Journal of Korea Multimedia Society Vol. 14, No. 3, March 2011(pp. 378-391)

인간의 주의시각에 기반한 시각정보 선택 방법

최경주[†], 박민철[††]

요 약

본 논문에서는 입력장치로 들어오는 수많은 시각정보 중 현 시점에서 가장 유용하다고 생각되는 정보를 인간의 상향식 주의시각에 기반하여 선택하는 시각정보 선택기법에 대해 소개한다. 제안하는 시스템은 색상, 명도, 방위, 형태 등 저수준의 공간특징 외에 시간특징으로서 움직임 정보와 3차원 정보인 깊이 정보를 추가적으로 사용함으로써 기존방법에 비해 정보 선택의 정확도를 높혔다. 움직임 정보 추출 시 발생할 수 있는 노이즈를 제거하기 위해 인간의 움직임 인지에 대한 연구결과를 이용하는 새로운 접근법을 사용하였으며, 입력 영상 내 객체들이 부분적으로 겹쳐있다거나 동일한 현저도를 가지고 있을 때에도 현저한 영역을 제대로 선택해낼 수 있도록 깊이 정보를 사용하여 유의미한 영역을 선별하고 우선순위를 부여하였다. 실험결과를 통해 제안하는 방법이 기존의 방법에 비해 높은 정확도를 가짐을 확인할 수 있었다.

Visual Information Selection Mechanism Based on Human Visual Attention

Kyung Joo Cheoi[†], Min-Chul Park[††]

ABSTRACT

In this paper, we suggest a novel method of selecting visual information based on bottom-up visual attention of human. We propose a new model that improve accuracy of detecting attention region by using depth information in addition to low-level spatial features such as color, lightness, orientation, form and temporal feature such as motion. Motion is important cue when we derive temporal saliency. But noise obtained during the input and computation process deteriorates accuracy of temporal saliency. Our system exploited the result of psychological studies in order to remove the noise from motion information. Although typical systems get problems in determining the saliency if several salient regions are partially occluded and/or have almost equal saliency, our system is able to separate the regions with high accuracy. Spatiotemporally separated prominent regions in the first stage are prioritized using depth value one by one in the second stage. Experiment result shows that our system can describe the salient regions with higher accuracy than the previous approaches do.

Key words: Visual Information Selection(시각정보선택), Attention Region(주의영역), Saliency(현저도), Depth(깊이), Motion(움직임)

※ 교신저자(Corresponding Author) : 최경주, 주소 : 청주시 흥덕구 개신동 12번기 충북대학교 전자정보대학 컴퓨터공학부, 전화 : 043)261-3487, FAX : 043)273-2265, E-mail: kjcheoi@cbnu.ac.kr
접수일 : 2009년 10월 5일, 수정일 : 2010년 6월 23일
완료일 : 2010년 11월 2일
[†] 정회원, 충북대학교 전자정보대학 소프트웨어학과 부교수
[††] 정회원, 한국과학기술연구원 포토닉스센서시스템센터 책임연구원
(E-mail : minchul@kist.re.kr)
※ 이 논문은 2009학년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었음(This work was supported by the research grant of the Chungbuk National University in 2009)

1. 서 론

시각은 환경정보를 수집하여 지식을 구축하고 그것을 토대로 적응적 행동을 할 수 있도록 해주는 가장 중요한 감각수단으로서 지능의 핵심요체라 할 수 있다. 기존의 시각연구는 서구 선진국의 경우 지각심리학, 신경생리학, 신경과학 등 인간의 시각원리에 대한 학제적인 연구결과를 바탕으로 일반성 있는 인공시각을 구현하려는 접근법을 사용한 반면, 우리나라는 전자공학 및 컴퓨터과학을 중심으로 시각연구가 진행되었고, 10여년전부터 생리학, 심리학 분야에서 인간의 시각피질과 신경세포간의 동작원리를 규명하는 연구, 시각 신경계에 있어서의 정보처리 메커니즘을 분석하려는 연구 등이 활발히 진행되고 있다. 이를 기반으로 생리학, 심리학 분야 등의 연구결과를 토대로 생물학적 시스템의 우수성이 증명되었으며, 이러한 메커니즘의 응용여부가 비상한 관심의 대상이 되고 있다.

컴퓨터시각(Computer Vision)이란 투시된 영상들로부터 주어진 장면에 관한 유용한 정보를 추출하여 물리적인 대상을 명확하고 의미있게 기술하도록 하는 과정이라 할 수 있는데 이러한 컴퓨터시각 시스템을 개발하는데 있어 가장 큰 문제점은 주어진 영상의 크기에 따라 계산의 복잡도가 달라진다는 것이다. 매 순간마다 입력장치를 통해 들어오는 수많은 시각정보를 모두 처리한다는 것은 불가능할 뿐만 아니라 제한된 자원을 효율적으로 사용한다는 측면에서도 불필요한 일이다. 이에 따라 시각정보 선택의 필요성이 증대되고 있으며, 영상입력장치의 파라미터를 동적으로 바꿔가면서 능동으로 시각정보를 지각하도록 하는 능동적 접근 방법에 대한 중요성이 강조되고 있다.

능동시각의 주요 특징은 사카드(saccade)를 생성해내는 시선 제어 방법에 있다. 이는 주어진 영상의 모든 부분을 다 처리하지 않고 의미있는 정보만을 선택적으로 받아들이도록 하는 것이다. 영상에서 의미있는 정보를 받아들이려면 의미있는 영역이 어디인지를 우선 알아야 한다. 여기서 의미있는 영역이란 다른 부분에 비해 상대적으로 중요한 정보, 많은 정보를 포함할 가능성이 큰 영역을 말하는 것으로, 입력되는 영상으로부터 의미있는 특정 몇몇 영역을 성공적으로 탐지할 수 있다면 전체 영상보다 상대적으로 좁은 선택된 영역에 자원을 집중하여 정보처리의 효율을 극대화시킬 수 있다.

일반적으로 인간의 시각주의 모형에 관한 연구는 의미있는 영역 선택에 대한 해결책을 제공해 주고 있다. 인간의 시각주의에 관한 연구결과[3]를 보면 인간은 보통 영상이 주어지면 그 영상에서 몇몇 영역에만 주의를 가한다고 한다. 영상을 보는 시간이 무제한으로 주어진다 하더라도 피험자들은 영상의 모든 부분을 세밀히 살피지 않고 계속 몇몇 영역에만 집중을 하여 본다는 것이다. 시각자극을 주는 부분은 일반적으로 유용한 정보를 포함하고 있을 가능성이 크며, 선택된 부분은 영상의 다른 부분에 비해서 시각적으로 현저하게 주의를 끄는 부분이라 할 수 있다. 시각주의란 이처럼 시각체계에 입력된 정보 중 일부를 선택하거나 여과하는 기제로서, 주의를 특정 영역이나 물체에 집중시킴으로써 시각정보의 처리능력을 극대화시키는, 현재 공학에서 주로 나누어서 처리되고 있는 전경과 배경의 분리와 인식의 문제를 효율적으로 결합할 수 있으나 아직 공학에서 충분히 활용되지 못하고 있는 매우 중요한 동물의 지적 능력이다.

본 논문에서는 이러한 인간의 시각주의 기능에 기본 바탕을 둔 시각정보 선택기법을 소개하고자 한다. 컴퓨터시각에서의 시각주의 시스템은 인간의 인지에 가능한 한 잘 부합하게 만드는 것을 목적으로 하고 있으며, 지금까지 인간의 시각시스템을 모방하여 영상으로부터 자동적으로 주의가 가는 부분을 선택해내는 연구가 많이 이루어져 왔다[1-9].

주의가 가는 영역(부분)은 영상에서 나머지 다른 영역들에 비해 현저한 주의를 일으키는 영역을 말하며, 이러한 현저함의 상대적 차이를 자료화 한 것을 현저도라 하고 이 현저도 값들을 2D 영상형태로 대응시킨 것을 현저도맵(Saliency map)이라 한다[1]. 현저도맵을 구하는 접근방법에는 영상의 특성을 어떻게 사용하느냐에 따라 달라진다. 영상의 색상, 명도, 형태, 방위, 깊이 등의 특징을 사용하는 공간현저도(Spatial Saliency), 연속된 영상으로부터 추출한 움직임 정보를 이용하는 시간현저도(Temporal Saliency), 이러한 2가지 특성을 결합하는 시·공간현저도(Spatiotemporal Saliency)가 바로 그것이다.

정지된 한 장의 영상으로부터 색상, 밝기, 방향 등의 공간적 특성을 분석해서 현저도를 계산해 내고 2D이미지 형태로 구축한 것을 공간현저도맵(Spatial

salienct map)이라 한다. 또한 인간의 시각시스템은 공간요소 뿐만 아니라 움직임을 가지는 물체에 대해서도 시간적 주의가 선택적으로 집중된다. 동영상에서 일어나는 움직임(motion)은 주의를 일으키는 중요한 요소가 되므로[2], 시간 t 동안에 픽셀들이 움직이는 수평/수직방향 성분들을 계산하여 모션벡터맵을 얻고, 이것을 분석하면 구성 요소들의 움직임을 알 수 있다. 영상 내에서 유의미한 모션이 존재하는 부분은 어떤 특정한 움직임이 있는 부분은 다른 부분보다 더 큰 주의를 일으킬 수 있다[3]. 시간주의 모듈에서 구한 현저도 값을 적절히 사용하면 앞에서 설명한 공간주의 모듈에서의 현저도 영역 탐지 결과를 바람직한 방향으로 보상할 수 있다[4-9].

공간현저도맵과 시간현저도맵을 각각의 과정을 통해서 계산해 내면, 최종적으로 인간의 시각주의를 현실적으로 잘 반영하는 하나의 현저도맵으로 통합해야 하는데, 이때 두 현저도맵이 보이는 특성을 고려하여 더욱 유용하게 사용될 수 있는 맵을 잘 선택해서 능동적으로 통합할 필요가 있다. 특히 영상 내에서 물체들간에 부분적 겹침이 일어난 경우, 또는 물체들이 비슷한 움직임과 형태를 보이는 경우 등 일반적으로 현저도의 강약 구분이 애매한 경우는 주의가 가는 영역을 제대로 선택해내기가 어렵다. 강한 현저도의 판단은 가중치 부여의 근거가 되며 많은 연구들에서 각각 다른 방식으로 시도 되어왔다. 영상의 구성이나 관찰자의 주관에 따라 판단은 달라질 수 있으며 아직까지 명백한 기준이 규명되지 않은 영역으로서 관련 분야의 연구가 이루어지고 있다[1, 4-9]. 또한 이러한 시각주의 시스템을 응용한 연구사례가 국내에서 이루어지고 있다[10].

본 논문에서는 위에서 기술한 시·공간 특징을 이용한 기존의 주의영역 선택 방식의 한계점을 보완하여 주의 영역으로 선택된 영역이 인간이 선택한 영역과 보다 더 잘 부합되게 할 수 있는 시각정보 선택방법을 제안한다. 이를 위해 시간특징과 3차원 공간특징을 2차원 공간특징과 더불어 통합하는 모델을 제안하고 그 효율성을 실험을 통해 증명한다.

2. 제안하는 시스템

제안하는 시스템을 통해 구성되어지는 맵의 전반적인 처리 흐름도는 그림 1과 같다. 입력되는 영상은 기본적으로 스테레오 영상의 좌우쌍(스테레오 영상의 기준 영상은 왼쪽 영상임)이며, 움직임을 가지는 동영상이다. 하지만 모든 영상이 스테레오 영상은 아니므로 2D 영상이 입력될 수도 있고, 정지영상이 입력될 수도 있다. 다시 말해 모든 종류의 영상을 입력받을 수 있도록 시스템은 설계되어 있다. 다만 입력

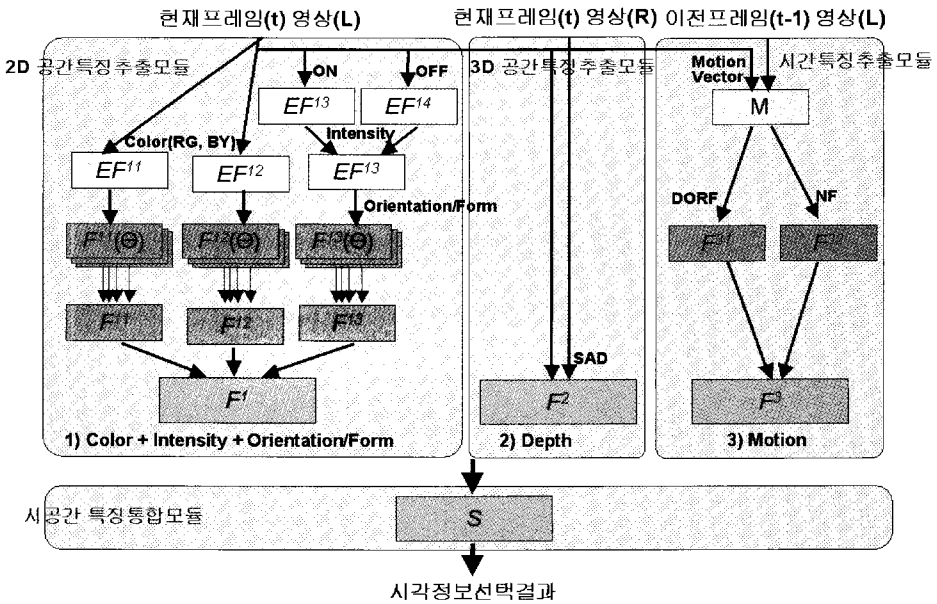

그림 1. 제안하는 시스템 구성도

되는 영상의 종류에 따라 추출되어지는 특징이 동적으로 달라진다. 예로 정지영상의 경우 시간특징모듈이 작동하지 않으며, 2D 영상의 경우는 3D 공간특징모듈이 작동하지 않는다.

입력되는 영상에 따라 2D, 3D 공간특징, 시간특징이 추출되어 기본특징맵이 생성되고, 생성된 기본특징맵이 재구성 및 통합되어 하나의 현저도맵이 생성된다. 이 현저도맵을 기반으로 우선수위에 따라 시각정보가 선택된다.

2.1 2D 공간특징 추출모듈

현재 프레임(t)의 스테레오 영상의 기준 영상인 왼쪽 영상을 대상으로 여러 초기 시각특징들을 추출하여 특징맵을 생성한다..

2.1.1 색상 및 명도대비특징

색상은 인간의 시각이 물체를 구분할 수 있는 가장 큰 특징 중의 하나로, 전-주의 단계에 있어 특징선택 작업에 큰 역할을 한다. 우리가 색 영상(color image)을 인지하는 방법은 3가지 지각변수인 색상(hue), 채도(saturation), 명도(lightness)의 3요소로 분류되는데, 우리가 색이라는 단어를 사용할 때 사용하는 '색'은 보통 '색상(hue)'을 말하는 것이다. 본 논문에서는 '색상(hue)'와 '명도(lightness)'를 특징으로 사용하였다. 생리학적인 실험에 의하면, 여러파장의 빛은 우리의 망막에 존재하는 수용기인 3가지 종류의 추상체(cone)에 의해 받아들여지는데, 이 3가지 추상체는 각각 단파(420nm), 중파(540nm)와 장파(560nm)에서 가장 민감하게 반응하도록 구성되어 있다[12]. 이 3가지의 추상체는 서로 중첩되어 반응하며, 이러한 특징은 시각시스템이 서로 다른 빛의 파장을 구분할 수 있게 해준다. 이들 추상체의 반응들은 다음 뉴런인 신경절 세포에 넘겨지고 바로 LGN을 거쳐 색상 정보를 뇌에 전달한다. 이때 색상 정보를 뇌로 전달해주는 신경회로는 3가지 추상체들에 의한 정보를 '적/녹', '청/황'의 반대쌍의 색상정보로 바꾸어 전달한다. 이렇듯 우리의 색상 지각 경험은 수용기 세포와 대립세포에서의 생리적 기제 모두에 의해 형성된다.

본 논문에서는 이러한 인간의 색상지각능력에 관련된 정보[12]를 토대로 색상대비(contrast)에 강하게 반응하는 동질(homogeneous) 형태의 수용야(receptive field)를 가진 색상대립세포인 R+G-('적/녹'대립) 세포와, B+Y-('청/황'대립) 세포를 모델링하여 색상특징맵으로 만들었다. 여기서 R+G- 세포의 경우 R에서는 활성화가 되고, G에서는 활성이 억제되어 R+G-세포가 흥분하면 적색을 지각하게 된다. 본 논문에서는 각각 순수한 색상에 조율된 *red*(적), *green*(녹), *blue*(청)을 생성하고[8], '적/녹'대립 *red-green*와 '청/황'대립 *blue-yellow*를 만들어 각각 EF^{11}과 EF^{12}로 만들었다.

명도특징으로는 색상을 제외한 명도(lightness)값을 사용하였는데, 어두운 곳에서는 밝은 곳을 찾기 위해 입력되는 명도값 그대로를 쓰고(ON 연산자 : EF^{13}), 밝은 곳에서는 어두운 곳을 찾기 위해 원래 값의 반전값 사용하도록 하였다(OFF 연산자 : EF^{14}). 기존의 연구에서는 이런 ON과 OFF 라는 특징을 사용하지 않고 입력되는 명도값 그대로를 쓰는 ON 명암도 특징맵만을 사용한 경우가 많았기 때문에 최악의 경우 밝은 배경에 검은 전경이 포함된 영상일 경우에는 OFF 특징을 사용하지 않는다면 밝은 배경 부분을 중요한 특징으로 추출하게 된다. 따라서 2개의 명암도 특징맵을 사용하였는데, 이 2개의 특징맵은 입력되는 영상에 따라 1개의 명암도 특징맵으로 통합되어진다. 본 논문에서는 특정 특징맵안의 특징값과 이웃하는 특징값 간의 통계적 정보를 바탕으로 활동량이 많은 부분을 그렇지 않은 부분에 비해 확연히 차이가 날 수 있도록 활성화시키고 전체적으로 독특한 부분이 없다면 전체적으로 균일한 부분을 억제시킨 후 2개의 특징맵을 선형조합하도록 하는 방법을 사용하다. 특징값이 있는 모든 지점에 대해서 전체 맵안에서 가장 큰 특징값과 평균 특징값을 비교하면 현재 지점의 특징값이 평균 특징값에 비해 얼마나 다른지 알 수 있게 된다. 이런 차이가 크면 클수록 해당 맵내의 특정 지점에서의 특징값이 다른 지점에 비해 두드러지다는 말이 되고, 결국 하나의 맵으로 통합될 때에 이러한 특징양상을 보이는 특징맵은 큰 우위를 차지하도록 하게 하면 된다. 하지만 만일 그 차이가 작다면, 해당맵은 별로 독특하지 양상을 보이는 특징값만을 가지고 있다는 말이 되므로 이러한 맵은 하나의 맵으로 통합될 때 별 영향을 미치지 않게 하면 되는 것이다. 따라서 생성된 각 EF^{13}과 EF^{14}마다 전역적인 최대값인 M^{13}, M^{14}와 과 열단위로 최대값들의 평균값인 m^{13}, m^{14}를 계산하여 EF^{13}의 특징

값은 $(M^{13}-m^{13})^2$으로 EF^{14}의 특징값은 $(M^{14}-m^{14})^2$으로 곱하여 업데이트 한 후, 각 EF^{13}과 EF^{14} 값을 0~1사이의 값으로 식(1)과 같이 정규화한다.

$$EF'^{k}_{x,y}=\frac{EF^{1k}_{x,y}-\min\left(EF^{13}_{x,y},EF^{14}_{x,y}\right)}{\max\left(EF^{13}_{x,y},EF^{14}_{x,y}\right)-\min\left(EF^{13}_{x,y},EF^{14}_{x,y}\right)},\ \text{단}\ k=3,4 \quad (1)$$

식 (1)에서 k=3,4가 되며, $\min(EF^{13}, EF^{14})$는 EF^{13}과 EF^{14}을 구성하는 특징값 중 가장 작은 값을, $\max(EF^{13}, EF^{14})$는 EF^{13}과 EF^{14}을 구성하는 특징값 중 가장 큰값을 뜻한다. 여기서, 정규화는 각각의 맵을 모두 일관되게 0~1 사이의 값으로 정규화시키는 것이 아니라, 해당되는 맵 뿐만 주변의 다른 맵의 값들도 고려한다. 이렇게 함으로써 다른 맵과의 관련성을 유지하게 된다. 이러한 과정을 반복하면 할수록 두드러지지 않는 특징값들은 그 수치가 낮아지게 된다. 따라서 이러한 과정을 4번 반복한 후 이렇게 처리된 모든 맵을 모두 선형조합한 후, 맵안의 특징값을 모두 0~1 사이의 값으로 정규화시켜 하나의 통합된 맵을 만든다. 여기서 4번이란 반복횟수는 가장 좋은 결과를 냈을 때의 반복횟수로 실험적으로 구한 수치이다.

2.1.2 색상 및 명도대비특징 별 방위/형태 특징

인간의 형태지각은 머리에서 일어나는 인식 이전의 과정으로서 물체를 형성하는 것이 아니라 덩어리나 경계선 등의 구분 정도를 나타낸다. 본 논문에서는 추출되어진 색상대비특징들로부터 방위를 탐지하고, 덩어리나 경계선 등의 구분정도를 확실히 드러내게 하기 위해 추출된 색상대비특징에 대하여 인간의 동심원 형태의 "ON-중심, OFF-주변" 수용야에서 보이는 세포 반응도를 모방한 중심-주변 연산을 수행한다. 이는 인간에게서 보이는 측면억제(lateral inhibition) 현상을 모방한 필터로써, 방위도 같이 탐지하게 하기 위하여 8가지 방위($\Theta\in\{0,\pi/8, 2\pi/8,\cdots,7\pi/8\}$)를 가진 ON-중심, OFF-주변 연산자인 $h(\Theta)$필터를 사용하였다.

각 방위별 $F^{11}(\Theta)\sim F^{13}(\Theta)$을 생성하는 방법은 식 (2)와 같다.

$$F^{1k}(\theta)=\left[\sum_{m,n}EF'^{1k}\cdot h_{x-m,y-n}(\theta)\right]^2 \quad (2)$$

식 (2)에서 k=1,..,3이며, 각각의 색상대비특징(EF^{11}, EF^{12}) 및 명도대비특징(EF^{13})은 $h(\Theta)$필터와 회선되어지고, 회선 후 회선된 영상의 대비를 증가시키기 위해 회선된 결과를 제곱한다. 여기서 $h(\Theta)$필터는 식 (3)과 같이 계산된다.

$$h_{x,y}(\theta)=|K_1\cdot G_{x,y}(\sigma,r_1\cdot\sigma,\theta) - K_2\cdot G_{x,y}(r_2\cdot\sigma,r_1\cdot r_2\cdot\sigma,\theta)\,RIGHT\,| \quad (3)$$

식 (3)에서, $G(\cdot,\cdot,\cdot)$은 방위를 가진 2차원 가우시안 함수이며, K^1, K^2는 양수, r^1은 2개의 가우시안의 이심률, r^2는 ON과 OFF 가우시안 간의 폭 비율을 뜻한다.

이 과정을 통하면 모두 24개의 특징맵이 생성된다. 24개의 색상/명도 별 방위/형태 특징맵은 3개의 특징맵으로 통합하여 F^{11}, F^{12}, F^{13}을 만든다. $F^{1k}(\Theta)(k=1,..3)$은 EF^{1k}에 식 (2)를 처리한 결과인데, 이는 위에서도 언급했듯이 8가지 방위와 함께 형태를 두드러지게 하는 연산이다.

이 연산을 통해 나온 각 맵 별 8가지 특징 중 가장 두드러지는 특징을 가지는, 즉 두드러진 방위를 나타내는 특징맵을 다음과 같은 방법으로 선택하였다. 먼저, $F^{1k}(\Theta)$의 최대값과, $F^{1k}(\Theta)$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. $F^{1k}(\Theta)$의 최대값에 비해 이 평균값이 작을수록, 강한값이 좁은 영역에 집중적으로 분포한다. 이러한 특징을 보이는 맵이 선별대상이 된다. 이 과정을 8개의 맵에 적용하여, 가장 특징이 두드러진 현저도가 높은 맵을 선택한다.

2.2 3D 공간특징추출모듈

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. 인간 두 눈의 기하학적 특성을 고려하여 3차원 공간상에 설치된 카메라로부터 얻어진 좌, 우 영상으로부터 3차원 정보를 획득하여 영상을 처리하게 된다. 일반적으로 스테레오 정합 결과에서 얻어진 정합점의 상대 거리를 변위(disparity)값으로 나타낼 수 있고, 이 값으로부터 물체까지의 거리(distance)나 깊이(depth)와 같은 3차원 정보를 구하

하게 된다. 최근 스테레오 매칭 알고리즘들은 크게 지역적인 방법(local methods)과 전역적인 방법(global methods)으로 분류할 수 있다. 일반적으로 전역적인 방법은 보다 높은 정확도의 변위 정보(disparity information)를 추정하는 데 반해 지역적인 방법은 비교적 낮은 정확도를 가지지만 전역적인 방법에 비해 보다 단순하고 빠른 방법으로 알려져 있다. 특히 지역적 방법의 성능은 종류에 상관없이 매칭 비용 또는 유사성을 계산하는 방법에 따라 좌우된다. 일반적으로 기존의 지역적 방법은 매칭 비용 계산을 위하여 픽셀 밝기 값에 대한 유사도 (pixel dissimilarity) 방법을 사용하고 있으나 최근 영상의 영역 분할을 이용한 방법[11]이나 적응적 가중치(Adaptive Support-Weight)를 이용한 방법[12] 등 전역적인 방법에 비해 보다 정확한 변위 정보를 추정할 수 있는 지역적 방법들이 제시되고 있다. 특히 [11]은 기존의 적응적 윈도우 방법과는 달리, 영역(support 또는window) 내의 중심 픽셀과 이웃 픽셀들 간의 컬러의 차이와 기하학적 차이를 이용한 가우시안 가중치를 사용함으로써 보다 좋은 결과를 얻을 수 있다. 그러나 이 방법은 좋은 결과에 비해서 많은 시간이 걸리거나 많은 메모리를 필요로 한다. 따라서 실시간 스테레오 매칭이 불가능하다.

본 논문에서는 계산량이 적은 지역적인 방법을 사용하였다. 영상의 깊이는 동일 축 상에 정렬된 좌/우 영상에서, 각 영상의 픽셀에 대응하는 다른 영상의 픽셀간의 위치의 차이(Disparity)로 정의되므로, 두 영상의 화소들의 상관관계를 파악하여 카메라 위치로부터 물체들의 상대적 거리(깊이)를 측정한다. 깊이값은 왼쪽 영상에 대하여 오른쪽 영상과의 에러값이 가장 작은 값을 찾아 계산한다. 에러값 연산에 있어서는 그 두 지점간의 거리간의 차의 절대값 합인 SAD(sum of absolute difference)를 이용하였다. 정확한 결과를 얻기 위해서는 좌/우 입력 영상이 정확히 수평으로 정렬되어 있어야 하고, 이미지의 각 픽셀은 주변의 픽셀들과 비슷한 깊이를 가져야 한다.

$$SAD(x,y,d) = \sum_{i=-\frac{1}{2}w}^{\frac{1}{2}w} \sum_{j=-\frac{1}{2}}^{\frac{1}{2}w} \left| L(x+i,y+j) - R(x+i+d,y+j) \right| \quad (4)$$

여기서 w값은 탐색할 영역의 범위를 결정하는 마스크 윈도우의 크기를 나타내며, L은 왼쪽 영상, R은 오른쪽 영상, d는 거리를 나타낸다. 깊이를 구하기 위해 좌영상의 x, y 좌표에 해당하는 윈도우의 값과 우영상의 x축으로 d만큼 떨어진 윈도우의 값의 차를 구하여 SAD볼륨 d에서의 SAD값을 결정하고, 이중에서 최소 SAD값을 찾는다. 최소 SAD값을 가지는 d는 좌,우 영상간 값의 차이가 해당 d에서 가장 작음을 의미한다. 이렇게 구한 d값이 작을수록 해당 x, y좌표의 픽셀은 카메라로부터의 깊이 값이 크고, d값이 클수록 깊이 값은 작다.

2.3 시간특징 추출모듈

현재 프레임(t)과 이전 프레임(t-1)으로 이루어져 있는 연속적인 명도 영상 시퀀스를 입력받아 시간특징인 움직임 정보를 추출한다[7]. 이 영상 시퀀스에서 블록매칭기법을 사용하여 모션벡터(motion vector)를 추출한 후 모션벡터맵 M을 구성한다. 본 논문에서는 정확도를 위해 전역탐색기법을 사용하였으며, 매칭함수로는 SAD(Sum of Absolute Difference)를 사용하였다. 모션벡터를 계산하는 것은 시간 현저도맵을 구성하기 위한 기본적인 과정이며, 이때 유의미한 모션들을 추출하는 것이 매우 중요한데 시간 현저도맵의 신뢰도가 높아질수록, 시공간현저도맵의 정확도가 향상되기 때문이다. 이것은 최종적인 결과의 향상에 유익한 바탕을 제공하게 된다. 움직임 정보를 얻을 때 발생할 수 있는 노이즈는 시간적 현저도 계산의 정확성을 떨어뜨린다. 본 논문에서는 움직임 정보로부터 노이즈를 제거하기 위하여 심리학적인 연구결과를 이용하였다[2].

제안하는 시스템에서는 모션벡터맵 구성시 발생할 수 있는 노이즈를 제거하기 위하여 인간의 대뇌시각피질 영역 중 MT 영역에서 보이는 이중대립수용야 반응과 노이즈 여과 특성을 모델링 한 결과를 반영하였다. 먼저 블록정합 방식을 사용하여 얻은 모션벡터맵 M을 구성하고 있는 모션벡터값들에 대해 MT 영역에서 보이는 수용야의 반응을 모델링 한 2개의 모듈에 의해 각각 처리되어 F^{31}과 F^{32}로 재구성된다. 이렇게 얻은 2개의 맵을 추후 하나의 모션특징맵 F^3로 통합한다.

1) 이중대립수용야 반응 모델링

많은 해부학적, 생리학적 연구들에 의하면 영장류

의 대뇌 시각피질 영역 중 MT(middle temporal area) 영역이 모션분석에 있어 주요한 역할을 하고 있다고 한다[2]. 이 MT 세포의 반응도는 모션자극이 MT 세포의 수용야 내부 영역과 주변 영역에 어떻게 주어졌는가에 따라 다르다((그림 3) 참조). 수용야 내부 영역에 수용야의 최적방향으로 모션자극이 주어졌을 때 MT 세포는 최대로 반응한다. 그런데 위와 같은 모션자극이 있는 상태에서 수용야의 주변영역에 모션자극이 최적방향의 반대방향으로 주어지면 MT 세포의 반응은 더욱 증대된다. 그러나 만일 이때 수용야 내부영역에 모션자극이 주어지지 않는다면 다시 말해 수용야 주변영역에만 최적방향의 반대방향으로 모션자극이 주어지고 수용야 내부영역에는 아무런 자극도 주어지지 않는다면 세포의 반응은 거의 없게 된다. 이러한 "이중대립수용야"의 반응 특성을 모션벡터맵 M에 적용함으로써 모션 자극의 방향성분에 따라 특정모션을 강조하거나 약화 하여 상대적으로 더 의미 있는 모션들을 추출할 수 있다.

이러한 MT 세포의 "이중대립수용야" 성질을 모방하여 모션벡터맵 M에 대하여 그림 2와 같이 주변부 모션자극쌍의 방향이 최적방향과 반대방향이고, 중심부 모션자극이 최적방향으로 주어졌을 때 최대로 반응하도록 모델링하였다. 각 픽셀의 반응값은 8개의 방향 ($\Theta \in (0, \pi/4, , ..., 7\pi/4)$마다 식 (5)에 의해 계산된다. 식 (5)에서 $V_c(\Theta)$는 관찰자 시점에서 본 수용야 중심부 모션벡터를 나타내고, $V_{s1}(\Theta+\pi)$ 와 $V_{s2}(\Theta+\pi)$는 주변부 모션벡터를 나타낸다.

$$F^{31}_{x,y} = \sum_{\theta}\left(\sum_{m,n}|R_{m,n}(\theta)|\right)$$

만일 $|V_c(\theta)| \neq 0$이면,

$$|R(\theta)| = \left|-\frac{1}{2}V_{s1}(\theta+\pi) + V_c(\theta) - \frac{1}{2}V_{s2}(\theta+\pi)\right|$$

그렇지 않을 경우,

$$|R(\theta)| = \left|-\frac{1}{4}V_{s1}(\theta+\pi) + V_{s2}(\theta+\pi)\right| \quad (5)$$

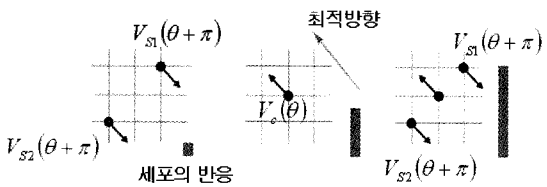

그림 2. MT 영역의 이중대립수용야 반응을 모델링한 결과

2) 노이즈 여과 특성 모델링

MT 영역의 세포는 수용야에 하나는 세포의 최적방향으로 움직이고 다른 하나는 최적방향의 반대방향으로 움직이는 2개의 점을 제시하면 세포의 반응도가 매우 약해지는 특성을 가지고 있다(그림 3 참조). 이러한 MT 세포의 특성을 모방하여 입력되는 모션벡터맵 M에 대하여 그림 3과 같이 4개의 방위 ($\Theta \in 2\pi/4, .., 5\pi/4$)를 가질 수 있는 주변부 모션벡터쌍의 방향이 서로 반대방향이 되었을 때 반응하도록 한다. 각 픽셀의 반응값은 식 (6)에 의해 계산된다. 식 (6)에서 $V_c(\Theta)$는 식 (6)에서와 마찬가지로 관찰자 시점에서 본 수용야 중심부 모션벡터를 나타내고, $V_{s1}(\Theta+\pi)$ 와 $V_{s2}(\Theta+\pi)$는 주변부 모션벡터를 나타낸다.

$$F^{32}_{x,y} = \sum_{\theta}\left(\sum_{m,n}|R_{m,n}(\theta)|\right) \quad (6)$$

만일 $|V_c(\theta)| \neq 0$이면, $|R(\theta)| = |V_c(\theta)|$

그렇지 않을 경우, $|R(\theta)| = |V_{s1}(\theta) - V_{s2}(\theta+\pi)|$

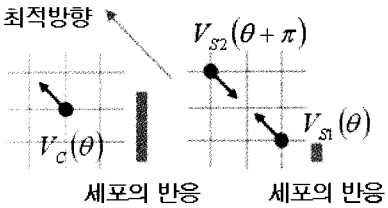

그림 3. MT 세포의 노이즈 여과 필터링 특성을 기반으로 모델링 한 결과

2.4 시공간특징 통합모듈

지금까지 기준영상으로부터 구성된 3개의 색상 및 명도대비 특징 별 방위/형태 특징맵과 좌우 영상으로부터 구성된 1개의 깊이특징맵, 이전 프레임(t-1) 왼쪽 영상과 현재 프레임(t) 왼쪽 영상으로부터 구성된 2개의 모션특징맵이 만들어졌다. 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 기반으로 주의 영역을 알아내기란 쉬운 일이 아니다. 즉 5개 각각의 특징이 제공해주는 정보는 모두 한 종류의 영상에 대한 것이므로, 각각의 특징마다 모두 각각 다른 영역을 주의 영역이라고 가이드해 줄 수도 있

다. 이러한 문제를 해결하기 위해서는 영상의 각 지점에 있어서의 서로 다른 특징값들이 모두 하나로 통합되어져서 영상의 각 지점마다 하나의 전역적인 중요도 측정값인 Saliency가 계산되어져야 한다.

기존의 시스템 [8]과 [9]는 움직임을 시간특징으로 사용하여 공간특징과 통합한 시스템이라는 점에서 제안하는 시스템과 비슷하다. 하지만 공간특징으로 깊이정보의 사용유무와 특징통합방법에서 큰 차이가 있다. [8]과 [9]에서는 깊이 정보를 사용하지 않아 3차원 스테레오 영상에 적용할 수 없으며, 깊이 정보를 사용함으로써 얻을 수 있는 잇점을 활용하지 못했다. 또한 [8]에서는 동영상일 경우에는 움직임 같은 시간특징만을, 움직임정보만 정지영상일 경우에는 공간특징만을 사용하여 주의영역을 탐지하였으며, [9]에서는 2가지 특징을 모두 사용하여 결합하였으나, 이 때 결합하는 방식이 특징 맵의 국부적인 영역에서의 픽셀들과 다른 주변영역의 픽셀들을 비교하여 차이가 크면 해당 픽셀들을 활성화시키고 그렇지 않으면 억제시키는 경쟁 메커니즘을 사용하여 두드러진 특정 특징맵 하나를 선택하여 결합하는 방식을 사용하였다. 그러나 이 방법은 주요한 하나의 특징맵만을 선택하고 다수의 다른 특징맵은 무시한다.

본 연구에서는 추출된 특징들이 모두 중요하다고 가정하였기 때문에 [9]에서의 방법처럼 두드러진 특정 특징맵만을 선택하지 않고, 특징의 두드러진 정도에 따라 적절한 가중치를 부여하여 결합하였다. 두드러짐 강도에 따라 높은 가중치를 부여하였고, 두드러짐의 정도가 차이가 나지 않을 경우에는 차이가 나지 않는 특징맵은 서로 같은 가중치를 부여하였다.

$F^{1k}(k=1,..,3)$를 F^1으로 통합하기 위해 각 맵의 특징값 분포의 두드러짐 정도가 큰 순서대로 5,3,1의 가중치를 부여하였고, $F^{3k}(k=1,2)$ 을 F^3으로 통합하기 위해 각 맵의 특징값 분포의 두드러짐 정도가 큰 순서대로 2,1의 가중치를 부여하였다. 특징값 분포의 두드러짐 정도는 [5]에서 사용한 특징통합방법에서 사용한 아이디어를 수정한 식 (7)을 사용하였다.

$$SI^k_{x,y} = \frac{MF^k_{x,y} - MinMF}{MaxMF - MinMF} \quad (7) \qquad (7)$$

여기서 $SI^k(k=1,..4)$는 4개의 특징맵인 F^k의 두드러진 정도를 나타내는 맵으로써, 4개의 특징맵의 통계적인 정보를 이용하여 업데이트한 것이다. $MaxMF$와 $MinMF$는 각각 MF^k를 구성하는 모든 특징값 중 가장 큰 값과 가장 작은 값을 뜻하며, MF^k는 식 (8)과 같이 계산되어질 수 있다. 여기서 $MaxF^k$ 는 각 F^k마다 이를 구성하는 특징값 중 최대값을, M은 입력되는 F^k의 크기를 뜻한다.

$$MF^k_{x,y} = F^k_{x,y} \times \left[MaxF^k - \frac{1}{M-1}\left[\left(\sum_{x,y} F^k_{x,y}\right) - MaF^k\right]\right]^2 \qquad (8)$$

여기서 $MaxF^k$는 F^k를 구성하는 뉴런들의 최대값을, M은 입력되는 F^k의 크기를 뜻한다.

F^1, F^2, F^3를 하나의 현저도맵으로 통합하기 위해서 생리학적인 연구결과도 사용하였다. 생리학자인 Gordon Walls은 움직이는 행위가 삶 및 생존에 밀접한 관계가 있기 때문에 움직임 지각은 일찍이 진화되어 왔으며, 색상 또는 깊이 지각이 형편없는 동물은 있어도 움직임을 지각하지 못하는 동물은 없다고 말하였다. 우리는 또한 시각적 주의에 대한 설문조사 결과도 사용하였는데, 어떤 특징이 시각적 주의를 일으킨다고 여기느냐에 대한 질문에서 설문조사 피험자들은 형태나 방위 또는 깊이보다는 움직임과 색깔이 주요한 역할을 한다고 대답했으며, 색깔보다는 움직임이 월등하게 중요하고, 형태나 방위, 깊이는 비등하다고 말하였다. 이러한 증거에 기반하여 F^3, F^2, F^1순으로 5,3,1의 가중치를 부여하였다.

3. 실험 및 결과

제안하는 시스템의 성능을 평가하기 위하여 다양한 실험영상 및 실험방법을 사용하여 시스템의 성능을 평가하였다.

먼저 실험영상으로는 색상, 강도, 방위, 형태, 깊이 등 모든 특징을 가지고 있는 스테레오 동영상상 외에도 명도값만을 가지는 명도 영상, 깊이를 가지지 않는 색상 영상, 정지영상 등 다양한 종류의 외부 실영상을 사용하였으며, 이와 함께 기존 심리학자들이 인간이 실제로 시각주의를 사용하여 어떻게 시선을 옮기는지를 알아보는 시각탐색 전략[13] 실험에서 사용했던 인공영상도 같이 사용하였다.

실험방법은 3가지를 수행하였는데, 첫 번째 실험은 시스템이 인간의 주의시각을 기반으로 한 바, 실제로 생물학적 기반을 가지는지 확인하기 위한 실험이 수행되었으며, 두 번째로 시스템의 전반적인 성능을 입력영상의 특징별로 분류하여 실험하고 이를 기

존방법과 비교하는 실험을 수행하였다. 마지막으로는 제안하는 시스템에서 추가한 3D 공간특징인 깊이 정보가 얼마나 효율적으로 사용되는지를 알아보기 위해 깊이정보를 특징으로 사용한 경우와 그렇지 않은 경우를 비교하여 실험하였다.

3.1 시스템이 생물학적 기반을 가지는가를 확인하기 위한 실험

첫 번째 실험으로 시스템이 생물학적 기반을 가지는지 확인하기 위하여 기존 심리학자들이 인간의 시각탐색 전략을 알아보는 실험에서 사용했던 영상을 대상으로 입력되는 영상에서 가장 먼저 시각주의가 가해지는 독특한 대상이 있는 영역을 찾아내는 pop-out 문제에 적용하는 실험을 수행하였다. 이 때 사용한 실험영상은 pop-out 심리실험에서 일반적으로 사용되는 color와 orientation pop-out의 2가지 부류의 192개의 막대바 영상을 만들어 사용하였다([그림 4] 참조).

'color pop-out' 실험을 위해, 모두 같은 방위를 가지지만 표적(target)은 적색의 막대바이고, 이외의 방해물(distractor)인 막대바는 모두 녹색인 영상을 사용하였다. 또한 'orientation pop-out' 실험을 위해 색상은 모두 적색으로 동일하지만, 표적에 해당되는 막대바는 다른 방해물(distrator)에 해당되는 막대바와 방위 면에서 약 90° 정도 차이가 나는 영상을 사용하였다. 이러한 크게 2가지 종류의 영상에서 각각의 영상에서 기본 방위는 임의적으로 주었고, 방해물의 방위는 -15°~15° 사이의 각도로 조금씩 변화를 주었다. 그리고 방해물의 갯수는 3개부터 시작하여 3개씩 증가시켜가면서 18개까지 되도록 영상을 만들었다. 또한 가우시안 분포 형태를 띈 20%의 색상 잡영을 추가하여 잡영에 대한 실험도 추가하였다.

그림 4는 pop-out 실험결과로, 그림 4의 왼쪽 2개의 영상은 각각 'color pop-out', 'orientation pop-out'에 대한 입력 영상 예와 해당 입력 영상에 대한 결과를 보여준다. 또한 오른쪽의 그래프는 각각의 입력 영상 속의 방해물의 증가에 따른 표적이 잘못 선택된 수를 그래프로 나타낸 것으로, x축은 방해물의 숫자를, y축은 표적이 선택되기 전까지의 잘못 선택된 수를 나타낸다. 여기서, y축의 표적이 선택되기 전까지의 잘못 선택된 수를 계산할 때에 각 방해물 숫자가 3개부터 시작했으므로 모두 16개의 영상을 사용한 것이 되므로 잘못 선택된 수는 방해물 수 별 전체 영상 16개에 대한 잘못 선택된 수를 평균 내어 사용하였다.

본 시스템에서 얻어진 2가지 형태의 'pop-out' 결과는 인간의 '시각탐색' 전략에서 보이는 결과와 동일한 결과를 나타내었다. 즉, 실험에 사용한 모든 실험 영상들이 방해물의 수의 증가에 관계없이 첫 번째에 표적을 찾아내었다. 이를 통해 제안하는 시스템

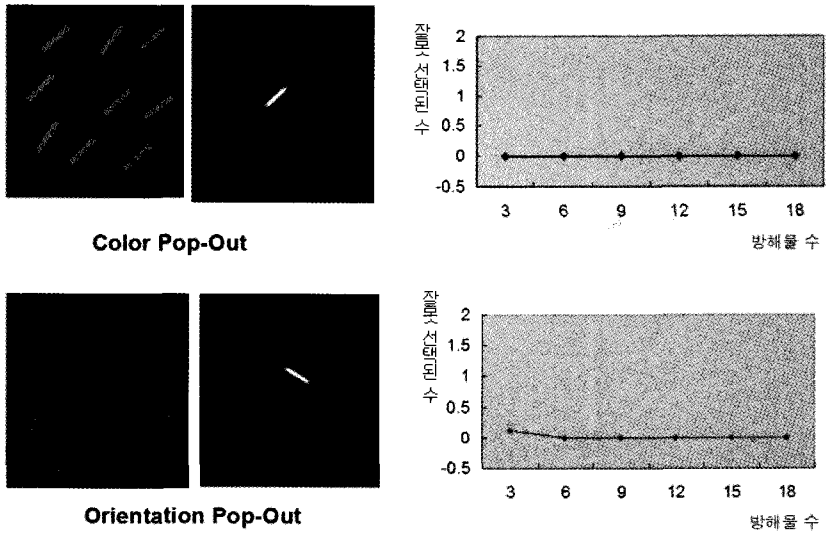

그림 4. 인간의 시각탐색전략 실험에 사용했던 영상에 대한 실험(Pop-out) 결과: 두드러진 특징이 1가지 존재 (상)색상, (하) 방위

은 인간의 pop-out 현상을 보인다고 할 수 있으며, 따라서 어느 정도 생물학적 기반을 가지고 있다고 할 수 있다.

3.2 입력영상의 특징 별 시스템 결과 및 기존 시스템과의 비교를 위한 실험

다양한 영상에 대란 시스템 결과를 살펴보고, 이를 기존 방법과 비교하는 실험을 수행하였다. 이때 영상은 정지영상, 동영상, 스테레오 동영상 등이 될 수 있는데, 제안하는 시스템은 영상에 따라 생성되는 특징맵이 동적으로 변할 수 있다. 2D 정지영상이 입력되었을 경우에는 움직임 정보와 깊이정보를 특징으로 사용할 수 없으므로 깊이 특징맵과 움직임 특징맵은 0가 된다.

3.2.1 입력영상의 특징 별 시스템 실험

1) 2D 정지영상 : 인공영상

그림 5에서의 실험영상은 검은 배경에 오렌지색의 막대바 23개와 적색의 막대바 1개가 있는 영상으로 색상과 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 5는 각각 방위, 색상, 모양 면에서 두드러진 단일 특징을 가지고 있다. 제안하는 시스템에서도 두드러진 특징을 잘 선택해냄을 알 수 있다.

그림 4와 그림 5에서도 확인할 수 있듯이 제안하는 시스템은 주의가 가는 부분을 제대로 선택해내는 동시에 영역의 형태로서 객체의 현저도 강도를 계산한다. 결과적으로 시각장면에서 얻게 되는 객체나 영역은 우리의 시스템에서 현저도에 따라 기술될 수 있다. 이는 주의가 갈만한 영역만을 희미하게 가이드해 주던 기존의 연구[1]와 달리 인식과도 같은 고차원적인 작업을 수행하기 위해서 해야 하는 후처리 작업이 매우 간단해졌다.

2) 3D 동영상 : 실영상

그림 6은 깊이 정보가 포함된 스테레오 동영상이

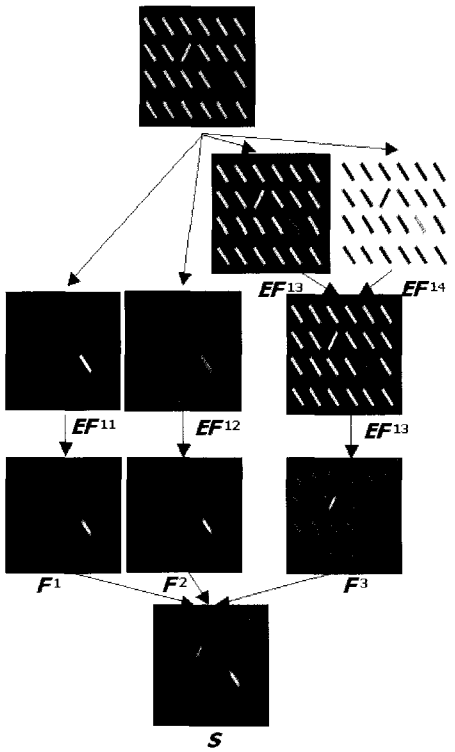

그림 5. 2D 정지영상의 실험 예 : 혼합특징-두드러진 특징이 2개 이상(색상 및 방위) 존재

입력되었을 때 시스템이 결과를 내기까지의 과정을 단계별 결과로 보여주고 있다. 스테레오 영상의 기준영상으로부터 색상, 명도강도, 형태, 방위 특징을 찾아내고, 좌우쌍 영상으로부터 깊이특징을 찾아낸다. 그리고 현재 프레임과 이전 프레임의 왼쪽영상으로부터 움직임 특징을 찾아낸다. 이렇게 구성되어진 모든 특징맵에 대하여 각 장소별 두드러짐에 기반한 가중치 결합을 수행하여 하나의 현저도맵을 생성하게 된다. 그림 6의 실험영상은 지나가는 자동차 안에서 고속도로를 달리면서 찍은 영상으로 가장 눈에 띄는 곳은 녹색 표지판이며, 그 다음으로는 중앙도로, 구름 등이다.

3.2.2 기존 시스템과의 비교 실험

기존 시스템 [9]에서는 추출되어진 다수의 특징맵을 하나의 현저도 맵으로 통합할 때 특징 맵의 국부적인 영역에서의 픽셀들과 다른 주변영역의 픽셀들을 비교하여 차이가 크면 해당 픽셀들을 활성화시키고 그렇지 않으면 억제시키는 경쟁 메커니즘을 사용하여 두드러진 특정 특징맵 하나를 선택하여 결합하

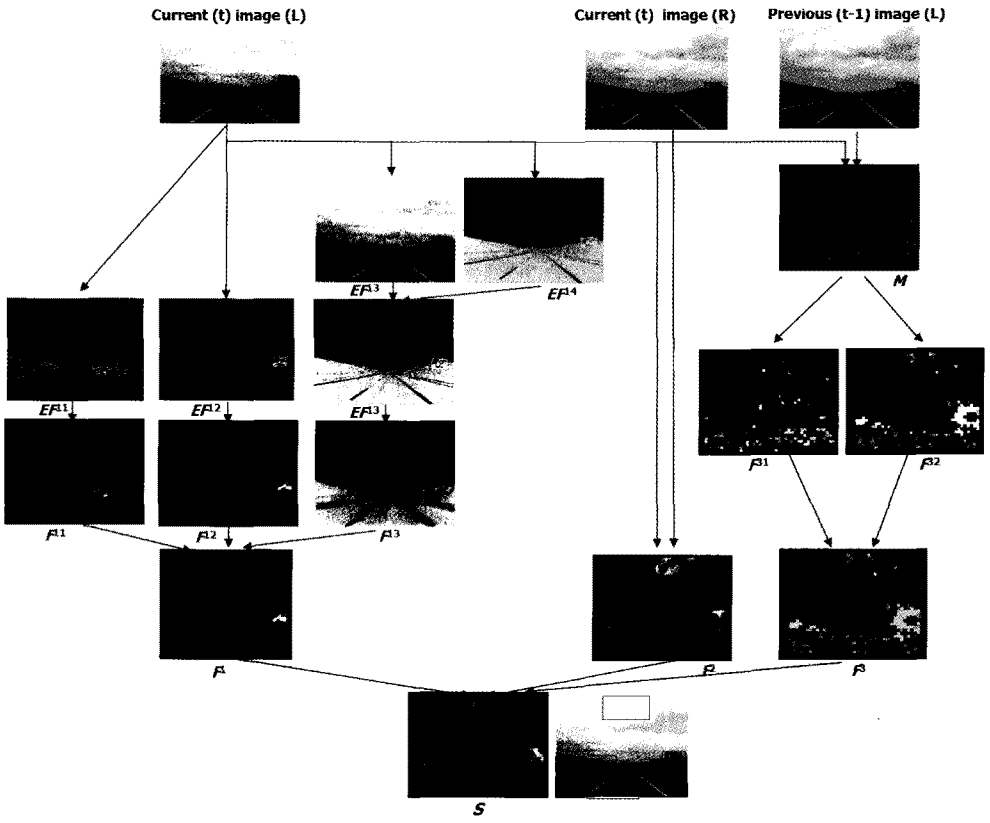

그림 6. 스테레오 동영상이 입력되었을 때의 구성되는 맵의 처리 흐름도

는 방식을 사용하였는데, 이 방법은 주요한 하나의 특징맵만을 선택하고 다수의 다른 특징맵은 무시하는 결과를 가져왔다. 이에 본 연구에서는 추출된 특징들이 모두 중요하다고 가정하였기 때문에 [9]에서의 방법처럼 두드러진 특정 특징맵만을 선택하지 않고, 특징의 두드러진 정도에 따라 적절한 가중치를 부여하여 결합하였다.

그림 7은 제안하는 시스템을 기존 시스템 [9]와 비교한 것이다. 기존 시스템에서는 두드러진 한 특징맵의 활동만을 부각시켜 주의강도가 가장 높은 한 영역만을 선택해내는 반면, 제안하는 시스템에서는 다수의 객체를 강도별로 선택할 수 있다.. 이를 기반으로 현저도 강도에 맞추어 시선처리하는 방안이 강구되고 있다.

3.3 3D 공간특징인 깊이 정보의 효율성에 대한 실험

전형적인 시각주의 시스템은 입력 영상내 객체들이 부분적으로 겹쳐있다거나 동일한 현저도를 가지고 있을 때 현저한 영역을 분리해내기가 쉽지 않은데, 본 논문에서는 시공간특징을 통해 나온 주의 영역을 깊이값을 사용하여 유의미한 영역을 선별하고 우선순위를 부여함으로써 주의 영역을 분리해 낼 수 있다.

그림 8은 스테레오 동영상을 입력으로 하였을 때 제안하는 시스템에서 깊이 정보를 사용한 결과와 그렇지 않은 결과를 비교한 것이다. 그림 8(a)는 움직이는 객체가 여러개로 겹침에 대한 문제와 한 객체가 여러 영역으로 쪼개지는 문제 등이 있고, 이를 깊이 정보를 사용함으로서 개선하였다. 그림 8(b)는 주요 물체가 하나라고 볼 수 있는 영상이다. 깊이 정보를 사용하지 않았을 때 물체의 단일성을 유지하지 못하였고, 인물 다음으로 큰 주의를 일으킨다고 볼 수 있는 부분을 선택함에 있어서 가까운 난간보다 멀리 있는 건물과 숲 쪽을 선택하였으나 깊이 정보를 사용함으로써 이러한 문제가 발생되지 않았다. 그림 8(c)의 경우 유사한 움직임을 보이는 객체들이 비교적 근거리에 모여있고, 객체 간의 겹침의 정도가 크다.

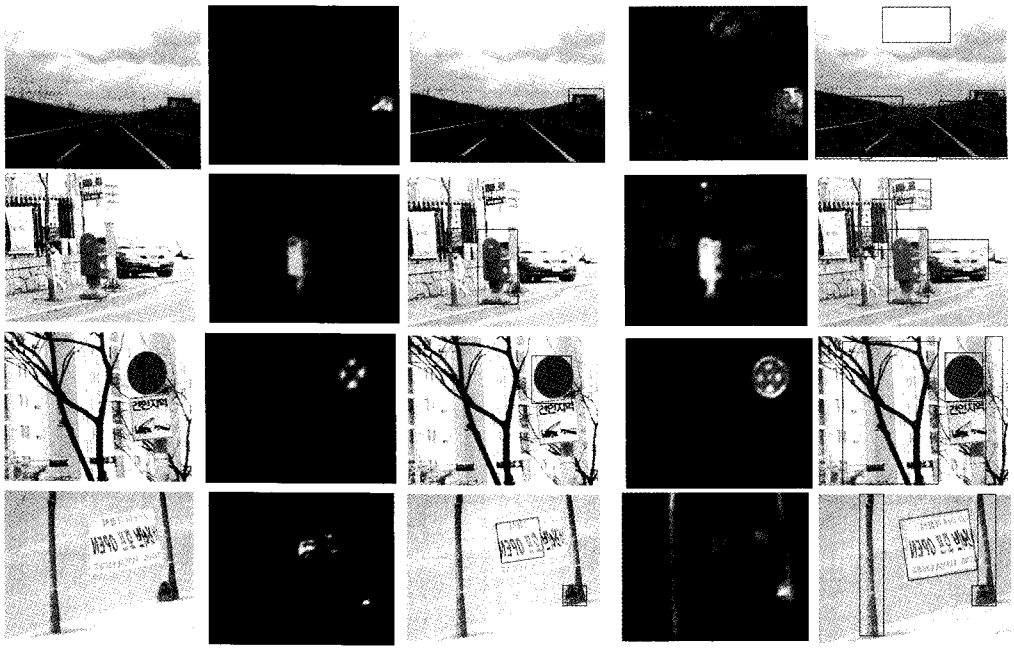

그림 7. 기존 시스템[9]과의 비교 결과

그림 8. 깊이 정보를 이용한 경우와 그렇지 않을 경우의 제안 시스템 결과

또한 물체의 색상과 배경의 색상이 상당부분 혼동되기 쉽다. 깊이 정보를 사용하지 않았을 때에는 주위에 비해 색이 두드러지는 의상의 하단부와 손의 움직임을 각각 탐지해내었으나, 물체의 연관성(단일성)이나 물체간의 위치관계를 알기 어려운 결과를 보였다. 그러나 깊이 정보를 사용함으로써 인물 하나에 주의 영역 하나를 할당한 형태를 유지하며 겹쳐진 인물들의 영역을 분리해내었다. 영상의 공간적 특징과 시간적 특징을 사용한 현저도맵을 만들고, 이것과 제안된 3차원의 깊이 정보를 활용하여 보상한 결과

를 비교하였다. 그 결과 현저도의 강약 구분이 애매한 영역들을 분리해내고 영역들에 우선순위를 부여할 수 있었다.

4. 결 론

본 논문에서는 입력장치로 들어오는 수많은 시각정보 중 현 시점에서 가장 유용하다고 생각되는 정보를 인간의 상향식 주의시각에 기반하여 선택하는 시각정보 선택기법을 제안하였다. 제안하는 시스템은 주의가 가는 부분을 보다 더 정확하게 선택할 수 있도록 하기 위해 영상의 공간특징과 시간특징을 사용, 통합하여 현저도맵을 만들고, 현저도맵의 정확도와 신뢰성을 향상시키기 위해 3차원의 깊이 정보를 활용하여 보상하였다. 또한 시간현저도를 구하는데 악영향을 미치는 노이즈를 효과적으로 제거하기 위해 기존의 심리학적 연구에 따라 시신경의 반응을 적용하여 국소영역에서의 노이즈를 제거하여 모션정보의 판단적 가치를 높일 수 있었다. 결과적으로 영상 내에서 물체들간에 부분적 겹침이 일어난 경우, 또는 물체들이 비슷한 움직임과 형태를 보이는 경우 등 일반적으로 현저도의 강약 구분이 애매한 경우에도 각 영역들을 명확한 기준에 의해 다른 현저도로 기술할 수 있음을 실험결과를 통해 확인할 수 있었다.

제안하는 시스템의 결과를 이용하면 어떠한 목표물 탐지 시스템에도 제안하는 시스템을 적용할 수 있다. 기존의 많은 목표물 탐지 알고리즘은 탐지하고자 하는 목표물에 대한 수많은 기반지식이 필요하기 때문에 실제적으로 특정 애플리케이션에만 적용되는 한계점이 있다. 하지만 제안하는 시스템을 이용하면 목표물에 대한 기반지식이 없더라도, 특정 애플리케이션에 튜닝된 목표물이 아닌 그 어떠한 목표물이라도 선택된 몇몇의 주의 영역을 통해 몇 번안에 탐지해낼 수 있다. 제안하는 시스템 결과를 통해 선택되어지는 몇몇 영역은 다른 영역에 비해 상대적으로 중요한 정보, 많은 정보를 포함할 가능성이 큰 영역으로 원하는 목표물을 탐지하고자 할 때 전체 영상보다 상대적으로 선택된 좁은 영역 순위별로 목표물 탐지 알리즘을 선택적으로 적용할 수 있다.

현재 스테레오 영상 분석 결과와 단일 영상 분석 결과와의 통합방법에 대한 연구 및 더욱 효과적인 깊이맵 구성방법에 대한 연구가 수행 중이며, 현저도맵을 통해 얻을 수 있는 주의 강도를 토대로 시선처리가 이루어질 수 있도록 하는 방안을 강구중이다.

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, pp. 1489- 1506, 2000.

[2] A. Hiroshi, "Visual Psychophysics (8) : Visual Motion Perception and Motion Pictures," *Journal of Image Information and Television Engineers*, Vol.58, No.8, pp. 1151-1156, 2004.

[3] E. Niebur and C. Koch. "Computational Architectures for Attention," *The Attentive Brain,* MIT Press, 1997.

[4] S. Kwak, B. Ko, and H. Byun "Automatic Salient-Object Extraction Using the Contrast Map and Salient Point," *LNCS*, Vol.3332, pp. 138- 145, 2004.

[5] S. Kwak , B. Ko, and H. Byun, "Salient Human Detection for Robot Vision," *Pattern Analysis and Applications*, Vol.10, No.4, pp. 291-299, 2007.

[6] K. Rapantzikos, Y. Avrithis, and S. Kollias, "Handling Uncertainty in Video Analysis with Spatiotemporal Visual Attention," The 14th IEEE International Conference on Fuzzy Systems, pp. 213-217, 2005.

[7] S. Li and M. Lee, "An Efficient Spatiotemporal Attention Model and Its Application to Shot Matching," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.17, No.10, pp. 1383-1387, 2007.

[8] M. Park and K. Cheoi, "An Adaptive ROI Detection System for Spatiotemporal Features," *J. of the Korea Contents Association*, Vol.6, No.1, pp. 41-53, 2006.

[9] M. Park and K. Cheoi, "Selective Visual Attention System Based on Spatiotemporal Features," *LNCS*, Vol.5068, pp. 203-212, 2008.

[10] 차명희, 김기협, 조경은, 엄기현, "시각적 특징

맵을 이용한 자율 가상 캐릭터의 실시간 주목 시스템," 한국멀티미디어학회논문지, v.12, no.5, pp.745-756, 2009.

[11] F. Tombari, S. Mattoccia, and L. Stefano, "Segmentation-Based Adaptive Support for Accurate Stereo Correspondence," In Proc. IEEE Pacific-Rim Symposium on Image and Video Technology, 2007.

[12] K. Yoon and I. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Tranansactions on PAMI*, Vol.28, No.4, pp. 650-656, 2006.

[13] A.Treisman and G.Gelad, "A Feature-integration theory of Attention," *Congnitive Psychology*, Vol.12, No.1, pp. 97-136, 1980.

최 경 주

1992년 3월~1997년 2월 충북대학교 컴퓨터과학과 학사
1997년 3월~1999년 2월 연세대학교 컴퓨터과학과 석사
1999년 3월~2002년 8월 연세대학교 컴퓨터과학·산업시스템공학과 박사
2002년 7월~2005년 2월 LG CNS 연구개발센터
2005년 3월~현재 충북대학교 전자정보대학 소프트웨어학과 부교수
관심분야 : 컴퓨터비전, 영상처리, 바이오컴퓨팅, 유비쿼터스컴퓨팅

박 민 철

1993년 홍익대학교 전자공학과 학사
1997년 일본 동경대학 전자정보공학과 석사
2000년 일본 동경대학 전자정보공학과 박사
2001년~현재 한국과학기술연구원 포토닉스센서시스템센터 책임연구원
관심분야 : 내용기반 영상검색, 멀티미디어, 3차원영상 디스플레이, 컴퓨터비전