

자동 윈도우 크기 결정 기법을 적용한 Minimum Entropy Clustering과 Iterative Over-Segmentation 기반 Semantic Segmentation

최형욱*, 송현승**, 손홍규***, 전문구*
*광주과학기술원 정보통신공학부
**국방과학연구소
***연세대학교 사회환경시스템공학부
e-mail:chu@gist.ac.kr

Semantic Segmentation using Iterative Over-Segmentation and Minimum Entropy Clustering with Automatic Window Size

Hyunguk Choi*, Hyeon-Seung Song**, Hong-Gyoo Sohn***, Moongu Jeon*
*School of Information and Communications, GIST
**Agency for Defense Development
***School of Civil & Environmental Engineering, Yonsei University

요 약

본 연구에서는 야외 지형 영상 및 항공 영상 등에 대하여 각각의 영역들의 속성을 분할 및 인식 하기 위해 minimum entropy clustering 기반의 군집화 기법과 over-segmentation을 반복 적용하여 군집화 하는 두 방법을 융합한 기법을 제안하였다. 이 기법들을 기반으로 각 군집의 대표 영역을 추출한 후에 학습 데이터를 기반으로 만들어진 텍스트론 사진과 학습 데이터 각각의 텍스트론 모델을 이용하여 텍스트론 히스토그램 매칭을 통해 매칭 포인트를 얻어내고 얻어낸 매칭 포인트를 기반으로 영역의 카테고리를 결정한다. 본 논문에서는 인터넷에서 얻은 일반 야외 영상들로부터 자체적으로 제작한 지형 데이터 셋을 통해 제안한 기법의 우수성을 검증하였으며, 본 실험에서는 영역을 토양, 수풀 그리고 물 지형으로 하여 영상내의 영역을 분류 및 인식하였다.

1. 서론

Semantic segmentation 연구는 입력 영상에서 각기 다른 의미의 물체들이나 영역 및 공간에 대하여 분할 및 인식하는 기술이며, 물체 인식 분야와는 다르게 찾고자 하는 대상과 배경으로 영상을 다루는 것이 아닌 영상에 있는 모든 오브젝트 들을 분류 및 인식하게 된다. 최근 감시 시스템, 정찰 시스템, 영역 특성화 속성 자료 수집 등의 다양한 지능형 영상 처리 분야에 연구되고 응용되는 분야이다. semantic segmentation 기반의 지형 분석의 목표는 지형 영상에 대해 각 영역의 지형적인 특징들을 기반으로 해당 지형이 어떠한 클래스에 속하는지를 분류하는 것이 목표이다. 정확도 향상을 위해서 레이더와 같은 다른 측정 장비를 이용하여 더 많은 정보를 얻을 수 있으나 시간과 비용이 많이 소요된다. 따라서 항공사진이나 무인항공기로부터 얻은 영상의 경우 직접 분석을 하게 될 경우 운영비용 및 시간을 많이 요구하므로 이러한 문제해결을 위해서 semantic segmentation 연구를 통한 지형 분석 기법이 필요하다.

본 논문에서는 over-segmentation된 영상에 대해

minimum entropy clustering(MEC)을 이용하여 군집화 하는 기법과 반복되는 Over-segmentation 알고리즘에 recoloring 기법을 더해서 군집화 하는 기법을 이용하여 군집화된 각 영역별 텍스트론 모델로 매칭 포인트를 추출한 뒤 해당 위치의 카테고리를 인식하는 방법을 제안하였다.

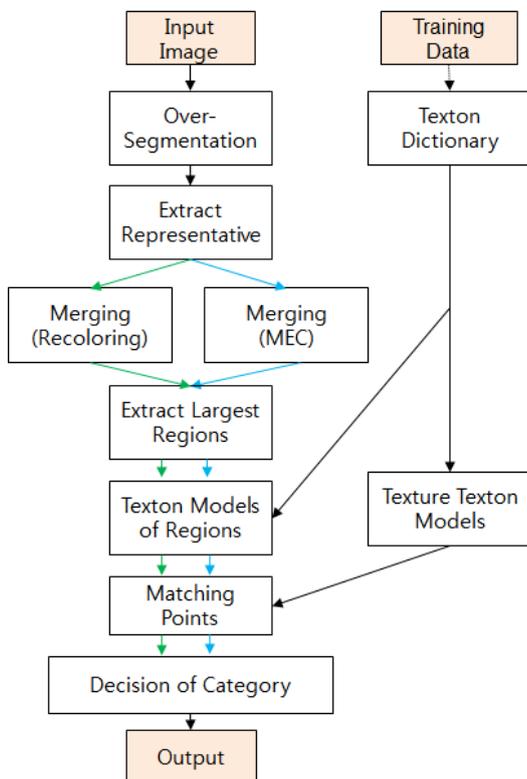
기존의 분할 방법들의 경우 군집화를 위해서 유사성을 계산할 때 계산을 위한 탐색 영역을 미리 정의해 주거나 군집의 개수 k 를 미리 직접 입력해주었다[1,5]. 그러나 본 논문에서 사용된 군집화중 하나인 MEC 기법은 자동 윈도우 사이즈 결정 알고리즘을 추가하여 탐색영역을 자동으로 결정해 주고 군집의 개수 k 는 미리 충분히 설정해 주면 입력 영상의 feature 분포에 맞게 k 를 자동으로 잡아 주게 되어 이러한 문제점을 해결 하였다.

또 다른 군집화 기법인 over-segmentation는 군집 개수 k 를 줄이면서 over-segmentation을 반복하여 서서히 군집화 하는 기법이다. 이 기법에서 보다 명확한 군집화를 위하여 반복 매 마다 영역별로 영역 내의 픽셀 값의 유사성(similarity)를 최대치로 만들었고 이를 위해 recolorization을 사용하였다.

본 논문의 구성은 2장에서 알고리즘의 특징과 기법의 설명을 논하고 3장에서는 실험 결과 및 분석 그리고 4장에서 결론을 논하는 식으로 구성된다.

2. Semantic segmentation기반 지형 분석 기법

본 논문에서 제안한 Semantic Segmentation기반 지형 분석 기법은 아래(그림 1)과 같다. 과정은 크게 학습데이터를 기반으로 클래스별 텍스처들의 텍스톤 히스토그램 모델을 생성하는 부분과 입력된 영상에 대해서 군집화 된 영역별로 텍스톤 히스토그램 모델을 생성하는 부분으로 나뉜다. 그리고 입력된 영상을 처리하는 부분은 내부적으로 minimum entropy clustering(MEC)을 이용하여 군집화 하는 부분과 iterative over-segmentation을 이용하여 군집화 하는 부분으로 나뉜다.



(그림 1) Semantic segmentation 기반 지형 분석 기법

인식을 위해 학습데이터로 각 텍스처별 텍스톤 모델을 만드는 과정에서 학습데이터는 3가지 종류(수풀, 토양, 물 지형)의 텍스처를 이용하였다. 텍스톤 사전을 만드는 과정에 있어서 각 텍스처별 텍스톤은 35개씩 하였으며 텍스톤 사전을 만들 때 사용된 Filterbank는 Gaussian, Laplacian of Gaussian(LoG), Derivative of Gaussian의 총 3가지 종류의 필터를 사용하였다. 채널은 CIELab 색공간을 상용하였으며, Gaussian 필터는 3개의 채널(L, a, b)에 대하여 각각 $\sigma = 1, 2, 4$ 를 적용하여 9개의 반응을 얻는다. LoG 필터의 경우에는 L채널에만 $\sigma = 1, 2, 4, 8$ 를 적용하여 4개의

반응을 얻고, Derivative of Gaussian 필터 역시 L채널에 대해서만 $\sigma = 2, 4$ 를 적용하지만 각각에 대해서 0° 와 90° 회전된 필터를 적용하여 4개의 반응을 얻는다. 따라서 텍스톤 사전의 요소들은 17차원의 벡터들로 이루어져있다 [2,7].

입력 영상에 대해 텍스톤 모델을 만드는 부분에서 첫 번째 단계는 입력 영상에 over-segmentation을 수행하는 단계이다[3].

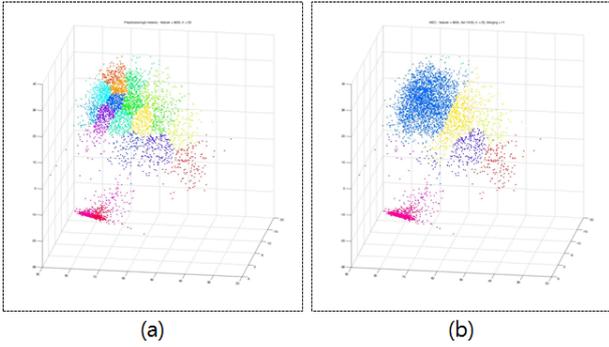
모델을 만든 뒤 대푯값 추출 단계에서 segment별로 가장 큰 정사각형 영역을 추출하고 그 영역 내부의 값들의 평균을 해당 segment의 대푯값으로 정한다. 이러한 과정의 이유는 segment의 경계에 있는 값들로 인한 노이즈의 영향을 없애기 위함이다. 추출된 feature들은 두 개의 군집화 과정을 거친다. 두 개의 군집화 과정인 recoloring을 적용한 iterative over-segmentation방법과 minimum entropy clustering방법은 다음 섹션에서 설명된다.

군집화 과정을 거친 결과들은 각각의 군집화 된 영역별로 영역 내 최대 정사각형 부분을 해당 영역의 대표 텍스처 영역으로 정의한 뒤 텍스톤 사전을 이용하여 영역의 텍스톤 모델을 생성한다. 학습데이터에 의해서 생성된 각 속성별 텍스톤 히스토그램 모델과 입력영상으로부터 얻은 영역별 텍스톤 히스토그램 모델로부터 매칭 포인트를 계산하여 최종적으로 카테고리를 결정하게 된다.

2.1 자동 윈도우 사이즈 결정 기법이 적용된 minimum entropy clustering

MEC의 가장 큰 장점은 feature의 데이터 분포 구조에 맞게 자동으로 클래스 개수 k 를 찾는다라는 점이다. 즉, 처음에 적당한 양의 k 값을 지정해 주면 그림 2와 같이 최적화된 클래스 k 의 값으로 만들어져서 군집화가 되는 효과를 볼 수 있다. 그림 2에서 (a)는 MEC를 적용하기 전 전처리 과정으로 K-means clustering을 적용한 결과이며 이때 k 값은 20으로 주었다. (b)는 (a)의 결과에 대해 MEC를 적용한 결과이다. K-means clustering과 같은 전처리 과정이 필요한 이유는 MEC가 supervised clustering이기 때문이다. 이때의 k 값은 6으로 데이터 구조에 맞게 군집화 되면서 자동으로 맞춰진 k 값이다.

이러한 장점에도 불구하고 MEC의 경우 entropy를 계산할 때 타깃 feature의 이웃들을 정의 할 윈도우 사이즈를 직접 결정해 주어야했다.



(그림 2) 자동 윈도우 사이즈 결정 기법이 적용된 MEC에 의해 군집화 된 결과

그러나 단순 2D 데이터와 달리 영상의 경우 feature가 3D이상이며 color space에 따라서 값의 범위가 천차만별이므로 그때마다 매번 윈도우 사이즈를 직접 입력하는 것은 비효율적이었다. 따라서 자동 윈도우 사이즈 결정 기법 개발하여 MEC에 적용시켰다.

- N = 입력 영상의 모든 데이터 개수
- K = 전처리 과정으로 얻은 결과에서 클래스 개수
- Assumption : 모든 클래스들은 각각 거의 같은 양의 데이터를 가지고 있다.
- 1. Compute $N/K = C'$: the number of elements in one cluster
- 2. Compute Distance each points (Euclidean Distance)
 - $D_1 = \{a_1 = \text{dist}(x_1, x_2), a_2 = \text{dist}(x_1, x_3), \dots, a_{N-1} = \text{dist}(x_1, x_N)\}$
 - $D_2 = \{b_1 = \text{dist}(x_2, x_1), b_2 = \text{dist}(x_2, x_3), \dots, b_{N-1} = \text{dist}(x_2, x_N)\}$
 - ⋮
 - $D_N = \{z_1 = \text{dist}(x_N, x_1), z_2 = \text{dist}(x_{20}, x_2), \dots, z_{N-1} = \text{dist}(x_N, x_{N-1})\}$
- 3. Sort in ascending order and Select C'
 - $V_1 = \text{Select}\{ \text{sort}(a_1, a_2, \dots, a_{N-1}) \} \rightarrow a_{c'}$
 - $V_2 = \text{Select}\{ \text{sort}(b_1, b_2, \dots, b_{N-1}) \} \rightarrow b_{c'}$
 - ⋮
 - $V_N = \text{Select}\{ \text{sort}(z_1, z_2, \dots, z_{N-1}) \} \rightarrow z_{c'}$
- 4. Bandwidth = Average(V_1, V_2, \dots, V_N)

(그림 3) 자동 윈도우 사이즈 결정 기법

그림 3에 설명된 윈도우 사이즈 결정 기법의 경우 모든 클래스들은 각각 거의 같은 양의 데이터를 가지고 있다는 가정을 둔다. 윈도우 사이즈가 결정 된 다음에 MEC 기법으로 군집화가 이루어지며 이때 계산되는 entropy의 변화 h 는 Shannon's entropy를 기반으로 다음과 같이 주어진다.

$$h = \sum_y [H(C|y) - H(C|y)] \quad (1)$$

y 는 현재 선택된 feature를 포함한 이웃한 feature들의 집합을 의미하며, H' 은 선택된 feature를 C_j 에 할당 시킨 후, H 는 할당시키기 전에 계산된 entropy를 의미한다. 자세히 쓰자면 다음과 같다.

$$h = - \sum_{i=1}^n \sum_{j=1}^m f(C_j, y_j) \quad (2)$$

$$f(C_j, y_j) = p'(C_j | y_i) \log p'(C_j | y_i) - p(C_j | y_i) \log p(C_j | y_i) \quad (3)$$

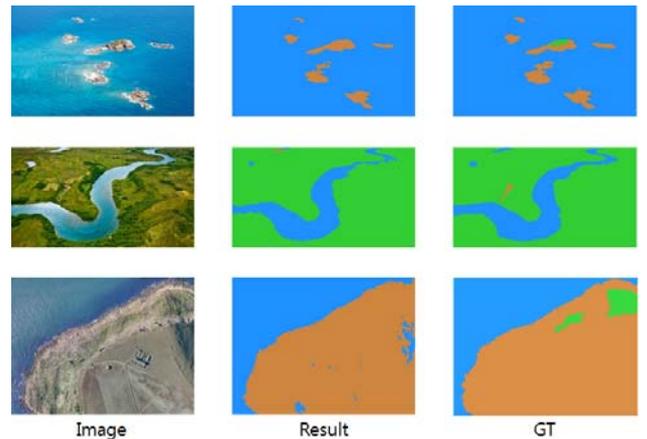
전체 feature들의 클래스는 주변의 이웃들을 고려하여 entropy가 낮은 방향으로 클래스 변경이 진행된다. 다시 말하자면 entropy의 변화 h 가 음수면 현재 선택된 feature의 클래스를 윈도우 내에서 가장 많이 차지하는 클래스로 변경시켜준다.

2.2 Iterative over-segmentation을 이용한 군집화

Iterative over-segmentation 기법은 over-segmentation의 클래스 개수 k 를 줄여 가면서 군집화 하는 과정을 의미한다[3]. 이 과정에서 명확한 군집화를 위하여 over-segmentation 이후에 영역 내부의 값들의 similarity가 최대치가 되게 하여 이후 반복되는 over-segmentation 과정에서 k 값이 줄어도 먼저 군집화 되었던 feature들은 계속 같은 클래스로 군집화가 되게 하였다. 이 과정에서 similarity를 최대화하기 위해 사용한 방법이 recoloring 기법인데 군집화 된 내부의 모든 pixel들을 그들의 평균값으로 바꾸어 주는 방법이다. 그렇게 되면 다음 over-segmentation 과정에서도 높은 similarity에 의해서 나뉘지 않고 계속 같은 클래스 안에 속해 있게 된다.

3. 실험 결과 및 분석

본 논문에서 실험한 데이터 셋은 항공 촬영 등의 야외 영상으로 데이터 셋은 수풀, 토양, 물 지형으로 구성 되어 있다.

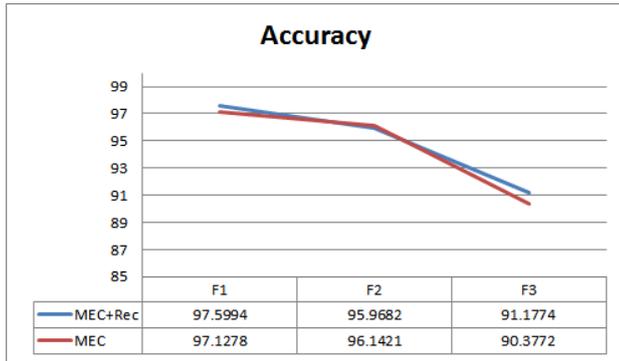


(그림 4) 지형 분석 결과

지형 영상의 경우에도 다른 일반 사물 영상과 마찬가지로 그림자와 빛에 영향을 많이 받는다. 그와 동시에 수풀, 토양, 물 지형을 정의하는 것도 쉽지 않다. 그림 4의 3번째 영상을 보면 검고 짙은 부분이 흙이나 토양 보다는 어둡고 푸른 물 지형과 더 유사성이 높아서 물 지형으로

분류 되는 것을 볼 수 있다.

<표 1> Ground Truth 대비 정확도



테스트한 영상에서 Ground Truth 대비 우수한 성능을 보이지만 F3과 같이 지형이 너무 어둡거나 불분명 할 경우 영역의 텍스처 히스토그램 모델의 인식이 어려운 모습을 볼 수 있다.

4. 결론

본 논문에서는 항공영상 및 야외 지형 영상에 대해 각 지형 영역을 분류하고 속성을 인식하였다. 지형을 분류한 다음 군집화 부분에 있어서 자동 윈도우 크기 설정 알고리즘을 추가 적용한 minimum entropy clustering 방법과 recoloring으로 similarity를 최대화 시켜서 군집화한 iterative over-segmentation 방법을 이용하였다. 분석을 위한 데이터 셋으로 수풀, 토양, 물 지형의 샘플 영상을 사용하였고 텍스처 사전으로 텍스처 히스토그램 모델을 만들어서 매칭 점수를 비교하는 방식으로 클래스 영역들을 인식하였다. 이 기법은 학습된 텍스처 샘플들에 대해서 우수한 성능을 보장하고 있다. 하지만 물과 유사한 풀, 토양과 유사한 잔디 등을 완벽하게 분석하기란 쉽지 않다. 향후 이러한 3가지 클래스의 지형 텍스처 이외에도 다른 샘플들을 학습 시켜 더욱 구체적인 지형 분석이 이루어지도록 하며 동시에 추출된 정사각형 영역을 텍스처 히스토그램 매칭을 통한 분석뿐만 아니라 다른 local feature 정보를 이용해서 분석의 정확도를 높일 것이다.

감사의 글

본 연구는 한국과학기술원 국방무인화기술 특화연구센터를 통한 방위사업청과 국방과학연구소 연구비 지원으로 수행되었습니다.

참고문헌

[1] Haifeng Li, Keshu Zhang and Tao Jiang, "Minimum Entropy Clustering and Applications to Gene Expression Analysis", Computational Systems Bioinformatics Conference(CSB), 2004

[2] Manik Varma and Andrew Zisserman, "A Statistical Approach to Texture Classification from Single Images", International Journal of Computer Vision(IJCV), Vol 62, Issue 1-2, pp 61-81

[3] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam and Rama Chellappa, "Entropy Rate Superpixel segmentation", Computer Vision and Pattern Recognition(CVPR), 2011

[4] Pablo Arbelaez, Bharath Hariharan, Chunhui Gu, Saurabh Gupta, Lubomir Bourdev and Jitendra Malik, "Semantic segmentation using regions and Parts", Computer Vision and Pattern Recognition(CVPR), 2012

[5] Yongjin Lee and SeungJin Choi, "Minimum Entropy, k-Means, Spectral Clustering", International Joint Conference on Neural Networks (IJCNN), 2012

[6] J. Shotton, J. Winn, C. Rother and A. Criminisi, "TextronBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and segmentation", European Conference on Computer Vision(ECCV), 2006, Vol 3951, pp 1-15

[7] Jamie Shotton, John Winn, Carsten Rother and Antonio Criminisi, "TextronBoost for Image Understanding: Multi-Class Object Recognition and segmentation by Jointly Modeling Texture, Layout, and Context", Vol 81, Issue 1, pp 2-23