

## 추출된 특징의 통계적 정보와 국부 경쟁력을 이용한 컬러영상에서의 돌출맵 모형

최경주<sup>1</sup>, 이일병<sup>2</sup>

<sup>1</sup>(주)LG CNS 기술연구부 연구개발센터, <sup>2</sup>연세대학교 컴퓨터과학과

### A Saliency Map Model for Color Images using Statistical Information and Local Competitive Relations of Extracted Features

Kyungjoo Cheoi, Yillbyung Lee

Information Technology Group Division, R&D Center, LG CNS Co., Ltd, Dept. of Computer Science, Yonsei University

**Abstract:** Biological systems appear to employ a serial strategy by which an attentional spotlight rapidly selects circumscribed regions in the scene, rather than attempting to fully interpret visual scenes in a parallel manner for further analysis. In this paper, we propose a biologically motivated saliency map model and applied our system to locating candidate regions of interest on various images for further detailed analysis. In proposed model, several basic features are extracted directly from visual stimuli, and these features are integrated based on their statistical information and local competitive relations. Through integration process, unnecessary features for detecting the salient object are spontaneously decreased while useful features are enhanced. The performance of the model is evaluated over various color images of synthetic and complex real images.

**Key words:** Visual Attention, feature, Saliency Map

**요약:** 생물학적인 시스템은 주어진 시각장면을 모두 병렬적으로 처리하기보다는, 보다 복잡한 고차원적인 처리를 위해 주의가 집중되는 일정한 영역들을 순간적으로 포착하여 그 부분만을 순차적으로 처리해가는 전략을 사용한다. 본 논문에서는 이러한 동물의 생물학적인 시스템의 정보처리 방식에 기본 바탕으로 두고, 컬러영상이 입력되었을 때 다른 부분과 현저히 다른 영역을 탐지해내는 돌출맵 모형을 제안하고, 이 모형을 향후 정밀한 분석이 필요한 몇몇 후보 영역을 찾아내는 문제에 적용하여 보았다. 제안하는 모형은 입력되는 영상정보가 여러 기본 특징으로 추출되고, 이러한 특징들이 통계적 정보와 국부적인 경쟁력에 바탕을 두어 서로 조합해 가는 과정에서 유용하지 않은 특징정보를 자연스럽게 걸러낸다. 모형의 평가를 위하여 간단한 인공영상부터 외부환경에서 흔히 볼 수 있는 복잡한 실영상에 이르는 다양한 컬러영상을 대상으로 실험한 후, 결과를 분석하였다.

**주요어:** 시각적 주의, 특징, 돌출맵

## 서론

현재의 영상분석 방법은 다소 비효과적이다. 대부분의 전형적인 방법들은, 전체 영상으로부터 고차원적인 특징을 먼저 추출하고, 추출된 각각의 특징집합들이 미리 저장된 대상(object)에 대한 프로토타입과 순차적으

로 비교되어진다. 따라서 이러한 방법들은 속도가 많이 느리며, 또한 특정 응용분야에 제한적인 성능을 가지고 있다. 이에 비해 인간의 시각 시스템은 실세계로부터 입력되는 시각정보 중, 응시할 영역을 선택적으로 결정하는 기능과 응시할 영역으로 신속하게 시선을 이동시키는 안구운동 기능을 통해 복잡한 시각정보를 효율적

교신저자: 최경주

(주)LG CNS 기술연구부 연구개발센터, 100-630 서울시 중구 회현동 2가 10-1 프라임타워 9층

Tel: 02-6363-3410

Fax: 02-6363-3499

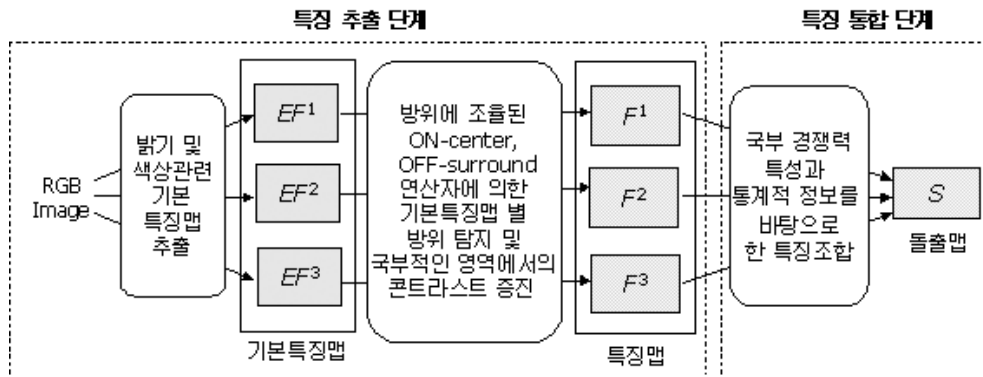
E-mail: choikj@lgens.com

으로 처리한다<sup>1</sup>. 전자는 안구운동이 없는 상태에서 일어나는 시각적 주의(covert visual selective attention)로, 후자는 안구운동(eye movement; overt visual selective attention) 기능으로 설명될 수 있는데, 안구운동(eye movement)은 주어진 장면 속에서 현 작업과 가장 “관련있는” 부분을 고해상도의 중심와에 가져다 놓는 작업으로(overt visual attention), 시각정보를 받아들이는 데 있어 안구운동이 없는 상태에서 일어나는 시각적 주의 기능(covert visual attention)과 서로 매우 밀접한 관계를 가지고 있다고 하겠다<sup>2</sup>. 즉, 인간은 실세계로부터 입력되는 시각정보 중, 응시할 영역을 시각적 주의 기능에 의해 선택적으로 결정하며, 도약 안구운동 기능을 통해 응시할 영역으로 신속하게 시선을 이동시켜 복잡한 시각정보를 효율적으로 처리한다<sup>3,4</sup>. 이러한 이유들 때문에, 컴퓨터 시각에서는 입력되는 시각정보의 양 및 이에 따른 계산의 복잡성 문제를 줄이기 위하여 입력되는 전체 영상정보 중에서 그 순간 현 작업에 가장 필요로 하는 원하는 부분만 찾아서 그 부분만을 분석하도록 가이드 해 주는 인간의 시각적 주의에 해당하는 기능을 수행하는 모듈을 필요로 하고 있다.

본 논문에서는 이러한 인간의 시각적 주의의 특성을 기반으로 하여 영상이 주어지면 전체 영상 중에서 다른 부분과 현저히 다른 돌출성이 있는 영역을 탐지해내는 돌출맵 모형을 제안한다. 여기서 돌출성이란 주변의 다른 여러 대상들에 비해 색상이나 밝기 또는 크기 면에

서 유달리 눈에 띄게 다른 성질을 가지고 있는 것이란 뜻으로 심리학적인 표현으로 흔히 알고 있는 pop-out이 되는 성질이 있는 것을 뜻한다. 제안하는 모형은 다른 자극들보다 두드러진 특정 부분에 대하여 집중하는 저차원적인 시각적 주의 기능인 상향식 방식에 기반을 두고 있으며, 그림 1과 같은 2가지 처리단계를 거친다. 먼저 특징 추출 단계에서, 입력된 영상은 서로 독립적인 기본특징맵(Elementary Feature Map)들인 1개의 무색상 관련 기본특징맵과 2개의 색상관련 기본특징맵으로 분할되고, 이렇게 분할된 기본특징맵마다 국부적인 콘트라스트(contrast)를 강조하고, 방위를 탐지하도록 하기 위해 각 기본특징맵을 구성하고 있는 모든 뉴런에 방위를 가진 중심-주변 연산을 수행하여 특징맵(Feature Map)으로 재구성된다. 이렇게 생성되어진 특징맵은 전 단계에서 계산된 기본특징맵의 국부적인 영역에 있어서의 중요도 값을 가지게 되고, 방위를 탐지하며 또한 잡음을 줄인다. 다음 두 번째 특징 통합 단계에서는, 생성된 모든 특징맵들을 구성하고 있는 뉴런들의 공간 경쟁력 특성 및 통계적 정보를 사용하여 특징맵들을 모두 통합하여 하나의 돌출맵(Saliency Map)을 생성한다. 그러면 시스템은 이러한 돌출맵을 기반으로 주의가 가해지는 영역을 탐지하게 된다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존의 연구에 대한 내용이 기술되며, 3장에서는 제안하는 모형이, 4장에서 실험결과가 기술된다. 마지막으로 5장



**Figure 1** 제안하는 모형의 전체적인 구성 : 크게 특징 추출과 통합의 2가지 단계로 구성되어 있는데, 먼저 특징 추출 단계에서 입력영상의 밝기 및 색상(‘적/녹’, ‘황/청’ opponency)에 관련된 3개의 기본특징맵이 입력영상으로부터 직접 추출되어지고, 이들이 각각의 맵으로부터 방위를 탐지하고, 주변과 크게 다른 영역을 부각시키기 위하여 방위에 조율된 ON-중심, OFF-주변 연산자에 의해 특징맵으로 업데이트되어 진다. 또한 특징 통합 단계에서 3개의 특징맵을 구성하고 있는 뉴런의 통계적인 정보와 이들간의 국부 경쟁력을 바탕으로 하나의 돌출맵으로 통합되어짐으로써 주의가 가해지는 현저한 영역을 탐지해낼 수 있게 된다.

에서 결론을 맺는다.

## 기존 연구

기존의 시각적 주의에 관련된 연구들은 주의가 가해지는 방식에 따라 크게 상향식(또는 데이터 기반) 방식과 하향식(또는 모델 기반) 방식으로 나뉘어 연구되어 왔다.

상향식 방식으로 설계된 시스템은 입력되는 시각 자극 전반에 걸쳐 분포되어 있는 여러 다양한 기본 특징을 추출함으로써 얻게 되는 상향식 단서(또는 현저함 단서)에 의해 주의가 가해지는 영역을 탐지하며<sup>1,6,7</sup>, 하향식 방식으로 설계된 시스템은 수행하는 작업 또는 찾고자 하는 물체 등의 이미 알고 있는 고차원 지식과도 같은 하향식 단서를 사용한다<sup>8</sup>. 이러한 상향식 단서와 하향식 단서를 모두 사용한 하이브리드 방식의 시스템이 제안되기도 하였다<sup>9,10,11,12,13</sup>.

상향식 방식으로 설계된 시스템은 주어진 작업에 관련된 어떠한 종류의 선지식도 사용하지 않기 때문에 기존의 시스템 구조를 크게 변화시키지 않고도 다양한 응용분야에 적절히 사용되어 질 수 있다. 한편, 하향식 방식으로 설계된 대부분의 기존 시스템들은 상향식 단서를 제대로 사용하지 않기 때문에, 일종의 “패턴 매칭(pattern matching)” 모듈과도 같은 역할을 하게 되는 문제가 있다. 이러한 경우, 주어진 작업에 따른 주의집중 과정을 학습할 수 있는 방법이 요구되며, 또한 인식 단계와의 부분적인 상호작용이 필요하게 되어 다른 문제에 적절히 확장시켜 적용시키기가 아주 어렵게 된다. 이러한 점들이 그동안 고차원적인 지식, 즉 하향식 단서에 대한 중요성이 계속 대두되어 왔지만, 이를 사용하는 시스템에 대한 연구가 많이 진행되지 않은 이유이기도 하다. 따라서 본 논문에서는 특정 분야에만 사용되지 않는 일반적인 시스템을 만들기 위하여 상향식 방식을 기반으로 두고 시스템을 설계하였다.

Treisman에 의해 연구된 “특징통합이론”(Feature Integration Theory)<sup>14</sup>은 시각적 주의에 관련된 연구에 상당한 영향을 끼친, 인간의 시각탐색 전략을 설명한 연구결과로써, 대부분의 시각적 주의에 관련된 연구들이 이에 기초를 두고 있다. 최초의 인간의 시각적 주의 계산적 모형은 ‘돌출맵’이라는 개념을 도입한 Koch와

Ullman의 모형<sup>1</sup>이라고 할 수 있으며, 지금까지 제안된 대부분의 성공적인 계산 모형들은 Treisman의 특징모듈들에 해당하는 특징맵을 계산하는 단계와 이들을 조합한 결과로 각 위치에서 특징들이 얼마나 현저히 변하는가를 나타내는 돌출맵을 계산하는 단계가 있다. 이들 모형들간에 다른 점이라 하면 입력되는 영상에서 특징을 추출하는 방법과, 돌출맵을 생성하는 방법의 차이이다.

컴퓨터 시각에서도 생물학적인 증거에 기반을 둔 많은 주의 시스템들이 개발되었는데, 이 중, Mozer<sup>15</sup>와 Sandon<sup>16</sup>의 시스템은 간단한 인공영상 및 알파벳과도 같은 아주 간단한 영상에만 시스템을 적용하여 다른 복잡한 컬러 실영상에 적용하기에는 한계가 있다. 또한 Itti<sup>7</sup>는 돌출맵과 승자전취망(winner-take-all network)으로 구성된 순수한 상향식 방식의 주의 모형을 제안하였는데, 이 모형에서는 가장 높은 주의 지점(location)을 찾아내고, 이 지점을 중심으로 일정한 크기의 원을 그려 주의가 가해지는 영역을 표시하였으며, 시간의 흐름에 따라 승자전취망으로 통해서 다른 영역으로 주의 영역이 옮겨지도록 하였다. 이때 주의가 가해지는 영역을 표시하는 방법에 있어서 영상에 따라 적절한 상수값으로 정해 영역을 지정하여, 컴퓨터 시각에서 실제적으로 인식과도 같은 고차원적인 작업을 수행하기 위해서 세그멘테이션 및 이진화 등과도 같은 후처리 작업에 상당한 작업이 필요하다. 이에 반해 Milanase의 제안된 모형<sup>12</sup>은 Itti의 모형에 비해 어느정도 주의가 가해지는 대상을 영역별로 찾아내지만, 모형의 성능을 분석하는데 사용한 실험 영상이 비교적 적으며 성능 분석 또한 미비하기 때문에 실제적인 컴퓨터 시각 문제에 적용하기엔 너무 미비한 점이 많다. 더욱이 Milanese의 모형에서는 상향식 단서와 하향식 단서를 동시에 사용하며 에너지를 최소화하는 기법인 반복적인 비선형적인 relaxation 방법을 사용하여 통합함으로써 시스템의 실행 속도가 느린 단점이 있다.

이렇듯 지금까지 고찰된 대부분의 시스템들에 대한 2가지 주목할 만한 사실은, 첫째로는 대부분의 모형들이 현재 시각적 주의에 대한 개념 정립 단계에 있으며, 실제적인 응용분야가 아닌 인간의 시각적 주의에 관한 이론적인 부분에 대한 분석에 치중해 있다는 것이며, 둘째, 대부분 시뮬레이트된 데이터나 간단한 영상에만 적용하였기 때문에 자연적인 컬러영상에 적용한 경우가

매우 드물다는 것이다. 이러한 2가지 사실을 바탕으로 할 때, 일반적인 시스템이라고 할 수 없으며, 따라서 특정 문제에 한해서만 동작하고 있는 경우가 많고, 아직은 실제적인 문제에 광범위하게 사용되지 못하고 있는 경우가 많다는 것이다. 이에, 본 논문에서 제안하는 시스템은 이러한 기존의 시스템, Itti<sup>7</sup>와 Milanese<sup>12</sup>에 의해 제안된 생물학적인 구조 및 성능을 확장시켜, 고차원적인 선지식을 사용하지 않고도 실영상을 사용하는 여러 응용분야에 효과적으로 적용될 수 있는 일반적인 시스템을 개발하는데 주 목적을 두고 설계되었으며, 다양한 종류의 실험 영상과 실험 방법을 통해 제안된 시스템의 성능을 확인하였다.

## 제안하는 시스템

제안하는 시스템은 사전지식 없이 주어진 시각 자극의 특징에 의해 목표물 탐지한다. 제안하는 시스템은 그림 1에서도 보았듯이 특징 추출과 통합의 2가지 처리 방식으로 되어 있으며, 이번 장에서는 이러한 2가지 처리방식에 대해 자세히 기술한다.

### 특징 추출 단계

인간의 시각적 주의에 영향을 미치는, 크게 무색상과 색상에 관련된 2가지 종류의 특징맵이 입력 영상으로부터 추출된다. 우선적으로, 무색상 특징맵으로, 입력 영상의 밝기값을 사용한 1개의 무색상 관련 기본특징맵이 추출되었고, 색상 특징맵으로, 인간의 색상 처리 과정에서 보이는 '색상 대립(color opponency)'을 바탕으로 한 2개의 색상관련 기본특징맵이 만들어졌다.

색상은 인간의 시각이 물체를 구분할 수 있는 가장 큰 특징 중의 하나이며, 정신생리학적인 결과에 의하면 색상이 전-주의 단계에 있어 특징 선택 작업에 큰 역할을 한다고 한다. 또한 밝기 정보는 시각장면의 현저한 영역을 선택하는데 유용한 특징이 될 수 있으며, 색상정보가 사용될 수 없는 경우에도 유용하게 사용될 수 있다. 인간의 컬러 지각 능력은 망막에 있는 세 가지 종류의 추상체(cone)들의 반응에 의해 발휘되는데, 각기 적색, 녹색, 청색의 빛 중 어느 하나에 반응하는 광색소를 가진다. 이 세 가지 형태의 추상체( $R$ ,  $G$ ,  $B$ )를 때로는 파랑(blue), 초록(green), 빨강(red)으로 취급하기도 하

는데, 이것은 잘못된 개념이다. 왜냐하면, 추상체의 최대 반응점은 이들 색깔과는 대응되는 파장과 정확히 일치하지 않기 때문이다. 이러한 세 가지의 추상체는 서로 중첩되어 반응하는데, 이러한 특징은 시각 시스템이 서로 다른 빛의 파장을 구분할 수 있게 해준다. 이들 추상체 세포는 다음 뉴런에 연결되며, 색상정보를 뇌로 전달하는데, 색상정보를 뇌로 전달해 주는 신경회로는 삼색 수용기의 정보를 '적/녹', '청/황'의 반대쌍의 색상 정보로 바꾸어 전달한다.

본 논문에서는 이러한 인간의 색상처리에 관련된 정보를 바탕으로 하여, '적/녹', '청/황'의 2가지 색상대립을 모델링하여 2개의 색상관련 기본특징맵을 만들고, 밝기정보를 사용하여 무색상 관련 기본특징맵을 만든다.

무색상 특징맵은 밝기정보를 이용하는데, 밝기정보는 컬러영상을 입력받았을 경우 색상 정보를 사용하여  $I=(R+G+B)/3$  과 같은 형식으로 추출할 수 있으며, 이  $I$ 가  $EF^1$ 이 된다. 여기서,  $R$ ,  $G$ ,  $B$ 는 각각 적색, 녹색, 청색을 나타내는 추상체 반응이라 가정한다.

색상에 관련된 기본특징맵을 만드는 순서는 다음과 같다. 먼저, 세 가지 추상체  $R$ ,  $G$ ,  $B$ 로부터 식(1)과 같이 다른 색상과 조금도 희석되지 않은 순수한 색상에 조율된 채널  $r$ ,  $g$ ,  $b$ ,  $y$ 를 생성한다.

$$\begin{aligned} r &= R - (G+B)/2, & g &= G - (R+B)/2 \\ b &= B - (R+G)/2, & y &= R+G-2(|R-G|+2) \end{aligned} \quad (1)$$

여기서  $r$ ,  $g$ ,  $b$ ,  $y$ 는 각각 순수한 색상에 조율된 적색, 녹색, 청색, 황색 채널을 나타내며, 각 채널은 흑색과 백색 자극에 대해 0을 출력한다. 이러한 채널을 사용하여 '적/녹' 대립과 '청/황' 대립을 사용한  $EF^2$ ,  $EF^3$ 를 만든다.

$$EF^2 = r - g, \quad EF^3 = b - y \quad (1)$$

이렇게 생성된 모든 기본특징맵들을 서로 다른 특징 추출기법에 기인한 across-modality 상이점을 배제하기 위해서 생성된 3개의 독립적인 기본특징맵들을 모두 0~1 사이의 값으로 정규화하고, 각각의 기본특징맵들로부터 방위를 탐지하고, 국부적인 영역에 있어서 주변과 크게 다른 영역을 부각시키는 반면 그 주변부는 억제시키는 효과를 내도록 하기 위하여 8가지 방위( $\theta \in (0, \pi/8, 2\pi/8, \dots, 7\pi/8)$ )를 가진 ON-중심, OFF-주변 연산

자인  $h(\theta)$  필터를 사용하여 식 (3)과 같이 계산한다.

$$F_{x,y}^k = \sum_{\theta} \left[ \sum_{m,n} E F_{m,n}^k \cdot h_{x-m,y-n}(\theta) \right]^2 \quad (3)$$

식 (3)에서, 각각의 기본특징맵은  $h(\theta)$  필터와 회전되어지며, 회전 후 회전된 영상의 콘트라스트를 증가시키기 위해 회전된 결과가 제공되어지고, 방위별로 나뉘어진 여러 지도들이 가장 반응이 하난의 방위로 통합되어진다. 여기서  $h(\theta)$  필터는 식 (4)와 같이 계산된다.

$$h_{x,y}(\theta) = |K_1 \cdot G_{x,y}(\sigma, r_1 \cdot \sigma, \theta) - K_2 \cdot G_{x,y}(r_2 \cdot \sigma, r_1 \cdot r_2 \cdot \sigma, \theta)| \quad (4)$$

식 (4)에서,  $G(\cdot, \cdot, \cdot, \cdot)$ 은 방위를 가진 2차원 가우시안 함수이며,  $K_1, K_2$ 는 양수,  $r_1$ 은 2개의 가우시안의 중심,  $r_2$ 는 ON과 OFF 가우시안 간의 폭 비율을 뜻한다.

### 특징 통합 단계

3.1절에서 계산된 특징맵들을 구성하고 있는 뉴런들은 입력된 영상의 각 지점마다 각각 특별한 의미를 가지고 있다. 그런데, 이를 기반으로 주의가 가해지는 영역을 찾아내기란 쉬운 일이 아니다. 그 이유는, 특징맵에 의해 제공받게 되는 영상에 대한 특징값들은 각각의 특징맵이 서로 다른 특징추출기법으로 생성되어진, 서로 독립적인 것이기 때문이다. 즉, 각각의 특징맵이 제공해 주는 정보는 모두 하나의 영상에 대한 것이므로, 각각의 특징맵마다 모두 다른 영역을 주의가 가해지는 영역이라고 가이드할 수도 있다. 이러한 문제를 해결하기 위해서는 영상의 각 지점에 있어서의 서로 다른 특징값들이 모두 하나로 통합되어져서 하나의 전역적인 중요도 측정값이 되도록 해야 한다. 따라서, 지금까지 계산된 특징맵들은 영상의 각 지점마다 하나의 중요도 측정치를 나타내주는 유일한 맵인 돌출맵으로 통합되어야 한다.

서로 독립적인 여러 개의 서로 다른 특징맵을 하나로 통합하는 것은 쉬운 작업이 아닌데, 대체로, 모든 특징맵에서 얻어지는 전체 뉴런 중 가장 큰 값을 나타내는 값을 취하거나<sup>9,17</sup>, 각각의 특징맵에 적절한 가중치를 두어 합하는 방법이 사용되기도 하였다<sup>6,18</sup>. 그러나, 첫 번째 방법에서는 주의가 가해지는 지점이 항상 높은 값을 가지고 있지는 않는다는 문제가 발생한다. 다시 말하면, 계산되어진 특징맵이 탐지하고자 하는 목표 영역에 대

한 정보가 들어있는 임의의 템플릿과의 상관관계로 얻어진 것이 아니라면 일반적으로 높은 값을 가진 장소가 낮은 값을 가진 장소보다 현저하다라는 근거가 없다는 것이다. 또한 두 번째 방법은 어떤 가중치를 적용할 것이냐에 따라 그 응용 분야가 제한되며, 또한 시스템의 성능이 현저히 달라진다는 단점이 있다.

본 논문에서 사용하는 특징 통합 방법은 특징맵을 구성하고 있는 뉴런들의 통계적 정보와 국부적인 경쟁력 특성을 이용한 아주 간단한 방법으로 그 수는 적지만 의미 있는 활동성을 보이는 특징이 함유된 맵이 강조되고, 그렇지 않은 맵은 억제한다. 여러 특징맵을 통합하여 하나의 돌출맵을 생성하는 방법은 다음과 같다.

가장 먼저, 3.1절에서 계산된 각각의 특징맵  $F^k(k=1,2,3)$ 를 입력받아 이를  $LoG$ 함수로 생성된 넓은 크기의 2차원 필터로 회전(convolution)한 후, 입력받은 원래의 영상과 더한다.

$$F_{x,y}^k = \sum_{m,n} (F_{m,n}^k \cdot LoG_{x-m,y-n}) + F_{x,y}^k \quad (5)$$

단, 사용한  $LoG$ 함수는 식 (6)과 같다.

$$LoG(x,y) = \left[ \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} \right] e^{-\left( \frac{x^2 + y^2}{2\sigma^2} \right)} \quad (6)$$

여기서,  $\sigma$ 는 가우시안의 너비를 뜻하며, 실험에서는 3.6으로 설정하였다. 이러한 처리과정은 입력되는 맵안의 뉴런들에 대해 좁은 범위에 있어서는 협동작용이, 넓은 범위에 있어서는 경쟁작용이 일어나는 효과를 내게 되는데, 특징맵의 국부적인 영역에서의 뉴런들과 다른 주변영역의 뉴런들을 비교하여 차이가 크면 해당 뉴런들을 활성화시키고 그렇지 않으면 억제시키는 경쟁 메커니즘이라 할 수 있으며, 이는 인간의 뉴런들이 측면억제에 의한 결합 방식을 띠고 있다는 실험적 증거를 모사하고 있다고 할 수 있다. 또한 Aguilar와 Ross의 흥미 있는 영역은 다른 것들과 가장 다른 것이 있는 영역이라는 연구결과<sup>19</sup>도 주목할 만하다. 이러한 과정이 일정 횟수 동안 반복되고, 결과적으로 계산된 맵  $F^k$ 는 식 (7)과 같이 계산되어 3개의 돌출맵을 만든다.

$$S_{x,y}^k = \frac{SF_{x,y}^k - MinSF}{MaxSF - MinSF} \quad (7)$$

$$\text{단, } SF_{x,y}^k = F_{x,y}^k \times (MaxF^k - AveF^k)^2$$

여기서,  $MaxF^k$ 는  $F^k$ 를 구성하는 특징값 가장 큰 값

을,  $AveF^k$ 는  $F^k$ 에서  $F^k$ 를 구성하는 특징값의 최대값인  $MaxF^k$ 을 제외한 값들의 평균값을 뜻하며,  $MaxSF$ 와  $MinSF$ 는 각각  $SF^1$ ,  $SF^2$ ,  $SF^3$ 를 구성하는 모든 특징값 중 가장 큰 값과 가장 작은 값을 뜻한다.

활동량이 있는 모든 지점에 대해서 전체 지도안에서 가장 큰 활동량과 평균 활동량을 비교하면 현재 지점의 활동량이 평균 활동량에 비해 얼마나 다른지 알 수 있게 된다. 이런 차이가 크면 클수록 해당 지도내의 특정 지점에서의 활동량이 다른 지점에 비해 두드러지다는 말이 되고, 그렇지 않으면 해당 맵은 별로 독특하지 활동양상을 보이는 값만을 가지고 있다는 말이 된다. 따라서 식 (7)에 의해 계산되어진  $S^k$ 맵은 전체적으로 평이한 특징값을 가지고 있던 특징맵들은 전체적으로 그 특징값이 저하되고, 그렇지 않은 맵은 예전에 비해 높은 값을 가지게 된다. 이를 통해 최종 목표 영역을 탐지하기 위해 필요 없는 정보를 가지고 있는 특징맵을 걸러낼 수 있다. 따라서, 3개의 돌출맵  $S^k$ 를 단순히 모두 합한 후, 0~1 사이의 값으로 정규화시켜 하나의 돌출맵  $S$ 를 만든다.

## 실험결과 및 분석

제안하는 시스템의 전반적인 성능을 분석하고자 인간의 시각탐색에 관한 심리학적인 실험에 사용되는 간단한 인공영상에서부터 외부환경에서 흔히 볼 수 있는 복잡한 실영상에 이르는 다양한 컬러영상 119개를 대상으로, 입력되는 영상에서 가장 먼저 주의가 가해지는 독특한 대상이 있는 영역을 찾아내는 문제에 적용하여 보았다. 복잡한 실영상의 경우, 각각의 영상은 간판, 신호등, 교통 표지판, 현수막, 우체통 등과 같은 목표물이 존재하며, 국부적인 영역에 있어서 조명이 균일하지 않거나, 목표물이 아닌 기타 다른 대상이 함께 존재하는 영상 94개를 사용하였다. 또한, 간단한 인공영상의 경우에는 주의 영역(attention region)이 주변의 방해물(distractor)들과 방위, 색상, 모양, 크기, 밝기면에서 틀린 영상 25개를 사용하였다.

그림 2는 청색의 교통 표지판을 탐지하는 문제에 모형을 적용한 것으로써 시스템의 전반적인 동작과정을 보여준다. 실험에 사용되는 컬러영상이 입력되고, 3개의 특징맵이 입력영상으로부터 추출되어진 후, 각 기본

특징맵들이 특징맵으로 업데이트 되어지고, 이어 하나의 돌출맵으로 통합되어져 결국 입력영상에서 가장 눈에 띄는 청색 교통 표지판을 제대로 탐지해 낸다. 또한 그림 3은 제안하는 시스템을 실제적인 여러 분야에서의 다양한 대상 탐지 문제에 적용한 예를 보여준다.

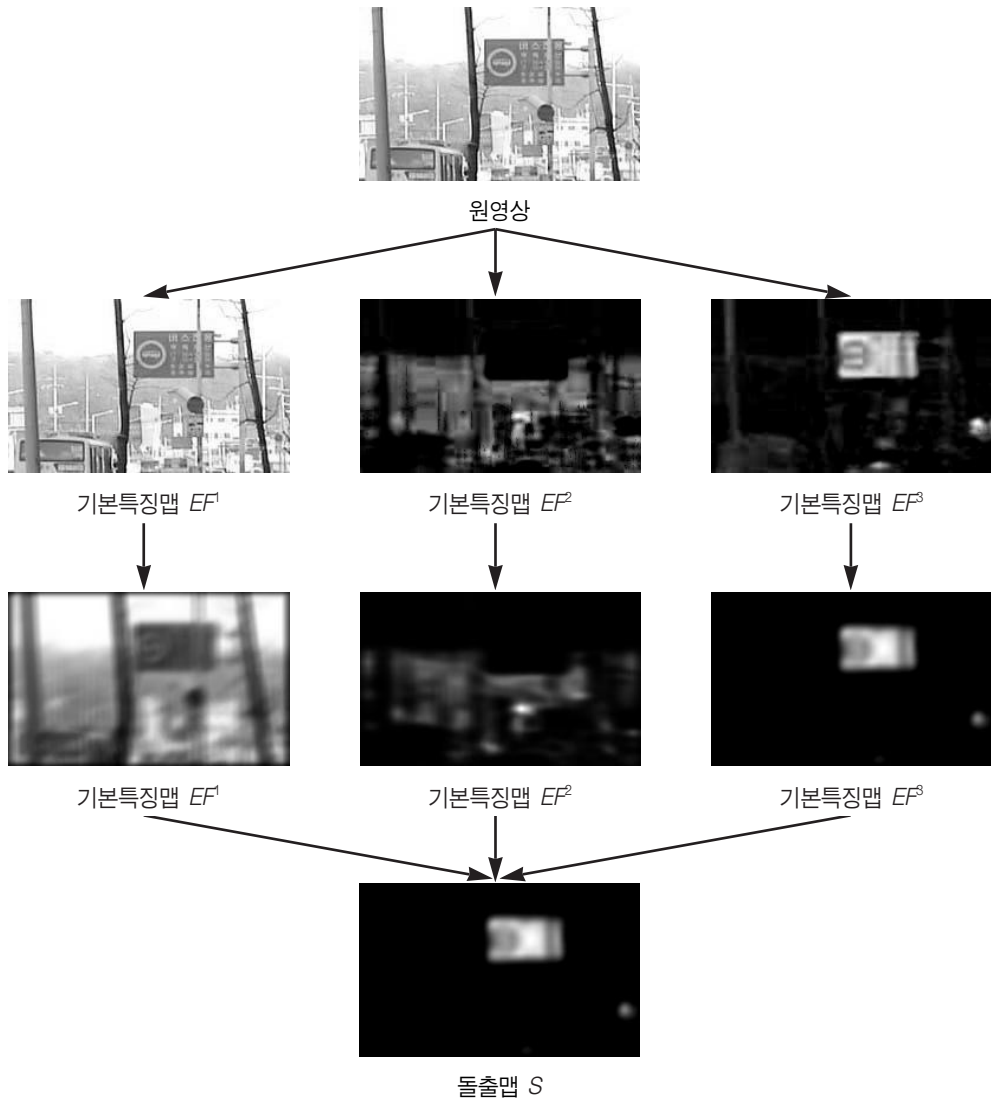
일반적으로 제안하는 시스템과도 같은 연구의 성능평가를 위한 표준화된 객관적인 성능분석용 데이터가(bench-mark data) 따로 존재하지 않기 때문에, 실험된 결과가 좋은지 나쁜지를 결정하여 시스템에 대한 평가를 내리는 것이 결코 쉬운 일이 아니다. 간단한 인공영상의 경우에는, 기존의 많은 심리학자들에 의한 인간의 시각 탐색 전략에 관한 연구결과를 토대로 어느 정도 분석 가능하다 할 수 있겠으나, 실영상인 경우에는 아주 난해하다. 이런 이유로 인해 본 논문에서는 선정된 실영상에 대해 인간의 행동을 측정하는 실험을 수행하였다. 12살의 초등학교 5학년 학생 40명(남학생 21명, 여학생 19명)이 실험에 참가하였으며, 실험에 사용되었던 전체 실영상 중, 눈에 확연히 드러나는 대상이 있는 영상을 제외한 총 70개의 영상을 선정하여 실험하였다. 실험방법은 피험자에게 선정된 영상을 보여주고, 주의가 가는 지점을 2개 이하로 한정하여 표시하라고 지시하였다. 또한 사람은 상향식 방식뿐만 아니라,

하향식 방식도 동시에 작용하는 고도의 인지능력을 보이기 때문에, 피험자로 하여금 영상을 보는 즉시 과제를 수행하도록 하였다. 실험 후, 각 영상마다 표시된 주의 지점에 대하여 표시된 횟수를 모두 합산하여 가장 많이 표시된 지점을 그 영상에 있어서의 최대 주의지점이라 가정하였다.

실영상에 대한 실험결과를 보게 되면, 제안하는 시스템은 인간이 주의를 주는 듯한 방식으로 주의 지점을 제대로 탐지함을 알 수 있다. 표 1은 여러 실영상에 대해 인간의 행동을 측정하여 얻게 된 각 영상의 최대 주의지점과 시스템에서 얻어진 결과가 얼마나 잘 부합되는지를 숫자로 표시한 것이다. 표 1을 보면, 제대로 탐지하지 못한 경우의 수가 18개로 나타났는데, 이는 인

**Table 1.** 실영상에 대한 인간의 행동 측정에 따른 시스템의 비교 평가.

실영상 전체 수	제대로 탐지한 경우의 수	제대로 탐지하지 못한 경우의 수
94	76	18



**Figure 2** 실영상에 대한 시스템의 전반적인 동작과정 예 : 3개의 특징맵이 입력영상으로부터 직접 추출되어지고, 이들이 각각의 맵으로부터 방위를 탐지하고, 주변과 크게 다른 영역을 부각시키기 위하여 방위를 가진 ON-중심, OFF-주변 연산자에 의해 처리되어 특징맵으로 업데이트 되어지며, 마지막으로 각 맵을 구성하고 있는 뉴런의 통계적인 정보와 이들간의 국부 경쟁력을 바탕으로 하나의 돌출맵으로 통합되어짐으로써 주의가 가해지는 현저한 영역을 탐지해낸다. 여기서 즉각적으로 주의가 가해지는 곳은 청색의 교통 표지판이며, 마지막으로 생성된 돌출맵을 보면 청색의 교통 표지판이 있는 영역을 제대로 찾아냄을 알 수 있다.

간이 주의를 가할 때 상황식 지식뿐만 아니라 인식에 필요한 하향식 지식도 사용하기 때문에 생기는 결과 때문인 것으로 보인다. 특히 주의를 끄는 영역이 여러 군데 겹쳐있는 영상일 경우에는, 심리학적 실험에서 아무리 빨리 주의 지점을 선택하라 요구했을지라도 어쩔 수 없이 하향식 지식이 사용될 수밖에 없으며, 이는 피험자들로 하여금 서로 다른 영역을 가장 현저한 영역이라

지정할 수밖에 없게 한다. 그러나, 이런 차이는 사람이 흔히 인식에 사용될 수 있는 흥미 있는 대상에 대한 특정 특징맵을 간단히 구현하여 시스템에 삽입시키면 이러한 오탐지의 경우의 수를 현저히 줄일 수 있다고 생각된다.

또한, 그림 4는 간단한 인공영상을 대상으로 실험한 결과를 보여준다. 그림 4와 같은 종류의 영상에 대한 실



**Figure 3** 다양한 분야의 대상 탐지 문제에 적용한 결과 : 각 영상에서 가장 먼저 주의가 가는 곳은 (a) 적색의 커다란 현수막(아주 복잡한 배경을 가진 조명이 불균일한 실영상) (b) 중간정도의 복잡도의 배경을 가진 실영상에 대한 실험결과로써, 주의가 가는 곳은 적색의 스키복을 입고 스키를 타는 사람(중간정도의 복잡도의 배경을 가진 실영상) (c) 보라색의 꽃잎이다(비교적 단순한 배경을 가진 실영상). (a)~(c)의 결과영상을 통해 시스템의 결과 또한 인간의 주의 지점과 동일함을 알 수 있다.

힘은 인간의 pop-out 현상이 본 모형에서도 보여지는지를 알아보기 위한 실험이었으며 이러한 종류의 실험을 통해 제안하는 시스템은 찾고자 하는 대상의 색상, 방위, 모양, 크기, 또는 밝기 등이 주변의 대상들과 현저한 차이를 보일 때, 찾고자 하는 대상이 있는 영역은 돌출맵을 형성해 가는 과정에서 다른 영역에 비해 유달리 튀게 되며, 돌출맵을 통해 한번에 해당 영역을 탐지하게 됨을 알 수 있었다. 이를 통해 제안하는 시스템은 인간의 pop-out 현상을 보인다고 할 수 있으며, 따라서 어느 정도 생물학적 기반을 가지고 있다고 할 수 있다.

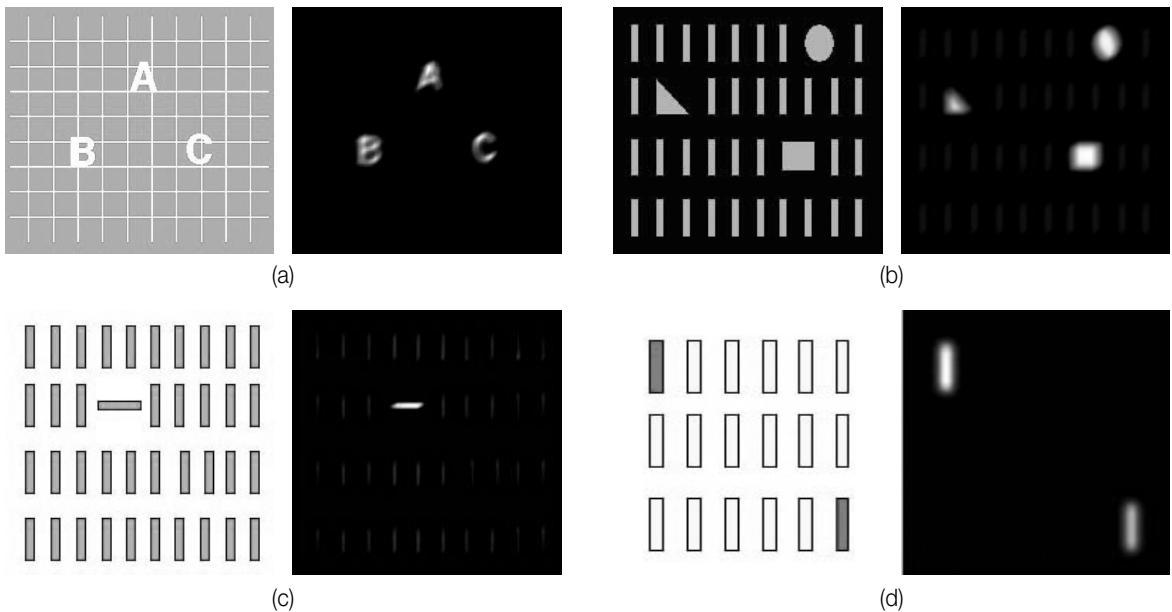
### 결론

본 논문에서는 컬러영상이 입력되었을 때 다른 부분과 현저히 다른 영역을 탐지해내는 돌출맵 모형을 제안하고, 이 모형을 향후 정밀한 분석이 필요한 몇몇 후보 영역을 찾아내는 문제에 적용하여 보았다. 제안하는 모형은 기본적으로 특징추출과 특징통합의 두 단계로 이루어져 있다. 특징추출 단계에서는 색상 또는 밝기 같은 아주 간단한 기본특징맵들이 입력 영상으로부터 먼

저 추출되고, 추출된 독립적인 기본특징맵들이 각각의 맵으로부터 방위를 탐지하고, 주변과 크게 다른 영역을 부각시키기 위하여 방위를 가진 ON-중심, OFF-주변 연산자에 의해 처리되어 특징맵을 구성한다. 이렇게 계산된 각각의 특징맵들은 특징통합 단계에서 각 특징맵을 구성하고 있는 특징값들의 국부적인 경쟁 관계와 통계적 정보를 이용하여 서로 통합되어지는데, 이러한 통합 과정을 통하여 목표로 하는 영역을 탐지하기에 유용하지 않은 특징은 자연스럽게 걸러지고, 그렇지 않은 정보는 부각되게 된다.

제안하는 시스템의 전체적인 성능을 평가하기 위하여 크게 2가지 부류의 컬러영상을 사용하였는데, 그 첫 번째가 서로 다른 범위에서 촬영된 외부환경에서 일반적으로 볼 수 있는, 우리의 주의를 끌만한 현저한 대상 외에 국부적인 조명변화 및 복잡한 배경을 함유하고 있는 실영상(real image)이며, 두 번째가 인간의 시각탐색에 관한 심리학적인 실험에 사용되는 간단한 인공영상(synthetic image)이다. 실험결과를 검증하기 위하여 실험에 사용한 실영상들에 대해 인간의 행동을 측정하는 실험을 추가적으로 수행하여 시스템에서 나온 결과와





**Figure 4** 인공영상에 대한 실험결과 예 : (a)와 (b)는 모양 및 크기가 다른 대상 탐지의 예를 보여주는데, 여기서 주의가 가는 곳은 (a)의 경우, 3개의 알파벳이며, (b)의 경우 다른 것들보다 크기나 모양이 다른 삼각형, 사각형, 원이다. 또한 (c)는 모양과 크기 및 색상은 모두 같지만, 방위가 현저히 다른 대상을 탐지하는 문제에, (d)는 방위는 같지만 색상에 차이가 나는 대상을 탐지하는 문제(distractor:황색 target:적색)에 본 모형을 적용한 경우이다. 특히 (a)와 (c)는 배경과 전경의 색이 서로 반대인 영상((a):녹색 배경, 흰 전경 (c):흰 배경, 녹색 전경)에 대한 실험으로써 결과를 통해 본 모형이 항상 높은 값을 가진 대상을 돌출성이 가장 큰 대상으로 간주하지 않는다는 것을 보여준다.

비교하였다. 실험결과를 통해, 시스템은 색상정보와 밝기정보만을 특징으로 사용하는 아주 간단한 구조를 가지고 있으면서도 복잡한 실영상에서도 영상내의 지각적으로 중요하다고 여겨지는 주요영역을 만족스럽게 탐지해 낼 수 있었다. 더욱 흥미로운 사실은 제안하는 시스템이 주어진 영상 안에서 돌출성이 있는 대상에 대한 어떠한 사전 지식을 사용하지 않고, 또한 조율이 되어 있지 않았는데도 다른 부분과 현저히 다른 대상을 제대로 탐지할 수 있다는 것이다. 특정 대상 탐지에 적합하도록 조율된 많은 컴퓨터 시각 알고리즘이 존재하지만 이러한 알고리즘들은 탐지하도록 조율된 대상 이외의 다른 형태를 가진 대상은 탐지하지 못한다. 이러한 내용을 바탕으로 한다면 제안하는 시스템은 제공되는 특징맵만을 쉽게 변형시키는 것만으로 외부환경의 다양하고도 복잡한 실영상을 대상으로 일반적인 대상 탐지기로 효율적으로 사용될 수 있다.

현재 제안된 시스템이 실질적인 여러 응용분야에서 효과적으로 적용할 수 있도록 보다 다양한 범위의 영상

에 대한 종합적이고 체계적인 성능분석을 수행하고 있으며, 컬러영상뿐 아니라 명암도 영상에서도 효과적으로 수행될 수 있는 일반적인 시스템으로 확장하는 작업을 수행 중이다.

## 참고문헌

1. Koch, C. & Ullman, S. Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry, *Human Neurobiology*, 4, 219-227, (1985).
2. Bear, M., Connors, B. W., & Paradiso, M.A., *Neuroscience exploring the brain*, Williams and Wilkins, USA, (1996).
3. Colby, C., *The neuroanatomy and neurophysiology of attention*, *Journal of Child Neurology*, 6, 90-118, (1991).
4. Yarbus, A.L., *Eye Movements and Vision*, Premium Press, New York, (1967).
5. Stewart, B., Reading, I., Thomson, M., Wan, C., & Binnie, T., Directing attention for traffic scene analysis, *Proc. of Intl. Conf. on Image Processing and Its*

- Applications, 801-805, (1995).
6. Chapman, D., Vision, Instruction, and Action, Ph.D. Thesis, AI Laboratory, Massachusetts Institute of Technology, (1990).
  7. Itti, L., & Koch, C., A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research*, 40(10-12), 1489-15068, (2000).
  8. Laar, P., Heskes, T., & Gielen, S., Task-Dependent Learning of Attention. *Neural Networks* 10, 6, 981-992, (1997).
  9. Cave, K., & Wolfe, J., Modeling the Role of Parallel Processing in Visual Search. *Cognitive Psychology*, 22, 225-271, (1990).
  10. Cave, K., The featuregate model of visual selection, *Psychological Research*, 62, 182-194, (1999).
  11. 김민식, 시각적 선택에 대한 신경망 모형 : FeatureGate 모형의 하향적 기제, *한국 인지과학회 논문집*, 10(3), 1-15, (1999).
  12. Milanese, R., Wechsler, H., Gil, S., Bost, J., & Pun, T., Integration of Bottom-up and Top-down Cues for Visual Attention Using Non-Linear Relaxation, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 781-785, (1994).
  13. Olivier, S., Yasuo, K., & Gordon, C., Development of a Biologically Inspired Real-Time Vis-ual Attention System. In: Lee, S.W., Buelthoff, H.-H., Poggio, T. (eds.): *BMCV 2000. Lecture Notes in Computer Science*, 1811. Springer-Verlag, Berlin Heidelberg New York, 150-159, (2000).
  14. Treisman, A.M., & Gelade, G.A., A Feature-integration Theory of Attention. *Cognitive Psychology*, 12, 97-136, (1980).
  15. Mozer, M., *The Perception of Multiple Objects : a Connectionist Approach*, MIT Press, Cambridge, MA, (1991).
  16. Sandon, P., Simulating Visual Attention, *Journal of Cognitive NeuroScience*, 2(3), 213-231, (1990).
  17. Olshausen, B., Essen, D., & Anderson, C., A neurobiological model of visual attention and Invariant pattern recognition based on dynamic routing of information, *NeuroScience*, 13, 4700-4719, (1993).
  18. Yagi, T., Asano, N., Makita, S., & Uchikawa, Y., Active vision inspired by mammalian fixation mechanism, *Intelligent Robots and Systems*, 39-47, (1995).
  19. Aguilar, M., & Ross, W., Incremental art: A neural network system for recognition by incremental feature extraction, *Proc. of WCNN-93*, (1993).