

분별력 있는 색상 채널의 선택을 통한 두드러진 영역 검출

채영수*, 김현철, 김희율

한양대학교 전자통신컴퓨터공학과

{yschae, hckim}@vision.hanyang.ac.kr, wykim@hanyang.ac.kr

Salient Region Detection by Distinctive Color Channel Selection

Young-Soo Chae*, HyunCheol Kim, Choi-Yul Kim

Department of Electronics and Computer Engineering, Hanyang University

요약

본 논문에서는 분별력 있는 색상 채널 선택을 통한 두드러진 영역 검출 방법을 제안한다. 제안하는 방법에서는 우선 분별력 있는 색상 채널의 선택을 위해 입력영상을 10개의 색상 채널로 변경하고, 각 채널을 $N \times N$ 블록으로 나눈다. 그리고 각 채널에서 나누어진 N 블록을 외각 블록, 중앙 블록으로 선정하고 중앙-외각 블록간의 대비와 외각 블록의 표준편차 정보를 이용하여 색상 채널 경쟁을 한다. 색상 채널 경쟁을 통해 선별된 K 개의 색상 채널을 이용하여 특징맵을 만들고 이를 조합하여 두드러진 맵을 얻는다. 실험에서는 제안된 방법을 총 1000장의 자연 영상에 적용하여 성능을 평가하였으며, 83%의 평균 정확도를 보임으로써 기존 방법들보다 성능이 뛰어남을 확인하였다.

1. 서론

시각적 주의(Visual attention)는 사람의 시각 시스템 특징 중 하나로써, 사람이 보고 있는 장면 중 중요한 일부만 선택적으로 처리하여 그 장면을 빠르고 정확하게 인지할 수 있도록 도와주는 기술이다. 시각적 주의 기술은 현재까지 그 연구가 활발히 진행되고 있으며, 자동영상 크로핑, 자동 영상분할 등 다양한 분야[1-5]에 응용되고 그 필요성 또한 증가되고 있다.

시각적 주의 기술은 크게 저단계 특징들의 다중 해상도 피라미드 조합을 이용하는 방법[6,7], 입력 영상의 주파수 특징을 이용하는 방법[8,9], 그리고 색상 대비를 이용하는 방법[10,11] 3가지로 나뉘질 수 있다. Itti는 저단계 특징들을 이용하여 다중 해상도 피라미드를 생성하고 특징 통합 이론(Feature integration theory)을 통하여 두드러진 영역을 검출한다[6,7]. Hou는 자연영상의 평균 로그 스펙트럼이 예측 가능한 중복된 특징을 가진다는 점을 이용하여, 중복된 정보를 제거함으로써 영상에서 두드러진 영역을 검출하였다. 최근에는 색상 대비를 이용한 두드러진 영역 검출 방법들이 제안되고 있다[10,11]. 이 방법들은 HSV 또는 Lab 색상 채널의 각 채널 별 명암 대비를 이용하여 두드러진 영역을 검출하는 방법이다. 본 방법은 [6-9] 보다 계산량이 낮아 검출속도가 매우 빠른 장점이 있으며 결과 영상의 해상도가 입력 영상의 해상도와 같기 때문에 두드러진 영역 검출 결과를 응용분야의 사용함에 있어서 별다른 후처리가 필요 없다는 장점이 있다.

본 논문에서는 분별력 있는 색상 채널의 선택을 통한

두드러진 영역 검출 방법을 제안한다. 분별력 있는 색상 채널의 선택을 위해, 먼저 입력영상을 이용하여 10개의 색상채널로 변환한다. 그리고 각 채널을 $N \times N$ 개의 블록으로 나눈 후 각 색상채널의 중앙 블록 평균값과 외각 블록 평균값과의 차이 및 외각 블록들의 표준편차를 구한다. 색상 채널의 분별력은 중앙-외각 블록의 평균값 차이와 외각 블록의 표준편차로 나눈 값으로 정의하며 가장 분별력 있는 K 개의 색상 채널을 이용하여 두드러진 맵을 생성한다. 실험에서는 1000장의 데이터 셋을 이용하여 기존 방법들과 성능 비교하였으며, 실험 결과 제안하는 방법이 기존 방법들 보다 성능이 뛰어남을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 두드러진 영역 검출연구에 대하여 설명하고, 3장에서는 제안하는 방법에 대해 상세히 기술하도록 한다. 4장에서는 제안하는 방법 및 기존 방법에 대한 실험 및 분석 결과를 제시하고, 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

본 절에서는 앞서 서론에서 언급한 기존의 대표적인 연구에 대해 좀 더 상세히 설명하도록 한다.

Itti는 특징 통합이론을 이용하여 두드러진 맵을 생성하고, 승자독식(Winner-take-all) 네트워크로 구성된 방법을 제안하였다[6,7]. 본 방법은 저단계 특징들의 선형조합을 이용하여 입력 영상에서 가장 높은 주의지점을 찾아낸다. 여기서 주의가 가장 높은 주의 지점은 두드러진 맵 결과영상의 누적밝기 값이 가장 높은 지점을 의미

한다. 그 지점을 중심으로 일정한 원으로써 주의 지점을 찾아내고 승자독식 네트워크를 이용하여 주의가 높은 지점에서 주의가 낮은 지점으로 순차적으로 주의가 옮겨지도록 한다.

Hou는 인간 시각 시스템이 영상에서 빈번히 발생하지 않는 변화에 민감하다는 특징을 이용하여 두드러진 영역을 검출하는 방법을 제안하였다[8]. 식 (1)은 자연 영상의 특징을 모델링한 것으로써, 여기서 $H(\text{Innovation})$ 은 새로운 부분을 의미하며, $H(\text{Prior knowledge})$ 는 중복된 정보를 의미한다.

$$H(\text{Image}) = H(\text{Innovation}) + H(\text{Prior Knowledge}). \quad (1)$$

통계적 접근법에서 자연영상은 변하지 않는 통계적인 특성을 가지고 있는데 이것이 중복된 정보이며 이 중복된 부분을 제거함으로써 얻을 수 있는 부분이 영상에서 두드러진 부분이다. Hou가 제안한 Spectrum Residual (SR) 방법은 식 (2)와 같이 로그 스펙트럼과 국부 평균 필터를 통해 변환된 영상과의 차이를 구한다.

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{A}(f) \quad (2)$$

여기서 $\mathcal{L}(f)$ 는 다운 샘플링 된 입력 영상의 로그 스펙트럼이며, $\mathcal{A}(f)$ 는 국부 평균 필터이다. 자연영상의 중복된 정보가 국부 선형특징을 가지고 있기 때문에 국부 평균 필터로 특징을 근사화 할 수 있다. 마지막으로 $\mathcal{R}(f)$ 를 역 푸리에 변환[12]하여 공간영역의 두드러진 맵을 생성한다.

최근에는 색상 대비를 이용하여 두드러진 영역 검출 방법들이 많이 제안되고 있다. Achanta의 방법[11]에서는 두드러진 영역 검출을 위해 가장 먼저 입력 영상을 Lab 색상 공간[13]으로 변환한다. 다음은 각 색상 별 평균값을 구하고 가우시안 필터링[13]된 영상과의 유클리디언(Euclidean) 차이를 통해 특징맵을 생성하고 선형조합을 통해 두드러진 맵을 생성한다. 본 방법은 매우 간단하지만 기존 연구 비교 성능이 매우 뛰어난 장점이 있다. 하지만 입력 영상의 평균 값과 두드러진 영역과의 색상 대비가 적을 경우 성능이 저하될 수 있으며, 영상에서 두드러진 영역 혹은 객체라 하더라도 크기가 클 경우 평균값에 많은 정보가 포함되기 때문에 잘못된 영역 검출결과가 나온다. 또한 색상채널의 평균값과의 편차가 큰 영상에서도 잘못된 영역을 검출하게 된다.

3. 제안하는 방법

본 논문에서는 배경과 관심영역과의 대비를 최대화 시키는 분별력 있는 색상 채널을 선택함으로써 기존의 문제점들을 해결하고자 한다. 그림 1은 제안하는 방법의 전체 흐름도를 보여준다. 본 방법에서는 먼저 입력영상을 이용하여 10개의 색상채널(R, G, B, H, S, V, L, a, b, I)로 변환한다. 변환된 각 색상채널은 $N \times N$ 개의 블록으로 나누는 다음 최외각 $3N$ 블록을 외각 블록(N_2)으로, 나머지 블록을 중앙 블록(N_1)으로 선정한다. 다음은 중앙 블록과 외각 블록의 대비를 구하기 위해 우선 각 블록의 산술평균값을 식 (3)를 이용하여 계산한다.

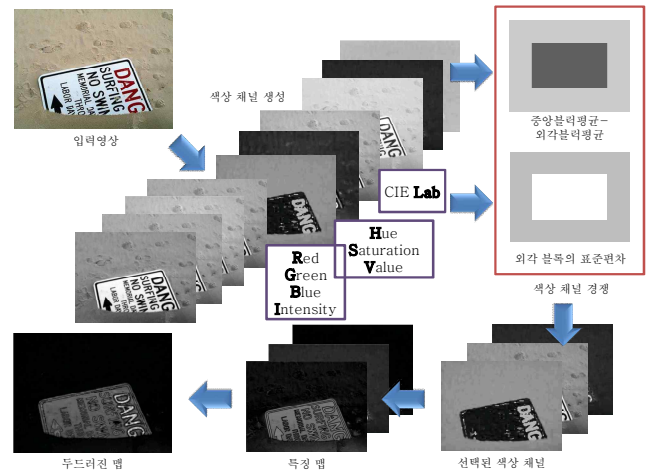


그림 1. 제안하는 방법의 전체 흐름도

$$\mu_{in}^k = \frac{1}{N_1^k} \sum_{x,y \in N_1^k} f^k(x,y), \quad \mu_{out}^k = \frac{1}{N_2^k} \sum_{x,y \in N_2^k} f^k(x,y) \quad (3)$$

여기서, k 와 $f^k(x,y)$ 는 각각 색상 채널, 색상 채널의 (x,y) 위치에서 화소값을 의미한다. 식 (3)을 통해 구해진 평균값의 차이가 클수록 배경과 두드러진 영역을 잘 분별할 수 있는 색상 채널이라 판단할 수 있다. 이는 사람의 시선이 영상의 중앙 부근에 집중할 가능성이 높다는 기존 연구 결과를 반영한 것이다[12]. 하지만 분산값이 큰 복잡한 영상에서는 중앙 블록과 외각 블록의 색상대비가 많이 일어나는 색상 채널을 선택하더라도 배경과 두드러진 영역을 분별하는 데에 문제가 있다. 이와 같은 문제를 해결하기 위하여, 중앙-외각 블록의 대비뿐 아니라 식 (4)의 외각 블록 표준편차 정보를 함께 고려한다.

$$\sigma_{out}^k = \sqrt{\frac{1}{N_2^k} \sum_{x,y \in N_2^k} (f^k(x,y) - \mu_{out}^k)^2}. \quad (4)$$

색상 채널의 분별력은 평균값 차이를 외각 블록의 표준편차로 나눈 값으로 정의하며, 10개의 색상 채널을 식 (5)를 이용하여 가장 분별력 있는 K 개의 색상 채널을 선택한다.

$$D^k = \frac{|\mu_{in}^k - \mu_{out}^k|}{\sigma_{out}^k}. \quad (5)$$

식 (5)의 채널 경쟁을 통해 선택된 색상 채널들은 10개의 색상 채널 중 가장 분별력이 있다고 판단되는 채널들이며, 이 채널들을 이용하여 식 (6)을 통해 각 색상 특징맵을 만든다.

$$F^k(x,y) = |\mu_{out}^k - f_g^k(x,y)| \quad (6)$$

여기서, f_g^k 는 선택된 색상 채널을 가우시안 필터링한 영상이다. 두드러진 영역 혹은 객체가 중앙에 위치할 가능성이 높기 때문에 외각 블록 정보만을 이용하여 특징맵을 만든다. 또한 선택된 색상 채널의 영상을 가우시안 필터링 함으로써 노이즈 제거뿐 영상의 부호화 인공잡음(Coding artifact)을 제거한다. 마지막으로 식 (7)을 통하여 특징맵들의 조합과 $[0,255]$ 정규화를 거쳐서 두드러진 맵 $S(x,y)$ 를 생성하며 결과영상은 그림 (3)과 같다

$$S(x, y) = \sum_{k=1}^K F^k(x, y). \quad (7)$$

두드러진 맵 영상은 흑백영상으로 표현되며 명암도가 높은 영역 일수록 두드러진 영역을 의미한다.



(a) 입력 영상 (b) 결과 영상

그림 2. 두드러진 맵 생성 결과

4. 실험결과

제안하는 방법의 성능 측정을 위한 실험은 Pentium core2 duo 2.66GHz 환경에서 수행되었으며 Visual C++ 2008을 이용하여 구현하였다. 그리고 성능 분석을 위한 실험 영상은 Ancjanta 데이터 셋 [11] 1000장과 실제값 영상을 이용하였다.

제안하는 방법의 성능을 비교 평가하기 위해 기존 방법들 중 Itti 방법[7], SR 방법[8], 그리고 IG 방법[11]을 선정하였다. Itti 방법과 SR 방법에서는 결과영상이 각각 입력영상의 1/256 과 64x64의 고정 해상도로 나오기 때문에 결과영상을 입력영상의 해상도와 동일하게 쌍입방(Bicubic) 보간법[13]으로 보정하였다. 실험을 위해 제안하는 방법에서는 나누어진 블록의 수(N)와 분별력있는 색상채널의 개수 K 를 각각 4, 3으로 설정하였다. 실험에서는 각 방법들의 결과 영상을 [0,255]사이의 이진화 영상으로 만들고 이를 실제값 영상과 비교하여 식 (8)의 정밀도(Precision), 재현율(Recall) 및 정확도(Accuracy)를 계산하였다.

$$\begin{aligned} Precision &= \frac{tp}{tp + fp}, & Recall &= \frac{tp}{tp + fn}, \\ Accuracy &= \frac{tp + tn}{tp + tn + fp + fn}, \end{aligned} \quad (8)$$

여기서, tp 와 fp 는 두드러진 영역이라 판단된 부분이 실제값과 일치하는 경우와 그렇지 않은 경우를 의미하며, fn , tn 은 두드러진 영역이 아니라고 판단된 부분이 실제값과 일치하는 경우와 그렇지 않은 경우를 의미한다.

그림 4와 표 1은 각 방법들의 정밀도-재현율 ROC 곡선과 평균 정확도를 보여준다. Itti의 방법에서는 재현율을 높으나, 배경 부분에서 잘못된 검출결과가 나타나서 낮은 정확도를 기록했다. SR 방법에서는 검출 결과가 두드러진 영역 전체를 표현하지 못하고, 경계영역에서만 나타나기 때문에 정확도는 높으나 낮은 정밀도와 재현율을 기록했다. IG 방법에서는 이전의 방법들 보다 높은 정확도와 재현율을 기록하였지만, 두드러진 영역과 배경과의 색상 대비가 적은 영상에서는 잘못된 영역을 검출 하였다. 또한 전체 평균값을 이용하기 때문에 분산값이 큰 영

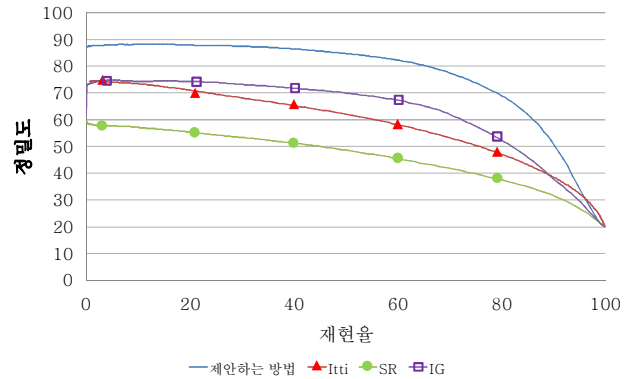


그림 4. 정밀도-재현율 ROC곡선

표 1. 평균 정확도

	Itti [7]	SR [8]	IG [11]	제안하는 방법
평균정확도	69%	75%	79%	83%

상에서는 배경 부분을 두드러진 영역이라 잘 못 판단하였다. 제안하는 방법에서는 분별력있는 색상 채널을 선택함으로써 기존 색상대비를 이용한 검출 방법의 문제점들을 개선하였고, 88%의 정밀도와 83%의 평균 정확도를 기록하며 기존 방법들보다 우수함을 증명하였다.

그림 5는 기존 방법들과 제안하는 방법의 두드러진 영역검출 결과의 예를 보여준다. Itti 방법은 대부분 실험영상에서 두드러진 영역 이외에 배경 부분에서도 높은 밝기값이 많이 검출되는 것을 확인할 수 있다. SR 방법은 영역 검출의 정확도는 높으나 검출결과가 영역 전체를 일정한 밝기값으로 표현을 못하며, 4, 5행의 입력영상과 같이 객체의 텍스처가 배경보다 약한 경우 배경 부분을 두드러진 영역이라 잘 못 판단하는 문제가 있었다. IG 방법은 고정된 색상 채널에서의 대비만을 이용하기 때문에 2, 4행의 입력영상과 같이 배경과 두드러진 영역과의 색상대비가 적은 경우 검출성능이 저하된다. 그리고 5-7행의 입력영상과 같이 두드러진 영역 혹은 객체의 크기가 큰 경우 관심영역의 색상 정보가 전체 평균값에 많이 분포가 되어서 배경 영역을 두드러진 영역으로 잘 못 검출하게 된다. 제안하는 방법에서는 배경과 두드러진 영역과의 색상대비가 적은 영상(2-4행), 두드러진 영역 혹은 객체의 크기가 큰 영상(5-7)에서도 기존 방법들보다 높은 검출결과를 보여주며 두드러진 영역 전체를 일정한 밝기 값으로 표현하는 것을 확인할 수 있다.

하지만 본 방법에서는 색상 대비만을 이용하기 때문에 8행의 입력영상 같이 배경부분이 두드러진 영역과 비슷한 색상 명암도를 가지는 영상에서는 정확도가 낮게 나타났다. 또한 9행의 입력영상과 같이 색상 대비는 약하지만 모양이나 텍스처의 대비가 일어나는 영상에서는 두드러진 영역을 잘 못 검출하게 된다.

입력영상	Itti [7]	SR [8]	IG [11]	제안하는 방법

그림 5. 기존 방법들과 제안하는 방법의 영역검출 결과

5. 결론

본 논문에서는 분별력 있는 색상 채널의 선택을 통한 두드러진 영역 검출 방법을 제안하였다. 제안하는 방법에서는 우선 분별력 있는 색상 채널의 선택을 위해 입력영상을 다수의 색상 채널로 변경하고, 각 채널을 $N \times N$ 블록으로 나눈다. 그리고 각 채널에서 나누어진 N 블록을 외각 블록, 중앙 블록으로 선정하고 중앙-외각 블록간의 대비와 외각 블록의 표준편차 정보를 이용하여 색상 채널 결정을 한다. 색상 채널 경쟁을 통해 선별된 K 개의 색상 채널을 이용하여 특징맵을 만들고 이를 조합하여 두드러진 맵을 만든다.

실험에서는 Anchanta 데이터 셋 1000장과 실제값 영상을 이용하여 기존 방법들과 제안하는 방법의 두드러진 영역 검출결과를 비교하였다. 실험 결과 기존 방법들보다 높은 정밀도와 재현율을 기록하였고, 83%의 평균 정확도

로써 기존 방법보다 더욱 효율적으로 두드러진 영역 검출하는 것을 증명하였다.

참조 문헌

- [1] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," *In CHI 06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 771-780, 2006
- [2] Y. Fu, J. Cheng, Z. Li and H. Lu, "Saliency Cuts: An Automatic Approach to Object Segmentation," *International Conference on Pattern Recognition*, 2008
- [3] S. Avidan and A. Shamir, "Seam Carving for Content-Aware Image Resizing," *ACM Transactions on Graphics*, Vol. 23, SIGGRAPH 2007
- [4] M. Rubinstein, A. Shamir and S. Avidan, "Improved Seam Carving for Video Retargeting," *ACM Transactions on Graphics*, Vol. 27, SIGGRAPH 2008
- [5] W. H. Chen, C. W. Wang and J. L. Wu, "Video Adaptation for Small Display Based on Content Recomposition," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 17, 2007
- [6] L. Itti and C. Koch, "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention," *Vision Research*, Vol. 40, 2000
- [7] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 1254-1259, 1998
- [8] X. Hou and Zhang, L, "Saliency Detection: A Spectral Residual Approach," *Proceeding of 2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007
- [9] Z. Wang and B. Li, "A Two Stage Approach To Saliency Detection in Images," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 965-968, 2008
- [10] R. Achanta, F. Estrada, P. Wils and S. Susstrunk, "Salient Region Detection and Segmentation," *International Conference on Computer Vision Systems*, Vol. 5008, pp. 66-75, 2008
- [11] R. Achanta, S. Hemami, F. Estada, and S. Susstrunk, "Frequency-tuned Salient Region Detection," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1597-1604, June 2009
- [12] T. Judd, K. Ehinger, F. Durand and A. Torralba, "Learning to Predict Where Humans Look," *International Conference on Computer Vision*, 2009
- [13] R. C. Gonzalez and R. E. Woods. *Digital Image Processing 2/E*. Prentice Hall, 2001, 2nd edition, 2002