韓國IT서비스學會誌
第16卷 第1號
2017年 3月, pp.55-72

Journal of Information Technology Services
https://doi.org/10.9716/KITS.2017.16.1.055

시각주의 탐색 시스템을 위한 새로운 성능 평가 기법*

최 경 주**

A New Performance Evaluation Method for Visual Attention System*

Kyungjoo Cheoi**

■ Abstract ■

Many of the studies of visual attention that are currently underway are seeking ways to make application systems that can be used in practice, and obtained good results using not only simulated images but also real-world images. However, despite that previous studies of selective visual attention are models intended to implement the human vision, few experiments verified the models with actual humans and there is no standardized data nor standardized experimental method for actual images. Therefore, in this paper, we propose a new performance evaluation techniques necessary for evaluation of visual attention systems. We developed an evaluation method for evaluating the performance of the visual attention system through comparison with the results of the human experiments on visual attention. Human experiments on visual attention is an experiments where human beings are instinctively aware of the unconscious when images are given to humans. So it can be useful for evaluating performance of the bottom-up attention system. Also we propose a new selective attention system that guides the user to effectively detect ROI regions by using spatial and temporal features adaptively selected according to the input image. We evaluated the performance of proposed visual attention system through the developed performance evaluation method, and we could confirm that the results of the visual attention system are similar to those of the human visual attention.

Keyword : Visual Attention, Evaluation, Human Experiment

Submitted : February 4, 2017 1st Revision : February 19, 2017 Accepted : February 23, 2017
* This work was supported by the research grant of the Chungbuk National University in 2014.
** 충북대학교 전자정보대학 소프트웨어학과 부교수

1. 서 론

컴퓨터 시각(Computer Vision)이란 투시된 영상들로부터 주어진 장면에 관한 유용한 정보를 추출하여 물리적인 대상을 명확하고 의미 있게 기술하도록 하는 과정이라 할 수 있는데, 이러한 컴퓨터 시각에 있어서의 주요 문제점은 주어진 영상의 크기와 이에 따른 계산의 복잡도가 늘어나는데 따른 제한된 능력에 있다고 할 수 있다. 실제로 컴퓨터 시각 시스템은 엄청난 양의 영상정보 또는 시각정보를 받아들이는데, 공학적인 측면에서 볼 때, 매 순간마다 영상 입력장치를 통하여 들어오는 많은 영상정보들을 모두 처리한다는 것은 불가능할 뿐만 아니라 제한된 인지 자원을 효율적으로 사용하는 측면에서도 불필요한 일이므로 다양한 입력정보 중에서 필요한 정보만을 계속 더 처리하고 그렇지 않은 정보는 걸러내는 정보선택 과정이 필요하다고 할 수 있다. 실제로 우리는 실세계로부터 입력되는 시각정보 중, 응시할 영역을 선택적 시각주의집중 기능에 의해 선택적으로 결정하며, 도약안구운동 기능을 통해 응시할 영역으로 신속하게 시선을 이동시켜 복잡한 시각정보를 효율적으로 처리한다. 그리고 입력되는 시각 영역 중 어떤 부분에 집중하고 안구를 움직이는가는 이들의 행동습성에 크게 좌우된다(Navalpakkam et al., 2005). 이렇게 주의를 특정 영역이나 물체에 집중시켜 시각정보의 처리능력을 극대화하는 작용을 선택적 시각주의(selective visual attention)라 하는데, 이러한 기능을 컴퓨터 시각에서 정보선택의 방법으로 도입하고자 많은 연구가 진행되고 있다.

인간의 시각주의 기제가 행하는 기능을 실질적으로 컴퓨터 시각 시스템이 갖도록 하게 하기 위해서는 시간상 또는 공간상으로 의미 있는 영역(ROI; region of interest)이 선정되어져야 한다. 영상에서 ROI 영역을 다른 말로 주의영역이라 할 수 있는데, 이 영역은 다른 부분에 비해 상대적으로 중요한 정보, 많은 정보를 포함할 가능성이 큰 영역을 말하는 것이다. 입력되는 영상으로부터 의미 있는 특정 몇몇 영역을 성공적으로 탐지할 수 있다면 전체 영상보다 상대적으로 좁은 선택된 영역에 자원을 집중하여 정보처리의 효율을 극대화시킬 수 있다. 시각주의 시스템은 이러한 주의영역 탐지해야 하는데, 기본적으로 인간의 시각을 모델링한다. 따라서 컴퓨터 시각에서의 시각주의 시스템은 인간의 인지에 가능한 한 부합하게 만드는 것을 목적으로 한다(Zhai and Shah, 2006). 일반적으로 인간의 시각주의 모델에 관한 연구는 의미 있는 영역 탐지에 대한 해결책을 제공해 주고 있다. 인간의 시각주의에 관한 연구결과를 보면, 인간은 보통 영상이 주어지면 그 영상에서 몇몇 영역에만 주의를 가하며, 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. 심리학자들이 연구한 바에 따르면 인간의 시각 시스템은 목표가 되는 지역과 주변의 차이에 민감하고 이 차이로 물체를 구분해낸다고 하였다. 이러한 연구를 통하여 많은 사람들이 사람의 시신경에 기반한 특징들을 찾는데 연구의 초점을 맞추었고, 다양한 시각주의 시스템이 연구되어지고 개발되어졌다.

그러나, 이러한 시각주의 시스템들이 인간의 시각을 구현하고자 한 모델임에도 불구하고 실제 인간을 대상으로 하여 검증한 실험이 거의 없으며, 이러한 실험에 대한 표준화된 성능평가기법도 없다는 것이다. 이에 본 연구에서는 시각주의 시스템의 객관적인 성능평가를 위해, 그리고 실제로 개발된 시각주의 시스템이 인간의 시각주의 시스템과 얼마나 비슷한가를 확인하기 위해, 실제 인간을 대상으로 한 시각주의 실험방법을 개발하고, 이 결과와의 비교를 통해 시각주의 시스템의 성능을 평가하는 새로운 평가방법을 제시하고자 한다. 또한 기존의 시공간 정보를 특징으로 하는 시각주의 연구들이 가지고 있는 여러 한계점들을 보완하여 시스템을 통

해 탐색된 영역이 인간의 시각 탐색 전략과 보다 더 유사하게 부합되는 상향식 시각주의(bottom-up visual attention) 시스템을 제안하고, 제안된 시각주의 시스템을 개발된 성능평가방법을 통해 성능을 검증함으로써 제안된 시각주의 시스템이 사람의 시각주의와 유사함을 증명한다.

다음 제 2장에서 기존의 시각주의 탐색에 관련된, 제안하는 시스템과 기본 틀을 같이 하는 기존의 연구들에 대해 설명하고, 제 3장에서 시스템에 인간을 대상으로 한 시각주의 실험 방법과 구축된 실험 영상들, 그리고 이를 이용한 시각주의 시스템의 성능평가기법에 대해 설명한다. 이어 제 4장에서 개발된 성능평가기법에 따라 성능평가를 수행할 대상인 '사람의 시각 주의와 유사한 향상된 시각주의 탐색 시스템'에 대해 설명하고, 제 5장에서는 제 3장에서 소개된 성능평가기법에 따라 제안된 시각주의 시스템의 성능을 평가하고 분석하며 제 6장에서 결론을 맺는다.

2. 관련 연구

최초의 인간의 시각주의에 대한 계산 모델(computational model)은 '현저도맵(saliency map)'이라는 개념을 사용한 Koch와 Ullman의 모델(Koch and Ullman, 1987)이라고 할 수 있다. 여기서 현저도맵이란 주어진 영상에서 나머지 다른 영역들에 비해 현저한 주의를 일으키는 영역(= 주의영역)들의 상대적 차이를 자료화 한 현저도 값들을 2D 영상의 형태로 대응시킨 것이다. 이 계산 모델에서는 인간은 영상으로부터 색, 밝기, 움직임 등과 같은 기본 특징들에 대한 특징맵을 추출하고, 이들의 가중치 합으로 현저도맵을 생성한 후, 가장 현저한 위치로 주의를 이동한다는 가정을 하고 있다. 이렇게 구축된 현저도맵에서 값이 높은 지역은 영상에서 주의가 높은 지역이며 시각주의 탐색은 현저도맵에서 값이 높은 부분부터 순차적으로 탐색한다. 제안된 대부분의 성공적인 계산 모델들 간의 차이점이라 하면 입력되는 영상에서 특징을 추출하는 방법과, 현저도맵을 생성해내는 방법의 차이라 할 수 있다.

Koch와 Ullman의 시각주의 모델(Koch and Ullman, 1987)은 현재 연구되고 있는 시각주의 모델의 초석이 된 모델이며, 상향식 시각주의 모델이다. 여기서 상향식 시각주의란 색상과 모양 움직임 같이 신경자극을 통하여 무의식적으로 사람의 주의가 가는 부분을 우선 찾는 방법을 말하고, 하향식 시각주의(top-down visual attention)란 찾고자 하는 물체의 우선순위가 머릿속에서 먼저 인지되어 있어 이 순서대로 물체를 찾는 방법을 말한다. 하향식 시각주의는 적응적 학습이 우선되어야 하며, 특정 애플리케이션에 국한되기 때문에 본 연구는 상향식 시각주의에 기초하여 연구되었다. 상향식 시각주의 연구는 이미 기술한 바와 같이 입력되는 시각장면의 특징을 추출하고 분석하여 현저도맵으로 구성한 후 이를 기반으로 시각주의 영역을 탐색하며, 제안하는 시스템도 이 기본 틀을 같이 한다.

Koch와 Ullman의 시각주의 모델(Koch and Ullman, 1987) 이후 많은 연구에 의해 모델은 조금씩 발전되어 왔고 현재의 많은 연구들이 이를 더 발전시켜, 이론에서 벗어나 실제 사용될 수 있는 응용시스템으로의 모색을 꾀하였으며, 시뮬레이트된 영상뿐만 아니라 자연영상을 이용하여 실험하여 좋은 결과를 얻었다(Cheoi and Lee, 2002; Cheoi and Park, 2011; Dhavale and Itti, 2003; Hong, 2006; Itti and Koch, 2000; Navalpakkam et al., 2005; Zhai and Shah, 2006).

그러나, 기존 연구들은 다음과 같이 몇 가지 개선되어야 할 필요성이 있다. 가장 중요한 것으로, 이러한 시각주의 시스템들이 인간의 시각을 구현하고자 한 모델임에도 불구하고 Cheoi 및 Hong의 연구(Cheoi and Lee, 2002; Hong, 2006)를 제외하고는 실제 인간을 대상으로 하여 검증한 실험이 없으며, 이러한 실험에 대한 표준화된 성능평가기법도 없다는 것이다. 또한 주의 영역을 탐지하기 위해 시각적 특징 요소들을 선정할 때의 기

준이 불분명하다는 것이다. 그 동안 색깔, 명암, 크기, 방위 등 여러 가지 시각요소들을 사용하긴 하였지만, 별다른 기준이 없어 구현하기 편하다거나 가장 눈에 띌 것 같은 요소들을 연구자의 역량에 의해 선별, 사용되어 왔었다. 움직임의 경우에는 픽셀 단위로 움직임을 탐색하다보니 움직이는 물체의 크기에 적응적으로 움직임을 탐색하지 못하였다. 또한 기존의 연구들은 단순히 주의가 가는 영역들을 추출해내는 것에 그쳤었는데, 여러 부분이 동시에 추출되거나 아예 추출이 되지 않거나 하는 부분도 있었고, 한가지만을 추출하도록 만든 경우에는 추출된 부분 외에 더 중요한 부분이 있었음에도 불구하고 걸러내지 못하는 경우도 발생하였다. 영상 내에 주의를 일으키는 물체들이 다수 존재하고, 이들이 비슷한 공간적 특성을 보이는 경우에는 현저도의 부여 기준이 애매해진다. 이러한 문제들을 해결하기 위해 시각주의 시스템에 실세계에 대한 더 많은 정보가 더해질 필요가 있다. 이러한 사실을 통해 본 논문에서는 특징을 선별하고 추출하는 방법과 추출된 다양한 특징을 입력되는 영상에 따라 동적으로 결합하는 새로운 방법을 제시하고 개발된 시각주의 시스템이 인간의 시각주의 시스템과 얼마나 비슷한가를 확인하기 위해, 실제 인간을 대상으로 한 시각주의 실험방법을 개발하고, 이 결과와의 비교를 통해 시각주의 시스템의 성능을 평가하는 새로운 평가방법을 제시하고자 한다.

3. 인간을 대상으로 한 주의집중 실험을 통한 성능 평가 방법

3.1 인간의 주의 집중 설문조사

인간을 대상으로 한 주의집중 실험은 인간에게 영상이 주어졌을 때 인간이 무의식적으로 주의가 가는 부분이 어디인지를 알아보기 위한 실험으로 시스템의 상향식 부분에 대한 성능평가를 할 때 유용하게 사용할 수 있는 방법이다. 개발된 시각주의 시스템의 결과와 인간실험 결과가 유사하게 나오면 개발된 시스템의 성능이 좋다고 할 것이다.

본 연구에서 개발된 인간을 대상으로 한 주의집중 실험은 설문조사를 통해 이루어지는데, 이 설문조사는 영상이 주어졌을 때 인간이 무의식적으로 주의가 가는 부분이 어디인지를 알아보기 위한 상향식 주의에 관한 설문조사이다. 설문조사 대상자와 방법은 다음과 같다.

3.1.1 설문조사 기간, 대상자

설문조사 기간은 한 달 동안 이루어지는데, 1주일 간격으로 진행하되 설문시간은 실험에 사용되어질 실험 영상의 총 갯수에 따라 유동적으로 결정하여 실시한다. 설문조사의 목적이 인간이 무의식적으로 주의가 가는 부분이 어디인지를 알아보기 위한 방법이기 때문에 설문 대상자는 초등학생 저학년 학생 또는 유아를 대상으로 한다.

3.1.2 설문조사 방법

설문 진행 방법은 설문 대상자가 화면에 보여지는 영상을 보고 눈에 띄거나 자연스럽게 머릿속에 떠오르는 것을 순서대로 설문지의 결과 안에 왼쪽부터 적도록 한다. 이 때 설문지 결과안의 빈칸을 모두 채울 필요는 없다. 설문지는 실험에 사용되는 영상의 총 개수 만큼의 문제에 대해 4개의 답란을 가지는 양식을 갖추고 있으며, 샘플 양식은 <Figure 1>과 같다. <Figure 1>은 총 4개의 영상에 대하여 설문하는 양식의 예이다.

gender :	school year :	( ) time

	1st Search	2nd Search	3rd Search	4th Search
1				
2				
3				
4				

〈Figure 1〉 Questionnaire form

설문 대상 영상은 컴퓨터에 연결된 TV를 통하여 설문 대상자들에게 보여준다. 정지영상의 경우 1초, 동영상의 경우 5초 이내로 보여준다. 설문 진행

시 영상에 대한 고차원적인 지식이 들어가면 안되기 때문에 실험 영상에 대한 어떠한 정보도 사전에 주지 않아야 하며, 영상은 반드시 한 번만 보여주어야 한다. 매회 영상을 보여줄 때마다 영상이 구성되는 영상 순서를 임의로 변경하고 동일한 영상을 사용하지 않도록 매 회에 사전 지식이 습득되지 않도록 한다. 화면에 영상이 보여지게 되면 매 영상 뒤에는 카운트다운 화면이 나와 10초 동안 영상에 대한 설문을 작성할 시간이 주어지므로 설문 대상자는 이때 설문지를 작성한다.

3.1.3 설문 결과의 분석

1) 설문지 결과를 통한 인간 실험 결과 분석 방법

시각주의 시스템에서 출력되는 결과와의 비교를 위하여 설문 결과를 정리하고 분석한다. 설문 결과에서 서로 다른 영역이 비슷하게 눈에 띌 수 있기 때문에 조사된 설문지의 각 문제의 결과를 순위에 상관없이 가장 많이 나온 결과 순으로 정리하여 순서가 높은 결과가 높은 주의를 가지는 것으로 분석한다. 예를 들면, 하나의 문제에 대하여 순위에 상관없이 바나나 4회, 사과 3회, 고추 2회, 딸기 2회, 오렌지 2회, 양파 1회, 무 1회, 키위 1회로 결과가 조사되었으면, 주의가 가장 높은 1순위는 바나나가 되며 2순위는 같은 횟수로 지목 당한 사과, 고추, 딸기, 오렌지가 되고 마지막으로 3순위는 양파, 무, 키위가 된다.

여기서 주의해야 할 점은 설문지에 답하는 문구가 실제로는 같은 영역인데도 설문 대상자마다 서로 다르게 적을 수 있다는 것이다. 이런 이유 때문에 유사하다고 생각되는 문구들은 하나의 문구로 통일하고, 이 문구들이 나타난 숫자들의 합을 계산하여 비교 자료로 사용한다. 예를 들어, 빨강, 빨간, 빨간색, 적색과 같은 단어는 의미는 모두 같은데, 표현이 여러 개이다. 따라서 이런 경우 이들 표현 중 하나의 표현으로 통일시켜야 옳은 결과를 낸다고 할 수 있다. 반면에 통합하지 않는 문구도 있는데, 예를 들면 초원에 말이 있는 영상의 경우 뒤의 초원을 보고 녹색, 초원, 풀 과같이 여러 가지 연상을 할 수 있다. 이런 확실하게 정의되지 못하는 문구들은 대다수의 사람이 어떤 문구를 더 떠올리는지 알기 위하여 통합하지 않는다.

2) 인간 실험 결과와 시각주의 시스템 출력결과와의 비교 분석 방법

인간 실험 결과와 시각주의 시스템 결과와의 비교를 통한 성능 분석은 설문조사 결과에서 정리한 결과의 상위 4개 결과와 시각주의 시스템에서 탐색한 결과의 일치율로 분석하였다. 일치율은 순위에 상관하지 않고 설문조사 결과와 시스템 결과에 대하여 일치하는 비율로 결정하는데 하나씩 틀릴 때마다 25%의 감소율을 가지도록 한다. 예를 들어 어떤 영상에 대한 시각주의 시스템의 탐색 순서가 바나나, 사과, 고추, 딸기이고, 정리된 설문 결과가 바나나, 사과, 고추, 딸기면 이는 100%의 일치율을 가진다고 평가한다. 하지만 동일한 시스템 결과에 대하여 정리된 설문조사 결과가 바나나, 사과, 오렌지, 고추가 되면 시스템 결과에 오렌지가 없으므로 75%의 일치율을 가진다고 평가한다.

3.2 성능 분석용 영상

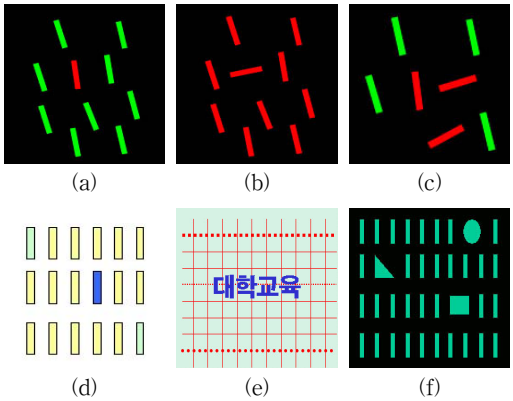
성능 분석용 영상은 실제 시각주의 시스템에서 성능평가를 위해 실험에 사용되는 영상으로 다양한 분야의 영상을 설문조사 영상으로 구축하기 위하여 정지영상과 동영상을 대상으로 여러 가지 기준을 수립하고 이 기준에 따라 영상을 구축하였다.

3.2.1 정지영상

정지영상으로는 글자, 도형 등과 같은 간단한 인공영상과, 외부에서 고해상도로 멀리에서 실제로 촬영한 실영상, 그리고 기타영상으로 인물사진, 포스터, 예술 작품 등 다양한 부류의 영상을 수집하여 사용하였다.

1) 인공영상

인공영상은 <Figure 2>에서 보이는 바와 같이



(a) Image Used in Color pop-out
(b) Image Used in Orientation pop-out
(c) Image Used in Conjunctive Search
(d) Image with Similar Color Saliency
(e) Image with Characters and Lines
(f) Image with Multiple Shapes

〈Figure 2〉 Some Examples of Artificial Images

6가지 종류로 분류하여 구축하였는데 전형적인 pop-out 문제에 사용되는 영상(<Figure 2(a)~(b)>), 결합탐색 문제에 사용되는 영상(<Figure 2(c)>), 비슷한 색상 현저도를 가진 영상(<Figure 2(d)>), 색 및 형태면에서 다양한 특이를 가지는 영상(<Figure 2(e)~(f)>)을 종류별로 분류하여 만들었다. <Figure 2(a)>는 색상 pop-out 문제에 사용되는 영상의 예로, 탐색을 방해하는 물체(방해자)는 녹색 막대이고, 찾고자하는 물체(목표물)는 적색 막대이다. <Figure 2(b)>는 방위 pop-out에 사용되는 영상의 예로, 모든 물체가 적색인 상황에서 방해자는 목표물과 방향이 다른 물체이다. <Figure 2(c)>는 결합탐색 문제에 사용되는 영상의 예로, 색상 pop-out과 방위 pop-out 영상이 합쳐진 영상이다. 방해자는 녹색의 다수의 주방향을 가진 것이며 목표물은 적색과 전체에서 소수의 반대방향을 가진 것이다. <Figure 2(d)>는 서로 비슷한 색상 현저도를 가진 영상에 대한 예로, 색상 pop-out의 방해자에 눈에 띄는 색이 아닌 목표물과 비교적 비슷한 눈에 띄는 색을 가지도록 하였다. <Figure 2(e)>와 <Figure 2(f)>는 형태 특이를 가지는 영상에 대한 예로, <Figure 2(e)>는 굵지만 청색으로 쓰여진 글자와, 얇지만 적색으로 그려진 선 중 어느 부분에 대해서 더 시선이 가는지에 대한 실험에 사용될 수 있는 영상의 예이다. 또한 <Figure 2(f)>는 색상은 동일한 상황에서 어떤 형태의 물체를 먼저 찾아내는지에 대한 실험에 사용될 수 있는 영상의 예이다.

2) 실영상

실영상은 외부에서 고해상도로 멀리에서 실제로 촬영한 영상으로, 영상을 구성하고 있는 특징 중 눈에 띄는 특징이 있는 영상의 경우로 제한하였고, 색상 및 배경의 복잡도가 낮은 영상(<Figure 3(a)>), 복잡도가 중간인 영상(<Figure 3(b)>), 복잡도가 높은 영상(<Figure 3(c)>)으로 구분하여 수집하였다. <Figure 3>은 구축된 실영상의 예를 보여준다. 복잡도가 낮은 영상은 영상에 눈에 띄는 특징이 하나가 있고 주변의 색상을 구성하는 색들의 종류가 단순한 영상이며 눈에 띄는 특징을 찾아내는 실험에 사용한다. 대부분 단일 표지판을 눈에 띄는 특징으로 선별하여 구성하였다. 복잡도가 중간인 영상은 눈에 띄는 특징과 주변의 색상의 종류가 비슷한 경우이거나 눈에 띄는 특징이 2개에서 3개의 가지고 있는 경우이다. 눈에 띄는 특징은 도로에 놓인 안전표시판이나 신호등, 차량, 신호 표지판을 함께 가진 영상으로 구성하였다. 복잡도가 높은 영상은 눈에 띄는 특징이 3개 이상이며 주변 배경의 색의 구성이 다양한 영상이다. <Figure 3(c)>처럼 간판이 2~3가지 있고 전체적으로 복잡한 구성을 가지고 있는 영상으로 구성하였다.

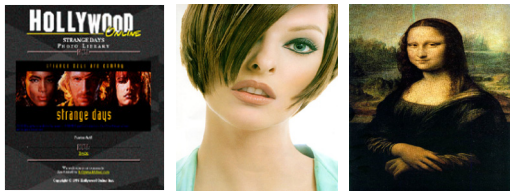
(a) Image with Low Complexity
(b) Image with Medium Complexity
(c) Image with High Complexity

〈Figure 3〉 Some Examples of Real-World Images

3) 기타영상

인공영상 및 실영상 외에 기타영상으로 포스터 및 예술 작품, 인물 사진 등 다양한 영상을 수집하였다. 포스터는 우리 주변에서 흔히 볼 수 있는 영화 포스터를 대상으로 하였다(<Figure 4> 참조).

포스터는 사람의 눈길을 끌기 위한 그림이기 때문에 눈에 띄는 부분을 1개 이상 가지고 있으며 사용된 색의 수가 적은 경우 수집하였다. 인물사진의 경우 사람의 얼굴과 목 부분을 중심으로 찍었고 입술이나 눈, 옷에 주의가 가는 영상들로 구성하였고, 예술 작품은 시대적으로 유명하여 사람들이 입에 오르내렸던 그림으로 구성하였다. 대표적인 것으로 모나리자가 있으며 이외에도 고전적으로 보이게 그린 그림들로 과일이 담긴 바구니나 중세 시대에 집에서 사람들이 만나는 모습을 그린 것들을 수집하였다.

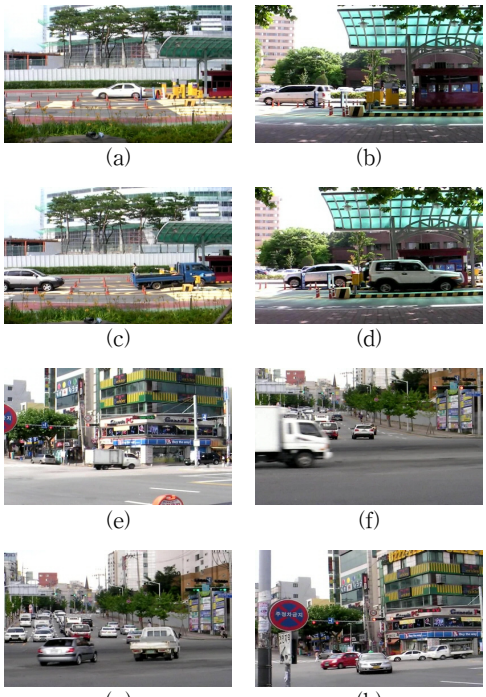


(a) (b) (c)

〈Figure 4〉 Some Examples of (a) Poster (b) Face Image (c) Painting

3.2.2 동영상

동영상은 대부분 도로변에서 흔히 볼 수 있는 영상을 촬영하여 수집하였다. 동영상은 영상 내 공간특징 복잡도와 시간특징 복잡도에 따라 다음과 같이 8가지로 구분하여 수집하였다. 공간특징 복잡도는 영상이 아주 복잡한 배경을 가지고 있어서 영상에서 눈에 띄는 배경의 공간특징이 매우 다양할 경우 높음으로, 그렇지 않은 경우는 낮음으로 구분하였다. 이렇게 공간특징 복잡도에 따라 나뉜 영상은 다시 시간특징 복잡도에 따라 낮고 높음으로 나누었는데 시간특징 복잡도가 낮다는 것은 영상 내 움직이는 물체의 개수가 1개이면서 움직이는 물체를 쉽게 찾을 수 있는 영상이고, 복잡하다는 것은 다수의 물체가 움직이는 경우라서 여러 단계에 걸쳐 시선처리를 해야 하는 경우를 말한다. 이렇게 본 연구에서는 동영상을 수집할 때 영상 내 움직이는 물체의 개수와 크기에 따라 영상을 구분하였는데, 먼저 움직이는 물체의 크기를 고려한 것은 해당 물체가 영상에서 어느 정도 큰 비중을 차지하여 움직임이 작더라도 눈에 띄게 되는



(a) (b) (c) (d) (e) (f) (g) (h)

(a) Video with Low Spatial Complexity in which Single Small Moving Object Exists
(b) Video with Low Spatial Complexity in which Single Big Moving Object Exists
(c) Video with Low Spatial Complexity in which Multiple Moving Objects Moves to Same Direction
(d) Video with Low Spatial Complexity in which Multiple Moving Objects Move to Different Direction
(e) Video with Low High Spatial Complexity in which Single Small Moving Object Exists
(f) Video with High Spatial Complexity in which Single Big Moving Object Exists
(g) Video with Low High Spatial Complexity in which Multiple Objects Move to Same Direction
(h) Video with Low High Spatial Complexity in which Multiple Objects Move to Different Direction

〈Figure 5〉 Some Examples of Videos

경우와, 반대로 영상에서 차지하는 비중이 작아 움직임이 큼에도 불구하고 바로 눈에 띄지 않는 경우를 고려하기 위함이다. 또 다른 경우로 움직이는 물체의 개수를 고려하여 다수의 움직이는 물체가 이동방향이 하나의 방향인 경우와 서로 다른 방향으로 이동하는 경우를 모두 고려하였는데, 이는 움직이는 물체들의 상대적인 속도가 현저도 변화에 미치는 영향을 고려하기 위함이다. <Figure 5>는 이렇게 구축된 영상의 일부를 보여준다.

4. 사람의 시각주의와 유사한 향상된 시각주의 탐색 시스템

본 연구에서 제안하는 시각주의 탐색 시스템은 공간과 시간 특징을 추출한 후 분석하여 현저도를 계산해내고 이를 토대로 주의 영역을 찾는 기존의 시스템에서 채택한 전형적인 상향식 방법에 기반을 하고 있지만, 특징을 추출하고 현저도를 계산하는 방법에 있어 기존의 연구에 비해 개선을 이루었다. 제안하는 시스템에서, 공간현저도는 입력되는 영상에 따라 특징이 적응적으로 선택되는 는 동적인 정보를 가지며, 시간현저도는 움직임 특징 분석을 통해 움직이는 객체마다 서로 연관성 있는 정보를 가진다. 또한 공간과 시간현저도를 결합할 때에는 입력되는 현저도의 활동량을 측정하여 활동량에 따라 동적으로 변하는 가중치에 계산하고 이에 따라 결합하는 방식을 채택하였다.

4.1 시공간 현저도 계산

인간의 눈은 모양, 색깔, 거리, 움직임, 명암, 질감, 기울기, 위치, 방향 등 주변 환경으로부터 다양하고 많은 정보들을 제공받는데, 그 수많은 정보에도 불구하고 체계적으로 분석하고 이해하는데 어려움을 겪지 않는 것은 빠르고 효율적인 시각정보처리체계를 가지고 있기 때문이다. 실제로 1987년에 Livingstone과 Hubel은 눈으로 들어온 여러 정보들 중 색, 형태, 깊이, 움직임과 같은 4가지 정보를 병렬적으로 처리하는 기능적 경로가 뇌 속에 있고, 그를 통해 시각처리과정이 일어난다고 주장하였으며(Livingstone and Hubel, 1988), 다양한 임상실험과 관찰로 이 주장이 옳음을 증명하고 있다. 깊이 특징을 추출하기 위해서는 깊이특징을 받아들이는 스테레오 카메라가 필수적으로 필요한데, 설치가 까다로우며 장비가 고가이기 때문에 현재 일반화되어 사용하기에는 이른 감이 있다고 판단되어, 본 연구에서는 깊이 특징을 제외한 색, 형태, 움직임의 3개의 특징을 주요 특징으로 정의하였다.

4.1.1 색과 형태 특징 추출 및 공간 현저도 계산

1) 색 특징 추출

색 특징으로 색상(hue) 2개(F^{11}, F^{12})와 명도(intensity) 1개(F^{13})를 사용하였으며, 적/녹(적+/녹-, 적-/녹+), 청/황(청+/황-, 청-/황+), 흑/백(흑+/백-, 흑-/백+)의 3개의 특징부류마다 상대적 특징맵을 모두 추출한 후 입력된 영상에 따라 2개의 상대적 특징맵 중 하나의 맵을 특징맵의 활동량을 비교하여 활동량이 큰 특징맵을 적응적으로 선택하도록 하였다. 특징맵의 활동량을 계산하는 방법은 특정 특징맵의 전역적인 최대특징값에서 국부적인 최대 특징값을 빼는 과정으로 계산한다. 전역적인 최대값은 해당맵의 특징값 중 가장 큰 값을 의미하고, 국부적 최대값은 해당맵의 x축과 y축의 라인별 최대 특징값의 평균 중 가장 높은 값을 선택하여 사용함으로써 x축과 y축 각각에 대한 변화량을 모두 고려할 수 있도록 하였다. 이 과정에 의하면 특징맵에서 특징값들의 차이가 많이 나면 해당 맵의 활동량은 높아지고, 반대로 특징값들이 차이가 적게 나면 해당 맵의 활동량은 낮아지게 된다.

2) 형태 특징 추출

형태 특징으로 인간의 “ON-중심, OFF-주변” 수용야(Receptive Field)에서 보이는 세포 반응도를 모방한 중심-주변 연산을 수행하여 추출하는데, 이 연산은 추출되어진 특징들 중 다른 주변의 특징들과 비교했을 때 현저히 차이가 나는 특징을 부각시

키고 그렇지 않은 특징들은 억제시키는 역할을 한다. 형태 특징은 입력영상에서 추출하지 않고 앞 단계에서 추출된 색상과 명도 특징맵에서 추출한다. 형태 특징을 추출하기 위하여 색상과 명도 특징맵 각각에 대하여 8개의 방위(Θ∈{0, π/8, 2π/8, …, 7π/8})에 조율된 8개의 DOOrG 연산자를 컨볼루션시킨 후, 가장 활동량이 큰 방위를 선택하여 형태 특징맵으로 추출한다. 이 과정을 통해 색상과 명도 특징맵인 F^{11}, F^{12}, F^{13}에 대하여 각각 8개의 방위 중 가장 큰 반응을 가지는 형태 특징맵이 선택되어 F^{21}, F^{22}, F^{23}이 생성된다.

3) 공간현저도 계산

윗 단계에서 계산된 3개의 형태 특징맵을 결합하여 공간현저도맵을 만든다. 본 연구에서 제안하는 결합 방법은 특징의 두드러짐 정도에 따라 적절한 가중치를 동적으로 부여하여 결합하는 방법으로 각각의 형태 특징맵에 대하여 식 (1)과 같은 방법으로 k번째 맵에 대한 가중치를 계산하여 각각 할당하고, 각 형태 특징맵에 각각 할당된 가중치를 곱하여 모두 합하는 방식으로 결합한다. 식 (1)에서 Act(F)는 특징맵 F의 활동량을 계산하는 함수이고, max(F)는 특징맵 F의 최대특징값을 나타낸다.

$$W_k = \frac{\sum_{k=1}^{n} Act(F^k)}{\sum_{k=1}^{n} Act(F^k) - \max(F^k)} \qquad (1)$$

4.1.2 움직임 특징 추출 및 시간현저도 계산

움직임 특징은 깊이 지각이나 색채 시각이 형편없는 동물들은 있어도 움직임을 지각하지 못하는 동물은 없다는 것에서도 그 중요성을 알 수 있다. 영장류에 있어 복잡한 움직임 처리는 2단계에 걸쳐 일어난다. 첫 번째 단계에서는 1차원의 국부적인 움직임이 분석되고, 두 번째 단계에서 국부적인 움직임이 전역적인 움직임으로 통합된다. 이러한 프로세스를 “움직임 통합”이라 하며, 본 연구에서는 이러한 기본지식을 바탕으로 움직임을 통합한다.

본 연구에서는 단위시간 동안 입력된 영상에서의 물체의 이동을 탐지하여 단위시간 동안 발생한 시간주의를 계산한다. 시간주의에서 어느 한 물체의 이동이 다른 물체들 보다 크게 일어난다면 이 물체는 다른 물체보다 더 높은 현저도를 가진다. 시간현저도를 계산하기 위하여 두 프레임 영상 사이에서 SIFT 알고리즘(Lowe, 2004)의 인자 및 다해상도 영상 구축 방법을 일부 변경하여 흥미점(F^{04}, F^{05})을 추출하였고, 추출된 흥미점을 기반으로 SIFT 매칭을 통해 흥미점들간의 서로 일치하는 점을 구한 후, 이들의 좌표값의 차이로 움직임 크기(이동량) 및 8개로 양자화시킨 방향을 계산하여 움직임 특징맵(F^{14})을 생성하고, 움직임 특징맵에서 움직임이 있는 영역이 어떤 부분인지 표현하여 시간현저도맵을 만든다. 시간현저도맵은 움직임 특징맵의 각 픽셀마다 같은 움직임 크기와 방향을 가지는 움직임 벡터 쌍들의 집합을 구하고 이 값들 중 최대와 최소 (x, y) 좌표를 선택한 다음, 선택된 좌표로 사각형 영역의 크기를 계산하여 이 영역의 움직임 크기를 해당 영역의 시간현저도 값으로 부여하여 만든다.

4.2 시공간현저도 결합

공간과 시간현저도 결합을 위해 본 연구에서는 인간은 공간적 특징보다 움직임에 더 민감하다는 사실과, 배경과 사람이 모두 움직일 경우에는 보통 배경보다 움직이는 사람에 대하여 더 주의를 가진다는 사실에 기인하여 가중치를 부여하여 결합하는 방법을 사용하였다. 그런데 일반적으로 움직임의 크기가 작을 때는 작은 움직임보다는 색상이나 밝기와 같은 공간현저도에 더 많은 주의를 주게 되므로 이러한 사실을 바탕으로 시간 특징이 높으면 시간현저도맵에 더 높은 가중치를 두고 반대의 경우에는 공간현저도맵에 더 높은 가중치를 주었다. 시공간현저도맵(S)은 식 (2)를 통해 계산되어진다.

$$S = (1 - K_t) * S_s + K_t * S_t \qquad (2)$$

식 (2)에서 K_t는 시간현저도맵(S_t)에 부여하는 가중치이며, 공간현저도맵(S_s)에 부여되는 가중치는 $1-K_t$로 계산된다. 시간현저도맵의 가중치 K_t는 움직임이 있는 블록의 크기, 움직임 블록간의 값의 차이, 움직임 블록간의 영역 차이 3가지를 고려하여 결정한다. 먼저, 움직임이 있는 블록의 크기는 전체 블록 중 움직임이 있는 블록을 찾아 영상에서 얼마만큼의 움직임이 존재하는지를 계산한다. 이 값이 클수록 시간현저도맵에 더 높은 가중치가 부여된다. 움직임 블록간의 값의 차이는 '1-블록차이값'으로 블록 차이 값은 영상 내 최대 움직임 값에서 '최대치를 제외한 나머지 블록의 평균값'을 뺀다. 블록 차이 값은 영상에서 최대값의 움직임만 존재할 경우 0이 되고 이외의 경우 0.4, 0.6, 0.8 등과 같은 수치가 나오게 된다. 블록 차이 값이 작을수록 시간현저도맵에 높은 가중치가 부여된다. 마지막으로 움직임 블록 간의 영역 차이는 '1-움직임블록차이값'으로 구한다. 이때 움직임 블록 차이 값은 '최대움직임을 제외한 블록/움직임이 존재하는 모든 블록'으로 구하는데 최대 움직임이 많을수록 이 비율은 낮아지게 되고 반대로 최대움직임이 상대적으로 적을 경우 이 비율은 높아지게 된다. 움직임 블록차이값이 낮을수록 시간현저도맵에 더 높은 가중치가 부여된다.

5. 인간의 대상으로 한 실험결과와의 비교 분석

제 4장에서 제안한 시각주의 시스템의 성능은 제 3장에서 제안된 인간을 대상으로 한 실험결과와 비교/분석함으로서 평가하였으며, 실험에 사용된 영상도 3장에서 제시된 성능분석영상을 사용하였다.

5.1 인간을 대상으로 한 실험

인간을 대상으로 한 실험은 제 3장에서 제시한 방법에 따라 설문조사를 통하여 결과를 얻었다. 설문조사는 1주일 간격으로 3회 실시하여 자료를 수집하였다. 설문 종류로는 설문 대상자에게 설문지를 나누어주고 설문 대상자가 직접 설문지에 결과를 기술하는 방식을 사용하였고, 설문조사 대상자는 서로 다른 지역에 사는 초등학생 총 72명(남 42명/여 30명)으로 3학년 초등학생 22명(남 11명/여 11명), 5학년 초등학생 22명(남 11명/여 11명)이다.

실험 영상은 초등학교 각 학급에 설치되어 있는 컴퓨터와 연결되어 있는 TV를 통하여 설문 대상자들에게 보여주었는데, 실험 영상 중 정지영상은 2장에서 설명된 인공영상과 실영상(복잡도 낮음, 복잡도 중간, 복잡도 높음), 기타 영상(예술작품, 인물사진 등)의 총 51개를 특징에 따라 고루 선택하였으며 1회에 인공영상 5개와 실영상 및 기타영상 12개씩 총 17개씩 보여주는 방식으로 총 3회에 걸쳐 실시하였다. 동영상은 8가지 특징의 동영상을 총 24개를 선택하여 1회에 특징마다 1개씩 총 8개를 보여주고 3회에 걸쳐 실시하였다.

5.2 인간을 대상으로 한 실험과 시스템의 결과에 대한 비교실험

5.2.1 정지영상에 대한 결과

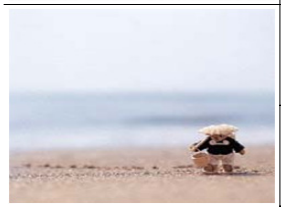
1) 복잡도가 낮은 실영상

복잡도가 낮은 실영상에 대한 전체 비교 분석 결과는 <Table 1>과 같다. <Table 2>의 첫 번째 영상은 <Table 1>의 9번 실험 영상으로 '컴퓨터과학과'와 'computer science'라는 글자가 회색 배경의 팻말에 청색으로 쓰여 있다. 팻말의 배경은 백색이다. 제안된 시스템에서는 '컴퓨터과학'이란 글자를 가장 먼저 찾고 그 다음으로 'computer science'라는 글자를 찾아냈다. 설문조사에서는 글자 33회, 영어 6회, 청색 5회, 백색 4회로 조사되었다. 이 실험에서는 시스템과 설문조사의 결과가 유사하게 나왔다. <Table 1>의 2번째 영상은 <Table 1>의 2번 실험 영상으로 시스템은 곰 인형을 가장 먼저 찾아냈고 그 다음으로 모래와 바다를 찾았다. 설문조사에서는 인형 32회, 물 25회,

〈Table 1〉 The Overall Results of Real-World Images with Low Complexity

rank / image no.	1st search	1~4th search
1	0	75
2	100	100
3	100	100
4	100	50
5	100	50
6	0	100
7	0	50
8	100	75
9	100	75
average	66.67	75.00

〈Table 2〉 Some Results on Image with Low Spatial Complexity

proposed attention system	human experiments	
	item	count
	character	33
	English	6
	blue	5
	white	4
	doll	32
	water	25
	ocean	16
	sand	12

바닷가 16회, 모래 12회로 조사되었다. 앞의 실험과 비교하였을 때 조사된 횟수의 총합이 많이 차이가 나는데 이는 한 영상에서 얻어낼 수 있는 정보량의 차이가 있기 때문이다. 첫 번째 영상에 대한 실험 결과는 단순한 정보만 유추할 수 있기 때문에 사람의 하향식 정보가 많이 반영되지 못하였는데 두 번째 영상에 대한 실험 결과에서는 바다, 물, 해변과 같은 정보가 여러 가지 표현으로 쓰일 수 있어 더 많은 총합이 나왔다.

복잡도가 낮은 실영상 전체에 대한 비교 분석 결과인 <Table 1>을 보면 1순위 일치율이 1-4순위 일치율보다 낮게 나왔지만 두 일치율에 큰 차이는 없었다. 이는 영상이 단순한 정보를 가지고 있기 때문에 선택할 수 있는 자료가 한정되어 결과의 차이가 적게 나온 것으로 보인다.

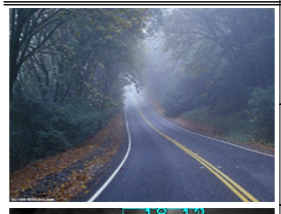
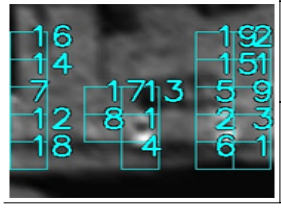
2) 복잡도가 중간인 실영상

복잡도가 중간인 실영상의 전체 비교 분석 결과는 <Table 3>과 같다. <Table 4>의 첫 번째 영상은 <Table 3>의 4번 실험 영상으로 가을에 낙엽이 내린 안개 낀 도로를 찍은 영상이다. 제안된 시스템에서는 도로의 황색 중앙선을 가장 먼저 찾고 다음으로 낙엽과 안개를 찾아냈다. 설문조사에서는 도로 30회, 나무 26회, 안개 18회, 낙엽 14회로 조사되었다. 도로라는 고차원적인 지식정보를 중앙선에 대입하면 시스템과 설문조사의 결과가 유사함을 알 수 있다. <Table 4>의 두 번째 영상은 <Table 3>의 7번 실험 영상으로 적색 벽돌로 이루어진 건물을 찍은 영상이다. 적색의 벽돌로 만들어진 벽 앞에는 백색 띠를 두른 적색의 안전콘(cone)이 있다. 시스템에서는 백색 띠를 두른 적색의 안전콘을 먼저 찾고 다음으로 적색의 벽돌로 된 벽을 찾아냈다. 설문조사에서는 집 13회, 적색 11회, 벽돌 9회, 벽 7회로 조사되었다. 설문조사 결과에서는 다소 고차원적 지식이 많이 반영 되었는데 전체적으로 보았을 때 적색과 벽에 주의를 두었다는 것을 알 수 있다. 이로 비추어 볼 때 제안된 시스템과 설문조사의 결과는 유사하다고 할 수 있다.

〈Table 3〉 The Overall Results of Real-World Images with Medium Complexity

rank / image no.	1st search	1~4th search
1	100	75
2	0	100
3	100	100
4	100	75
5	0	50
6	0	75
7	0	100
8	100	25
9	0	50
average	44.44	72.22

〈Table 4〉 Some Results on Image with Low Spatial Complexity

proposed attention system	human experiments	
	item	count
	road	30
	tree	26
	fog	18
	leaves	14
	building	13
	red	11
	brick	9
	wall	7

복잡도가 중간인 실영상 전체에 대한 비교 분석 결과인 <Table 3>을 보면 1순위 매칭율은 44%, 1~4순위 매칭율은 72%가 나왔다. 1순위 매칭율이 복잡도가 단순한 영상보다 낮게 나온 이유는 설문조사 결과에 집과 같은 고차원적 정보가 많이 들어갔기 때문으로 보인다. 또한 1~4순위의 결과를 보았을 때 비슷하지만 낮은 일치율을 보이는데, 이는 선택할 수 있는 자료의 량이 많아지면서 자료간의 구분이 명확해졌지만 아직도 적은 정보량의 영상이기 때문에 단위 시간 동안 한 물체에 대한 여러 가지 고차원지식이 반영되었기 때문으로 보인다.

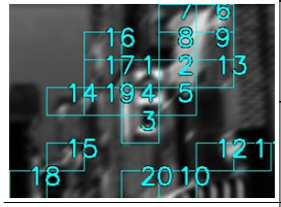
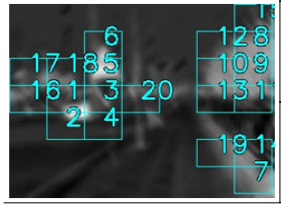
3) 복잡도가 높은 실영상

복잡도가 높은 실영상의 전체 비교 분석 결과는 <Table 5>와 같다. <Table 6>의 첫 번째 영상은 <Table 5>의 4번 실험 영상으로 여러 집들이 있는 장소에서 적색과 청색으로 이루어진 간판을 찍은 영상이다. 시스템에서는 적색의 간판을 먼저 찾고 다음으로 적색의 벽돌과 벽을 찾아냈다. 설문조사에서는 가게 29회, 집 14회, 가게 간판 중 스마일이라는 글자에 11회, 전체 가게 간판 11회로 조사되었다. 설문조사에서 가게나, 집, 간판과 같은 고차원적 지식이 많이 반영되긴 하였지만 전체적으로 시스템이 찾은 위치와 설문조사 결과는 유사하다고 할 수 있다. <Table 7>의 두 번째 영상은 <Table 1>의 7번 실험 영상으로 아파트를 배경으로 자동차와 표지판을 찍은 영상이다. 시스템은 황색의 자동차를 먼저 찾고 다음으로 적색의 표지판, 하늘색의 울타리를 찾았다. 설문조사에서는 자동차 13회, 표지판에 쓰여진 '천천히'라는 글자 10회, 집 7회, 표지판 7회로 조사되었다. 원 영상에서 보면 집이라는 존재는 잘 찍히지 않았는데 설문 대상자들은 영상의 단서를 보고 유추해 낸 것으로 보이며, 신호등은 그 크기가 작아 잘 찾아내지 못한 것으로 판단된다. 하지만 자동차와 천천히라는 부분에서 일치함을 보여 유사한 결과를 낸다고 할 수 있다. 복잡도 높은 실영상 전체에 대한 비교 분석 결과인 <Table 5>를 보면, 1순위의

〈Table 5〉 The Overall Results of Real-World Images with High Complexity

image no. \ rank	1st search	1~4th search
1	100	100
2	100	75
3	100	75
4	100	100
5	0	75
6	100	100
7	100	75
8	100	100
9	100	50
average	88.89	83.33

〈Table 6〉 Some Results on Image with Low Spatial Complexity

proposed attention system	human experiments	
	item	count
	store	29
	house	14
	sign (character)	11
	sign (whole)	11
	car	13
	characters in the traffic sign	10
	house	7
	traffic sigh	7

일치율과 4순위의 일치율이 88%로 높게 나왔다. 이는 영상의 정보량이 많아지면서 단위 시간동안에 입력되는 고차원적 지식이 많이 줄어들었기 때문으로 보인다.

4) 기타 실험 영상

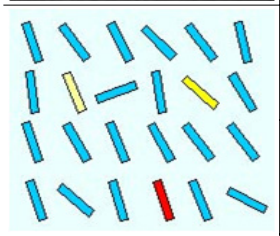
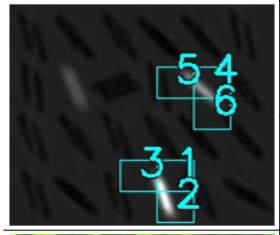
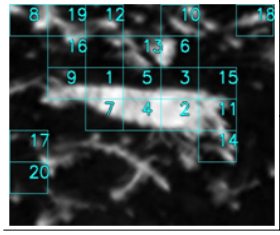
기타 실험 영상의 전체 비교 분석 결과는 <Table 7>과 같다. <Table 8>의 첫 번째 영상은 <Table 7>의 18번째 실험 영상으로 흑색 배경에 백색 글씨와 황색 글씨, 적색 빛 사람이 그려져 있는 포스터이다. 시스템은 사람, 영어 순으로 찾아냈고, 설문조사에서는 사람 20회, 영화 12회, 영어 12회, 할리우드 7회로 조사되었다. 영화라는 고전적인 지식을 빼면 제안하는 시스템에서 찾은 영역과 설문조사에서 주의를 둔 영역이 일치함을 확인 할 수 있다. <Table 8>의 두 번째 영상은 <Table 7>의 21번 실험 영상으로 하늘색 배경에 청색, 황색, 적색의 막대가 배치되어 있는 인공 영상이다. 시스템은 적색을 가장 먼저 찾고 다음으로 황색, 하늘색을 찾아냈다. 설문조사에서는 황색 15회, 적색색 11회, 막대바 10회, 하늘색 9회로 조사되었다. 1순위로 찾아낸 부분에서는 차이를 보이지만 1~4순위까지 고려한 경우에서 직사각형이라는 고차원적 지식을 배제하면 시스템과 설문조사 결과는 일치함을 확인 할 수 있다. <Table 8>의 3번째 영상은 <Table 7>의 22번 실험 영상으로 녹색 수초와 적색의 새우를 찍은 영상이다. 시스템에서는 새우를 먼저 찾고 다음으로 녹색 수초를 찾았다. 설문조사에서는 새우 25회, 초록색 5회, 풀 5회, 나무 5회로 조사되었다. 풀과 나무라는 고차원적 지식을 반영하면 제안하는 시스템과 설문조사 결과는 일치하는 것을 확인할 수 있다. 실영상이 아닌 인공영상 및 그 외 영상 전체에 대한 비교 분석 결과인 <Table 7>을 보면 1순위의 일치율은 66%, 1~4순위의 일치율은 94%가 나왔다. 1~4순위의 일치율이 복잡도가 낮은, 중간, 높은 영상보다 높게 나왔는데, 이는 영상이 목적을 가지고 만들어져 정보구성이 명확하였기 때문에 설문조사에서 결과가 여

러 개로 분산되지 않았기 때문으로 판단된다. 설문조사의 경우, 동영상을 보는 시간이 지남에 따라 학습된 지식이 들어가기 때문에 정확한 비교가 어려운 점이 있다.

〈Table 7〉 The Overall Results of Simple Artificial and Other Images

rank / image no.	1st search	1~4th search
1	100	100
2	100	75
2	0	100
3	0	75
4	100	75
5	100	100
6	0	100
7	100	100
8	100	100
9	100	100
10	0	75
11	0	100
12	100	100
13	0	75
14	0	50
15	0	100
16	100	75
17	100	75
18	100	100
19	100	100
20	100	50
21	0	100
22	100	100
23	100	100
3	100	75
4	100	100
5	0	75
6	100	100
7	100	75
8	100	75
9	100	75
10	100	100
11	0	75
12	100	100
13	100	75
14	100	100
15	100	75
16	100	75
17	100	100
18	0	75
19	100	100
20	100	75
21	100	75
22	100	75
23	100	100
24	100	100
average	66.67	91.67

〈Table 8〉 Some Results on Simple Artificial and Other Images

proposed attention system	human experiments	
	item	count
	man	20
	movie	12
	english	12
	hollywood	7
	yellow	15
	red	11
	bar	10
	sky blue	9
	shrimp	25
	green	5
	grass	5
	tree	5

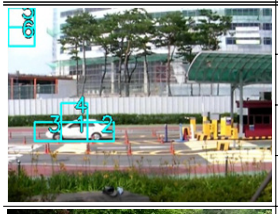
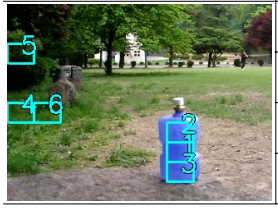
<Table 10>의 첫 번째 영상은 <Table 9>의 3번 영상으로 적색 톨게이트를 지나는 백색 차량을 찍은 영상이다. 시스템은 움직이는 차를 먼저 찾고 다음으로 하늘, 톨게이트, 안전물을 찾아냈다. 설문조사에서는 자동차 23회, 집 12회, 도로 7회, 황색 3회로 조사되었다. 1순위 탐색은 자동차로 두 결과 모두 일치하였고, 1~4순위 탐색은 도로를 제외하고는 일치한 결과를 보였다. <Table 10>의 두 번째 영상은 <Table 9>의 20번 영상으로 녹색 나무들 앞에 청색 병이 있으며, 나무 사이에서 사람들이 보이지 않을 만큼 작게 움직이고 있다.

〈Table 9〉 The Overall Results of Videos

spatial complexity	temporal complexity	image no.	1st search	1~4th search
low	small single moving object	1	100	75
		2	100	100
		3	100	75
	big single moving object	4	100	100
		5	100	75
		6	100	100
	multiple moving object (same direction)	7	100	75
		8	100	50
		9	100	50
	multiplemoving object (multiple directions)	10	100	50
		11	100	50
		12	100	10
high	small single moving object	13	100	75
		14	50	50
		15	100	50
	big single moving object	16	100	75
		17	50	50
		18	100	50
	multiple moving object (same direction)	19	100	75
		20	100	75
		21	100	50
	multiple moving object (multiple directions)	22	50	75
		23	100	50
		24	100	50
average			93.75	63.96

시스템 결과는 청색 병에 시선을 집중하는 것으로 나타났다. 다수의 작은 움직임이 있고 고정된 눈에 띄는 물체가 있는 경우, 즉 시간현저도가 공간현저도보다 작은 경우 공간현저도가 높은 물체에 먼저 시선을 집중하고 있음을 알 수 있다. 기존의 대부분의 연구는 항상 시간현저도에 우선을 두어 작은 움직임이라도 발생하면 움직임에 먼저 시선을 집중시키는 반면 제안된 시스템은 시간과 공간현저도를 결합할 때 현저도 크기에 따라 동적으로 결합하여 기존의 연구에서 발생했단 한계를 극복한다. 설문조사의 경우 청색이 13회, 사람 5회, 돌 2회, 바위 2회로 조사되었다. 1순위 탐색은 청색 병으로 두 결과 모두 일치하였다.

〈Table 10〉 Some Results on Videos

proposed attention system	human experiments	
	item	count
	car	23
	house	12
	road	7
	yellow	3
	blue	25
	person	5
	stone	2
	rock	2

5.3 비교 분석 결과 정리

제안된 시스템은 정지영상 실험 중 영상의 복잡도에 따라 나누어진 영상에서, 영상의 복잡도가 높은 영상에서 더 좋은 결과를 보였는데 이는 영상의 정보가 많아짐으로써 실험자의 고차원적 지식이 적용되는 시간이 줄어 부족하였기 때문으로 보인다. 또한 영상의 목적이 뚜렷한 기타 종류의 영상은 영상의 복잡도에 따른 실영상보다 높은 일치율을 보였는데 이러한 영상은 사람의 인지를 위

하여 정보가 구성되었기 때문으로 생각된다.

전체적으로 보았을 때 설문조사결과와의 일치율이 다소 낮게 나온 이유는 설문조사에서 답안을 자유롭게 쓰도록 하였는데 같은 물체에 대하여서도 여러 가지 다른 문구가 나와, 보다 보편적인 결과를 위하여 주관적인 정리를 할 수 밖에 없어 일정부분의 오차를 가지고 들어가야만 했다. 이러한 한계점을 극복하기 위해서는 설문 대상자들이 주의가 가는 부분을 문구로 쓰는 것이 아닌, 영상에 표시를 하거나 실험 영상에 대하여 기대되는 일정 문구를 사전에 지정해 주는 방법을 사용해야 할 것이다. 동영상의 경우 제안하는 시스템과 설문조사의 탐색 문구가 다소 다른데, 이는 제안하는 시스템은 현재와 이전의 단기간의 시간의 자료만 취합하며 고차원적 지식을 사용하지 않았기 때문이다. 사람의 경우 차를 한번 보고나면 학습이 되기 때문에 다른 차를 보는 것보다 주변의 물체를 더 많이 탐색하며 고차원적 인지를 수행한다. 이러한 오차 범위를 주관적인 분류로 수정을 하였을 때는 더 좋은 일치율을 보였다. 위 결과들을 토대로 제안된 시스템이 인간의 시각 인지와 유사함을 확인할 수 있다.

6. 결 론

현재 진행되고 있는 많은 시각주의에 대한 연구들은 이론에서 벗어나 실제 사용될 수 있는 응용 시스템으로의 모색을 꾀하고 있으며, 시뮬레이트된 영상뿐만 아니라 자연영상을 이용하여 실험하여 좋은 결과를 얻고 있다. 그러나, 이러한 시스템들이 인간의 시각을 구현하고자 한 모델임에도 불구하고 일부 시스템을 제외하고는 실제 인간을 대상으로 하여 검증한 실험이 거의 없다. 또한 화면상에서 사람들이 무의식적으로 어디에 집중하는지 알기 위해 시선에 대한 추적(tracing) 기법이 있으나, 그 결과를 어떤 방법으로 분석하여 시각주의 시스템 성능평가에 적용하는지에 대한 연구는 없다. 이에 본 연구에서는 실제 인간을 대상으로 한 시각주의 실험 결과와의 비교를 통해 시각주의 시스템의 성능을 평가하는 새로운 평가방법을 제시하였다. 또한 '사람의 시각 주의와 유사한 향상된 시각주의 탐색 시스템'을 제안하고, 제안된 시각주의 시스템은 개발된 성능평가방법을 통해 성능을 평가하였다. 실험 결과, 우선적으로 주의를 가지는 부분이 설문조사와 제안하는 시스템이 대부분 일치하여 사람이 주의를 가지는 부분과 유사한 결과를 내는 것을 확인 할 수 있었다.

설문조사는 고차원적 지식을 최대한 줄이기 위해 초등학생들을 대상으로 실시하였고, 다양한 환경의 결과를 고려하기 위하여 다른 3개 지역에 사는 다른 학년의 학생들에게 3회에 걸쳐 매 회 다른 영상과 영상 구성으로 설문조사를 실시하였다. 설문조사를 통한 시스템 성능 비교는 정지영상, 동영상을 대상으로 수행되었다. 정지영상은 실영상과 기타영상으로 나누어졌는데 실영상의 경우 영상의 복잡도가 높은 수록 일치율이 좋았으며, 기타영상의 경우 일정한 목적으로 만들어진 영상이어서 실영상보다 좋은 일치율을 보였다. 동영상의 경우 시간이 지남에 따라 설문자가 학습하는 부분이 생겨 결과에 고차원적 지식이 드러난 부분이 많았으나 대부분 제안하는 시스템과 일치하는 결과를 냈다. 하지만 설문조사 시 고차원적 지식을 최대한 배제하기 위하여 해당 문제에 대한 일정한 답을 주지 않았던 것 때문에 주관적인 분류를 해야 하여 일정부분의 오차를 감수해야만 하였고 이 부분은 조금 더 객관적으로 진행할 수 있도록 하여 보완하여야 할 것이다.

본 연구를 통해 추후 수행되어져야 할 과제가 2가지가 있다. 첫 번째로 설문조사 시 고차원적 지식이 들어간 문제의 답을 주관적으로 분류하였는데, 이러한 언어로 된 결과 수집이 아닌 영상에 직접 위치를 표시하는 실험을 수행하여 보다 객관적인 자료를 획득하여 비교하여야 한다. 마지막으로 영상 내 현저함 영역을 탐색 할 때 점(point) 위주로 현저함을 탐색하여 같은 물체를 여러 번 탐색하였는데, 현저함 영역을 묶는 작업을 하여 탐색의 효율을 높여야 한다. 마지막으로 설문분석결과

와 단순히 빈도수를 기술하는 기술통계를 이용하였는데 이를 발전시켜 통계적 검증 방법을 이용할 수 있도록 검증방법을 개선해야 한다.

References

Cheoi, K.J. and Y.B. Lee, "A Method of Extracting Objects of Interest with Possible Broad Application in Computer Vision", *Lecture Note on Computer Science*, Vol. 2525, 2002, 331-339.

Cheoi, K.J. and M.C. Park, "Visual Information Selection Mechanism Based on Human Visual Attention", *Journal of Korea Multimedia Society*, Vol.14, No.3, 2011, 378-391.
(최경주, 박민철, "인간의 주의시각에 기반한 시각정보 선택 방법", *멀티미디어학회논문지*, 제14권 제3호, 2011, 378-391.)

Dhavale, N. and L. Itti, "Saliency-based multifoveated MPEG Compression", *Proceedings of IEEE International Symposium on Signal Processing and its Applications*, Vol.1, 2003, 229-232.

Hong, H.M., "Vision Information Processing System Based On Visual Attention Using Four Vision Path In Human", Master Degree of Yonsei University, 2006.
(홍혜민, "인간의 네 가지 시각경로를 이용한 주의시각 기반 시각정보처리체계 시스템의 구현", 연세대학교 석사학위논문, 2006.)

Koch, C. and S. Ullman, *Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry*, L.M. Vaina(edt), *Matters of Intelligence*, Reidel Publishing, 1987, 115-141,

Itti, L. and C. Koch, "A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention", *Vision Research*, Vol.40, No.10-12, 2000, 1489-1506.

Livingstone, M. and E. Hubel, "Segregation of form, Color, Movement, and Depth : Anatomy, Physiology, and Perception", *Science*, Vol.240, No.4853, 1988, 740-749.

Lowe, D., "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, Vol.60, No.2, 2004, 91-110.

Navalpakkam, V., M. Arbib, and L. Itti, "Attention and Scene Understanding, Neurobiology of Attention", San Diego, CA : Elsevier, 2005, 197-203.

Zhai, Y. and M. Shah, "Visual Attention Detection in Video Sequences Using Spatiotemporal Cues", *Proceedings of the 14th annual ACM international conference on Multimedia*, 2006, 815-824.

About the Authors

Kyungjoo Cheoi (kjcheoi@chungbuk.ac.kr)

Professor Kyungjoo Cheoi received the B.S. degree in Computer Science from Chungbuk National University in 1996 and M.S. and Ph.D. degrees in Computer Science from Yonsei University in 1998 and 2002, respectively. During 2002~2005, she worked as a research engineer in TI-specialist-Tech. of LG CNS, Korea. She is currently an associate professor of the Department of Computer Science in Chungbuk National University. Her research interests are in the field of computer vision and machine learning.