0	-452.1	192375.0	421875.0	1.25
7	132.4	-357.8	192675.5	424777.3	1.50
8	72.2	-465.3	190158.5	421956.2	1.94
9	247.5	-446.7	195382.5	421296.8	0.83
10	107.5	-936.8	187742.2	407675.3	0.65
11	210.5	-8.39.8	191518.3	409821.1	0.90
13	101.6	-681.4	189425.0	425300.0	0.71
14	96.0	-850.6	188021.9	410336.5	0.77
15	44.3	-388.3	189778.4	424427.1	1.80



그림 4. Hyperion 영상 (좌) 전체영상, (우) 기하보정된 연구대상지역 (rgb:50,27,17)

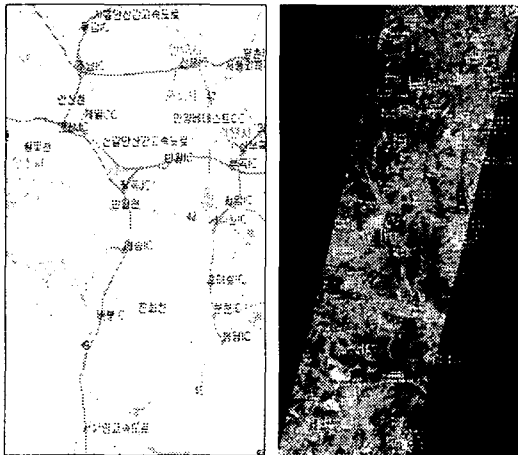


그림 5. (좌) 대상지역 특성, (우) 트레이닝 선정

#### 4.2. 초기 영상 분석

355.59nm~2577.08nm 파장범위에서 연속적으로 영상이 취득되었기 때문에 지표면 에너지가 흡수되는 파장범위에서 영상정보가 나타나지 않았고, 노이즈가 심한 밴드도 많이 존재하였다. 위의 파장영역은

지형지물 분류를 위한 정보를 담고 있지 않거나 노이즈로 작용하므로 통계량 분석 없이 이후 영상처리에서 제거하였다(표 3 참고),

표 3. 영상의 지형정보량 유무 구별

단계	밴드	범위	히스토그램	시각*
정보 포함	10~56 82~97 101~119 131~164 182~184 187~218	유	유	양호
노이즈 적음**	8~9 57 77~81 98~100 120 128~130 165, 181 219~220	유	유	구분가
노이즈 심함	121~127 166~180 185~186 221~224	유	무	불가
정보 없음	1~7 58~76 225~242	무	무	불가

\* : 시각적인 판단.

\*\* 노이즈가 심하거나 화소값 대비가 떨어짐.

#### 4.3. 영상 분류

대상지역의 복잡한 지형지물 특성과 영상 취득시기 농지의 특성으로 인하여 클래스의 조합과 정확도 평가에 어려움이 있었다. 실험은 다음과 같은 경우의 조합으로 이루어졌으며, 사용된 분류자는 MLC, SAM, ECHO(Extraction and Classification of Homogeneous Objects)이다.

- ① 초기 추출된 151개 밴드
- ② DBFE(Decision Boundary)에 의한 60개 밴드
- ③ 151개 밴드에서 클래스 분리도를 고려한 최적 4개 밴드 조합으로 12클래스는 (19 31 144 145)이고 6클래스는 (6 27 31 37)임.
- ④ DBFE의 밴드에서 클래스 분리도를 고려한 최적 4개 밴드 조합 : 2.6,12,15
- ⑤ Barren1, Barren2, Rice Field, Wet Rice Field, Dry Field, Grass, Urban, Forest, Shadow

Forest, Water, Background  
 ⑥ Barren, Field, Urban, Vegetation, Water, Background

표 4는 트레이닝 영역에 대한 분류정확도 결과를 나타낸 것이며, 이후 1:25,000 지형도에서 정확도 평가를 위한 참조데이터는 추출하여 평가할 예정이다.

표 4. 12 클래스 분류 결과(Overall Accuracy: %)

	MLC		SAM		ECHO	
	⑤	⑥	⑤	⑥	⑤	⑥
①	98.0	97.8	60.9	56.8	98.2	98.4
②	91.6*	95.4*	59.0	58.8	93.8	97.8
③	80.1	80.4	43.4	49.8	82.7	90.2
④	62.7	×	62.0	×	77.2	×

\* : 60개 밴드의 경우 Singular가 발생하여 40개 밴드의 분류 결과임.

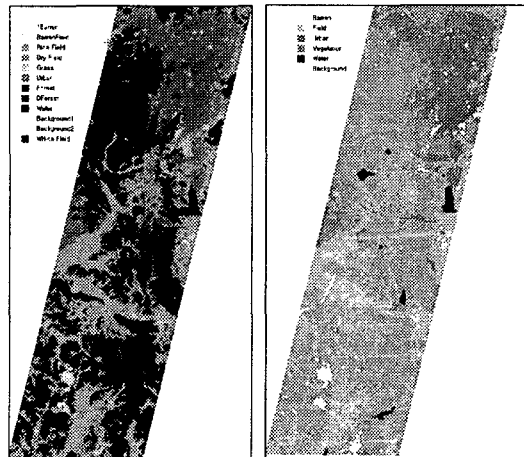


그림 6. 분류결과: (좌) ①+⑤+ECHO 조합, (우) ①+⑥+ECHO 조합

MLC와 ECHO비슷한 정확도를 보였지만 ECHO가 약간 우수한 결과를 나타내었다. 하지만 ECHO는 개개의 화소값과 공간적 분포를 고려하기 때문에 특정 클래스가 확대되는 단점을 보였다. SAM은 정확도가 상대적으로 낮았고, 시각적으로도 많은 분류오류를 생성하였다. 본 실험과 같은 고분광영상에서 MLC를 적용할 경우 트레

이닝 화소가 작으면 분류를 수행할 수 없으며, 공분산 행렬을 적당히 추정해서 분류하는 단점을 지니고 있다. 클래스 정보를 바탕으로 최적의 밴드를 찾아내는 분리도 평가는 고분광영상에서 많은 처리시간이 필요하므로 실용적이지 않았다.

### 5. 결론

첫째, 특정 목적에 고분광영상을 사용할 경우 목적에 맞는 유효 정보밴드 파장대역에 대한 폭넓은 연구가 필요하다.

둘째, 밴드추출로 생성된 영상의 유효밴드 개수에 대한 검증이 이루어져야 하며, 적절한 개수로 축소될 경우, 분리도 평가를 통해 클래스별 최적 밴드를 선정할 수 있다.

셋째, 영상의 노이즈 제거, 대기 보정 등 전처리 과정이 제대로 수행되어야 이후 밴드 추출 및 분류 작업의 효율성을 기대할 수 있다.

### 참고 문헌

어양담, "위성영상 분류를 위한 트레이닝 정규화 알고리즘과 클래스 분리도 측정기법 개발", 서울대 대학원 박사학위논문, 1999.

David Landgrebe, Larry Biehl, Multispec Tutorial, An Introduction to Multispec version.5.2001.

ENVI Tutorial, Research Systems.

Freek D. Van Der Meer, etc.,2001, "Imaging Spectrometry", Kluwer Academic Publishers.

David Landgrebe, "Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem," (Invited), Special Issue of the IEEE Signal Processing Magazine, Vol 19, No. 1 pp. 17-28, January 2002.