

주의 기반 시각정보처리체계 시스템 구현을 위한 스테레오 영상의 변위도를 이용한 새로운 특징맵 구성 및 통합 방법

박민철* · 최경주**

요약

인간의 시각 주의 시스템은 주어진 시각장면을 모두 다 처리하기보다는 주의가 집중되는 일정한 작은 영역들을 순간적으로 선택하여 그 부분을 순차적으로 처리함으로써 복잡한 시각장면을 단순화시켜 쉽게 분석할 수 있는 능력을 가지고 있다. 본 논문에서는 주의 기반 시각정보 처리체계 시스템 구현을 위한 새로운 특징맵 구성 및 통합 방법을 제안한다. 제안하는 시스템에서는 시각특징으로서 색상, 명도, 방위, 형태 외에 2개의 스테레오 영상 쌍으로부터 얻을 수 있는 깊이 정보를 추가하여 사용하였다. 실험결과를 통해 깊이 정보를 사용함으로써 주의 영역의 정탐지율이 개선됨을 확인하였다.

키워드 : 주의영역, 스테레오 영상, 깊이 정보

A Novel Feature Map Generation and Integration Method for Attention Based Visual Information Processing System using Disparity of a Stereo Pair of Images

Min-Chul Park* · Kyungjoo Cheoi**

ABSTRACT

Human visual attention system has a remarkable ability to interpret complex scenes with the ease and simplicity by selecting or focusing on a small region of visual field without scanning the whole images. In this paper, a novel feature map generation and integration method for attention based visual information processing system is proposed. The depth information obtained from a stereo pair of images is exploited as one of spatial visual features to form a set of topographic feature maps in our approach. Comparative experiments show that correct detection rate of visual attention regions improves by utilizing depth feature compared to the case of not using depth feature.

Keywords : ROI(Region Of Interest), FOA(Focus Of Attention), Depth Information, Stereo Pair of Images

1. 서론

인간은 효율적인 시각정보처리체계 덕분에 색깔, 움직임, 방향, 모양, 깊이감 등과도 같은 다양하고 복잡한 정보를 외부환경으로부터 끊임없이 제공받는데도 불구하고 이를 실시간으로 해석하고 이해하는 데 큰 어려움을 겪지 않는다. 실제로 인간은 실세계로부터 입력되는 시각정보 중 응시할 영역을 선택적으로 결정하는 기능과 응시할 영역으로 신속하

게 시선을 이동시키는 안구운동 기능을 통해 입력되는 복잡한 시각영상정보를 효율적으로 처리한다고 알려져 있다[1]. 인간에게서 보여지는 이러한 선택적 주의 능력은 영상압축, 인덱싱, 목표물 탐지 시스템, 네비게이션 보조 시스템, 로봇틱스 컨트롤 시스템, 영상검색 등 다양한 분야에 응용되어져 왔다.

선택적 주의에 관한 연구로는 크게 분류하면 인간의 선택적 주의에 대한 심리학적인 계산 모형[2-3], 어느 정도의 선까지만 생물학적인 증거에 기반을 두고, 실질적인 데이터를 가진 응용 분야에 적용하기 위한 컴퓨터 시각에서의 시각적 주의 시스템[1, 4-9] 등이 있다. 초기 선택적 주의 모형은 심리학, 신경생리학, 정신물리학 분야에서 우선 연구되어져 왔다. Treisman이 1980년대 제안한 특징통합이론(Feature

* 이 논문은 2008학년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.
† 정 회 원 : 한국과학기술연구원 포토닉스센서시스템센터 선임연구원
** 중신회원 : 충북대학교 전자정보대학 컴퓨터공학부 조교수(교신저자)
논문접수 : 2009년 8월 13일
수정일 : 1차 2009년 10월 8일, 2차 2009년 11월 16일, 3차 2010년 1월 4일
심사완료 : 2010년 1월 4일

Integration Theory)은 시각탐색(visual search)에 관련된 몇몇 실험을 통해 얻어진 것으로 영상의 각 영역은 색, 모양, 등의 기본 특징들의 조합으로 표현되며 이를 이용하여 영역으로의 inhibition 및 주의 집중이 발생된다고 하였다[10]. 여기서, 특징들은 병렬적이고 자동적으로 추출되며, 지각적 개체가 특징들을 형성하는 순서로 특별한 공간적 자취(locus)에 주의(attention)가 가해진다는 기본적인 가정을 하고 있다. 또한 Itti는 인간이 가지고 있는 선택적 주의 기능을 수행하는 데 있어서 다른 영역과 구별되는 현저한 장소에 무의식적으로 자연스럽게 주의가 집중되는 방식인 상향식(bottom-up) 방식과, 대상에 대한 정보를 미리 알고 있어서 그 대상으로 주의가 의식적으로 집중하게 되는 방식인 하향식(top-down) 방식이 동시에 사용된다고 말하였다[11].

기존의 컴퓨터시각에서 사용되는 많은 선택적 주의 시스템은 상향식(bottom-up) 방식과 하향식(top-down) 방식으로 설계되었다. 상향식 방식으로 설계된 시스템은 설계된 시스템의 구조를 크게 변화시키지 않고도 다양한 응용분야에 적절히 사용되어질 수 있다. 하지만 하향식 방식으로 설계된 대부분의 시스템들은 주어진 작업에 따른 주의집중 과정을 학습할 수 있는 방법이 요구되며, 인식단계와의 부분적인 상호작용이 필요하여 특정 목적의 응용분야에만 효과적으로 사용될 수 있다. 이러한 점들로 인해 하향식 단서에 대한 중요성은 계속 대두되었지만 이를 사용하는 시스템에 대한 연구가 많이 진행되지 못하였다.

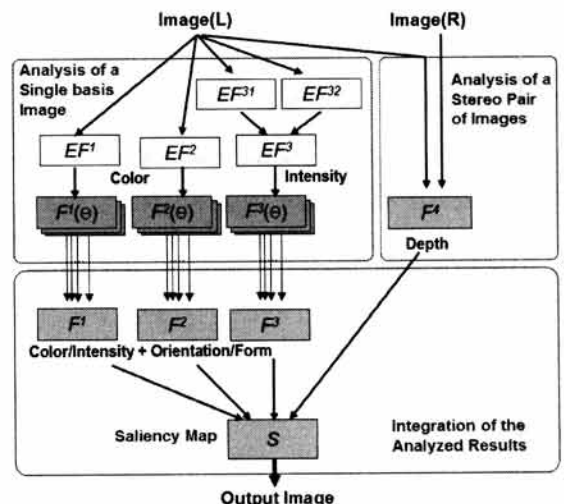
기존의 전형적인 선택적 주의 시스템은 상향식 방식과 특징통합이론에 기반을 둔 것이다. 이러한 시스템들은 주의 영역을 찾기 위해 대뇌의 1차 시각피질에서 나타나는 저수준(low-level)의 영상 특징을 사용한다[1]. 뇌의 V1 영역의 세포들은 방위, 공간 주파수, 움직임 방향, 색상에 선택적으로 반응한다고 알려져 있으며, 이러한 특징이 저수준의 영상 특징이 될 수 있다. 주의 영역은 정지된 한 장의 영상으로부터 색상, 밝기와도 같은 공간적 특징을 분석하여 현저도를 계산해내고 이를 2차원 영상 형태로 구축한 공간 현저도맵(Spatial Saliency Map)을 토대로 찾을 수 있다. 상향식 방식으로 설계된 시스템 중 대표적인 것이 Itti가 개발한 돌출맵과 승자전취망(winner-take-all network)[11]으로 구성된 시스템이다. 이 시스템에서는 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 지정하였다. 하지만 이 시스템은 주의가 가는 영역을 회피하게 가이드 하는 역할만을 수행하는 것이어서 컴퓨터시각에서 실제적으로 인식과도 같은 고차원적인 작업을 수행하기 위해서는 영역분할과도 같은 후처리 작업에 상당한 노력이 필요하다. 이러한 선택적 주의 시스템은 시간특징 중 하나인 모션정보를 상향식 단서로 사용하여 동영상에서도 사용할 수 있도록 확장되기도 하였다[9].

영상 내에 주의를 일으키는 물체들이 다수 존재하고, 이들이 비슷한 공간적 특성을 보이는 경우에는 현저도의 부여 기준이 애매해진다. 또한 주의를 일으키는 객체들 간에 부분적 겹침이 발생할 경우, 기존의 방식으로 얻은 공간 현저도맵은 개별 물체들을 하나의 영역으로 탐지하게 된다. 이는 실제 인간의 시각반응과는 상이한 결과이므로, 비슷한 현저도값을 보인다 하더라도 다소간의 겹침 때문에 각 주의 영역의 개별성을 잃는 상황을 피해야 한다. 이러한 문제들을 해결하기 위해 시각주의 시스템에 실세계에 대한 더 많은 정보가 더해질 필요가 있다. 본 논문에서는 추가적인 정보로서 영상의 3차원적 깊이를 활용한다. 2차원적인 공간 현저도 결과를 객체들간의 3차원적 위치관계에 대한 정보를 활용하여 보상함으로써 현저도맵이 실제 환경을 더욱 잘 반영할 수 있다. 보다 높은 차원의 3차원 정보를 이용함으로써 인간의 주의 영역에 보다 부합되는 주의 영역을 추출해낼 수 있다. 깊이 특징이 양한 거리지각에 있어 기반이 된다는 것은 이미 알려진 사실이지만, 주의 시스템에서 특징으로 사용되어진 예는 거의 없다. 3차원 정보를 사용하게 되면 3D TV와도 같은 3차원 입체 영상을 대상으로 한 여러 응용분야에도 사용될 수 있을 것이다.

2장에서 제안하는 시스템에서 대해 설명하고, 3장에서 실험 및 결과를, 4장에서 결론을 맺는다.

2. 제안하는 시스템

제안하는 시스템에서 전반적인 동작과정은 (그림 1)과 같이 크게 '분석'과 '통합'의 두 부분으로 구성되어 있다. '분석' 단계에서 인간의 지각처리방식에 근거하여 색상, 명도, 방위, 형태, 깊이에 반응하는 서로 독립적인 특징이 추출되어 기본특징맵이 구성되고, '통합'단계에서 기본특징맵이 재구성 및 통합되어 하나의 현저도맵이 만들어진다. 이 현저도맵이 제안하는 시스템에서의 출력이 된다. 대부분의 경우 왼쪽 영상이 스테레오 영상의 기준 영상이 된다.



(그림 1) 전반적인 시스템 흐름도

2.1 기준영상의 분석

스테레오 영상의 기준 영상을 대상으로 여러 초기 시각특징들을 추출한다. 각각의 초기 시각 특징에 대한 분석 및 통합은 각 특징과 관련된 기존의 심리학적인 연구결과를 바탕으로 이루어진다.

(1) 색상 및 명도대비특징

색상대비특징은 인간의 색상지각능력에 대한 연구결과 [12]를 바탕으로 하였다. 색상대비(contrast)에 강하게 반응하는 동질(homogeneous) 형태의 수용야를 가진 색상대립세포 R+G-와 B+Y-를 모델링하였는데, R+G-기체의 경우, R에서는 활성화가 되고 G에서는 활성이 억제되어 R+G-세포가 흥분하면 적색을 지각하게 된다. 명도대비특징으로는 어두운 곳에서는 밝은 곳을 찾기 위해 입력되는 명도값 그대로를 쓰고, 밝은 곳에서는 어두운 곳을 찾기 위해 원래 값의 반전값을 사용하도록 하였다. 특징 추출방법을 개괄적으로 설명하면 다음과 같다. R, G, B를 각각 적색, 녹색, 청색을 나타내는 추상체 반응이라 가정하면, 3가지 추상체 R, G, B로부터 식(1)과 같이 다른 색상과 조금도 희석되지 않은 순수한 색상에 조율된 채널 red, green, blue, yellow, lightness를 생성한다.

$$\begin{aligned} red &= R - (G+B)/2, & green &= G - (R+B)/2, & blue &= B - (R+G)/2 \\ yellow &= R+G-2 \cdot (|R-G|+B) \\ lightness &= (R+G+B)/3 \end{aligned} \quad (1)$$

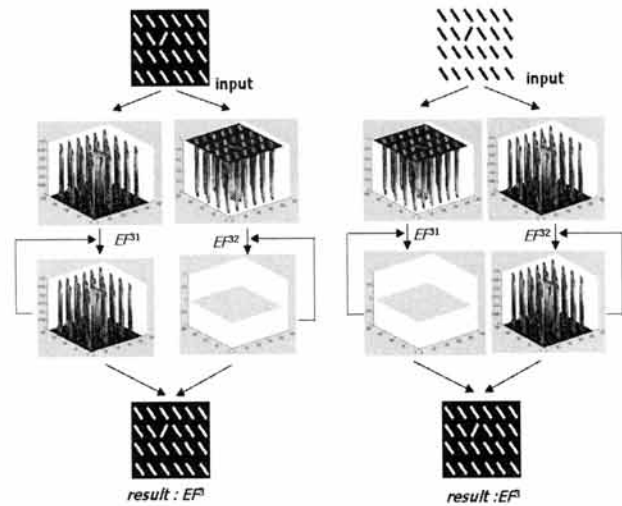
여기서 red, green, blue, yellow는 각각 순수한 색상에 조율된 적색, 녹색, 청색, 황색 채널을 나타내며, 각 채널은 흑색과 백색 자극에 대해 0을 출력한다. 이러한 채널을 사용하여 '적/녹'대립 red-green와 '청/황'대립 blue-yellow를 만든다. red와 green, blue와 yellow는 보색관계에 있기 때문에 red-green, blue-yellow의 2가지 값을 통해 특이한 색상이 나타나는 부분을 찾을 수 있다. EF^1 은 적녹보색대비특징맵을, EF^2 은 청황보색대비특징맵을 나타낸다. 명도대비특징은 어두운 곳에서 밝은 곳을 찾는 ON 연산자 EF^{31} 과, 밝은 곳에서 어두운 곳을 찾는 OFF 연산자 EF^{32} 를 사용한다. EF^{31} 과 EF^{32} 은 다시 하나의 명도대비특징맵인 EF^3 으로 구성되는데 그 방법은 다음과 같다. 먼저, 생성된 각 $EF^{3k}(k=1,2)$ 마다 전역적인 최대값인 M^{3k} 와 열 단위로 최대값들의 평균값인 m^{3k} 을 계산한다. 계산된 값을 이용하여 각각의 EF^{3k} 의 특징값을 식 (2)과 같은 방법으로 업데이트시킨다.

$$EF^{3k} = EF^{3k} \times (M^{3k} - m^{3k})^2 \quad (2)$$

마지막으로 업데이트 된 각각의 EF^{3k} 을 정규화하는데 이때 정규화는 자기 자신의 맵 뿐만 아니라 다른 맵의 정보도 고려하여 정규화된다. 다시 말해 EF^{31} 을 정규화하기 위해 EF^{31} 을 구성하는 특징값의 최대값과 최소값만을 사용하지

않고, EF^{31} 과 EF^{32} 를 구성하고 있는 모든 특징값 중 최대값과 최소값을 사용한다. 이러한 과정을 3번 반복하여 얻은 EF^{3k} 를 모두 선형조합하여 EF^{3k} 을 만든다.

이 방법은 좁은 영역에서의 활동량이 많은 특징값들을 부각시키고, 그렇지 않은 특징값들은 억제시키는 효과를 주며, 같은 맵이 아닌 다른 맵에서의 특징값도 고려하여 정규화시키는 과정을 포함하였기 때문에 두드러지지 않는 특징값들은 자연스럽게 걸러지는 효과가 있다. 이러한 방식으로 통합한 결과 예가 (그림 2)에 나타나있다. 배경이 어두운 영상이 입력되었을 경우에는 밝은 부분이, 배경이 밝은 영상이 입력되었을 경우에는 어두운 부분이 명도대비특징맵으로서 채택된다.



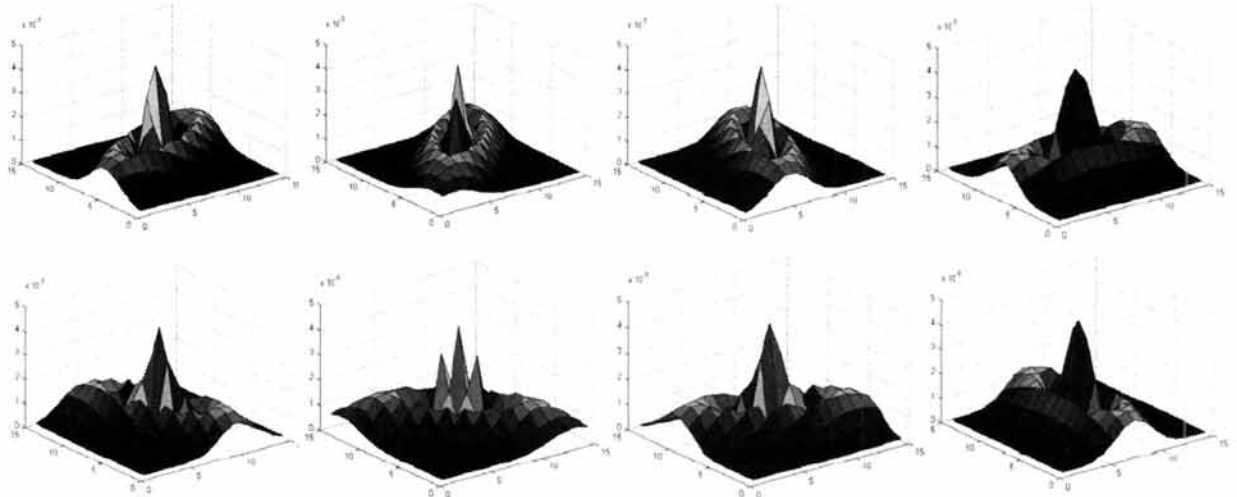
(그림 2) 예제를 통해 본 명도대비특징맵의 통합 처리 과정

(2) 색상 및 명도대비특징 별 방위 및 형태 특징

인간의 형태지각은 머리에서 일어나는 인식 이전의 과정으로서 물체를 형성하는 것이 아니라 덩어리나 경계선 등의 구분 정도를 나타낸다. 본 논문에서는 추출되어진 색상대비 특징들로부터 방위를 탐지하고, 덩어리나 경계선 등의 구분 정도를 확실히 드러내게 하기 위해 색상 및 분석과정 중 추출된 특징에 대하여 인간의 동심원 형태의 "ON-중심, OFF-주변" 수용야(Receptive Field)에서 보이는 세포 반응도를 모방한 중심-주변 연산을 수행한다[12]. 본 논문에서는 8가지 방위($\theta \in \{0, \pi/8, 2\pi/8, \dots, 7\pi/8\}$)를 가진 ON-중심, OFF-주변 연산자인 $h(\theta)$ 필터를 사용하였다. 각 방위별 $F^k(\theta)$ ($k=1, \dots, 3$)를 생성하는 방법은 식 (3)과 같다.

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

식 (3)에서, 각각의 색상대비특징맵(EF^1, EF^2) 및 명도대비특징맵(EF^3)은 $h(\theta)$ 필터와 회선되어지며, 회선 후 회선된 영상의 대비를 증가시키기 위해 회선된 결과를 제공한다. 여기서 $h(\theta)$ 필터는 식 (4)와 같이 계산되며, 식 (4)를 통해 생성된 8개의 $h(\theta)$ 필터는 (그림 3)과 같다.



(그림 3) 생성된 8개의 h(θ) 필터

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

식 (4)에서, $G_{x,y}(\cdot, \cdot, \cdot)$ 은 방위를 가진 2차원 가우시안 함수이며, K_1, K_2 는 양수, r_1 은 2개의 가우시안의 이심률, r_2 는 ON과 OFF 가우시안 간의 폭 비율을 뜻한다.

이 과정을 통하면 모두 24개의 특징맵이 생성된다. 이 특징맵이 특징통합과정을 거쳐 하나의 현저도맵으로 재구성된다.

2.2 스테레오 영상 분석

인간의 시각체계는 서로 다른 위치에서 획득된 2개 영상을 적절히 정합함으로써 거리정보를 얻는 것으로 알려져 왔다. 스테레오 정합은 인간 시각 체계의 거리 추출 능력을 자동화하기 위한 컴퓨터시각 분야 중 하나이다. 본 논문에서는 3차원 깊이 정보를 간단히 얻기 위하여 SAD(sum of absolute difference)를 이용하였다. (그림 4)는 사용되어진 좌우영상과 이로부터 얻어진 깊이 특징맵이다.



(그림 4) 사용된 좌우영상과 이를 통해 생성된 깊이 특징맵

2.3 분석된 결과 통합

지금까지 기준영상인 왼쪽영상으로부터 구성된 24개의 색상/명도 별 방위/형태 특징맵과 좌우 영상으로부터 구성된 깊이 특징맵이 구성되었다. 특징맵을 구성하고 있는 특징값들은 모두 입력된 영상의 각 지점마다 각각 특별한 의미를 가지고 있지만 각각의 특징들은 모두 서로 다른 특징추출방법으로 추출되어진 서로 독립적인 것들이기 때문에 이를 기

반으로 주의 영역을 알아내기란 쉬운 일이 아니다. 즉 24개 각각의 특징이 제공해주는 정보는 모두 한 종류의 영상에 대한 것이므로, 각각의 특징마다 모두 각각 다른 영역을 주의 영역이라고 가이드해 줄 수도 있다. 이러한 문제를 해결하기 위해서는 영상의 각 지점에 있어서의 서로 다른 특징값들이 모두 하나로 통합되어져서 영상의 각 지점마다 하나의 전역적인 중요도 측정값인 Saliency가 계산되어져야 한다.

먼저 (그림 1)에서처럼 24개의 색상/명도 별 방위/형태 특징맵을 3개의 특징맵으로 통합하여 F^1, F^2, F^3 를 만든다. $F^k(\theta)(k=1, \dots, 3)$ 은 EF^k 에 식 (3)을 처리한 결과인데, 이는 위에서도 언급했듯이 8가지 방위와 함께 형태를 두드러지게 하는 연산이다. 이 연산을 통해 나온 각 맵별 8가지 특징 중 가장 두드러지는 특징을 가지는, 즉 두드러진 방위를 나타내는 특징맵을 선택한다. 이 과정은 다음과 같이 설명된다. $F^k(\theta)$ 의 최대값과, $F^k(\theta)$ 를 구성하는 각각의 수직선별 최대값의 평균을 구한다. 수직선별 최대값의 평균이 크다면, 맵에 전체적으로 0이 아닌 값이 고르게 분포되어 있는 것으로 판단할 수 있다. 역으로 평균값이 작다면, 값을 나타내는 부분은 맵 전체 면적 중 일부에 국한되어있다고 볼 수 있다. $F^k(\theta)$ 의 최대값에 비해 이 평균값이 작을수록, 강한값이 좁은 영역에 집중적으로 분포한다. 이러한 특징을 보이는 맵이 선별대상이 된다. 이 과정을 8장의 맵에 적용하여, 가장 특징이 두드러진 현저도가 높은 맵을 선택할 수 있다.

이렇게 하여 통합된 특징맵 F^1, F^2, F^3 과 깊이 특징맵 F^4 는 다시 최종적인 하나의 현저도맵 S로 통합되어져야 한다. 본 연구에서는 추출된 특징들이 모두 중요하다고 가정하였기 때문에 [7]에서의 방법처럼 두드러진 특징 특징맵만을 선택하지 않고, 특징의 두드러진 정도에 계산하여 이에 따라 가중치를 다르게 부여하여 결합하였다. 두드러짐 강도에 따라 높은 가중치를 부여하였고, 두드러짐의 정도가 차이가 나지 않을 경우에는 차이가 나지 않는 특징맵은 서로 같은 가중치를 부여하였다. 두드러짐의 강도는 다음과 같은 식 (5)와 같이 구한다.

$$SI_{x,y}^k = \frac{MF_{x,y}^k - MinMF}{MaxMF - MinMF} \quad (5)$$

여기서 $SI^k(k=1,\dots,4)$ 는 4개의 특징맵인 F^k 의 두드러진 정도를 나타내는 맵으로써, 4개의 특징맵의 통계적인 정보를 이용하여 업데이트한 것이다. $MaxMF$ 와 $MinMF$ 는 각각 $MF^1 \sim MF^4$ 를 구성하고 있는 특징값 중 가장 큰 값과 가장 작은 값을 뜻한다. $MF^k(k=1,\dots,4)$ 는 식 (6)과 같이 계산되어 질 수 있다. 여기서 $MaxF^k$ 는 각 F^k 마다 이를 구성하고 있는 특징값 중 최대값을, M 은 입력되는 F^k 의 크기를 뜻한다.

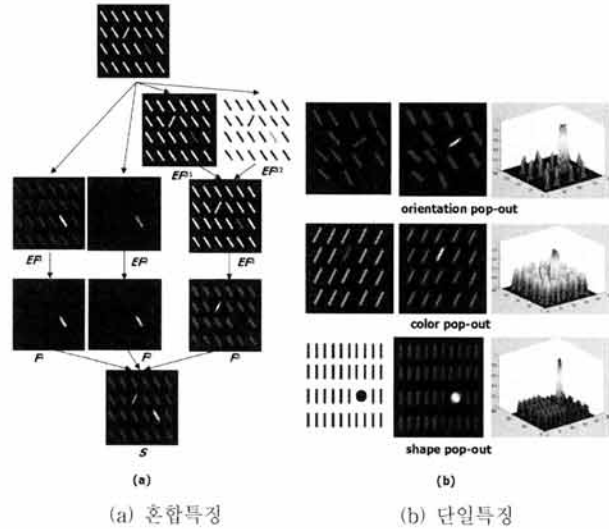
$$MF_{x,y}^k = F_{x,y}^k \times \left[MaxF^k - \frac{1}{M-1} \left[\left(\sum_{x,y} F_{x,y}^k - MaF^k \right) \right] \right] \quad (6)$$

3. 실험 및 결과

제안하는 시스템의 성능을 평가하기 위하여 색상, 강도, 방위, 형태, 깊이 등 모든 특징을 가지고 있는 스테레오 영상 외에도 명도값만을 가지는 명도 영상, 깊이를 가지지 않는 색상 영상 등 다양한 종류의 외부 실영상을 대상으로 실험하였다. 제안하는 시스템의 성능 평가에서 중요한 것은 얼마나 다양한 영상을 대상으로 실험을 했는가도 중요하지만 영상이 주어졌을 때 시스템이 인간이 지각적으로 중요하다고 생각되는 영상의 영역들을 얼마나 잘 알려주는가에도 달려 있다고 볼 수 있다. 이러한 이유로 인해 본 논문에서는 시스템의 전반적인 동작과정을 확인하기 위하여 비교적 간단한 인공 영상(synthetic image) 뿐만 아니라, 외부환경에서 일반적으로 볼 수 있는 복잡하고도 다양한 실 영상(real image)들에 더욱 초점을 맞추어 실험을 수행하였다.

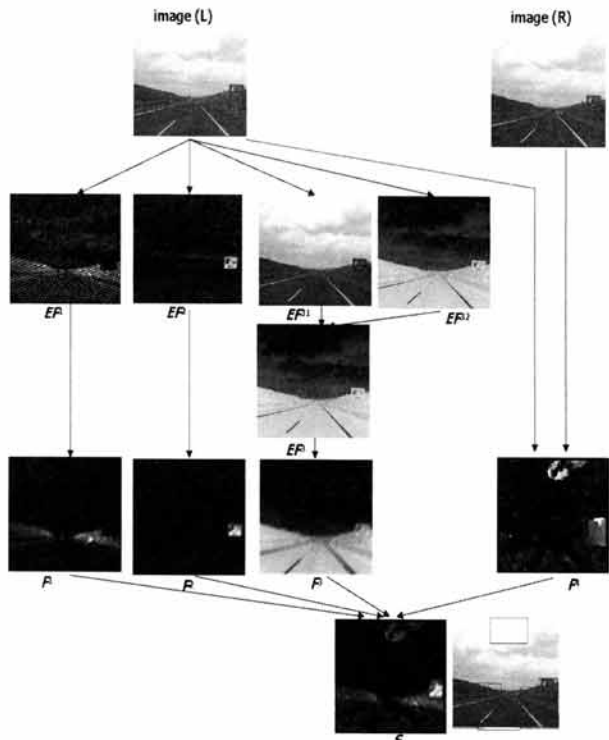
간단한 인공영상은 깊이 정보가 없는 단일영상을 사용하였는데, 기존의 심리학자들에 의해 인간의 '시각탐색' 전략을 알아보기 위한 실험에서 주로 사용하는 'Color Pop-Out', 'Orientation Pop-Out' 문제에 관련된 영상을 사용하여 본 시스템도 인간에게서 보이는 색상 및 방위 pop out이 보이는지 실험하였다. 또한 이러한 영상 외에도 색상, 크기, 밝기 등등에 변화를 준 다른 종류의 영상도 실험영상으로 사용하여 'Shape Pop-Out' 및 여러 가지 특징이 결합된 상태에서의 주의 영역 탐지에 대한 부가 실험을 수행하였다. 본 논문에서는 방위 면에서 $0^\circ \sim 360^\circ$ 사이의 8가지 형태의 방위 변화를 준 영상을, 색상 면에서 적색, 청색, 녹색, 황색, 흑색, 백색 계열의 색상 변화를 준 영상을 사용하였으며, 크기에 변화를 준 영상, 색상정보가 없는 명암도 영상, 배경이 표적보다 어두운 영상, 또는 이와 반대로 배경이 표적보다 밝은 영상, 그리고 위의 특징들을 혼합해 놓은 영상 등을 선정하여 사용하였다.

(그림 5) (a)는 깊이 정보가 없는 간단한 단일 인공영상에 대한 실험 예이다. 이 영상은 검은 배경에 오렌지색의 막대바 23개와 적색의 막대바 1개가 있는 영상으로 색상과



(그림 5) 깊이정보가 없는 간단한 단일 인공영상의 실험 예

방위면에서 두드러진 특징을 가지고 있다. 인간을 대상으로 실험한 결과 이 영상의 경우 30명의 피험자 중 25명이 가장 눈에 띄는 부분은 적색바이며, 그 다음이 방위가 틀린 오렌지 막대바이라고 대답하였다. 시스템의 처리과정으로 보면 색상 특징맵에서 두드러진 색상 특징을 잡아내고, 색상을 모두 없앤 명도특징맵에서 방위를 찾아낸다. 두드러짐의 강도는 색상이 우선하고 그 다음 방위임을 알 수 있다. (그림 5) (b)는 각각 방위, 색상, 모양 면에서 두드러진 단일 특징을 가지고 있다. 제안하는 시스템에서도 두드러진 특징을 탐지함을 알 수 있다. (그림 5)에서도 확인할 수 있듯이 제



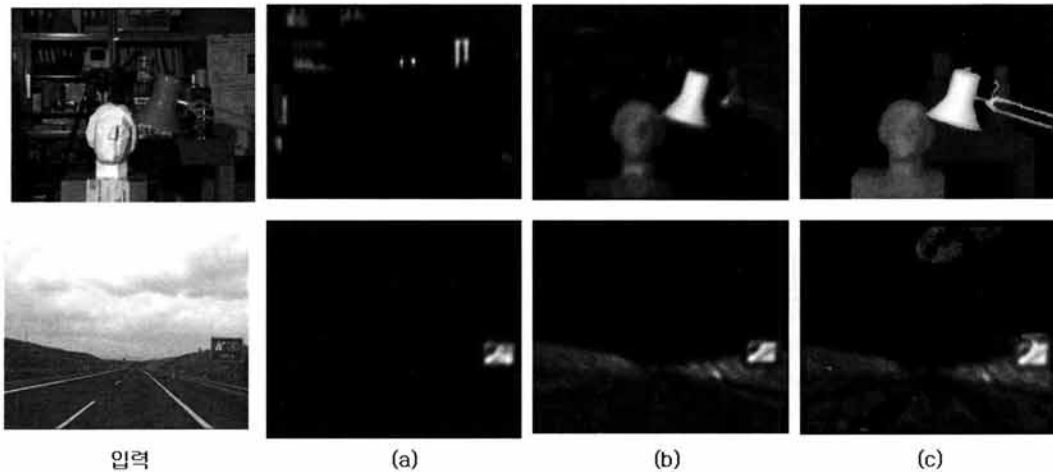
(그림 6) 스테레오 영상이 입력되었을 때의 구성되는 맵의 처리 흐름도

안하는 시스템은 주의 영역을 찾아내는 동시에 영역의 형태로서 객체의 현저도 강도를 계산한다. 사실 몇몇 객체나 영역은 그 자체의 현저도만으로도 탐지될 수 있다. 결과적으로 시각장면에서 얻게 되는 객체나 영역은 우리의 시스템에서 현저도에 따라 기술될 수 있다. 이는 주의가 갈만한 영역만을 희미하게 가이드해주는 기존의 연구[1]와 달리 인식과도 같은 고차원적인 작업을 수행하기 위해서 해야 하는 후처리 작업이 매우 간단하다.

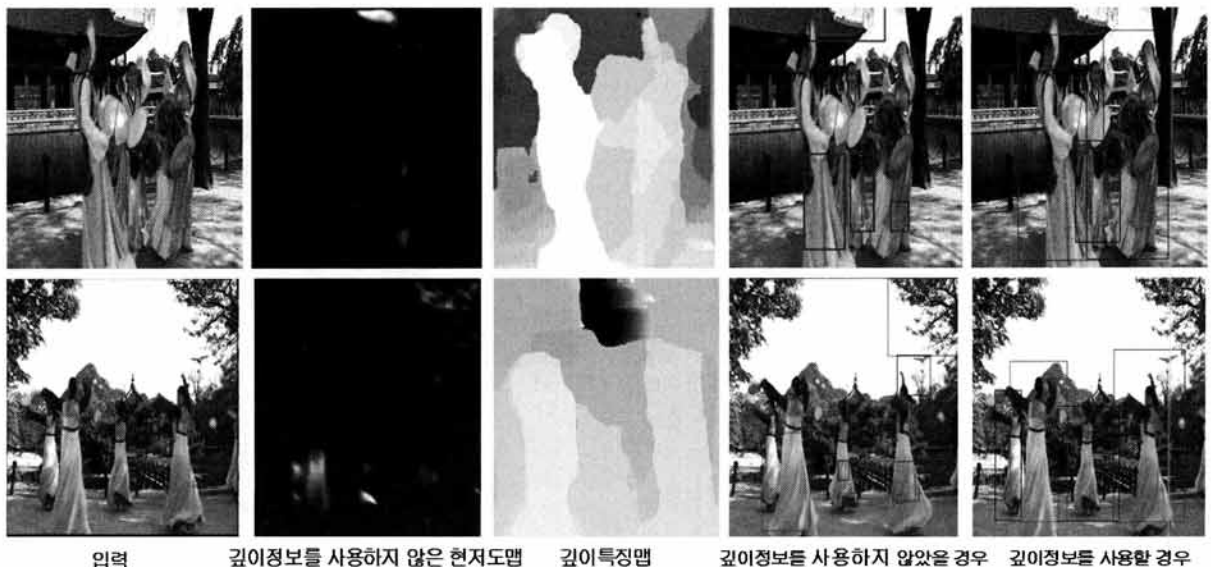
(그림 6)은 깊이정보가 포함된 스테레오 영상이 입력되었을 때의 시스템에서 구성되는 맵의 처리 흐름도를 나타낸 것이다. 단일영상인 왼쪽 영상으로부터 색상/강도, 형태/방위 특징을 찾아내고, 스테레오영상인 좌우쌍 영상으로부터 깊이특징을 찾아낸다. 구성되어진 모든 특징맵으로부터 각 장소별 두드러짐에 기반한 가중치 결합을 수행하여 하나의 현저도맵을 구성한다.

(그림 7)은 스테레오 영상에 대한 실험 예를 기존 연구 [7]과 비교한 것이다. (그림 7) (a)는 기존연구 [7]의 결과로, 두드러진 한 특징맵의 활동만을 부각시켜 주의강도가 가장 높은 한 영역만을 탐지한다던가, 혹은 아예 탐지하지 못하는 경우이다. 또한 (그림 7) (b)~(그림 7) (c)는 제안하는 시스템의 결과인데, (그림 7) (b)는 깊이 특징을 사용하지 않았을 경우의 제안한 시스템의 결과이며, (그림 7) (c)는 깊이 특징을 사용했을 경우의 제안하는 시스템의 결과이다. 깊이 특징을 사용함으로써 탐지되지 못했던 영역을 탐지할 수 있으며, 겹쳐지는 부분을 효과적으로 분리할 수 있음을 알 수 있다.

(그림 8)을 통해 깊이 특징맵이 시스템 결과에 미치는 영향을 알 수 있다. (그림 8)의 첫 번째 실험영상에서 객체들은 모두 부분적으로 겹쳐있으며, 모여있는 객체간의 거리는 크지 않다. 또한 두 번째 실험영상에서 영상의 2차원



(a) 기존 연구 [7]에서의 결과
 (b) 깊이 특징을 사용하지 않은 제안시스템의 결과
 (c) 깊이 특징을 사용한 제안시스템의 결과
 (그림 7) 스테레오 영상에 대한 실험 예



(그림 8) 깊이 특징을 이용한 경우와 그렇지 않을 경우의 제안 시스템 결과

평면상의 좌측에 두 객체가 일부 공간적으로 중복된 부분 (이하 겹침이라 표현)이 존재한다. 겹침의 정도는 크지 않고 객체간에는 분명하게 인지할 수 있을 정도의 거리(깊이)차가 있다. 그러나 깊이특징을 사용하지 않는 현저도 맵이 나타내는 결과로는 (그림 8)에서 알 수 있듯이 겹침이 있는 현저도 영역의 분리 여부가 불투명하다. 그러나 깊이 특징 맵으로부터 유의미한 Blob들을 선별하고 제안한 방법에 따라 현저도맵을 구성한 결과, 확실하게 주의영역을 분리해 내었다.

4. 결 론

본 논문에서는 스테레오 영상의 변위도 정보를 이용하여 성능을 개선한 주의 기반 시각정보처리체계 시스템을 제안하였다. 현저도맵은 영상의 색상이나 명도 등의 특징들을 다양한 방식으로 활용해서 구하는데, 이러한 저수준(low level)의 특징만으로는 영상 내의 구성요소들간의 의미적 관계를 파악하기 어렵고, 이 때문에 실제 인간의 인지반응과는 상이한 결과를 보일 수 있다. 색상, 명도, 방향 등의 정보는 영상 내의 물체들의 배치 형태, 전경, 배경 등에 관계없이 광범위 하고 무차별적으로 분포할 수 있으며, 영상의 구성이 복잡해질수록 그 경향은 강화되며 그에 따라 저수준의 정보들의 활용도는 필연적으로 낮아진다. 이러한 상황에서 저수준의 특징들을 사용하여 추출한 현저도 값의 신뢰도는 높다고 보기 힘들다. 이 때문에 순수한 현저도맵이 나타내는 결과만으로는 사실적인 주의영역을 탐지해 내기 어렵다.

본 논문에서는 시각특징으로서 색상, 명도, 방향, 형태 외에 2개의 스테레오 영상 쌍으로부터 얻을 수 있는 깊이 특징을 추가하여 사용하였다. 주의 영역을 추출하기 위한 토대가 되는 공간 현저도맵의 정확도와 신뢰성을 향상시키기 위해 3차원의 깊이정보를 활용하여 보상하는 방법을 제안하였다. 결과적으로 영상 내에서 물체들간에 부분적 겹침이 일어난 경우, 또는 물체들이 비슷한 형태를 보이는 경우 등 일반적으로 현저도의 강약 구분이 애매한 경우에도 각 영역들을 명확한 기준에 의해 다른 현저도로 기술할 수 있게 되었다.

현재 스테레오 영상 분석 결과와 단일 영상 분석 결과와의 통합방법에 대한 연구 및 더욱 효과적인 깊이맵 구성방법에 대한 연구가 수행중이며, 현저도맵을 통해 얻을 수 있는 주의 강도를 토대로 시선처리가 이루어질 수 있도록 하는 방안을 강구중이다. 또한 스테레오 영상이면서 움직임이 있는 동영상에도 사용가능하도록 시스템을 확장하고 있다. 현재 이루어진 성능 분석을 보강하는 작업도 별도로 수행되어야 한다.

참 고 문 헌

[1] L. Itti and C. Koch, "A saliency-based search mechanism

for overt and covert shifts of visual attention," *Vision Research*, Vol.40, No.10-12, pp.1489-1506, 2000.

- [2] M. S. Kim and K. R. Cave, "Top-down and Bottom-up Attentional Control : on the Nature of Interference from a Salient Distractor," *Perception and Psychophysics*, Vol.61, No.5, pp.1009-1023, 1999.
- [3] N. J. Cepeda, K. R. Cave, N. P. Bichot and M. S. Kim, "Spatial Selection via Feature-driven Inhibition of Distractor Locations," *Perception and Psychophysics*, Vol.60, No.5, pp.727-746, 1998.
- [4] Q. Tian, Y. Wu and T. S. Huang, "Combine user defined region-of-interest and spatial layout for image retrieval," *IEEE Int. Conf. on Image Processing*, Vol.3, pp.746-749, 2000.
- [5] M. Seo, B. Ko, H. Chung and J. Nam, "ROI-Based Medical Image Retrieval Using Human-Perception and MPEG-7 Visual Descriptors," *LNCS*, Vol.4071, pp.231-240, 2006.
- [6] J. Tsotsos, S. Culhane, Y. Winky, L. Yuzhong, N. Davis and F. Nufflo, "Modeling Visual Attention via Selective Tuning," *Artificial Intelligence*, Vol.78, pp.507-545, 1995.
- [7] M. C. Park and K. J. Cheoi, "Selective Visual Attention System Based on Spatiotemporal Features," *LNCS*, Vol.5068, pp.203-212, 2008.
- [8] J. Han, K. Ngan, L. Mingjing and H. Jhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.16, No.1, pp.141-145, 2006.
- [9] G. Boccignone, A. Chianese, V. Moscato and A. Picariello, "Foveated shot detection for video segmentation," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.15, No.3, pp.365-377, 2005.
- [10] A. M. Treisman and G. Gelade, "A Feature-integration Theory of Attention," *Cognitive Psychology*, Vol.12, No.1, pp.97-136, 1980.
- [11] L. Itti, C. Koch, and E. Niebur, "A Model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp.1254-1259, Vol.20, No.11, 1998.
- [12] L. Spillmann and John S. Werner, 'Visual Perception : the neurophysiological foundations,' Academic Press Inc., 1990.



박민철

e-mail : minchul@kist.re.kr

1993년 홍익대학교 전자공학과(공학사)

1997년 일본 동경대학 전자정보공학과
(공학석사)

2000년 일본 동경대학 전자정보공학과
(공학박사)

2001년~현 재 한국과학기술연구원 포토닉스센서시스템센터
선임연구원

관심분야: 멀티미디어, 3차원 영상 디스플레이, 컴퓨터비전 등



최 경 주

e-mail : kjcheoi@chungbuk.ac.kr

1996년 충북대학교 컴퓨터학과(이학사)

1999년 연세대학교 컴퓨터학과(공학석사)

2002년 연세대학교 컴퓨터과학·산업시스템공학과(공학박사)

2002년~2005년 LG CNS 연구개발센터

2005년~현재 충북대학교 전자정보대학 컴퓨터공학부 조교수

관심분야: 컴퓨터비전, 뇌정보처리, 유비쿼터스컴퓨팅 등