

논문 2014-51-7-18

돌출영역 분할을 위한 대립과정이론 기반의 인공시각집중모델 (An Artificial Visual Attention Model based on Opponent Process Theory for Salient Region Segmentation)

정 기 선*, 홍 창 표*, 박 동 선**

(Kiseon Jeong, Changpyo Hong, and Dong Sun Park[©])

요 약

본 논문에서는 자연영상에 대한 돌출영역을 자동으로 검출하고 이를 분할하기 위한 새로운 인공시각집중모델을 제안한다. 제안된 모델은 인간의 생물학적 시각인지 기반이며 주된 특징은 다음과 같다. 먼저 영상의 강도특징과 색상특징을 사용하는 대립과정이론 기반의 새로운 인공시각집중모델의 구조를 제안하고, 돌출영역을 인지하기 위해 영상의 강도 및 색상 특징채널의 정보량을 고려하는 엔트로피 필터를 설계하였다. 엔트로피 필터는 높은 정확도와 정밀도로 돌출영역에 대해 검출 및 분할이 가능하다. 마지막으로 최종 돌출지도를 효율적으로 구성하기 위한 적응 조합 방법 또한 제안되었다. 이 방법은 각 인지 모델로부터 검출된 강도 및 색상 가시성지도에 대하여 평가하며 평가된 점수로부터 얻어진 가중치를 이용해 가시성 지도들을 조합한다. 돌출지도에 대해 ROC분석을 이용한 AUC를 측정한 결과 기존 최신의 모델들은 평균 0.7824의 성능을 나타낸 반면 제안된 모델의 AUC는 0.9256으로서 약 15%의 성능 개선을 보였다. 또한 돌출영역 분할에 대해 F-beta를 측정한 결과 기존 최신의 모델은 0.5178이고 제안된 모델은 0.7325로서 분할 성능 또한 약 22%의 성능 개선을 보였다.

Abstract

We propose an novel artificial visual attention model that is capable of automatic detection and segmentation of saliency region on natural images in this paper. The proposed model is based on human visual perceptions in biological vision and contains there are main contributions. Firstly, we propose a novel framework of artificial visual attention model based on the opponent process theory using intensity and color features, and an entropy filter is designed to perceive salient regions considering the amount of information from intensity and color feature channels. The entropy filter is able to detect and segment salient regions in high segmentation accuracy and precision. Lastly, we also propose an adaptive combination method to generate a final saliency map. This method estimates scores about intensity and color conspicuous maps from each perception model and combines the conspicuous maps with weight derived from scores. In evaluation of saliency map by ROC analysis, the AUC of proposed model as 0.9256 approximately improved 15% whereas the AUC of previous state-of-the-art models as 0.7824. And in evaluation of salient region segmentation, the F-beta of proposed model as 0.7325 approximately improved 22% whereas the F-beta of previous state-of-the-art models.

Keywords : entropy filter, salient region detection, salient region segmentation, opponent process theory

* 학생회원, ** 정회원, 전북대학교 전자공학과

(Department of Electronic Engineering, Chonbuk National University)

© Corresponding Author(E-mail: dspark@jbnu.ac.kr)

※ 본 연구는 미래창조과학부 및 정보통신산업진흥원의 산학협력 특성화 지원사업의 연구결과로 수행되었음 (NIPA-2013-H0803-13-1003).

접수일자: 2014년02월03일, 수정일자: 2014년06월02일, 수정완료: 2014년07월09일

I. 서 론

일상생활에서 인간의 시각 시스템은 막대한 양의 시각 데이터($10^8 - 10^9$ bits)를 매초마다 눈을 통해 받아들인다. 입력된 시각 데이터 정보는 뇌로 전달되고 이를 처리함으로써 인간이 이해할 수 있다. 이러한 시각 데이터 정보의 전달과정과 처리과정에는 각각 실시간 처리에 효율적인 메커니즘이 적용된다^[1~2].

먼저 첫 단계인 전달과정에서는 입력된 시각 데이터 정보들을 뇌로 전달한다. 이 때 망막(retina)의 세포중 하나인 신경절 세포(ganglion cell) 무리들에 의해 병렬적으로 전달됨으로써 실시간 전달을 달성할 수 있다^[3]. 다음으로 눈의 망막은 포빌(foveal), 포비아(fovea), 파라포빌(parafoveal), 페리포빌(perifoveal)의 4가지 영역으로 구분되며 각각의 영역은 객체검출(object detection)과 객체인식(object recognition)에 있어 중요한 역할을 수행한다. 그리고 망막의 각 영역들은 망막의 중심에서 바깥으로 갈수록 시각 데이터의 해상도는 지수(exponential)의 형태로 감소한다^[4]. 이러한 눈의 구조는 처리과정에서 입력된 시각 데이터의 중요한 정보 혹은 눈에 띄는 정보는 자세히 관찰하는 반면 그 이외의 데이터는 함축된 관찰을 함으로써 효율적으로 데이터를 처리할 수 있다.

이러한 메커니즘은 인간의 시각 시스템뿐만 아니라 컴퓨터비전 및 로봇틱스의 인공시각 분야에서도 필요하다. 특히 객체 인식이나 객체추적(object tracking)과 같은 응용분야에서는 CCD(charge coupled device)카메라를 통해 입력된 많은 양의 영상 데이터에 대한 실시간 처리가 요구된다. 이에 따라 지난 수 십 년 동안 이를 효과적으로 해결하기 위해 컴퓨터비전 및 로봇틱스의 인공시각 분야의 연구자들은 데이터 처리의 메커니즘인 돌출(saliency)라는 개념을 모방하였다. 돌출이란 영상에서 관찰자에게 주변 환경에 비해 눈에 띄는 객체 또는 영역을 의미하며 이를 검출하기 위한 시스템 또는 모델을 인공시각집중모델(artificial visual attention model)이라 한다^[5~6].

또한 돌출을 이용한 메커니즘은 데이터의 실시간 처리뿐만 아니라 돌출영역을 관심영역으로서 사용하여 데이터를 처리함으로써 낮은 에러율을 달성할 수 있다. 이와 같이 시각적으로 독특한 돌출정보는 영상분할(image segmentation)^[7], 객체 검출(object detection)

^[8~10], 객체 인식(object recognition)^[11~13], 객체 추적(object tracking)^[14~15], 로봇 지역화(robot localization)^[16~17], 그리고 로봇 항해술(robot navigation)^[18~19] 등과 같은 응용분야에서 매우 유용하게 사용된다. 이 중에서 본 논문에서는 돌출영역에 대한 영상분할을 위한 인공시각집중모델을 제안한다.

인간의 시각 시스템은 장면(scene)이나 영상(image)에 대해 관찰할 시 자유 관찰(free viewing) 또는 탐색 관찰(search viewing)을 한다. 만약 영상에서 사람이나 동물, 차량 등 탐색에 대한 사전정보가 주어진다면 탐색 관찰이며 이와 반대는 자유 관찰이다. 이를 모방하여 인공시각집중모델은 상향식(bottom-up)모델과 하향식(top-down)모델로 구분된다. 상향식모델은 주어진 영상의 강도(intensity), 색상(color), 방향(orientation), 질감(texture) 등과 같은 영상 특징(stimulus driven) 기반의 모델이며 하향식모델은 사람의 얼굴(face), 보행자(pedestrian), 문자(text), 차량(car) 등에 대한 검출기를 이용한 특정 업무(task driven) 기반의 모델이다.

기존의 상향식모델은 필터에 따라^[20~25] 구분할 수 있다. 이러한 기존의 인공시각집중모델은 영상의 돌출영역에 대해 검출이 가능하지만 영역 전체를 검출하지 못하는 단점이 있기 때문에 영상 분할에는 적합하지 않다. 예를 들어 그림 1에서 대부분의 돌출지도는 지상검증자료(GT - ground truth)와 비교했을 때 돌출영역 전체를 검출하지 못하고 에지(edge)영역만 검출된다. 이 외에도 돌출지도의 품질에 대한 단점으로서 다중 스케일 분석(multi-scale analysis)으로 인한 번짐 효과(blurring effect), 배경과 비슷한 영상 특징으로 인한 돌출영역의 모양왜곡(shape distortion) 그리고 돌출영역에 대하여 실시간으로 검출(real-time detection)이 가능해야 한다.

이러한 단점을 극복하기 위해 본 논문에서는 인간시각인지능력 기반의 상향식모델을 제안하며 이는 대립과정이론(opponent process theory)^[26]에 따라 영상의 강도특징과 색상특징을 사용한다. 또한 각 특징의 정보량을 고려하기 위한 엔트로피 필터를 설계 및 제안하였다. 제안된 모델은 각 특징을 분석하기 위한 강도 인지 모델(intensity perception model)과 색상 인지 모델(color perception model)로 구성되어 있으며 이에 따른 가시성지도(conspicuous map)를 이용하여 최종 돌출지도를 구성하기 위한 적응조합방법을 제안한다. 제안된 모델

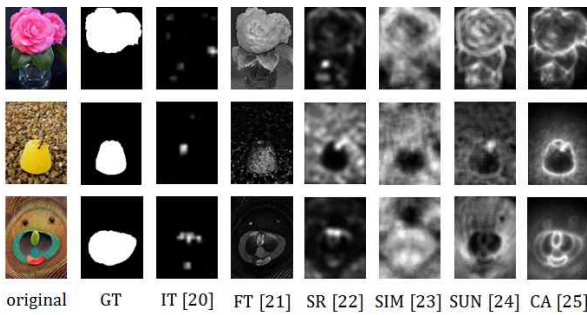


그림 1. 기존의 인공시각집중모델에 따른 돌출지도 예시
Fig. 1. Examples of previous artificial visual attention models.

은 돌출영역 전체에 대해 검출이 가능하며 이는 높은 분할 정확도의 성능을 보인다.

II. 관련 연구

인공시각집중모델은 필터에 따라 구분할 수 있다. 필터의 종류에는 CSD필터, DoG필터, FFT필터, DWT필터, ICA필터 등이 있으며 필터를 통해 돌출영역을 검출하여 돌출지도를 구성한다.

1. 필터 기반의 모델

가. CSD필터 기반의 모델

Itti et al. (모델 IT)^[20]은 영상의 강도, 색상, 방향에 대한 특징채널을 사용하였다. 또한 인간 시각 세포 수용장(receptive field)을 모방한 CSD(Center-Surround Difference)필터를 제안하여 각각의 특징채널에 대한 돌출영역을 검출하였다. 마지막으로 CSD필터에 의해 검출된 각각의 특징채널의 가시성지도를 선형적으로 조합하여 최종 돌출지도를 구성하였으며 이에 승자독식(winner take all) 알고리즘을 적용하여 가장 돌출된 영역만을 검출하였다.

나. DoG필터 기반의 모델

Achanta et al. (모델 FT)^[21]은 주어진 영상에서 돌출영역을 검출하기 위해 DoG(Difference of Gaussians)필터를 각각의 특징채널에 적용하여 돌출영역을 검출하였다. 이 때 모델 FT는 Lab 색상 영역에 대해 DoG필터를 적용하여 돌출영역을 검출하였다.

다. 도메인 변환 필터 기반의 모델

Hou, Zhang (모델 SR)^[22]과 Murray et al. (모델 SIM)^[23]은 각각 FFT(Fast Fourier Transform)와 DWT(Discrete Wavelet Transform)를 이용한 인공시각집중모델이며 도메인 변환 후 돌출영역을 검출하고 역 변환을 적용하여 최종 돌출지도를 검출한다.

라. ICA필터 기반의 모델

L. Zhang et al. (모델 SUN)^[24]은 베이지안 구조와 정보이론에 따른 확률기반의 인공시각집중모델을 제안하였다. 모델 SUN은 ICA(Independent Component Analysis)를 이용한 접근방법을 이용하여 최종 돌출지도를 검출하였다.

마. 문맥인지 기반의 모델

그림 2의 첫 번째 행과 같은 영상들이 주어졌을 때 사람들은 어떻게 이해하겠는가? 아마 대부분의 사람들은 각 영상들에 대해 웃고 있는 소녀, 꽃밭에 서있는 한 남자, 올림픽에 참가한 역도선수와 같이 돌출영역의 주변 정보 즉, 문맥 인지(context-aware) 정보와 돌출영역의 정보를 동시에 이해할 것이다.

이에 따라 S. Goferman et al. (모델 CA)^[25]은 문맥인지 기반의 돌출 개념을 제안하였고 이를 검출하기 위

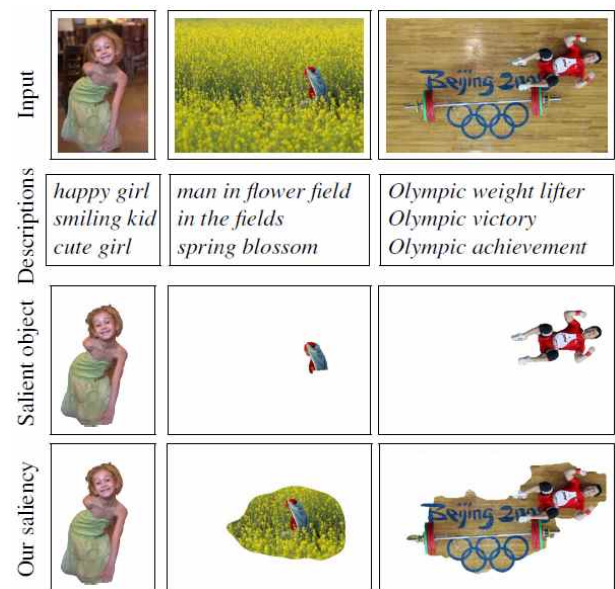


그림 2. 자연영상에서 문맥 정보에 따른 돌출영역 예시^[25]

Fig. 2. Examples of salient region according to context information in natural images^[25].

한 인공시각집중모델을 설계하였다. 또한 사전정보로서 얼굴 검출기(face detector)를 추가적으로 사용하여 모델의 성능을 개선시켰다.

2. 대립과정이론에 따른 시각인지

인간의 시각 시스템은 주어진 영상에 대해 대비(contrast)가 큰 영역에 대해 시각적으로 집중 하며 특히 대립과정이론에 따른 대립색상 쌍(opponent color pair)들의 대비에 집중 한다. 삼원색 이론(trichromatic theory)은 눈의 망막에 존재하는 3가지 종류의 추체(cone)들에 의해 시각 정보를 인지하는 반면 대립과정 이론은 추체들로부터 시각정보를 어떻게 받아들이고 어떻게 처리되는지에 대한 메커니즘을 그림 3과 같은 구조로 설명한다^[26].

먼저 눈을 통해 입력된 시각 정보에 대한 스펙트럼은 L-추체, M-추체, S-추체에 의해 각각 빨강, 초록, 파랑에 대한 색상 정보에 대한 스펙트럼을 받아들이고 간상체(rod)에 의해 명암정보를 받아들인다. 이러한 정보를 이해하기 위해 눈의 시각 뉴런들은 빨강-대-초록(red versus green), 파랑-대-노랑(blue versus yellow), 검정-대-하양(black versus white)의 대비정보를 시신경(optic nerve)을 통해 뇌로 전달시켜 영상을 이해하게 된다. 이에 따라 본 논문에서는 영상의 강도특징과 색상특징을 사용하며 이에 대한 가시성지도를 검출하고 조합하여 최종 돌출지도를 구성한다.

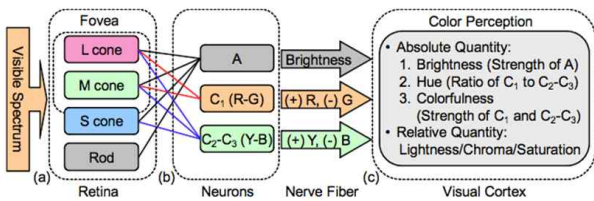


그림 3. 대립과정이론에 따른 시각인지의 흐름도^[26]
 Fig. 3. Flowchart of visual perception based on opponent process theory^[26].

3. 심리학적인 요인에 따른 시각인지

인간은 주어진 영상을 관찰할 때 무의식적 또는 의식적으로 영상의 중앙에 시선을 집중하려는 심리학적 경향이 있다^[27-29]. 이러한 이유로 사람이 사진을 찍을 때 관심영역이나 관심객체를 중앙으로 위치시키는 것이

다. 이러한 심리학적인 시각인지 정보를 돌출영역의 위치에 대한 사전정보로써 사용한다면 인공시각집중모델의 성능을 개선시킬 수 있다.

III. 제안된 인간시각인지 기반의 인공시각집중모델

인간의 시각 시스템은 관찰하는 영상에서 눈에 띄는 돌출영역을 사전정보 없이 자연스럽게 검출이 가능하다. 특히, 대립과정이론에 따른 대립색상 쌍들의 대비에 대해 시각적으로 집중하기 때문에 본 논문에서는 대립과정이론 기반의 생물학적 인공시각집중모델을 제안하며 이는 그림 4와 같다. 제안된 모델의 구조는 크게 영상의 강도와 색상에 대한 특징을 분석하기 위한 강도인지모델과 색상인지모델로 구성되어있다. 그리고 각 인지모델에 의해 검출된 가시성지도는 적응 조합 방법에 의해 조합되어 최종 돌출지도로 구성된다.

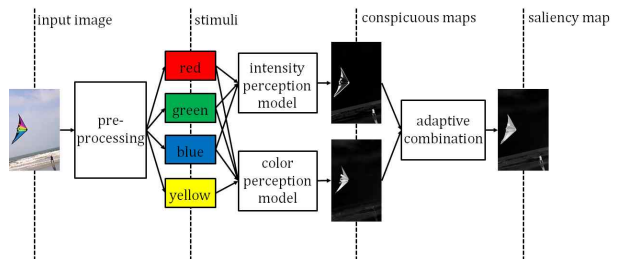


그림 4. 대립과정이론 기반의 제안된 인공시각집중모델의 구조
 Fig. 4. Framework of proposed artificial visual attention model based on opponent process theory.

1. 전처리 과정(Preprocessing)

전처리 과정은 입력된 RGB 색상공간의 영상 잡음(noise)을 완화시키기 위한 필터링 단계와 화소 값의 범위를 일정하게 만들기 위한 정규화 단계로 구성되어 적용된다. 먼저 필터링 단계에서는 원본영상에 대해 반복적인 평균이동필터(iterative Mean-Shift filter)와 미디언 필터(median filter)를 적용한 후 식 (1)과 같이 정규화를 적용한다.

$$I \leftarrow \frac{I}{\|I\|_{inf} + \epsilon}, \tag{1}$$

where $\|I\|_{inf} = \arg \max \{I_{0,0}, I_{0,1}, \dots, I_{H-1,W-1}\}$

다음으로 전처리가 적용된 영상은 빨강, 초록, 파랑,

그리고 노랑의 시각적 스펙트럼으로 분리되어 각각의 인지모델에 입력된다. 이 때 노랑 색상의 스펙트럼은 빨강과 초록 색상의 평균 스펙트럼이다. 또한 각각의 인지모델에는 사람의 심리학적인 시각인지인 중앙편향 정보가 입력되어 사전정보로 사용된다. 이는 다음절에서 자세히 설명하겠다. 이 중앙편향정보는 가우시안 커널(gaussian kernel)을 이용하여 식 (2)와 같이 중앙편향지도(center bias map)로 모델링된다.

$$CBM(x, y; \sigma_x, \sigma_y) = \exp\left(-2\left(\left(\frac{x-x_c}{\sigma_x W}\right)^2 + \left(\frac{y-y_c}{\sigma_y H}\right)^2\right)\right) \quad (2)$$

이 때 W 와 H 는 각각 입력된 영상의 너비와 높이 정보이며, σ_x 와 σ_y 는 각각 x 축과 y 축으로의 가우시안 커널의 너비(breadth)를 조정하기 위한 변수로서 0에서 1사이의 값을 가진다. 예를 들어 입력된 영상의 너비와 높이가 267×400 일 경우 σ_x 와 σ_y 에 따른 중앙편향지도는 그림 5와 같다.

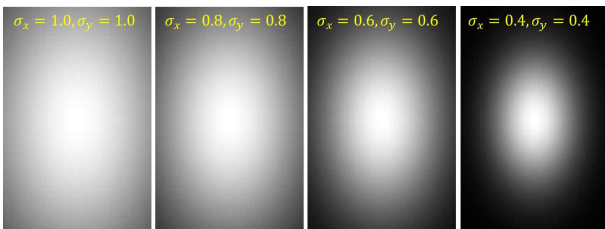


그림 5. σ_x 와 σ_y 에 따른 중앙편향지도
Fig. 5. Center bias maps according to σ_x and σ_y .

2. 엔트로피 필터(Entropy filter)

인간의 시각 시스템은 돌출정보를 이해하기 위해 확률적으로 희소성이 높거나(high rarity) 대비가 높은(high contrast) 영역에 집중한다. 예를 들어 어두운 한밤에 공원에서 산책 중에 있다고 상상해보자. 그러던 도중 밝은 손전등 빛이 그림 6의 왼쪽 그림과 같이 시야에 들어온다면 대부분의 사람들은 높은 손전등 빛에 시각적으로 집중할 것이다. 또한 그림 6의 오른쪽 그림에선 주변 환경에 비해 대비가 높은 빨간 꽃에 시각적으로 집중할 것이다. 이와 같이 본 논문에서는 희소성과 대비가 높은 영역을 돌출영역이라 가정하고 이를 검출하기 위한 엔트로피 필터를 제안한다.



그림 6. 높은 희소성(좌)과 높은 대비(우)에 따른 돌출영역의 예
Fig. 6. Examples of salient regions by high rarity(left) and high contrast(right)

이를 달성하기 위한 엔트로피 필터는 영상의 각 화소에 대한 유사성(similarity)을 이용하여 돌출 값을 계산할 수 있다. 즉, 주변 환경과 유사성이 높은 화소 값(low entropy)은 돌출 값 또한 낮다. 이와 반대로 주변 환경에 비해 유사성이 낮은 화소 값(high entropy)은 돌출 값이 높다. 이를 식 (3)과 같이 SAD(Sum of Absolute Difference)를 이용하여 간단하게 연산하면 다음과 같다.

$$\begin{aligned} C(0,0) &= \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |F(0,0) - F(x,y)| \\ C(0,1) &= \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |F(0,1) - F(x,y)| \\ &\dots \\ C(W-1,H-1) &= \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |F(W-1,H-1) - F(x,y)| \end{aligned} \quad (3)$$

이를 다시 정리하여 쓰면 다음과 같다.

$$C(x,y) = \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \left(\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} D(F(i,j), F(x,y)) \right), \quad (4)$$

where $D(v_1, v_2) = |v_1 - v_2|$

이 때 $F(x,y)$ 와 $C(x,y)$ 는 각각 특징채널의 영상과 그에 따른 가시성지도이고 $D(v_1, v_2)$ 는 SAD를 이용하여 유사성을 계산하기 위한 함수이다. 이와 식은 연산이 비교적 간단하지만 $O(N^4)$ 의 복잡도를 가지기 때문에 실시간 시스템에 적용하기 힘들다. 따라서 히스토그램을 이용하여 바꾸면 식 (4)는,

$$\begin{aligned}
 c(x, y) &= \sum_{l=0}^{N-1} E(k, l), \\
 &= \sum_{l=0}^{N-1} (H(k) \times D(k, l)), \\
 &\text{where } k = F(x, y)
 \end{aligned}
 \tag{5}$$

위의 식과 같이 바꿀 수 있다. 여기서 $H(k)$ 는 특징채널 영의 영상에 대한 정규화 히스토그램으로서 희소성에 대한 항이고 $D(v_1, v_2)$ 는 대비에 대한 항이다. 이러한 식은 앞에서 가정한 조건을 만족하는 식이며 엔트로피 필터의 최종 식이다. 그리고 히스토그램을 이용함으로써 $O(N)$ 의 복잡도로 계산량이 개선되었다. 또한 영상에 존재하는 모든 화소 값에 따른 순람표(LUT - look up table)를 미리 작성해둔다면 모든 돌출 값을 일정한 시간(constant time)내에 계산할 수 있을 것이다.

가. 중앙편향지도에 따른 엔트로피 필터의 개선된 희소성 항

본 논문에서 제안된 인공시각집중모델은 대립과정 이론에 따라 강도인지모델과 색상인지모델로 구성되어 있으며 이에 따라 영상의 강도와 색상의 정보가 각각의 모델에 입력된다. 또한 이 외에도 인간의 심리학적인 시각인지인 식 (2)의 중앙편향지도를 돌출영역 위치의 사전정보로서 사용한다. 이에 따라 희소성 항은 중앙편향지도에 따른 가중치 히스토그램(weighted histogram)으로 표 1과 같이 다시 만들 수 있다.

나. Weber법칙에 따른 엔트로피 필터의 개선된 대비 항

Weber법칙 기반의 기존 많은 영상처리 시스템에 따르면 인간의 시각 시스템은 영상의 정보를 로그(logarithm)의 형태로서 인지한다^[34~36]. 이를 엔트로피 필터에 적용하면 식 (5)의 $D(k, l)$ 을 다음과 같이 다시 쓸 수 있다.

$$D(k, l) = \begin{cases} \sum_{i=k}^l (-\log(H_n(i) + \varepsilon)), & k \leq l \\ \sum_{i=l}^k (-\log(H_n(i) + \varepsilon)), & k > l \end{cases}
 \tag{6}$$

ε 은 $\log 0$ 을 막기 위한 아주 작은 상수이며 $H_n(i)$

표 1. 가중치 히스토그램을 만들기 위한 의사 코드
Table 1. Pseudo code for making weighted histogram.

```

● 가중치 히스토그램  $H_w(k)$ 의 메모리를 0으로 초기화
● for  $x = 0, \dots, W - 1$ :
  ● for  $y = 0, \dots, H - 1$ :
    1.  $H_w(k) = H_w(k) + (1 - CBM(x, y; \alpha, \beta))$ , where  $k = F(x, y)$ 
  ● end
● end
● 가중치 히스토그램  $H_w(k)$ 의 모든 값에 대해 0~1사이로 정규화
    
```

는 0과 1사이의 값을 지닌 정규화된 히스토그램이다. 이러한 엔트로피 필터는 확률적으로 희소성이 아주 높은 소금, 후추 잡음(salt and pepper noise)에 아주 민감하게 반응할 수 있다. 이러한 이유로 전처리 과정의 필터링 작업은 아주 중요한 요소로 작용된다.

3. 강도 인지 모델(Intensity perception model)
엔트로피 필터를 이용한 강도인지모델의 흐름은 그

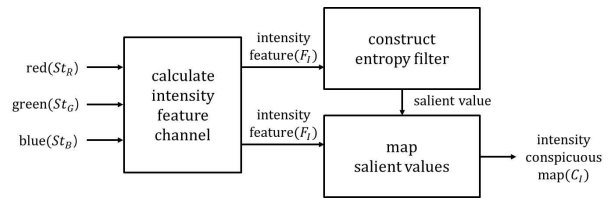
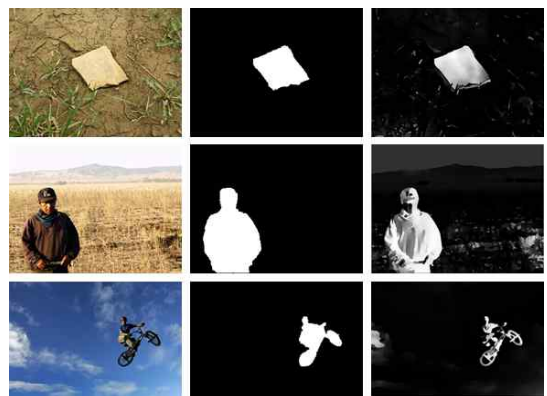


그림 7. 엔트로피 필터를 이용한 강도 인지 모델의 흐름

Fig. 7. Flowchart of intensity perception model using the entropy filter



(a) (b) (c)

그림 8. 강도 인지 모델에 의한 강도 가시성 지도, (a) 원본영상, (b) 지상검증자료, (c) 강도 가시성 지도

Fig. 8. Intensity conspicuous map from intensity perception model, (a) original image, (b) ground truth, (c) intensity conspicuous map

림 7과 같다. 먼저 빨강, 초록, 파랑의 색상 스펙트럼의 평균을 계산하여 강도 영상을 다음과 같이 구성한다. 구성된 강도 영상에 대해 식 (8)와 같이 엔트로피 필터를 적용하면 강도특징채널에 대한 돌출영역이 검출되며 그 후 0~1사이의 값으로 정규화를 적용한다.

$$F_l(x, y) = (St_R(x, y) + St_G(x, y) + St_B(x, y)) / 3 \quad (7)$$

$$C_l(x, y) = \sum_{l=0}^{N-1} E(F_l(x, y), l) \quad (8)$$

4. 색상 인지 모델(Color perception model)

제안하는 색상 인지 모델은 대립과정이론에 따라 대립색상 쌍(opponent color pair)들의 대비를 이용하여 색상 특징채널에 대한 돌출영역을 검출한다. 먼저 빨강, 초록, 파랑, 노랑 색상에 대한 스펙트럼을 입력받아 식 (9-13)과 같이 대립색상에 대한 대비영역인 대립지도(opponent map)를 검출한다.

$$OM_{rg}(x, y) = \text{clamp}(St_R(x, y) - St_G(x, y)) \quad (9)$$

$$OM_{gr}(x, y) = \text{clamp}(St_G(x, y) - St_R(x, y)) \quad (10)$$

$$OM_{by}(x, y) = \text{clamp}(St_B(x, y) - St_Y(x, y)) \quad (11)$$

$$OM_{yb}(x, y) = \text{clamp}(St_Y(x, y) - St_B(x, y)) \quad (12)$$

$$\text{clamp}(v) = \begin{cases} v, & v > 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

구성된 대립지도 영상에 대해 엔트로피필터를 적용하면 색상 특징채널에 대한 돌출영역이 검출되며 가시성 지도(color conspicuous map)를 생성하기 위해 식 (14)와 같이 선형조합을 적용한다. 마지막으로 색상특징에 대한 가시성지도 $C_c(x, y)$ 는 0~1사이의 값으로 정규화가 적용된다.

$$C_c(x, y) = OM_{rg}(x, y) + OM_{gr}(x, y) + OM_{by}(x, y) + OM_{yb}(x, y) \quad (14)$$

4. 최종 돌출지도를 위한 적응 조합 방법

(Adaptive combination for final saliency map)

최종 돌출지도를 구성하기 위해서는 강도 인지 모델과 색상 인지 모델에 의해 검출된 가시성 지도를 조합해야한다. 일반적으로 선형조합을 적용하지만 영상에 따라 강도 및 색상 가시성 지도가 최종 돌출지도를 구성하기 위한 기여가 다를 수 있다. 따라서 본 논문에서는 두 가시성 지도에 대해 평가를 하고 이에 따라 가중치 조합(weighted combination)을 적용하여 최종 돌출지도를 구성한다.

III장의 1절에서 언급했듯이 사람은 영상을 관찰할 때 중앙에 집중하려는 심리학적인 요인이 있다. 이에 따라 제안된 적응 조합 알고리즘은 중앙편향지도를 이용하여 식 (15)과 같이 각각의 가시성 지도에 대해 평가 및 조합을 적용한다. 또한 구성된 돌출지도는 식 (16)과 같이 최종적으로 정규화를 적용한다.

$$S(x, y) = \frac{C_l(x, y) + C_c(x, y)}{|C_l(x, y) - CBM(x, y; \alpha, \beta)| + |C_c(x, y) - CBM(x, y; \alpha, \beta)|} \quad (15)$$

$$S(x, y) = \frac{S(x, y)}{\arg \max(S(x, y)) + \epsilon} \quad (16)$$

IV. 실험 및 성능 평가

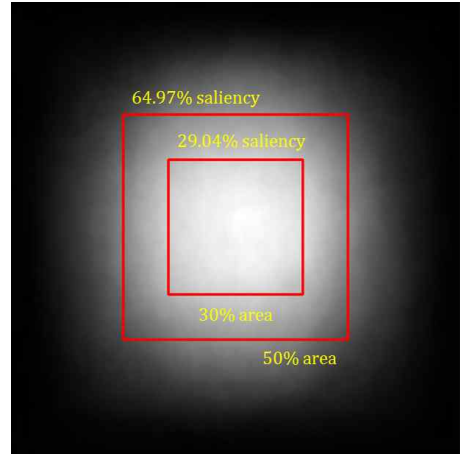
본 논문에서는 자연영상의 돌출영역에 대한 영상분할을 위한 인공시각집중모델을 제안하였다. 이를 평가하기 위해 돌출영역 영상분할을 위한 EPFL 벤치마크 데이터베이스를 사용하였다. 이 데이터베이스는 그림 9와 같은 자연영상 1000장과 이에 따른 지상검증자료로서 돌출영역에 대한 이진영상(binary image)이 포함되어 있다. 이러한 지상검증자료는 사람에 의해 수동으로 레이블링이 적용된 이진영상이다.

제안된 인공시각집중모델을 평가하기 위해 2가지 측면에서의 성능을 측정한다. 먼저 ROC분석을 통하여 강도 인지 모델과 색상 인지 모델에서의 성능을 평가하고 기존 최신의 6가지 모델들과 최종 돌출지도에 대해 성능을 평가한다. 다음으로 영상의 돌출영역 분할에 대한 성능을 평가한다. 이는 F-beta분석을 통하여 성능을 평가한다.

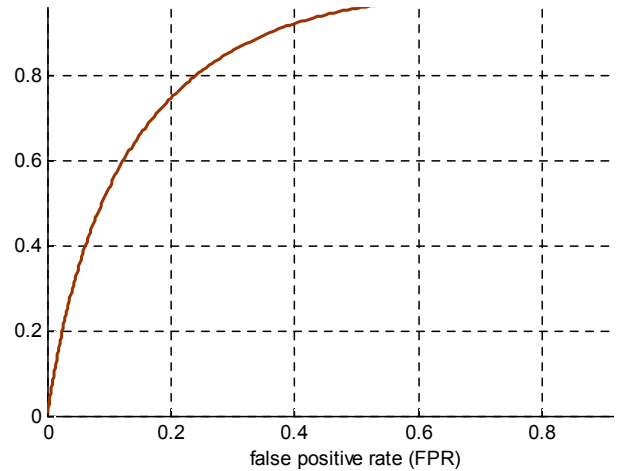


그림 9. EPFL 데이터베이스에 포함된 자연영상과 그에 따른 지상검증자료

Fig. 9. Natural images and ground truth on EPFL database



(a)



(b)

그림 11. EPFL 데이터베이스 영상에 따른 돌출영역의 중앙 편향성

(a) 돌출영역의 평균영상,

(b) 중앙편향지도의 ROC성능

Fig. 11. Center bias of salient regions according to natural images on EPFL database

(a) Average image of salient regions

(b) ROC performance of center bias map

1. 데이터베이스 분석

EPFL 데이터베이스 영상의 돌출영역 면적의 분포도는 그림 10과 같으며 영상 전체의 10~40%에 해당되는 면적이 주로 분포되어 있으며 특히 20% 이하의 작은 면적의 돌출영역이 가장 많다. 그리고 돌출영역의 위치에 대한 특성을 분석하기 위해 EPFL 데이터베이스 영상의 돌출영역에 대한 평균영상을 계산한 결과 그림 11의 위쪽 그림과 같다. 구체적으로 평균영상의 30%의 중앙 사각영역에서 29.4% 그리고 중앙 50%의 사각영역에서 64.97%의 돌출영역이 편향되어 있음을 확인할 수 있었다. 이에 따라 본 논문에서 제안된 중앙편향지도의 ROC성능을 측정한 결과 그림 11의 아래쪽 그림과 같이 0.8573의 높은 AUC를 나타냈다.

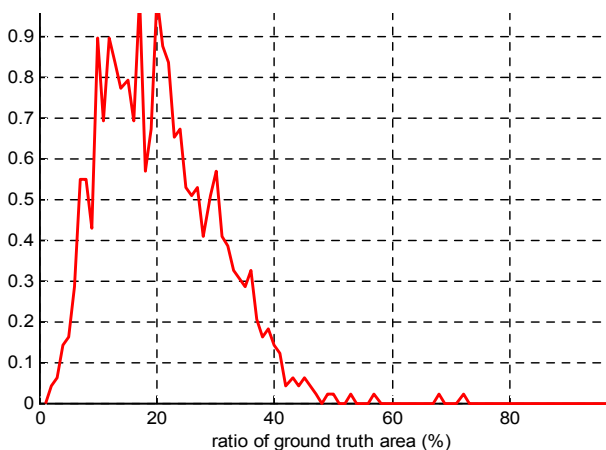


그림 10. EPFL 데이터베이스에 포함된 자연영상의 돌출영역 면적에 대한 분포도

Fig. 10. Area tendency of salient region of natural images on EPFL database

2. ROC 분석을 통한 돌출지도 평가

일반적으로 인공시각집중모델의 평가방법으로는 Kullback-Leibler Divergence(KLD), Normalized Scanpath Saliency(NSS), Area Under Curver(AUC), Linear Correlation Coefficient(LCC) 등이 있으며 그 중에서도 ROC분석을 통한 AUC는 가장 대중적인 방법이다. 따라서 본 논문에서도 ROC분석을 통한 AUC를 통해 각각의 인공시각집중모델들에 의해 검출된 돌출 지도를 성능 평가하였다.

먼저 인공시각집중모델에 의해 검출된 돌출지도는 일반적으로 8비트 영상으로 정규화가 적용되며 이에 따라 0부터 255까지의 다양한 문턱치(threshold)에 대한 이진분류기(binary classifier)를 통해 돌출지도에 대한 이진영상을 만든다. 이러한 다양한 문턱치에 대한 이진영상들은 지상검증자료와 모든 화소를 비교하여 식 (17,18)과 같이 TPR(true positive rate)과 FPR(false positive rate)를 계산하여 ROC를 그리게 된다.

$$TPR = \frac{TruePositive}{TruePositive + FalseNegative} \quad (17)$$

$$FPR = \frac{FalsePositive}{FalsePositive + TrueNegative} \quad (18)$$

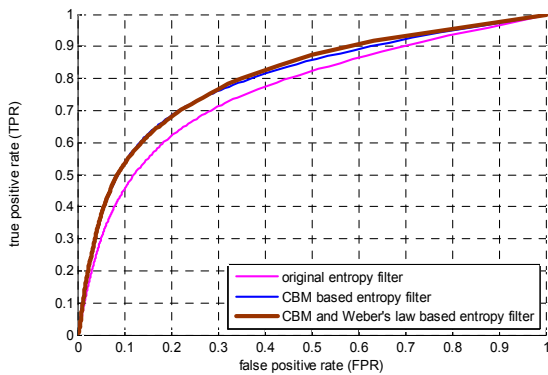


그림 12. 강도 인지 모델에서 중앙편향지도와 Weber법칙을 이용한 엔트로피 필터의 효과
Fig. 12. Effect of entropy filter based on center bias map and Weber's law in intensity perception model

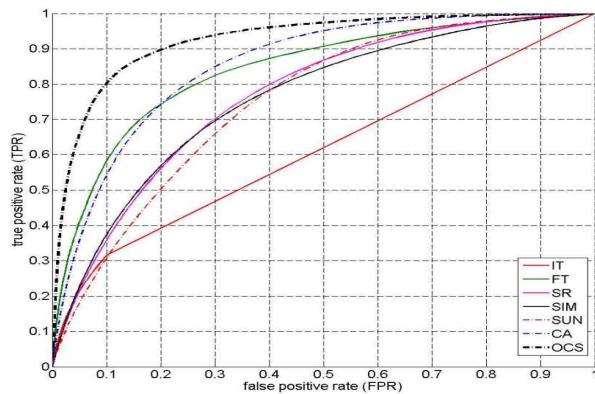


그림 13. EPFL 데이터베이스에 대한 인공시각집중모델들의 ROC성능
Fig. 13. ROC performance of artificial visual attention models in EPFL database.

이러한 과정을 통해 강도 인지 모델에서 Weber법칙과 중앙편향지도에 따른 성능을 평가하였다. 평가한 결과 그림 12와 같이 강도 인지 모델에서 엔트로피 필터를 이용한 AUC는 0.7617이며 중앙편향지도를 이용한 엔트로피 필터와 중앙편향지도와 Weber법칙을 동시에 이용한

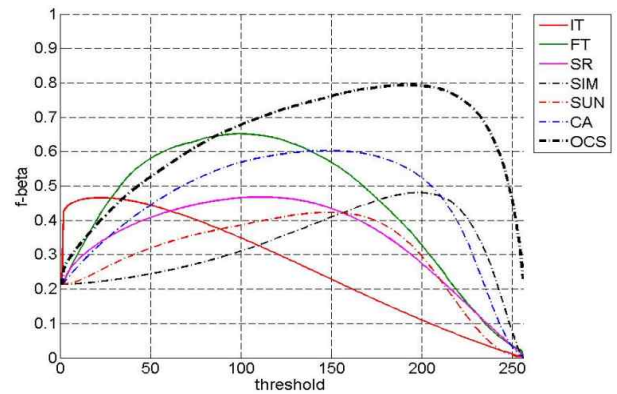


그림 14. 인공시각집중모델들의 돌출지도에 대한 문턱치에 따른 F-beta 그래프
Fig. 14. F-beta graph according to threshold by saliency map of artificial visual attention models.

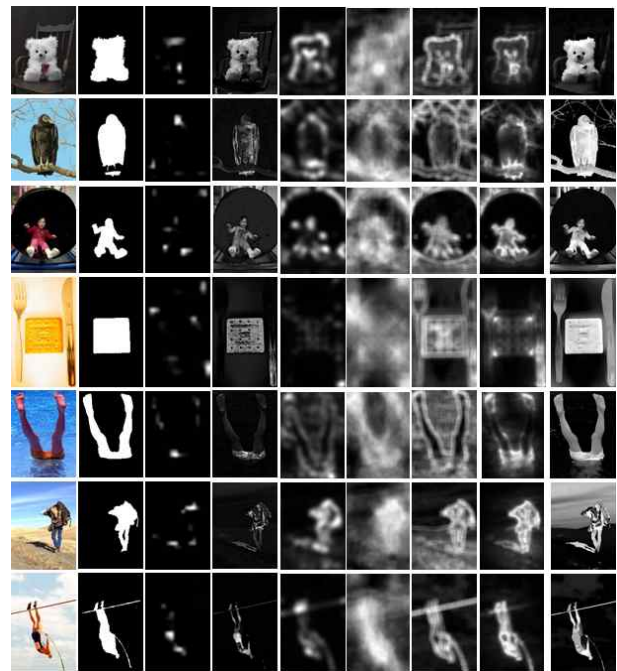


그림 15. 인공시각집중모델에 따른 돌출지도의 예시 (왼쪽부터 오른쪽으로 원본영상, 지상검증자료, 모델 IT^[20], FT^[21], SR^[22], SIM^[23], SUN^[24], CA^[25], 제안된 모델)
Fig. 15. Examples of saliency maps by artificial visual attention models (left to the right, original image, ground truth, IT^[20], FT^[21], SR^[22], SIM^[23], SUN^[24], CA^[25], proposed model).

엔트로피 필터의 AUC는 각각 0.7974, 0.8035로서 강도 인지 모델의 성능이 약 3~4%정도 개선이 되었다.

마지막으로 제안된 모델은 대립과정이론에 기반하여 돌출영역을 검출하기 때문에 OCS(opponent contrast based saliency)모델이라 부르고 기존 최신의 6가지 인공시각집중모델들과 EPFL 데이터베이스에 대한 돌출지도의 성능을 평가한 결과 그림 13과 같았다. ROC분석을 한 결과 제안된 모델의 성능이 우수함을 확인할 수 있었고 특히, FPR이 10%인 경우 기존의 모델들은 약 70%이하의 TPR 성능을 보인 반면 제안된 모델은 약 80%의 TPR 성능을 보였다. 구체적으로 AUC가 0.9256으로서 1에 가까운 높은 성능을 보이며 이는 돌출영역에 대한 분할 정확성이 높다는 것을 의미한다.

2. F-beta 분석을 통한 돌출영역 분할 평가

돌출영역 분할에 대하여 평가하기 위해 앞의 결과 유사하게 다양한 문턱치에 따른 F-beta를 측정하였다. F-beta 측정 식은 다음과 같고 본 논문에서는 β 의 가중치는 0.3을 사용하였다.

$$F_{\beta} = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \quad (21)$$

$$\text{where Precision} = \frac{TP}{TP + FP}, \text{ Recall} = \frac{TP}{TP + FN}$$

이에 따른 측정 결과는 그림 14과 같다. 제안된 모델은 돌출지도에 대한 문턱치가 높은 경우에 높은 F-beta 성능을 보인다. 이러한 이유는 앞서 언급했듯이 돌출영역에 대한 분할 정확성이 높다는 것을 의미한다. 모델 SIM을 제외한 나머지 모델 대부분은 문턱치가 약 150

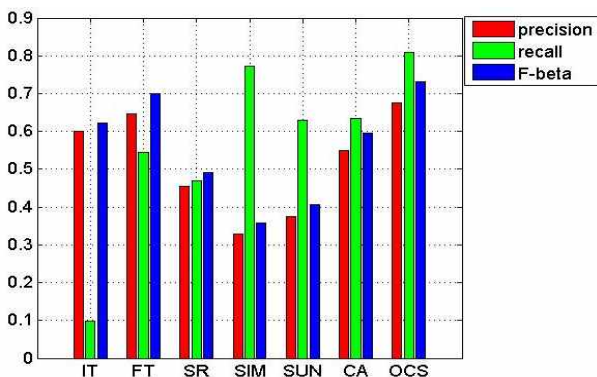


그림 16. 오츠방법을 통한 돌출영역 분할 성능

Fig. 16. Performance of salient region segmentation using Otsu method.

일 때 가장 높은 F-beta 성능을 나타내며 문턱치가 높아질수록 성능이 낮아지므로 돌출영역 분할에 대한 성능이 좋지 않다는 것을 의미한다. 또한 기존 최신의 6가지 인공시각집중모델과 제안된 모델에 의해 검출된 최종 돌출지도에 대해 오츠(Otsu)방법을 사용하여 돌출영역에 대한 영상분할을 적용한 결과 그림 16과 같이 가장 높은 정확성(precision)으로 인해 제안된 모델의 F-beta(0.7325) 성능이 가장 좋았다.

V. 결 론

본 논문에서는 주어진 영상의 돌출영역 분할을 위한 인공시각집중모델을 제안하였으며 이는 생물학적 관점인 대립과정이론에 기반 된다. 이에 따라 제안된 모델은 강도 인지 모델과 색상 인지 모델로 구성되며 각각의 가시성 지도에 적응 조합 방법을 적용하며 최종 돌출지도를 구성한다. 이 때 각 특징채널에 대한 가시성 지도를 검출하기 위해 엔트로피 필터를 설계하였으며 이 필터에 의해 강도 인지 모델에서는 중앙편향지도와 Weber법칙에 따라 성능이 개선됨을 확인할 수 있었고 색상 인지 모델에서는 엔트로피 필터를 적용하여 성능을 극대화 시켰다.

본 논문에서는 성능 측정을 위해 EPFL 데이터베이스를 사용하여 돌출지도와 그에 따른 영상분할 성능을 측정하였으며 그 결과 제안된 모델은 높은 정확성으로 인해 기존 최신의 6가지 모델들보다 성능이 ROC분석을 이용한 AUC를 측정한 결과 기존 최신의 모델들보다 제안된 모델의 AUC는 약 15%의 성능 개선을 보였다. 또한 돌출영역에 오츠방법을 통하여 영상분할을 적용한 결과도 기존 최신의 모델보다 제안된 모델의 분할 성능이 약 22%의 성능 개선을 보였다.

REFERENCES

- [1] K. Koch, J. McLean, R. Segev, Michael A. Freed, Micheal J. Berry, V. Balasubramanian, and P. Sterling, "How Much the Eye Tells the Brain," *Cutt. Biol.*, Vol. 16, no. 14, pp. 1428-1434, 2006.
- [2] L. Itti, "Models of Bottom-Up and Top-Down Visual Attention," PhD Thesis, California Inst. Of Technology, 2000.

- [3] http://en.wikipedia.org/wiki/Visual_system
- [4] A. Alexandre, O. Raphael, V. Pierre, "FREAK: Fast Retina Keypoint," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 510-517, Rhode Island, Providence, USA, 2012.
- [5] A. Borji, L. Itti, "State-of-the-art in Visual Attention Modeling," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 35, no. 1, pp. 185-207, 2013.
- [6] K. Duncan, S. Sarkar, "Saliency in images and video: a brief survey," IEEE Trans. Computer Vision, IET, Vol. 6, Issue 6, pp. 514-523, 2012.
- [7] Ajay K. Mishra and Y. Aloimonos, "Active Segmentation," Int'l J. of Humanoid Robotics, Vol. 6, pp. 361-386, 2009.
- [8] S. Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search, Springer, 2006.
- [9] P. Viola and Michael J. Jones, "Robust Real-Time Face Detection," Int'l J. Computer Vision, Vol. 57, Issue 2, pp. 137-154, 2004.
- [10] N. J. Butko and Javier R. Movellan, "Optimal Scanning for Faster Object Detection," in Proc. of Conf. on Computer Vision and Pattern Recognition, Miami, FL, USA pp. 2751-2758, 2009.
- [11] Albert A. Salahh, E. Alpaydin, and L. Akrun, "A Selective Attention-Based Method for Visual Pattern Recognition with Application to Handwritten Digit Recognition and Face Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, no. 3, pp. 420-425, 2002.
- [12] D. Walther and C. Koch, "Modeling Attention to Salient Proto-Objects," Neural Networks, Vol. 19, no. 9, pp. 1395-1407, 2006.
- [13] S. Frintrop, "General Object Tracking with a Component-Based Target Descriptor," in Proc. of IEEE Int'l Conf. Robotics and Automation, Anchorage, AK, pp. 4541-4536, 2010.
- [14] S. Mitri, S. Frintrop, K. Pervolz, H. Surmann and A. Nuchter, "Robust Object Detection at Regions of Interest with an Application in Ball Recognition," in Proc. of IEEE Int'l Conf. of Robotics and Automation, pp. 125-130, 2005.
- [15] V. Mahadevan and N. Vasconcelos, "Saliency-Based Discriminant Tracking," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1007-1013, Miami, FL, USA, 2009.
- [16] C. Siagian and L. Itti, "Biologically Inspired Robot Vision Localization," IEEE Trans. Robotics, Vol. 25, Issue 4, pp. 861-873, 2009.
- [17] N. Oeurhani, and H. Hagli, "Visual Attention-Based Robot Self-Localization," in Proc. of IEEE Int'l Symp. Computational Intelligence in Robotics and Automation, pp. 308-314, 2005.
- [18] S. Baluja and Dean, A. Pomerleau, "Expectation-Based Selective Attention for Visual Attention Monitoring and Control of a Robot Vehicle," Robotics and Autonomous Systems, Vol. 22, pp. 329-344, 1997.
- [19] C. Scheier and S. Egner, "Visual Attention in a Mobile Robot," in Proc. Int'l Symp. on Industrial Electronics, pp. 48-53, Guimaraes, 1997.
- [20] L. Itti, C. Koch, and E. Nieber, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, Issue 11, pp. 1254-1259, 1998.
- [21] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned Salient Region Detection," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1597-1604, Miami, FL, USA, 2009.
- [22] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in Proc. of Conf. on Computer Vision and Pattern Recognition, pp. 1-8, 2007.
- [23] N. Murray, M. Vanrell, X. Otazu, and C. Alejandro Parraga, "Saliency Estimation Using a Non-Parametric Low-Level Vision Model," in Proc. on Conf. IEEE Computer Vision and Pattern Recognition, pp. 433-440, Providence, RI, 2011.
- [24] L. Zhang, T. K. Marks, M. H. Tong, H. Shan, G. W. Cottrell, "SUN: A Bayesian Framework for Saliency Using Natural Statistics," Journal of Vision, Vol. 8, no. 7, 2008.
- [25] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-Aware Saliency Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 34, Issue 10, pp. 1915-1926, 2012.
- [26] http://en.wikipedia.org/wiki/Opponent_process
- [27] Benjamin T. Vincent, R. Braddely, A. Correani, T. Troscianko and U. Leonards, "Do We Look at Lights? Using Mixture Modeling to Distinguish low- and high-level Factors in

- Natural Image Viewing,” Visual Cognition, Vol. 17, Issue 6-7, pp. 856-879, 2009.
- [28] B. W. Tatler, “The Central Fixation Bias in Scene Viewing, Selecting an Optimal Viewing Position Independently of Moto Biases and Image Feature Distribution,” Journal of Vision, Vol. 7, Issue 14, pp. 1-17, 2008.
- [29] M. Bindemann, “Scene and Screen Center Bias Early Eye Movements in Scene Vewing,” Vision Redearch, Vol. 50, no. 23 pp. 2577-2587, 2010.

— 저 자 소 개 —



정 기 선(학생회원)
 2012년 전북대학교 전자정보공학부
 학사 졸업.
 2014년 전북대학교 전자공학과
 석사 졸업.
 <주관심분야 : 컴퓨터비전, 패턴
 인식>



홍 창 표(학생회원)
 2014년 전북대학교 전자정보
 공학부 학사졸업
 2014년~현재 전북대학교
 전자공학과 석사과정
 <주관심분야 : 패턴인식, 영상처
 리>



박 동 선(정회원)
 1979년 고려대학교 전기전자
 공학과 학사 졸업.
 1984년 Missouri-Columbia
 공학석사 졸업.
 1991년 Missouri-Columbia
 공학박사 졸업.

1991년~현재 전북대학교 전자정보공학부 교수
 <주관심분야 : 신경망, 패턴인식, 영상처리, 디지
 털 시스템 설계>