

준 지도 학습에 기반한 중요 객체 검출 방법

황인성, 이상화, 박재성, 조남익

서울대학교 전기정보공학부

coee5@ispl.snu.ac.kr, lsh529@snu.ac.kr, jason79.park@ispl.snu.ac.kr, nicho@snu.ac.kr

Saliency Detection Based on Semi-Supervised Learning

Insung Hwang, Sang Hwa Lee, Jae Sung Park, and Nam Ik Cho

Dept. of Electrical and Computer Engineering, Seoul National University

요약

본 논문에서는 준 지도 학습에 기반한 중요 객체 검출 방법을 제안한다. 첫째, 색상과 공간 정보를 활용하여 이미지를 분할한 후, 분할된 영역을 색상의 유사도로 연결하여 그래프를 만든다. 둘째, 색 대비 및 가장자리 사전 지식을 활용하여 중요 객체에 해당하는 씨앗 노드와 배경에 해당하는 씨앗 노드를 추출한다. 끝으로, 중요 객체 및 배경 씨앗 노드를 이용하여 준 지도 학습 기법에 적용함으로써 이미지 전체 노드의 중요도를 계산한다. 실험 결과, 제안한 알고리즘이 최신의 다른 알고리즘보다 높은 재현율 구간에서 높은 정밀도를 보임을 확인할 수 있고, 시각적으로도 좋은 성능을 보임을 확인할 수 있다.

1. 서론

사람들은 모든 대상들에 대하여 동일한 비중을 두고 집중하는 것이 아니라, 각 객체에 대해 우선 순위를 매기고 시간 및 노력을 달리 할애한다. 이러한 점에서 컴퓨터 비전 연구 분야에서도 이미지 내의 중요한 객체를 검출하려는 연구가 최근에 많이 진행되고 있다. 중요 객체 검출은 눈에 띄는 영역 또는 객체를 찾는 것을 뜻하는데, 이미지 분할, 분류, 검색, 리타겟팅, 압축 등 이미지 프로세싱, 컴퓨터 비전의 다양한 응용 분야에 활용될 수 있다는 점에서 중요한 연구로 주목받고 있다.

중요 객체를 검출하는 많은 방법들은 bottom-up 접근 방법에 기반하고 있는데, 이는 데이터에 따른 처리 방식이며, 중요 객체에 대한 사전 지식을 검출에 활용하는 것을 의미한다. Bottom-up 방식을 활용하는 방법들은 활용하는 사전 지식에 따라 다양한 방법으로 제안되었다. 중앙 사전(center prior) 지식이란 이미지 중심에 중요 객체가 위치할 가능성이 크다는 사실로서, 기존에 다양한 연구에서 활용되었다 [1]. Yang et al.[2]은 단순히 이미지 중앙이 아니라 객체를 둘러싸는 convex-hull을 찾아 객체의 중심으로 규정하여 중앙 사전 지식을 발전시켜 적용하였다. 이와는 반대로 가장자리 사전 지식은 중요 객체는 이미지의 가장자리에 거의 위치하지 않는다는 점과 가장자리의 영역은 배경 영역과 상당한 유사성이 있다는 점이다 [3, 4, 5]. 한편, 색 대비에 따른 사전 지식이란 사람들이 일반적으로 높은 색 대비가 존재하는 영역에 주목을 한다는 사실로 중요 객체 검출에서 가장 널리 이용되는 특징이다 [1-6].

최근에는 중요 객체를 보다 균일하게 검출하는 방법과 배경 영역을 효과적으로 억제하여 검출하는 방법에 초점을 맞춰 연구가 진행되고 있으며, 그래프를 활용하는 방법이 대표적인 예이다. Yang et al. [2]은 색 대비 및 convex-hull 사전 지식을 활용하여 얻은 중요 객체

맵을 그래프 평탄화 방법에 적용하였고, Jiang et al. [5]은 무작위 행보의 관점에서 접근하였는데, 마르코프 연쇄가 거듭되면서 모든 무작위 행보자가 흡수 노드에 모두 흡수되는데 걸리는 시간을 측정함으로써 중요 객체를 검출한다. Yang et al. [4]은 가장자리 사전 지식을 활용하여 배경에 해당하는 영역을 추출하고, 이와 유사도를 그래프 랭킹 매니폴드 방법을 활용하여 계산함으로써 중요 객체를 검출한다.

본 논문에서는 준 지도 학습을 활용하여 중요 객체 검출의 성능을 높이는 방법을 제안한다. [4]의 방법과는 다르게, 제안하는 방법에서는 가장자리에 의한 사전 지식 뿐만 아니라 색 대비 사전 지식을 동시에 활용한다. 중요 객체 및 배경에 대한 두 가지 종류의 씨앗 노드를 모두 추출한 후, 준 지도 학습에 적용함으로써 검출 성능을 높인다 [7]. 이는 둘 중의 하나의 사전 지식을 활용하는 것보다 효과적이며, 검출 성능을 향상시킬 수 있다. 실험 결과를 통해 제안하는 알고리즘이 높은 재현율 구간에서 기존의 다른 방법들에 비해 높은 정밀도로 중요 객체를 검출하는 것을 확인할 수 있고, 시각적으로도 배경을 효과적으로 억제하면서 중요 객체를 균일하게 검출하는 것을 확인할 수 있다.

본 논문은 다음과 같이 구성이 되어있다. 2장에서는 제안하는 방법에 대해 구체적으로 설명을 하고, 3장에서는 제안하는 방법의 실험 결과와 성능이 뛰어난 최근의 논문들의 실험 결과를 비교하고, 4장에서 결론으로 논문의 끝을 맺는다.

2. 제안하는 알고리즘

준 지도 학습에 기반한 중요 객체 검출 과정은 그림 1과 같이 그래프 생성, 씨앗 노드 추출 그리고 준 지도 학습의 세 가지 단계로 나눌 수 있다. 2장에서는 각 과정을 나누어 제안하는 알고리즘을 구체적으로 기술한다.

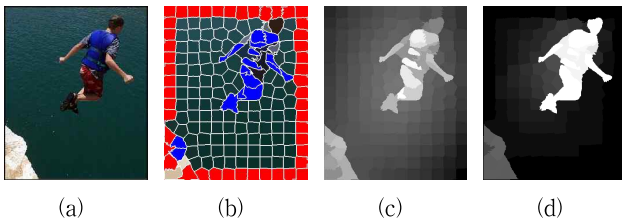


그림 1. 중요 객체 검출 과정의 예. (a) 입력 이미지, (b) 씨앗 노드 추출 : 중요 객체 배경객체 씨앗 노드는 각각 파란색, 빨간색으로 표시, (c) 색 대비[6], (d) 제안하는 알고리즘의 결과.

2.1 그래프 생성

그래프를 생성하기에 앞서, 빠르고 메모리 관점에서도 효율적인 SLIC 방법을 이용하여 이미지를 과 분할된 영역으로 분할한다 [8]. 분할된 영역을 노드(V)로 하고, 이들을 에지(E)로 연결하여 그래프 $G=(V, E)$ 를 생성한다. 노드 v_i 와 v_j 를 연결하는 에지 e_{ij} 는 두 노드간의 유사도를 가중치(w_{ij})로 하여 연결된다. 색상의 유사도를 이용하여 노드간의 유사도를 나타내었는데, 이는 아래의 식 1과 같이 지수 함수 형태로 기술된다.

$$w_{ij} = e^{-\frac{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)}{2}} \quad (\text{식. 1})$$

$$\Sigma = \frac{1}{N(E)} \sum_{e_{ij} \in E} (x_i - x_j)(x_i - x_j)^T \quad (\text{식. 2})$$

여기서 x_i, x_j 는 CIE Lab 색상 공간상의 벡터로, 각 노드의 영역 평균 색상 값을 의미하고, $\Sigma, N(E)$ 는 각각 색상 조절 행렬과 그래프의 모든 에지의 개수를 뜻한다. 고정된 파라미터를 사용하는 기존의 방법들과는 달리 [3, 4, 5], 우리는 이미지에 따라 적응적으로 변하는 Σ 를 사용하였다. 그림 2에서 확인할 수 있듯이, 고정된 값을 사용할 경우에 비해 색 대비가 낮은 이미지에서는 배경 영역을 덜 검출하고, 색 대비가 높은 이미지에 대해서는 중요 객체를 보다 잘 검출한다.

그래프의 노드는 해당 노드와 위치가 인접한 노드끼리만 연결하였고, 예외적으로 가장자리에 위치한 노드는 모두 서로 연결하였다. 이는 지역적인 특성을 고려함과 동시에 가장 자리 노드는 서로 관련성이 깊다는 사실을 이용하였다 [3]. 최종적으로 그래프는 다음과 같이 유사도 행렬로 표현할 수 있다.

$$W = \begin{cases} w_{ij}, & i \in n(j) \text{ or } i, j \in B \\ 0, & \text{otherwise} \end{cases} \quad (\text{식. 3})$$

여기서 $n(j)$ 와 B 는 각각 i 에 인접한 노드와 가장자리 노드이다.

2.2 씨앗 노드 추출

준 지도 학습 방법에 적용을 위한 씨앗 노드 추출을 위해서, 주요 객체 씨앗 노드와 배경 씨앗 노드를 추출하였다. 주요 객체 씨앗 노드를 추출하기 위해서 영역 간 색 대비를 활용하였는데 [6], 이는 각 노드와 다른 모든 노드와의 색 대비를 거리에 가중치에 두어 계산하는 방법이다. 이미지 분할을 그래프 생성 과정에서 수행했기 때문에 조금의 추가적인 계산만으로도 이를 쉽게 계산할 수 있다. 색 대비가 가장 높

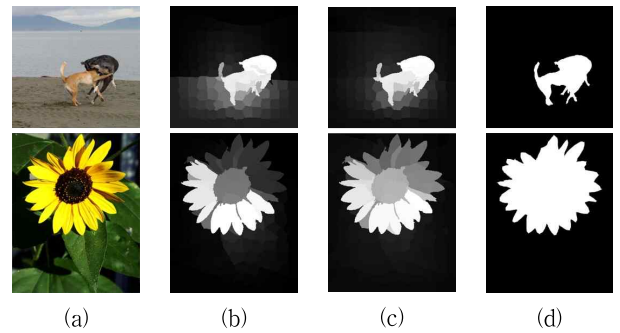


그림 2. 고정된 색상 파라미터와 적응적인 색상 파라미터와의 비교. (a) 입력 이미지, (b) 고정된 파라미터, (c) 적응적인 파라미터, (d) ground-truth.

은 $100 \times \gamma$ %개의 노드를 추출함으로써 간단히 주요 객체 씨앗 노드를 추출할 수 있다. 추출한 노드 중 색 대비가 높은 배경 노드가 포함될 수도 있고, 이는 최종적인 주요 객체 검출 성능에도 영향을 줄 수 있다. 하지만, 이러한 노드들의 수가 크지 않다면, 최종적인 결과에 큰 영향을 주지 않는다.

한편, 배경 씨앗 노드를 추출하기 위해서는 가장자리 사전 지식을 활용하였다. 일반적으로 사진을 찍을 때 중요 객체를 이미지의 가장자리에 위치시켜 취득하는 경우가 거의 없다는 사실로부터 이미지 가장자리를 배경으로 간주하고, 배경 씨앗 노드로 사용하였다. 하지만, 가장자리 노드가 높은 색 대비를 갖게 될 경우에는 중요 객체일 가능성이 높기 때문에 배경 씨앗 노드로 포함하지 않았고, 또한 중요 객체 씨앗 노드에서도 제외시켰다.

요약하면, 중요 객체 씨앗 노드는 가장자리 내부에 위치하는 색 대비가 높은 노드이고, 배경 씨앗 노드는 가장자리에 위치한 색 대비가 낮은 노드이다. 씨앗 노드이면 1, 그렇지 않으면 0의 값을 갖는 씨앗 벡터를 객체 및 배경에 대해 각각 생성한다 (y_s, y_b).

2.3 준 지도 학습

위에서 추출한 주요 객체 및 배경 씨앗 노드를 이용하여 준 지도 학습을 수행하였다. Zhou et al. [7]는 그래프에 기반하여 레이블된 샘플과 레이블되지 않은 샘플을 학습하는 방법을 제안하였는데, 이는 그래프를 학습하는 문제로도 해석할 수 있다 [9]. 학습된 그래프는 다음과 같은 식으로 표현된다.

$$W_L = (1 - \alpha)(D - \alpha W)^{-1} \quad (\text{식. 4})$$

여기서 α 는 학습 균형 파라미터이고, $D = \text{diag}\{d_1, d_2, \dots, d_N\}$ 는 W 의 정도 행렬(degree matrix)로 대각 성분이 $d_i = \sum_j w_{ij}$ 로 구성된 행렬이다. W 의 k 번째 열 w_L^k 은 k 노드와 다른 노드 간의 학습된 유사도를 나타내며, 또한 다음의 에너지 함수를 최소화시키는 벡터로 표현된다.

$$w_L^k = \arg \max_z \sum_{i=1}^N d_i \left| z_i - \frac{c_{ik}}{d_i} \right|^2 + \lambda \sum_{i,j=1}^N w_{ij} |z_i - z_j|^2 \quad (\text{식. 5})$$

여기서 c_{ij} 는 $i=k$ 일 때 1, $i \neq k$ 일 때 0이며, $z = [z_1, \dots, z_N]^T$ 와 λ

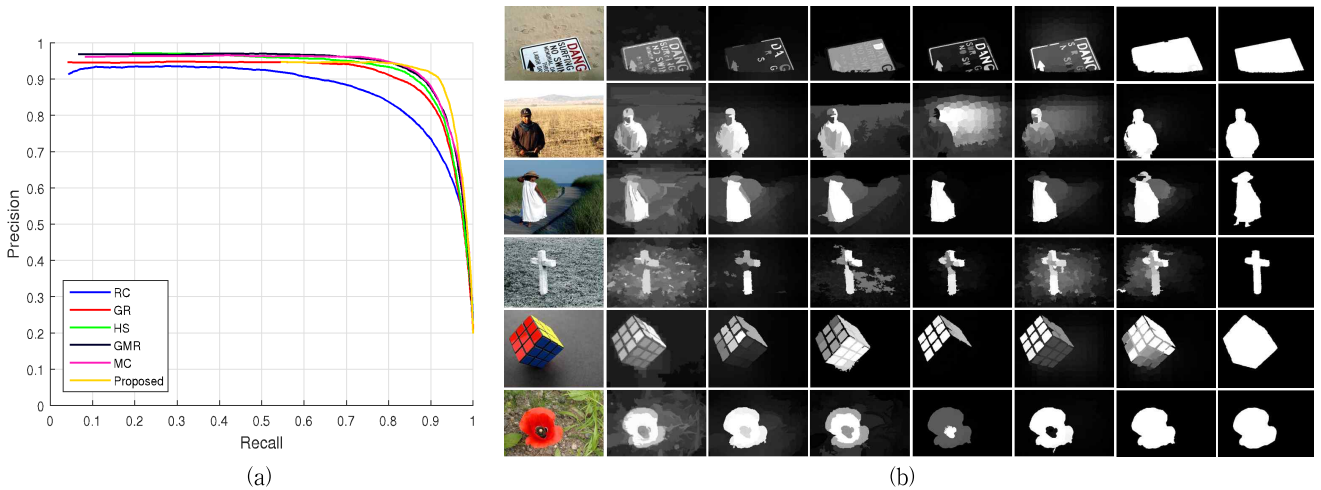


그림 3. 실험 결과 비교. (a) 정밀도-재현율 곡선, (b) 정성적인 비교 (입력 이미지, RC, GR, HS, GMR, MC, 제안하는 알고리즘, ground-truth)

는 각각 유사도 벡터와 균일 파라미터이다. 식 5의 우변 첫 번째 식은 데이터 피팅을 위한 식이고, 두 번째 식은 유사 노드간 균일함을 위한 식이다.

이를 중요 객체 검출 모델에 적용하기 위하여, 씨앗 노드로부터의 전체적인 유사도(overall affinity)를 구하였다. 중요 객체 및 배경 씨앗 노드와의 전체적인 유사도는 각각 아래와 같이 적을 수 있다.

$$f_s = W_L y_s = \sum_{i \in S_s} w_L^i \quad (\text{식. 6})$$

$$f_b = W_L y_b = \sum_{i \in S_b} w_L^i \quad (\text{식. 7})$$

여기서 S_s, S_b 는 각각 중요 객체, 씨앗과 배경 씨앗의 집합이다. 이로부터 최종적으로 중요 객체 정도를 다음과 같이 측정한다.

$$s = (f_s - f_b) / (f_s + f_b) \quad (\text{식. 8})$$

위의 식에서 $/$ 는 벡터의 원소별 나누기 연산자이다. 위 식의 분자는 객체의 중요 정도를 나타내며, 분모는 노드간의 균형을 맞추는 역할을 한다. 최종적으로 중요 객체 정도는 $[0, 1]$ 의 범위로 정규화하여 나타낸다.

3. 실험 결과

실험에는 중요 객체 검출 분야에서 가장 널리 사용되는 ASD 데이터 셋을 사용하였다 [10]. ASD 데이터 셋은 명확한 중요 객체가 존재하는 1,000장의 이미지와 이에 해당하는 1,000장의 ground-truth 이미지로 구성된다. ASD 데이터 셋을 이용하여 성능이 우수하다고 알려져 있는 최신의 방법들과 제안하는 알고리즘을 정량적, 정성적으로 비교하였다.

3.1 실험 환경 설정

실험에는 이미지 과 분할 개수, 씨앗 취득 비율 및 학습 균형 파라미터, 총 3가지의 파라미터가 사용되었다. 이미지를 과 분할하는 개수를 $N=200$ 으로 설정하였는데, 이는 이미지에 따라 SLIC 알고리즘

에 의해 200보다 조금 많거나 적게 분할된다. 또한, 씨앗 취득 비율을 전체 노드의 개수의 $\gamma=0.1$ 배 설정하였으며, 학습 균형 파라미터는 $\alpha=0.99$ 로 설정하였다.

3.3 비교 실험

제안하는 알고리즘을 색 대비 사전 지식 또는 가장자리 사전지식을 활용하는 기존의 다양한 방법들과 비교를 하였다 (RC [6], GR [2], HS [1], GMR [4], MC [5]). RC, GR, HS는 색 대비를 활용한 방법이고, GMR과 MC는 가장자리 사전지식을 활용한 방법이다.

그림 3(a)의 정밀도-재현율 곡선을 보면 제안하는 알고리즘이 낮은 재현율 구간에서는 HS, GMR, MC 보다 낮은 정밀도를 보이고 있다. 이는 제안하는 알고리즘이 주요 객체와 배경에서의 검출값이 크게 차이가 남으로써 발생하는 것으로, 높은 문턱값을 적용했을 경우에도 오검출되는 경우가 생기기 때문에 나타나는 현상이다. 하지만, 중요 객체 검출 방법이 많은 응용 분야에 실제로 활용되는 높은 재현율 구간에서는 기존의 다른 방법들 보다 좋은 성능을 내는 것을 확인할 수 있다.

정량적인 비교뿐만 아니라 정성적인 비교도 수행하였다. RC는 중요 객체를 균일하게 검출하지 못하고, 배경에서도 많은 부분 오검출되는 것을 확인할 수 있다. GR과 GMR은 배경에서 발생하는 오검출은 크게 줄였으나 객체를 덜 검출하는 경향이 있고, HS와 MC의 경우에는 배경의 오검출을 많이 줄였지만, 만족스러운 수준은 아니었다. 이에 반해 제안하는 방법은 이 두 가지 문제점을 어느 정도 만족스럽게 해소시켰다.

4. 결론

이미지를 과 분할하여 그래프를 생성하고, 색 대비 및 가장자리 사전 지식을 활용하여 씨앗 노드를 추출한 후, 이를 준 지도 학습에 적용함으로써 중요 객체를 검출하는 방법을 제안하였다, 두 종류의 씨앗 노드를 모두 활용함으로써 중요 객체 및 배경과의 상대적인 유사 정도를 측정하였다. 높은 재현율 구간에서 기존의 방법들 보다 좋은 성능을 보였는데, 높은 재현율 구간에서 다양한 컴퓨터 비전 응용이 이용된다. 이는 점에서 의미가 있다. 정성적인 결과에서도 보다 중요객체가 균일하게 검출하는 것을 확인할 수 있었다.

참고문헌

- [1] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1155~1162.
- [2] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," IEEE Signal Process. Lett., vol. 20, no. 7, pp. 637~640, 2013.
- [3] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in Proc. Europ. Conf. Comput. Vis., 2012, pp. 29~42.
- [4] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 3166~3173.
- [5] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 1665~1672.
- [6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 409~416.
- [7] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 321~328.
- [8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274~2282, 2012.
- [9] T. H. Kim, K. M. Lee, and S. U. Lee, "Learning full pairwise affinities for spectral segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 7, pp.1690~1703, 2013.
- [10] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp.1597~1604.