

초분광 이미지 픽셀 분류를 위한 이산 범위 기반의 최적 밴드 선택 연구

**Optimal Band Selection based on Discrete Range
for Hyperspectral Image Pixel Classification**

한성대학교 컴퓨터공학과
20211302 장 두 혁

CONTENTS

1. 논문 요약
2. 초분광이미지/ 데이터 세트
3. 관련 연구
 - Simulated Annealing
 - 상호 정보량
4. 기존 기법의 문제점
5. 구현 내용
 - 이산 범위별로 대표하는 밴드추출(DI)
 - Mutual Information & Classification accuracy(MC)
6. 실험 및 결과
7. 향후 방향

요약

- **초분광 이미지는** 일반 이미지와 달리 전자기 스펙트럼을 파장에 따라 수많은 밴드로 나누어 촬영된 것으로 고용량 고해상도 이미지이다.
- 일반 이미지보다 정보량이 많아 물체나 물질 탐사에 활용된다. 처리할 초분광 이미지의 정보량을 줄이기 위해 **밴드 선택(band selection)기법**을 활용한다.
- 기존 밴드 선택기법들은 통계를 바탕으로 하는 **휴리스틱 기법**으로, 소요 시간이 길며, 일반성과 보편성이 떨어지는 경우가 많다.
- 이를 보완해 본 논문에서는 양자화(Quantization) 개념을 활용하여, **이산 범위(Discrete Range)를 통해 범위별로** 대표적인 밴드 선택한 후보밴드 집합으로 사용한다.
- 실험 결과를 통해 제안 기법이 기존 밴드 선택 방식보다 **수행 시간이 매우 빠르며 밴드 수를 1/10~1/7로 줄였음에도 원본과 성능 정확도가 유사함**을 보였다.


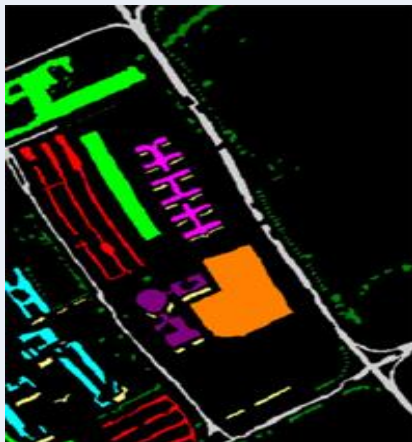
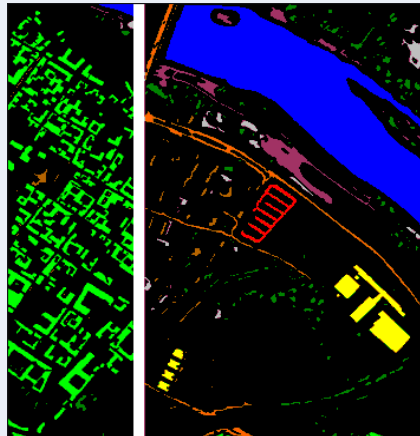
초분광 이미지/ 실험 데이터 세트

❖ 초분광 영상

수백 개의 분광 채널을 통해서 대상 물체의 특성을 나타내는 연속적인 스펙트럼을 수집하는 센서를 활용하여 촬영한 영상이다.

❖ 초분광 이미지 처리

공간 정보에 분광 기술을 더한 것으로 전자기파의 스펙트럼 밴드에 따른 2차원적인 영상정보를 초분광 큐브 형태로 피쳐(feature)를 추출하는 기술이다.

Indian Pines(IP)	Pavia Centre (PC/PU)	Pavia University (PU)
<p>파장 범위 0.4~2.510-6m의 (145,145) 픽셀과 220개의 스펙트럼 반사 대역 밴드</p> 	<p>파비아 센터는 (1096, 715) 픽셀로 구성되고, 파비아 대학은 (610,340)으로 구성되어있다. 그리고 9개의 클래스로 구성</p> 	

Dimensionality Reduction Method

초분광이미지는 넓은 스펙트럼을 가진 고해상도 고용량 이미지이므로
classification의 class에 필요한 밴드들만 가지고 모델 Input으로 넣어 차수를 줄이는 의미

1. Feature Extraction

original 데이터를 projected된 공간으로 변환하여 feature을 가져오는 방법

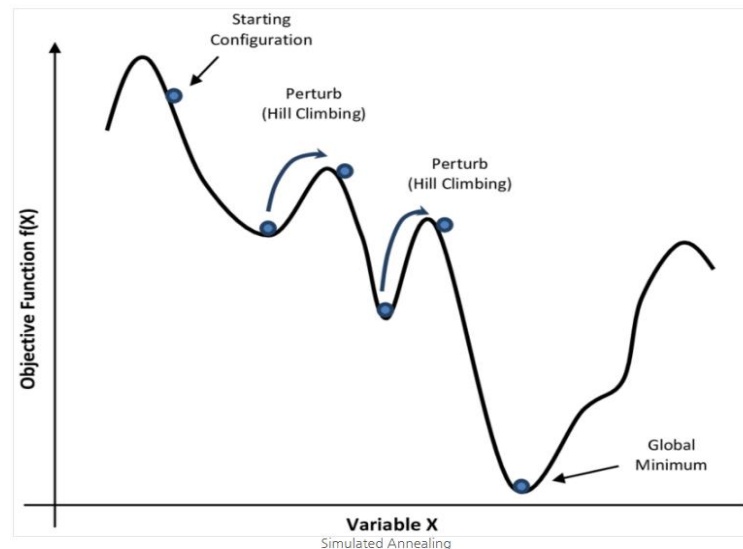
2. Feature Selection

데이터 기본특성을 포함하는 형상의 SubSet을 식별하는 방법

시뮬레이티드 어닐링(Simulated annealing)

시뮬레이티드 어닐링(Simulated annealing)

- 탐색공간에서 주어진 함수나 문제에 대해 전역 최적점에 대한 가까운 근삿값을 찾아, 점진적으로 해에 가까운 방향으로 이동한다.
- 해가 예상되는 곳으로, 접근하는 확률적 휴리스틱 방식이다. 밴드 선택의 후보 개수와 밴드 크기에 따라 소요 시간이 발생한다.

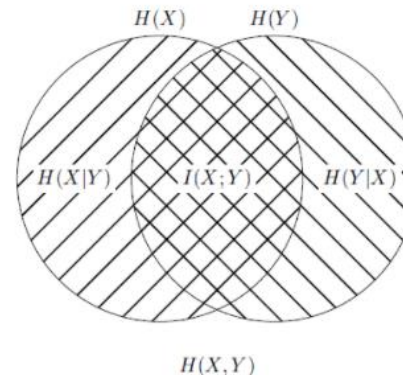


상호 정보량(Mutual Information)

상호 정보량(Mutual Information)

- Mutual Information은 두 확률 변수들이 얼마나 독립한지를 측정하는 방법을 의미
- 직관적으로 보면, 상호 정보량을 벤 다이어그램으로 나타내면 아래 그림과 같은 형태

$$I(A, B) = \sum_{A \in \mathcal{A}, B \in \mathcal{B}} p(A, B) \log \frac{p(A, B)}{p(A)p(B)}$$

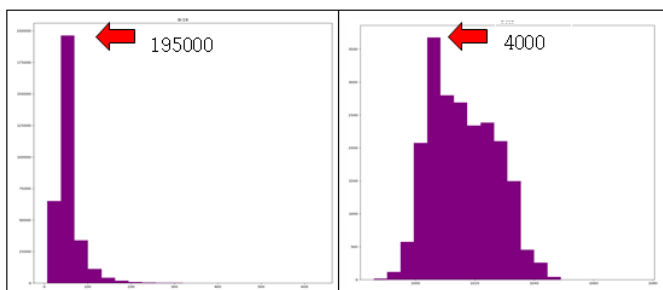


1. 첫 번째, 실제로 대상 물체에 대해서 잘 추출되는 파장을 분석하여, 단시간으로 피쳐 추출하기가 어렵다.
2. 두 번째, 밴드 선택(band selection)이 어떤 한 지역에 대해서 적용되는 것뿐만 아니라, 다른 지역에 대해서도 일반성과 보편성을 띠면서 적용되기 어렵다.
3. 세 번째, 모든 스펙트럼 밴드를 기계학습과 딥러닝 모델에 적용하여 실시간 분석이 효율적이지 않다.
4. 어떤 스펙트럼이 어떠한 클래스를 잘 classification하는 지 알 수 없는 경우

이산 범위별로 대표하는 밴드추출(DI)

• 양자화(Quantization) 개념 활용

1. 제한된 범위 안에 비트 수에 따라 양자화 스텝(Quantization step)의 범위를 정한다.
2. 양자화 레벨(Quantization level)을 설정한다.
3. 실숫값과 정숫값에 스케일링 연산을 통해 대응한다.



1. 초분광 이미지 밴드별로 유사한 피처를 가지고 있는 밴드는 분포구성 또한 유사
2. 밴드별로 최솟값과 최댓값을 이산적인 범위로 밴드를 분류[나]
3. 분류된 밴드집합 내에서 픽셀과 GT간의 상호 정보량이 높은 밴드를 선택[가]

가

$$I(A, B) = \sum_{A \in A, B \in B} p(A, B) \log \frac{p(A, B)}{p(A)p(B)}$$

나

$$R_n = \{b_1, b_2, b_3, \dots, b_m\}$$

$$\Delta r_n = \partial((P_{i_{Max}} - P_{i_{Min}})/10^a, 3) * 10^a$$

Mutual Information & Classification accuracy(MC)

참고 논문 기법

Reference Paper : Band selection based on optimization approach for hyperspectral image classification

- Objective Function으로 구성하여 Simulated Annealing 기법으로 classification 분류에 최적화된 밴드를 선택

$$E = accuracy(N) + E_{relevance} \longrightarrow E_{relevance}(F) = \frac{\sum_{i=1}^n f_i \times I(X_i; Y)}{\sum_{i=1}^n I(X_i; Y)}$$



$$\begin{cases} accuracy(N) = \frac{\sum_{i=1}^N assess(p_i)}{N}, n_i \in N \\ assess(p_i) = \begin{cases} 1 & \text{if } classify(p_i) = c \\ 0 & \text{otherwise} \end{cases} \end{cases}$$

Mutual Information & Classification accuracy(MC)

본 논문에서 개선

Objective Function으로 구성하여 Simulated Annealing 기법으로
classification 분류에 최적화된 밴드를 선택



DI 기법을 통해 상호정보량의 최댓값을 뽑지 않고,
참조 논문의 목적함수의 최댓값을 분류된 밴드별로 선택

Why ? Simulated Annealing 기법을 적용 시간이 데이터셋 픽셀별로 시간소요가 큼

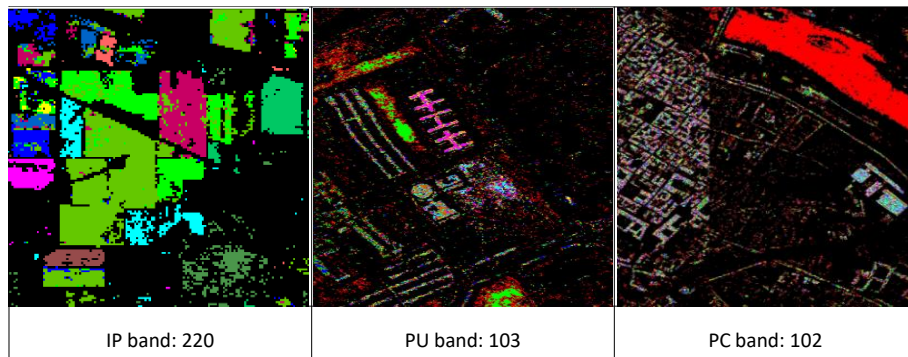
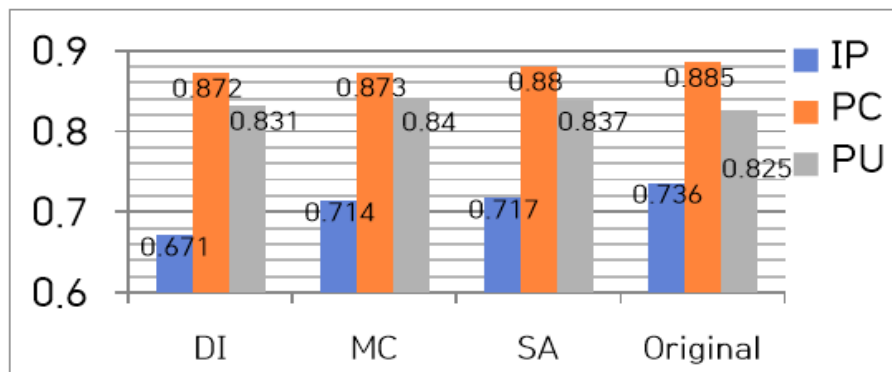
$$E = accuracy(N) + E_{relevance} \longrightarrow E_{relevance}(F) = \frac{\sum_{i=1}^n f_i \times I(X_i; Y)}{\sum_{i=1}^n I(X_i; Y)}$$



$$\begin{cases} accuracy(N) = \frac{\sum_{i=1}^N assess(p_i)}{N}, n_i \in N \\ assess(p_i) = \begin{cases} 1 & \text{if } classify(p_i) = c \\ 0 & \text{otherwise} \end{cases} \end{cases}$$

05 개선방향 구현 내용3 및 결과

실험 결과



PU	Original	band selection(DI)
not labeled	0.90	0.90
asphalt	0.46	0.47(+0.01)
Meadows	0.61	0.62(+0.01)
Gravel	0.40	0.52(+0.12)
Trees	0.27	0.31(+0.04)
metal sheets	0.83	0.82(-0.01)
Bare Soil	0.28	0.33(+0.05)
Bitumen	0.64	0.66(+0.02)
blocking bricks	0.44	0.47(+0.03)
shadows	0.36	0.46(0.10)

향후 방향

- 본 연구에서는, 기존 밴드 선택기법 중 SA 알고리즘의 적용 시간이 많이 소요되는 점을 보완한 기법으로, 밴드별로 픽셀의 최댓값과 최솟값 차이를 통해, 밴드를 분류한다.
- 분류된 밴드 집합별로 대표적인 값을 선택하는 방식을 제안했다.
- MC 알고리즘에서 집합별로 대표 밴드를 선택할 때의 연산을 멀티 스레드 사용하여, 40분보다 40* 1/집합 수로 감소할 것이 예상되며, 앞으로는 멀티 스레드 적용 및 임베디드 기기에 알고리즘 적용 전력 소모와 시간 소요 시간을 상세하게 비교하는 실험을 동시에 진행할 예정이다.

감사합니다.