

관심영역 검출을 위한 상향식 현저함 모델 기반의 선택적 주의 집중 연구

김종배
서울디지털대학교 컴퓨터공학부
e-mail:jbkim@sdu.ac.kr

Detection of ROIs using the Bottom-Up Saliency Model for Selective Visual Attention

Jong-Bae Kim
Dept of Computer Engineering, Seoul Digital University

요 약

본 논문은 상향식 현저함 모델을 이용하여 입력 영상으로부터 시각적 주의를 갖는 영역들을 자동으로 검출하는 방법을 제안한다. 제안한 방법에서는 인간의 시각 시스템과 같이 사전 지식 없이 시각정보의 공간적인 분포에 근거하여 장면을 해석하는 상향식 현저함 모델 방법을 입력 영상에 적용하여 관심 물체 영역을 검출하는 연구이다. 상향식 현저함 방법은 Treisman의 세부특징이론 연구에서 제시한 바와 같이 시각적 주의를 갖는 영역은 시각정보의 현격한 대비차이를 가지는 영역으로 집중되어 배경에서 관심영역을 구분할 수 있다. 입력 영상에서 현저함 모델을 통해 3차원 현저함 맵을 생성한다. 그리고 생성된 현저함 맵으로부터 실제 관심영역들을 검출하기 위해 제안한 방법에서는 적응적 임계치 방법을 적용하여 관심영역을 검출한다. 제안한 방법을 관심영역 분할에 적용한 결과, 영역 분할 정확도 및 정밀도가 약 88%와 89%로 제시되어 관심 영상분할 시스템에 적용이 가능함을 알 수 있다.

1. 서론

최근 들어 보안, 모니터링, 감시 등의 자동적인 수행 요구에 따라 CCTV의 급속한 설치 요구가 증대하고 있다. 이로 인해 방대한 양의 기록된 영상물이 저장되고 있으며 이를 검색하기 위해 많은 시간이 소요되고 있다. 이처럼 입력 영상에서 관심 물체를 검출하기 위해서는 인간의 시각적인 방법을 실제 검출을 위한 시스템에 적용한다면 빠른 시간에 물체를 검출할 수 있을 것이다. 인간의 시각적 주의 집중 방법은 입력된 모든 시각정보를 처리할 수 없어 선택적 주의 집중 방법으로 처리한다. 인간의 시신경에 수백만개의 수용기가 자극되고 그 수용기의 신호들은 망막 밖으로 나와 배외측 슬상핵(lateral geniculate nucleus:LGN)과 피질로 전달한다[1]. 인간의 시각시스템은 망막으로부터 전달된 신호가 방대하여 모든 신호를 처리하게 된다면 시신경의 급격한 과부하가 발생하게 된다. 따라서 인간의 시각 시스템은 처리하고 분석할 수 있는 시각 정보들을 선택적으로 처리하도록 설계되어 있어 관심이 있는 영역들에 집중할 수 있다. 선택적으로 영역을 집중하는 방법으로는 작업에 관련된 정보, 환경, 지식 등에 의해 좌우 될 수 있을 것이다. 이처럼 일반적인 관심물체 검출 연구에서도 대상에 관련된 물체에 대해 정보들을 사전에 학습하는 과정이 필요하다. 하지만, 사전에 관심 대

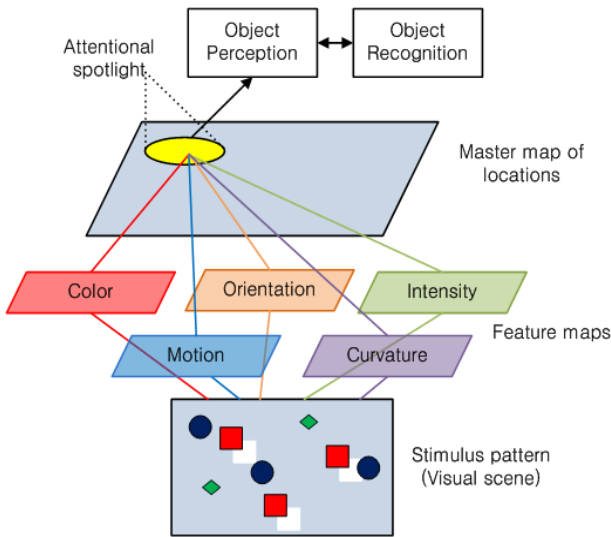
상에 대해 학습하는 과정은 관심 대상의 수와 다양한 환경에 따라 많은 학습 시간이 소요되는 것이 일반적이다. 이를 해결하기 위한 방법으로 관심 대상에 대한 사전 지식없이 입력된 시각 정보들의 조합으로 관심 대상을 구분할 수 있는 방법이 있다면 관심물체 검출에 상당한 처리 비용을 줄 수 있을 것이다.

따라서 본 연구에서는 상향식 현저함 모델을 사용하여 관심 대상에 관한 사전지식 없이 입력 영상으로부터 관심영역을 검출하는 방법을 제안한다. 제안한 방법을 관심 대상에 대한 어떠한 정보도 요구하기 않기 때문에 간단하고 빠르게 처리할 수 있는 방법이다. 입력 영상들에서 다양한 시각 정보들의 대비차이를 계산하고 차이 값이 두드러지는 정보를 나타낸 3차원 지도 형태로 결과를 제시한다. 그리고 3차원 현저함 지도에서 실제 관심영역을 검출하기 위해 적응적 임계치 방법[2]을 적용하여 기준값이 이상인 현저함 영역들을 검출하는 방법을 제안한다.

2. 현저함 모델(Saliency model)

기존에 시각 장면에서 유용한 시각 정보들만을 선별하여 시각적 주위에 대한 연구는 크게 2부류 나누어져 있다 [3]. 첫 번째는 시각적 주위가 장면에 포함되어 있는 다양한 정보들의 사전 지식 없이 공간적인 분포에 근거하여

장면을 해석하는 상향식 방식이다. 이 방식은 어떤 장면이 다양한 시각 정보들의 기본 조합으로 이루어져 있다는 그림 1과 같은 Treisman의 세부특성이론 연구[4]에 따라 시각적 주의를 갖는 영역은 시각 정보들의 조합으로 집중될 수 있는 방식이다. 두 번째는 시각 주의 방식이 사전 경험을 통해 알고 있는 주어진 지식 기반의 혹은 작업 혹은 목표 기반의 장면을 해석하는 하향식 방식이다. 이 방식은 주어진 작업에 관련된 사전 지식, 예를 들어 작업과 연관된 물체의 색상 혹은 모양, 질감 등을 이용하는 것이 더 빠른 영상 해석이 가능하다는 선호에서 시작한다.



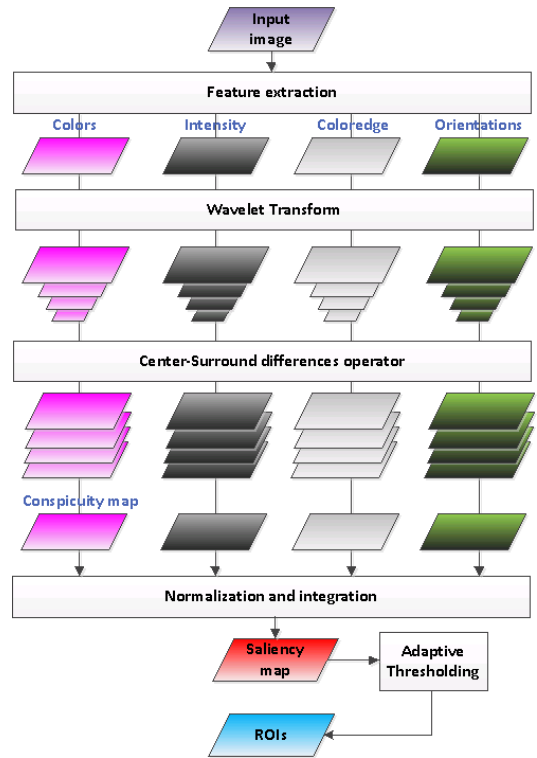
(그림 1) Treisman의 세부특성이론[4]

3. 제안 연구

제안한 연구는 입력 영상에서 칼라, 밝기, 칼라에지, 방향성 정보들을 이용하여 시각적 주의를 갖는 현저함 지도를 생성하고 적응적 임계치 방법을 적용하여 관심영역을 검출한다. 제안한 관심영역 검출 과정은 그림 2와 같다. 그림 2(가)는 Itti et al., [5]이 제안한 computational model을 바탕으로 관심 물체에 고유칼라에 대한 특징 지도와 웨이블릿 변환을 사용한 다해상도 영상 표현 방법을 적용하여 현저함 지도를 생성하였다. 그리고 생성된 3차원의 현저함 지도에서 적응적 임계치 방법을 적용하여 적응적 임계치 이상이 영역들만을 관심영역으로 검출한다. 제안한 방법은 입력 영상의 다양한 특징 정보들의 조합을 사용한 상향식 시각 집중 방식이다.

가. 특징 지도

입력 영상에서 관심 영역들을 검출하기 위해 영상의 특징정보들의 분석이 필요하다. 입력 영상에서 관심을 갖는 영역들은 일반적으로 배경과 비교하여 두드러진 색상



(그림 2) 제안한 모델

과 질감으로 구성되어 있는 특징을 가지고 있다. 따라서 이러한 특징을 잘 표현할 수 있는 특징 정보들을 선택하는 것이 중요하다. 이러한 특징정보들은 다양한 조건 하에서도 관심영역들을 잘 묘사할 수 있는 칼라나 모양과 같은 global 특징정보이다. 본 연구에서는 입력 영상에서 RGB 칼라 색상, 밝기, 에지, 방향성 정보들의 대비차이를 이용하여 특징 지도를 생성한다. 입력된 rgb 색상공간은 여러 색상공간에 겹쳐서 표현된 색상공간이다. 따라서 순수한 색상을 확보하기 위해 적색은 $R=r-(g+b)/2$, 녹색은 $G=g-(r+b)/2$, 푸른색은 $B=b-(r+g)/2$, 그리고 노란색은 $Y=(r+g)/2-|r-g|/2-b$ 의 순수한 색상공간으로 표현하고 음수 색상값은 0으로 설정한다. 이때 입력 영상에 대한 칼라 특징정보는 인간의 시각 시스템에서 색상간의 대비 차에 강하게 반응하는 점을 이용하여 반대 색상끼리 대비차를 가지게 한 색상 특징정보를 입력으로 사용한다. 즉, 적색과 녹색 그리고 청색과 황색과 같은 반대 색상공간간의 대비차이를 색상 특징지도 생성에 사용한다. 그리고 칼라에지 특징 지도를 검출하기 위해 한 픽셀이 RGB 칼라값의 3차원 벡터라 간주하고, 이 벡터의 최대 기울기 값을 계산함으로써 에지의 강도를 검출할 수 있다. 여기서 칼라 벡터의 최대 기울기값을 계산하기 위해 Jacobian 행렬의 가장 큰 고유값(eigenvalue)을 계산함으로써 가능하다. 그리고 방향성 특징 지도는 입력 영상의 0, 45, 90, 135도 각도로 회전한 영상의 특징 지도를 생성한다.

나. 중요도 지도

특징 지도로부터 중요도 지도를 생성하기 위해 본 단계에서 Haar 웨이블릿 변환을 이용하여 다해상도 특징 지도로 표현한다. 웨이블릿 변환에 의한 영상의 특징 추출 방법은 영상을 주파수 대역별로 분할하고, 분할된 부 밴드들을 이용하여 한 영상을 대표하는 특징 벡터를 얻는다. 분할된 저대역(lowpass) 부밴드들은 원 영상이 가지는 에너지의 대부분을 소유하고 있으며, 그리고 고대역(highpass) 부밴드(sub-band)는 집중도가 떨어지지만 인간 시각에 민감한 영상의 에지 정보를 보유하고 있다[6]. 본 단계에서는 특징 지도를 입력으로 2-level 웨이블릿 변환을 통해 저대역 부밴드들을 사용하여 다해상도 영상을 표현한다. 그리고 다해상도 특징 지도를 구성하고 있는 각 픽셀들이 주변 픽셀들과 가진 차이를 계산하여 중요도 지도를 생성한다. 주변과 비교하여 두드러진 픽셀들을 검출하기 위해 중심차분연산(Center-surround difference operation)을 이용한다[5]. 우선 서로 다른 크기의 다해상도 영상을 bilinear interpolation을 통해 같은 크기로 정규화한다음 중심차분연산을 두드러진 픽셀을 검출하였다. 그리고 검출된 픽셀들을 강화하고 그 외 픽셀들은 제거하기 위해 L. Itti와 C. Koch의 global-mean local maxima 비선형 정규화[5]를 수행하였다. 하나의 특징정보에 해당하는 다해상도 특징 지도로부터 하나의 중요도 지도를 획득하기 위해, 크기 및 값이 정규화된 특징 지도들 간의 평균 교차 합 계산을 통해 중요도 지도를 획득한다.

다. 현저함 지도

중요도 맵들에 포함된 특징값에 의해 현저함 지도를 생성한다. 중요도 지도는 영상에 포함된 특징정보들이 두드러지게 나타난 영역을 중요한 영역으로 간주하여 표시한 3차원 지도이다. 따라서 각 중요도 지도는 서로 다른 특징정보들 기반으로 독립적으로 생성되었기에 이를 현저함 지도에 표현할 경우 상당히 많은 영역들이 현저함 영역으로 표시될 수 있다. 따라서 본 단계에서는 각각의 중요도 지도에 가중치를 두어 합하는 방법으로 현저함 지도를 생성한다. 그리고 생성된 현저함 지도의 높이 정보가 가우시안 분포를 따른다는 가정하여 두 개의 가우시안 분포가 만나는 지점을 임계치 값으로 설정하여 임계치 값 이상인 영역들을 관심 영역으로 구분한다.

4. 실험 결과

제안한 방안을 평가하기 위해서 마이크로소프트사의 MSRA salient object 데이터베이스 영상들을 실험에 사용하였다[7]. 실험데이터베이스에는 원영상과 3명의 사람이 영상에서 시각적 주의를 갖는 물체가 있는 영역을 포함하는 시각형 위치를 표시하였다. 제안한 실험에서는 3명의 위치 값들 중 최대 및 최소값을 제외한 가운데 값을 선택을 시각적 주의 영역을 표시한 위치로 간주하고 다른 연

구자들의 결과 위치와 비교하였다. 성능 비교 기준은 MS사의 레이블된 결과 영상이 제안한 검출 결과와 다른 연구자들의 검출 결과를 식(1)을 사용하여 비교하였다. 식(1)은 정확도(precision)와 recall(정밀도)를 측정하여 비교 평가 하였다.

$$P = \frac{\sum_x g_x a_x}{\sum_x a_x}, R = \frac{\sum_x g_x a_x}{\sum_x g_x}, a_x, g_x \in \{1, 0\} \quad (1)$$

여기서 a_x 는 x 위치에 시각적 주의 영역이 존재하는지를 의미하고 g_x 는 ground-truth 방법에 의해 검출된 시각적 주의 영역을 의미하는 이진 마스크이다. 표 1에서 실험 영상(실험1: 산 영상, 실험2: 동물 영상, 실험3: 합성 이미지)에서 관심물체 검출 성공(DA), 검출 실패(DF), 그리고, 오인식(FA)를 측정하였다. 검출 성공은 ground-truth 방법과 제안한 방법에서 두 영역간의 정확도가 20% 오차범위 내에서 동일하게 검출한 경우, 검출 실패는 ground-truth 방안의 검출 결과와 달리 해당 영역을 검출하지 못한 경우, 그리고 오인식은 픽셀의 겹침 정도가 80% 이상 서로 다른 픽셀 영역을 검출한 경우로 한정하였다. 실험한 결과, 시각적 주의영역 검출 성공률이 약 88.4% 이상을 차이를 나타내었고 오인식은 주로 배경이 복잡하거나 배경과 다른 색상을 가진 질감 영역이 존재하는 경우가 관심 영역으로 오인식되는 경우가 발생하였다. 정밀도의 경우는 검출 성공이 발생한 영역들에서 ground-truth 방안에 검출된 영역들 간의 픽셀들의 겹침 정도를 측정하였다. 실험결과 89%의 검출 정밀도를 제시하였다. 그림 3은 시각적 주의영역 검출의 성능 비교를 위해 IT[5], GA[8], FT09[9], 제안한 방법과 실험 DB에서 사람이 표시한 관심물체 영역들과 각 방법들의 의해 검출된 영역들을 비교한 결과이다. 실험 결과, 다른 검출 방법들의 대부분은 관심영역을 검출하였으나 그 실제 물체 영역들과의 정확도 면에서 약 16% 정도 떨어지는 경향이 발생함을 알 수 있다.

<표 1> 정확도와 정밀도 실험 결과

| 실험 이미지 (%) | 정확도 | | | 정밀도 |
|-------------------|-------|-------|-------|-----|
| | DA | DF | FA | |
| 이미지 수 / 관심물체 수 | | | | |
| 실험 1 23 / 41 | 38 | 3 | 7 | 91% |
| 실험 2 27 / 32 | 26 | 6 | 6 | 87% |
| 실험 3 18 / 22 | 20 | 2 | 3 | 89% |
| 평균 | 88.4% | 11.5% | 16.8% | 89% |

5. 결론

본 논문에서는 입력 영상에서 시각적 주의를 갖는 영역을 자동적으로 검출하는 방법을 제안한다. 제안한 방법에서는 상향식 현저함 모델을 사용하여 현저함 지도를 생성하고 이 현저함 지도에서 적응적 임계치 방법을 적용하여 임계치 값을 사용하여 현저함이 높은 영역들을 검출한다. 제안한 방법을 실험한 결과, 관심물체 영역의 검출 정확도는 약 88%, 그리고 정밀도는 약 89%의 성능을 제시하였다. 이러한 결과를 바탕으로 제안한 방법은 영상 압축, 영상 분할, 물체 추적 등의 시스템에 적용할 수 있을 것이다. 실험 환경에서는 MS사의 실험 데이터베이스 일부만을 사용하여 실험한 결과로써 제안한 방법의 전체 성능을 대표할 수 없는 문제점이 있다. 따라서 향후 연구에서는 일반 실험 이미지 하에도 실험을 실시하도록 하고 MS사의 전체 이미지 DB에서도 실험을 실시하도록 할 것이다.

참고문헌

[1] Butko, N. J., Lingyun Z., Cottrell. G. W. Movellan, J. R., "Visual saliency model for robot cameras", Proc. of IEEE Int. Conf. on Robotics and Automation, pp.2398-2403, 2008

[2] Kim J. B. and Kim H. J., "Efficient Region-Based Motion Segmentation for a Video Monitoring System", Pattern Recognition Letter, Vol. 24, No. 1, pp. 113-128, 2003.

[3] Joel B. and Pedro, P., "Visual attention driven image to video adaptation", Graduation Projects, Technical University of Lisbon, Portugal, 2005.

[4] Treisman A. E., "A feature-integration theory of attention", Cognitive Psychology, Vol. 12, pp. 97-136, 1980.

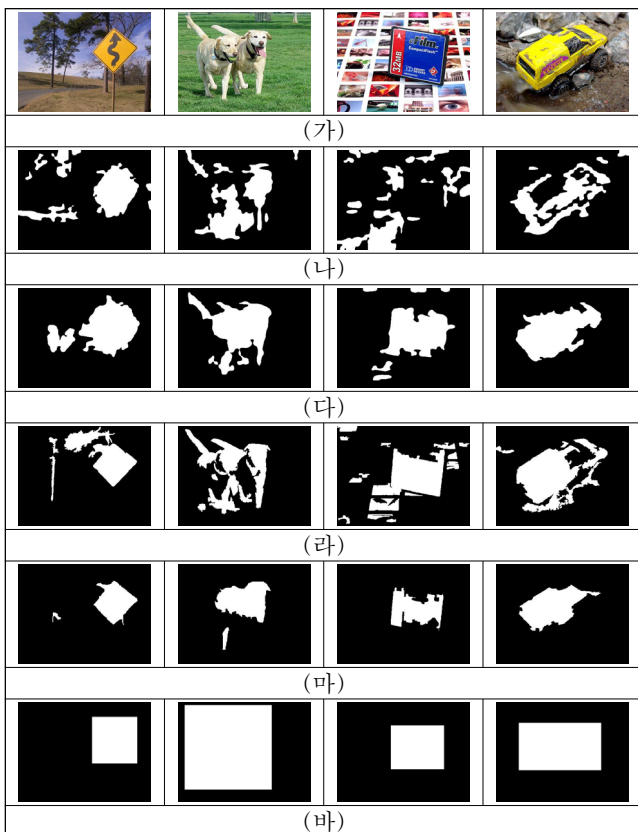
[5] Itti L., Koch C., Niebur E., "A model of saliency-based visual attention for rapid scene analysis", IEEE Trans. on PAMI, Vol. 20, No. 11, pp. 1254-1259, 1998.

[6] Kim J. B. and Kim H. J., "Multiresolution-Based Watersheds for Efficient Image Segmentation", Pattern Recognition Letter, Vol. 24, No. 1, pp. 473-488, 2003.

[7] MSRA Salient Object Database: http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient_object.htm

[8] Harel J., Koch C., Perona P., "Graph-based visual saliency", Advances in Neural Information Processing Systems, Vol. 19, pp.545-552, 2007.

[9] Achamta R., Hemami S., Estrada F., Susstrunk S., "Frequency-tuned salient region detection", Proc. of IEEE conf. CVPR., pp.1597-1604, 2009.



(그림 3) 관심영역 검출 결과

Acknowledgment

This research is supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology(2011-0026551) in 2011.