Journal of the Ergonomics Society of Korea
Vol. 31, No. 6 pp.705-713, December 2012 http://dx.doi.org/10.5143/JESK.2012.31.6.705

Modeling the Visual Target Search in Natural Scenes

Daecheol Park[1], Rohae Myung[1], Sang-Hyeob Kim[2], Eun-Hye Jang[2], Byoung-Jun Park[2]

[1]School of Industrial Management Engineering, Korea University, Seoul, 136-713
[2]BT Convergence Technology Research Department, Electronics and Telecommunications Research Institute, Daejeon, 305-700

ABSTRACT

Objective: The aim of this study is to predict human visual target search using ACT-R cognitive architecture in real scene images. **Background:** Human uses both the method of bottom-up and top-down process at the same time using characteristics of image itself and knowledge about images. Modeling of human visual search also needs to include both processes. **Method:** In this study, visual target object search performance in real scene images was analyzed comparing experimental data and result of ACT-R model. 10 students participated in this experiment and the model was simulated ten times. This experiment was conducted in two conditions, indoor images and outdoor images. The ACT-R model considering the first saccade region through calculating the saliency map and spatial layout was established. Proposed model in this study used the guide of visual search and adopted visual search strategies according to the guide. **Results:** In the analysis results, no significant difference on performance time between model prediction and empirical data was found. **Conclusion:** The proposed ACT-R model is able to predict the human visual search process in real scene images using salience map and spatial layout. **Application:** This study is useful in conducting model-based evaluation in visual search, particularly in real images. Also, this study is able to adopt in diverse image processing program such as helper of the visually impaired.

Keywords: ACT-R, Cognitive architecture, Visual search, Visual attention, Real scene image

1. Introduction

컴퓨터의 성능이 급격히 향상됨에 따라 비교적 많은 양의 데이터 처리가 요구되는 영상(Image) 정보의 생성과 유통이 활발하게 이루어지고 있다. 이에 따라 최근에는 컴퓨터가 자동으로 영상 정보를 인식할 수 있게 하는 영상 처리 기술이 많이 연구되고 있다. 특히 제시된 영상 정보 중에 특정한 목표의 존재 여부를 확인하기 위해서는 그 목표의 특성과 부합된 영상 처리 기술이 필요하며, 주로 색상, 질감, 형태의 3가지 특성을 각각 식별하기 위한 개별적인 기술이 많이 연구되었다. 그러나 자연에 존재하는 많은 사물들에는 어떤 한 가지 특성만이 아니라 여러 가지 특성이 복합적으로 내재되어 있기 때문에, 특정한 영상 특성만을 식별할 수 있는 단일 기술만으로는 대상 목표를 탐색해내기 어려운 문제점이 있다.

한편 사람은 일상생활에서 입력되는 수많은 지각 정보들을 동시에 모두 처리할 수 없기 때문에 선택적 주의 과정을 통해 실제 입력되는 시각 정보 중 필요한 정보만을 선별하여 처리하게 된다. 즉, 인간은 실제 입력된 시각 정보 중에서 응시할 영역을 시각적 주의 기능에 의해 선택적으로 결정하며, 도약 안구운동 기능을 통해 응시할 영역으로 신속하게 시선을 이동시켜 정보를 효율적으로 처리한다(Koch & Ullman, 1985; Choi & Lee, 2002). 이러한 사람의 선택적 주의 집중

Corresponding Author: Rohae Myung. Department of Industrial Management Engineering, Korea University, Seoul, 136-713.
E-mail: rmyung@korea.ac.kr
Copyright@2012 by Ergonomics Society of Korea(pISSN:1229-1684 eISSN:2093-8462). All right reserved.
ⓒThis is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. http://www.esk.or.kr

능력은 복잡하고 용량이 큰 정보를 선택적으로 처리할 수 있게 하여, 영상기반 감시시스템이나 로봇의 인공시각 시스템 등 다양한 시각정보처리 시스템에 적용되어 효율적 정보처리 기능을 제공해 줄 수 있다(Lee, 2009). 컴퓨터 모델링을 통해 다양한 특징을 가지는 사물들이 포함된 실제 영상 또는 이미지로부터 특정한 목표를 탐색하기 위해서는 다수의 이미지 패턴 인식 기술과 많은 연산 처리가 요구된다.

본 논문에서는 컬러 이미지가 입력되었을 때 다른 부분과 현저하게 다른 영역을 탐지해내는 Saliency map 모형과 학습을 통해 얻어진 맥락 정보를 이용해 사람의 주의를 이동할 영역을 가이드해주는 모델링 방법론을 연구하고자 하였다. 특히 사람의 시각적 탐색 과정을 모델링 하는데 많이 사용되어 온 ACT-R(Adaptive Control of Thought-Rational) 인지아키텍처를 기반으로 하여 실제 영상 정보로부터 특정한 시각 목표를 탐색할 수 있는 시각 목표 탐색 모델링 방법론을 제안하고자 한다. 이를 위하여 두 종류의 실제 이미지에서 목표 대상을 탐색하는 실험을 실시하였고, 실험데이터를 제안한 모델의 결과와 비교 분석하였다.

2. Related Works

2.1 Visual search

그 동안 시각적 주의에 대한 연구들은 크게 상향식(Bottom-up) 방식과 하향식(Top-down) 방식으로 나뉘어 연구되어 왔다. 상향식 방식은 시각 정보에 분포되어 있는 다양한 기본 특징들을 추출함으로써 얻어진 현저한 단서에 의해 주의가 가해지는 영역을 탐지하며, 하향식 방식은 찾고자 하는 대상 등의 이미 알고 있는 지식과 같은 하향식 단서를 사용한다. 두 방식이 각각 연구되기도 하지만, 이 두 방식 모두 사용한 혼합 방식의 시스템에 대해서도 많은 연구가 이루어졌다(Choi & Lee, 2002).

Treisman과 Gelade(1980)이 제안한 특성통합이론(feature integrated theory)에 따르면, 여러 개의 속성들로 이루어진 대상(object)은 전주의 단계를 거쳐 구분된 "feature map"을 가지게 되며, 주의는 집중 단계를 거쳐 각각의 속성들을 통합해 전체 하나의 대상으로 지각하게 된다. 이 이론은 대부분의 시각적 주의에 관련된 연구들에 영향을 주었다. 최초로 시각적 주의의 계산적 모델인 'Saliency map'이 등장하였으며, 이 모델은 관찰자의 주의를 끄는 지역을 어느 정도 예측할 수 있도록 해준다(Torralba et al., 2006). Itti et al.(1998)은 밝기, 색, 방향으로 이루어진 Saliency map을 구성하여 상향식 방식의 주의 모형을 제안하였다.

Saliency map 모형은 이미지의 각 속성값 계산을 통해 이루어지므로 특정한 과제나 작업에 따라 달라지지 않는 상향식 방식이다. 하지만 실제 이미지에서는 어떤 장면에 의미있는 대상들이 같이 존재하며, 이들 대상들의 관계에 의해서 어디에 주의를 주어야 하는지 알 수 있게 되는 경우가 많다(Biderman et al., 1982). 가령 이미지에서 의자를 찾는다고 하는 경우, 의자는 책상이나 테이블 근처에 있는 경우가 많으므로 두 대상의 관계에 의해 탐색 대상의 위치를 어느 정도 파악할 수 있다. 또한 이미지 전체 속성의 통계적 분포나 의미적 맥락, 3차원의 기하학적 맥락 등을 통해 주의를 줄 곳을 찾을 수도 있으며(Divvala et al., 2009), 실제 이미지의 공간적인 배치(Spatial layout)에 대해서 사전 지식을 가지고 있거나 미리 알려주는 경우에도 사람은 빠르게 목표 대상을 찾을 수 있다(Sanoki & Epstein, 1997).

본 연구에서는 사람이 컬러 이미지 정보를 입력 받았을 때, 처음으로 주의를 이동시키는 영역과 이어서 발생하는 목표 대상 탐색 과정을 Itti의 Saliency map과 학습된 이미지의 공간적인 배치의 조합으로 설명하고자 한다. 사람의 시각 주의는 시각 장면들의 다양한 측면들 중에서 가장 관련성 있는 지역을 선택하게 되므로(Meurm & Callet, 2009), 이 두 방식이 혼합된 형태를 사용해 시각 탐색 과정을 시뮬레이션 하고자 하였다.

2.2 ACT-R

ACT-R 인지아키텍처는 사람의 인지와 수행에 대해 설명하고 예측할 수 있도록 한 컴퓨터 시뮬레이션이라고 할 수 있다. 인지아키텍처는 인간의 인지와 지각 및 운동을 많은 실험을 통해 얻은 파라미터(parameter)와 제약(constraints)을 가지고 설명하며, 이를 통해 사람의 행동이나 수행을 예측할 수 있도록 해준다(Fleetwood & Byrne, 2006). 인지아키텍처에는 ACT-R(Anderson, 1983), SOAR(Laird et al., 1987), EPIC(Kieras & Meyer, 1997) 등이 있으며, 이들은 인간의 행위를 비교적 낮은 수준의 세부행위까지 자세히 묘사할 수 있다. 이 중에서 ACT-R은 인간의 인지 과정을 세부적이고 정량적으로 정확히 예측할 수 있는 것으로 인정받고 있다. 따라서 본 연구에서는 ACT-R을 이용해 입력된 실제 이미지를 처리하는 사람의 시각 탐색 수행을 분석하고 예측하고자 하였다.

ACT-R은 여러 개의 모듈로 이루어져 있으며, 이 모듈들과 두 가지 기억(Memory) 형태인 Declarative Memory과 Procedural Memory의 상호작용을 통해 인간의 인지적 행동을 처리한다. Declarative Memory는 어떤 사실에 대해 기억하고 있는 것과 지각된 것을 정보의 최소 묶음이라고 할 수 있는 청크(chunk) 형태로 저장하고 있다. Procedural

Memory는 감각 정보와 서술적 지식을 IF-THEN의 production rule에 따라 처리하는 역할을 한다.

이 논문에서 중요하게 다루게 될 모듈은 Vision Module로, 이 모듈은 사람의 눈과 같은 역할을 한다. 이 모듈은 사람 두뇌의 시각 피질처럼 정보가 '무엇'인지를 해석하는 what system과 '어디'에 있는지 해석하는 where system으로 구성된다. 먼저, where system에 찾고자 하는 대상이 어디에 위치하는지 알려주게 되며, 이 위치 정보를 받아 what system은 시각 주의를 이동하게 된다. ACT-R에서는 기본적으로 주의를 이동하는데 135ms가 소요된다(Anderson, 1997). 50ms는 규칙이 활성화되는데 걸리는 시간이며, 85ms는 시각 주의를 이동시키고 대상을 처리하는데 걸리는 시간이다.

ACT-R을 이용한 사람의 시각 탐색 연구는 여러 연구자들에 의해 많이 이루어져왔다. Anderson et al.(1995, 1997)에 따르면 사람이 어떤 영역을 주목하면, 그 시각 영역 안에 특정한 속성들이 있다는 것을 알게 되지만 그 영역에 주의를 주어야만 속성들을 통합해 대상을 인식할 수 있다고 설명하며 이를 ACT-R 인지아키텍처로 표현하였다. Salvucci (2000)는 EMMA(Eye Movement & Movement of Attention) 모델을 개발해 사람의 안구 운동과 주의 이동을 통합하여 표현하였다. 최근에는 Fleetwood와 Byrne(2002, 2006)이 ACT-R의 vision module을 이용하여 아이콘 탐색 모델을 제안하였으며, 아이콘의 집합 크기와 질, 복잡도에 의해 시각 탐색 수행이 달라진다는 것을 설명하였다. ACT-R은 이처럼 많은 연구의 결과로 인간의 시각처리와 이에 따른 인지 과정을 잘 표현할 수 있는 장점을 가지므로, 본 논문에서도 ACT-R을 이용해 사람이 입력된 이미지에서 목표 대상 탐색 과정을 설명하고자 한다.

3. Experiment

본 연구에서는 도시의 바깥 전경 이미지에서 사람을 찾는 실험과 실내의 침실 이미지에서 벽에 걸린 그림을 찾는 실험을 실시하였다.

3.1 Participants

본 연구의 실험은 고려대학교에 재학중인 20대의 남, 여 총 10명(남자 6명, 여자 4명)의 대학생과 대학원생을 대상으로 하였다. 이들은 본 연구에 대한 설명을 듣고 자발적으로 연구에 참여하였다.

3.2 Apparatus

본 연구의 실험에서는 Java script로 제작된 프로그램을 사용하여 컬러 이미지를 제시하였으며, 이를 시현하기 위해 24인치 와이드 모니터(Samsung SyncMaster B2430HD)와 PC가 사용되었다. 제작된 프로그램은 이미지가 제시될 때부터 피실험자가 숫자패드를 누를 때까지 걸리는 시간을 1/1,000초 단위로 저장할 수 있도록 하였으며, 피실험자가 누른 숫자를 확인할 수 있도록 만들어졌다. 이를 통해 피실험자가 실제 이미지에서 목표 대상을 찾는데 걸리는 시간과 수행의 에러를 측정하였다.

3.3 Design and procedure

실험은 10명의 피실험자를 대상으로 within-subject design으로 설계되었다. 실험에 사용된 이미지는 크게 실외 이미지와 실내 이미지로 구분되며, 실외 이미지는 도심의 바깥 전경을 사용하였고 실내 이미지는 침실의 전경을 사용하였다. 각 이미지의 크기는 800×600pixel로 모니터 중앙에 제시되었다. 이미지는 그 동안 이미지 인식 연구에 많이 사용되었던 LabelMe(Russell et al., 2008)와 SUN09 (Torralba et al., 2010) 데이터베이스에서 추출하여 사용하였으며, 추가적으로 구글 검색을 통해 이미지를 수집하였다. 본 실험에 사용된 이미지들은 카테고리에 따라 유사한 전경과 공간적 배치를 가지고 있다.

Figure 1. Diagram of the experiment

실험은 연습 단계와 실험 단계로 구분되는데, 연습 단계에서는 바깥 전경과 실내 전경 이미지에서 각각 120장의 이미

지를 제시하였으며, 실제 실험에서는 각각 20장의 이미지를 사용해 실험을 진행하였다. 피실험자로 하여금 외부 이미지에서는 사람을, 실내 이미지에서는 액자를 찾도록 과제를 설정하였다. 입력되는 이미지는 Figure 1과 같이 모니터 화면 가운데에 나타나게 되며, 피실험자는 목표 대상 탐색을 모두 마쳤을 경우 최종 목표 대상의 숫자를 키보드의 숫자패드를 통해 누르게 된다. 키보드를 누르게 되면 1초 후에 다음 이미지가 나타나게 되며, 제시되는 이미지는 임의의 순서로 설정되었다.

4. Model Development

본 연구에서는 피실험자를 통한 실험과 동일한 조건으로 사람의 수행을 예측하는 모델을 수립하였고, 이를 실험 결과와 분석하였다.

4.1 Guide of regions for visual search

먼저, 이미지가 입력되었을 때 처음으로 주의를 이동시킬 가능성이 높은 영역을 Figure 2처럼 Itti의 Saliency map 모형과 학습을 통한 공간적 배치를 통해 추출하였다. Walther (2006)는 Itti의 이론을 이용하여 이미지에서 Saliency map을 추출할 수 있는 Matlab Saliency toolbox를 제공하였으며, 본 연구에서는 이를 이용하여 입력된 이미지에서 응시할 영역을 선정하였다. 다른 영역에 비해 현저한 차이를 보이는 영역을 선정하기 위해서 밝기(intensity), 색(color), 방향(orientation)의 feature map을 구성하였고, 각 속성은 동일한 가중치를 가지도록 설정하였다. 또한 실험을 진행하기 전에 연습하였던 120장의 이미지들에서 장면과 대상들의 공간적인 배치를 추출해 목표 대상이 있을 것으로 예상되는 전체 이미지의 20% 영역을 제시하였다. 사람은 시각 탐색을 할 때 위아래로 주의를 옮기기 보다는 좌우로 탐색을 실시하기 때문에 수평적 영역을 선택하였다. 실제로 이미지를 분석한 결과 바깥 이미지는 도로 위를 걷는 보행자를 찾는 탐색 과정이기 때문에 도로 위, 즉 이미지의 가운데 영역을 주로 찾게 됨을 알 수 있었으며, 실내 이미지에서 그림은 주로 벽에 걸려 있기 때문에 이미지의 위쪽 영역을 찾게 됨을 알 수 있었다.

4.2 Model Hypothesis for visual strategy

본 연구의 ACT-R 모델에서는 시각 탐색 전략을 하나의 대상을 찾아 주의를 옮기는 Exhaustive Search(Nilsen, 1991)와는 다르게 대상을 1개에서 2개까지 보도록 정의하였다. 사람 눈의 중심와(fovea)의 각도는 2°이며 모니터와 피실험자 사이의 거리가 약 55~60cm 정도 떨어져 있었으므로, 계산했을 때 한 눈에 볼 수 있는 길이가 약 2cm이므로 이 범위에 속한 대상이 있을 경우에는 한 번만 주의를 이동시키도록 전략을 수립하였다.

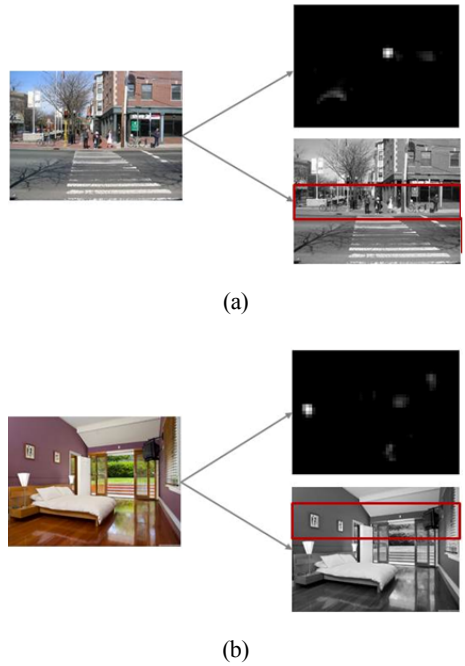

Figure 2. Saliency map and spatial layout

또한 모델은 이미지가 주어졌을 때 Saliency map 모형을 통해 현저한 특징을 가지는 지역이 어디인지 빠른 시간 내에 알게 되며, 학습으로 인해 대상이 위치할만한 공간적 배치 정보를 사전에 알고 있다고 가정하였다. 그러므로, 모델이 처음 시작되었을 때 이 두 영역 중에서 모델이 임의로 한 영역을 선택하게 되며, 어떤 영역을 선택하느냐에 따라 수행에 걸리는 시간이 달라지게 된다.

4.3 Procedure

본 연구의 ACT-R 모델은 실제 피실험자가 입력된 이미지에서 목표 대상에 시각적 주의를 기울이는 전략을 사용하였다. 먼저 ACT-R 모델의 Declarative memory에 목표 대상에 대한 정보를 청크(Chunk) 형태로 저장해 두었고, 이를 바탕으로 ACT-R 모델은 Production rule에 따라 시각적 영역을 옮겨가면서 주의를 기울이게 된다. Declarative memory에 대상 목표의 개수를 셀 수 있도록 설정하였으며, 초기값은 0으로 설정하였다. 모델은 이미지가 주어졌을 때

Saliency map 모형으로부터 얻어진 영역과 공간적 배치 영역 사이 중 한 곳에서 위치를 찾고, 주의를 옮기게 된다. 주의를 기울인 영역의 대상이 Declarative memory에 저장되어 있는 목표 대상의 속성과 일치할 경우, 모델은 대상 목표의 개수를 하나씩 추가하여 Declarative memory에 저장한다. 모델은 대상 목표의 개수를 저장하는 동시에 다음 대상에 대한 기억을 인출하는 것을 병렬적으로 수행하게 된다. 최종적으로 목표 대상을 모두 탐색하였을 경우에는 Declarative memory에 저장된 숫자에 맞는 키를 누르게 된다. 모델을 구성하는 기본적인 Production rule의 흐름은 Figure 3과 같다.

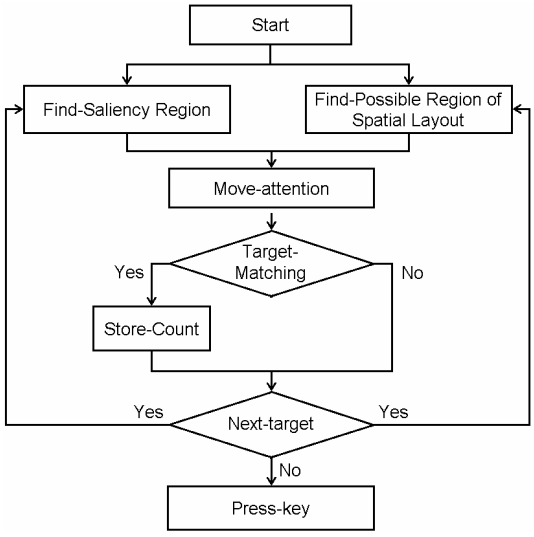

Figure 3. Process of ACT-R model

5. Results

실제 이미지에서 10명의 피실험자를 통한 시각 탐색 수행과 ACT-R 모델을 10번 simulation하여 얻은 탐색 수행의 평균시간과 표준오차는 Table 1과 Table 2와 같다. Table 1은 바깥 이미지에서 사람을 찾는 수행을 한 결과이며, Table 2는 실내 이미지에서 그림을 찾는 수행을 한 결과이다.

결과를 보면 모델과 실험 데이터의 평균이 비교적 일치하며, 표준오차도 유사하게 나타남을 알 수 있다. 이를 정량적이고 구체적으로 알아보기 위해 각각의 이미지마다 t-test를 실시하였으며, 일부를 제외하고는 유의수준 0.05보다 높은 값을 가지는 것을 확인할 수 있었다. 또한 인지 모델과 피실험자의 결과의 적합도를 알아보기 위해 회귀분석을 통

Table 1. Comparison of performance time in outdoor image

Image No.	Model Avg. (msec)	Model STD	Data Avg. (msec)	Data STD
1	1,564	89.0	1,525	110.5
2*	1,023	31.3	1,225	61.6
3	1,493	31.3	1,477	97.8
4*	953	39.2	1,086	23.8
5	1,141	50.2	1,136	76.7
6	1,595	96.7	1,632	213.3
7	1,399	38.4	1,330	58.9
8*	1,376	70.5	1,133	58.0
9	1,399	52.0	1,337	50.0
10	1,282	89.0	1,288	30.1
11	1,423	39.2	1,407	52.0
12	1,423	72.2	1,304	53.9
13	1,634	94.0	1,596	82.8
14	1,916	52.0	1,973	190.9
15	2,198	31.3	2,234	142.4
16	1,963	31.3	2,115	137.8
17*	1,864	50.2	1,655	70.8
18	2,571	36.2	2,752	295.8
19	2,357	52.0	2,254	272.6
20	2,146	39.2	2,186	197.5

*p-value < 0.05

Table 2. Comparison of performance time in indoor image

Image No.	Model Avg. (msec)	Model STD	Data Avg. (msec)	Data STD
1	1,117	31.3	1,077	102.6
2*	1,235	78.7	1,013	17.1
3	1,070	78.3	1,010	40.6
4*	1,211	87.2	1,017	27.4
5	1,235	61.2	1,186	114.3
6	1,376	61.2	1,378	121.6
7*	878	23.4	8,04	38.1
8	951	26.0	911	40.0
9	1,282	89.0	1,225	75.7
10*	934	52.0	820	39.4
11	1,211	52.0	1,121	66.1
12	992	43.1	965	66.8
13	953	39.2	939	41.5
14	1,015	53.1	936	32.9
15	1,282	65.1	1,226	67.8

Table 2. Comparison of performance time in indoor image (Continued)

Image No.	Model Avg. (msec)	Model STD	Data Avg. (msec)	Data STD
16	1,211	62.7	1,127	73.6
17	1,587	31.3	1,541	121.0
18	1,393	36.1	1,318	98.1
19	1,294	56.1	1,232	53.1
20	2,527	76.8	2,471	118.6

* p-value < 0.05

해 R^2를 계산하였고, 정확한 대응의 정도를 산정하기 위해 RMSE(Root Mean Square Error)를 계산하였다. 회귀분석의 결과 바깥 이미지의 경우 R^2 값은 0.949(RMSE=99.5)이고, 실내 이미지의 경우 0.979(RMSE=54.1)로 모델의 예측 값과 피실험자의 데이터 간에 높은 상관관계가 있음이 나타났다(Figure 4, Figure 5). 이를 통해 ACT-R 모델이 전체 수행 시간과 각 과제별 수행 시간에서 피실험자의 행동 양식을 정확히 묘사하고 있음을 확인할 수 있었다.

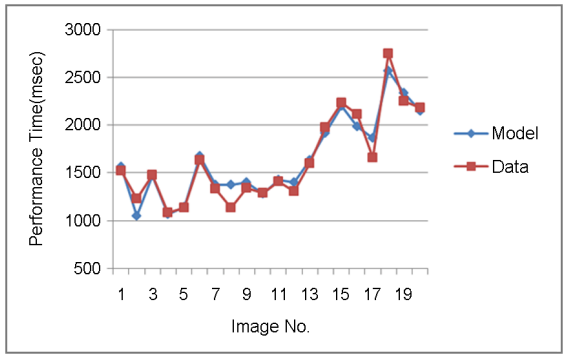

Figure 4. Graph of the results of outdoor image

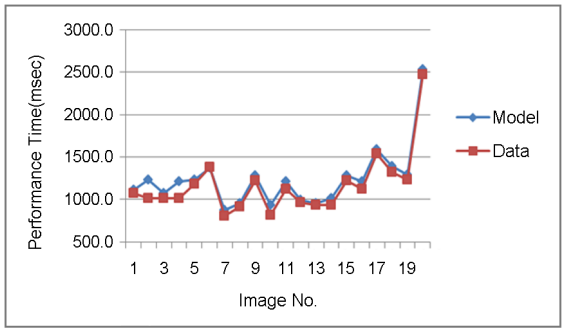

Figure 5. Graph of the results of indoor image

추가적으로 목표 대상의 개수에 따라 평균을 구한 결과, 목표 대상이 없는 경우에는 목표 대상이 1~2개 존재할 때보다 수행 시간이 오래 걸리는 것을 알 수 있었다(Figure 6). 이는 목표 대상이 없는 경우 대상을 재탐색하게 되므로 주의를 반복해서 주어야 하기 때문이다. 또한 실내 이미지의 경우에는 벽에 위치한 그림(액자)들이 가까운 위치에 모여 있는 경우가 많아 한꺼번에 처리를 하는 경우가 많으므로 목표 대상이 1개부터 3개 사이에는 수행 시간이 유사하게 나타남을 알 수 있다(Figure 7). 또한 Error의 경우에는 목표 대상의 크기가 아주 작은 경우에만 발생하였다.

Figure 6. Performance time of outdoor image

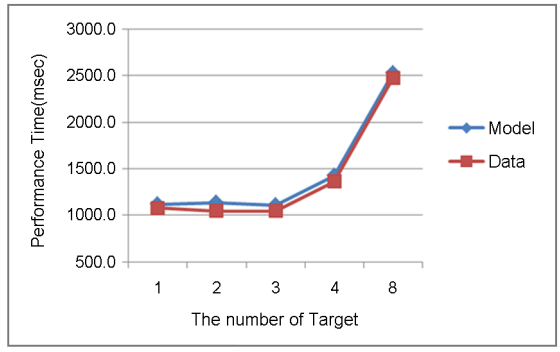

Figure 7. Performance time of indoor image

6. Discussion

본 연구에서는 실제 이미지에서 목표 대상을 탐색하는 사용자의 시각 탐색 과정을 묘사할 수 있는 ACT-R 모델을 수립하였다. ACT-R 모델(Allegro Lisp 8.2)과 실험(Java script)에서 수행된 프로그램은 다르지만 동일한 조

건으로 맞추어 모델링을 진행하였으므로 결과에 다른 영향이 없도록 하였다. 실험에서 키보드의 숫자패드를 누르도록 하였으므로, ACT-R 모델 또한 Motor module의 'press-key' 움직임을 통해 키보드를 누르는 행위를 표현하였다. 또한 피실험자와 동일하게 오른쪽 검지 손가락이 키보드 숫자패드 '1'에 위치하도록 초기 설정을 하였다. 모델과 실험 데이터의 비교 결과, 바깥 이미지와 실내 이미지의 두 조건에서 ACT-R 모델이 피실험자를 바탕으로 한 실험 데이터와 유사한 결과가 도출되었다. 이는 곧 ACT-R 모델이 실제 이미지에서 사람의 시각 탐색 과정을 잘 표현했다고 말할 수 있다.

위의 결과를 보면 탐색 목표 대상의 개수가 같더라도 실내 이미지의 수행이 바깥 이미지의 수행보다 빠름을 알 수 있다. 여기에는 세 가지 이유가 있는 것으로 보인다. 먼저, 실내 이미지의 경우 상대적으로 찾기 쉬운 대상이 목표이기 때문이다. 침실에서 그림 액자는 보통 벽에 걸려있는 경우가 많기 때문에 이미지에서 위쪽 영역만 살펴보면 된다. 두 번째 이유는 그림 액자의 경우 비슷한 영역 가까운 위치에 모여 있는 경우가 많기 때문에 몇 개의 대상이 한꺼번에 사람의 중심와(fovea)에 들어와 한번에 처리할 수 있다. 세 번째 이유는 바깥 이미지에 비해 상대적으로 깊이(Depth)가 얕기 때문에 목표의 크기 차이가 많이 나지 않으며, 깊이가 주어질 경우 탐색하는데 더 오랜 시간이 걸리게 된다.

본 연구의 실험 결과에서는 실험데이터와 모델 평균이 통계적으로 차이가 있는지 검증하고자 실험데이터와 모델의 평균값을 이용해 t-test를 실시하였다. 그 결과 실외 이미지에서는 2, 4, 8, 17번, 실내 이미지에서는 2, 4, 7, 10번을 제외하고는 유의한 차이를 보이지 않았다. 실외 이미지 2, 4번의 경우에는 실험데이터가 모델 평균보다 수행 시간이 길게 나타나는데, 이 두 이미지는 목표의 크기가 다른 이미지들과 비교해 작고 상대적으로 깊이가 깊은 곳에 위치한 목표를 포함하고 있는 이미지 형태이다. 그렇기 때문에 피실험자가 목표를 찾을 때 깊이 있고 크기가 작은 목표가 더 있는지 탐색하기 위해서 1~2차례 시선을 더 움직이는 형태를 보이기 때문으로 추정할 수 있다(Figure 8의 왼쪽). 실외 이미지의 8, 17번의 경우에는 실험데이터가 모델 평균보다 수행 시간이 짧게 나타나는데, 이 두 이미지는 상대적으로 목표가 크고 앞쪽에 위치하고 있으며, 또한 목표들이 가까운 위치에 몰려 있는 경향이 있어서 주의를 한번만 주고도 목표 탐색이 가능했던 것으로 추정할 수 있다(Figure 8의 오른쪽).

실내 이미지의 경우에는 모델의 평균이 실험데이터보다 수행 시간이 오래 걸리는 경향을 보이고 있다. 이는 실내 이미지 2, 4, 7, 10번의 경우에서는 피실험자가 중심와(fovea) 뿐만 아니라 부중심와(parafovea) 영역까지 활용하기 때문으로 예상된다. 특히 이들 이미지 중에서 4번은 p-value가

Figure 8. Examples of outdoor images(#4, #17)

0.043, 7번은 0.049로 유의수준과 유사하게 나타나므로, 아주 조금의 차이가 있는 것으로 보인다.

Figure 9을 보면 왼쪽 그림에서는 왼쪽 영역에 액자 3개가 근접한 위치에 몰려 있는 것을 알 수 있으며, 오른쪽 그림의 경우에도 한쪽에 몰려있는 것을 알 수 있다. 이런 경우에 모델에서는 목표가 한 곳에 몰려있지만 중심와를 벗어나므로 주의를 한번 더 이동하는 것으로 설명하였지만, 실험데이터에서는 중심와와 부중심와를 이용해 한번에 파악할 수 있었던 것으로 파악할 수 있을 것이다.

Figure 9. Examples of indoor images(#2, #10)

두 결과를 놓고 살펴보았을 때, 깊이가 깊을 경우에는 주의를 더 많이 옮기는 경향을 보이며, 상대적으로 평면적인 구조에서 타겟의 크기가 클 경우에는 더 빠르게 탐색하는 경향을 보이는 것으로 나타남을 알 수 있다. 이러한 모델 평균과 실험데이터 평균의 차이는 추후 모델의 수정을 통해 좀 더 예측력을 높일 수 있을 것으로 예상된다.

본 연구에서는 이미지가 주어졌을 때 다른 영역에 비해 현저한 차이를 보이는 지역을 Saliency map을 이용해 구분하였고, 여기에 학습으로 얻어진 공간적인 배치에 대한 지식을 조합해 주의를 옮길 지역을 안내해 주는 모델을 제시하였다. 이를 통해 좀 더 객관적으로 사람이 주의를 주는 곳을 찾고자 시도하였다. 이제까지 컴퓨터과학 분야나 영상처리 분야에서는 visual attention을 정량적으로 모델링하기 위해 계층적 모델, 통계적 모델(Walther & Koch, 2007), 베이지안 모델(Elazary & Itti, 2010) 등 다양한 방법을 제시하고 있다. 본 논문에서 사용한 방법 이외에도 위에 제시한 다양

한 모델들을 이용해 ACT-R 시각 탐색 모델을 수립하는 것도 좋은 시도라고 여겨진다.

본 논문에서 제안한 ACT-R 모델이 상당히 정확한 예측을 보여주고 있지만, 실제로 사람이 논문에서 제안한 과정대로 시각 탐색을 진행하는지는 Eye-tracking을 통해서 정확히 살펴볼 필요가 있으며, 그 결과에 따라 추가적인 연구가 필요할 것으로 예상된다.

7. Conclusion

본 연구에서는 ACT-R 인지아키텍처를 이용해 사람이 실제 이미지에서 목표 대상을 찾는 시각 탐색 과정을 묘사하였다. 본 연구에서 제안한 ACT-R 모델은 이미지 자체가 가지고 있는 속성들을 이용한 Salience와 사용자가 학습을 통해 이미지의 공간적 배치에 대해 가지고 있는 지식을 조합하여 수립되었다. 이번 연구에서는 두 가지의 카테고리로 나눌 수 있는 실외 이미지와 실내 이미지를 사용하였으며, 연구 결과 두 이미지 모두에서 사람의 시각 탐색 과정을 적절히 표현하는 것으로 나타났다. 연구 결과 목표 대상이 없는 경우에는 소수의 목표 대상이 존재할 때보다 수행 시간이 오래 걸림을 알 수 있었고, 가까운 대상들이 모여있는 경우 여러 개가 동시에 처리될 수 있다는 사실도 알 수 있었다. 본 연구에서 제안한 ACT-R 모델은 사람의 시각 탐색 과정과 인지 과정을 묘사하고 있으므로, 실제 사람의 인지구조를 반영한 영상처리 등 다양한 분야에 활용될 수 있을 것으로 예상된다.

Acknowledgements

This research was supported by the Converging Research Center Program through the Converging Research Headquarter for Human, Cognition and Environment funded by the Ministry of Education, Science and Technology (No. 2012K001333).

References

Anderson, J. R., The Architecture of Cognition. Cambridge, *MA: Harvard University Press*, 1983.

Anderson, J. R., Matessa, M. and Douglass, S., "The ACT-R theory and visual attention.", *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, (pp. 61-65), Hillsdale, NJ: Lawrence Erlbaum, 1995.

Anderson, J. R., Matessa, M. and Lebiere, C., ACT-R: A theory of higher level cognition and its relation to visual attention, *Human-Computer Interaction*, 12(4), 439-462, 1997.

Biderman, I., Meaanotte, R. J. and Rabinowitz, J. C., Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143-177, 1982.

Choi, K. J. and Lee, Y. B., A Saliency Map Model for Color Images using Statistical Information and Local Competitive Relations of Extracted Features, *Korean Brain Society*, 2(1), 69-78, 2002.

Divvala, S. K., Hoiem, D., Hays, J. H., Efros, A. A. and Hebert, M., An empirical study of context in object detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp.1271-1278), 2009.

Elazary, L. and Itti, L., A Bayesian model for efficient visual search and recognition, *Vision Research*, 50(14), 1338-1352, 2010.

Fleetwood, M. D. and Byrne, M. D., Modeling icon search in ACT-R/PM, *Cognitive Systems Research*, 3(1), 25-33, 2002.

Fleetwood, M. D. and Byrne, M. D., Modeling the visual search of displays: a revised ACT-R model of icon search based on eye-tracking data, *Human-Computer Interaction*, 21(2), 153-197, 2006.

http://www.saliencytoolbox.net/index.html (retrieved March 2, 2012)

http://act-r.psy.cmu.edu (retrieved January 10, 2011)

Itti, L., Koch, C. and Niebur, E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259, 1998.

Kieras, D. E. and Meyer, D. E., An overview of the EPIC architecture for cognition and performance with application to human-computer interaction, *Human-Computer Interaction*, 12, 391-438, 1997.

Koch, C. and Ullman, S., Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, *Human Neurobiology*, 4, 219-227, 1985.

Laird, J. E., Newell, A. and Rosenbloom, P. S., SOAR: An architecture for general intelligence, Artificial Intelligence, 33, 1-64, 1987.

Lee, M. H., Artificial vision system of selective attention, *Journal of the Korean Institute of Electronics Engineers*, 36(11), 52-65, 2009.

Meur, O. L. and Callet, P. L., What we see is most likely to be what matters: visual attention and application, *ICIP2009*, pp.3085-3088.

Nilsen, E. L., Perceptual-motor control in human-computer interaction, *Technical Report, No. 37, Ann Arbor, University of Michigan, Cognitive Science and Machine Intelligence Laboratory*, 1991.

Russell, B. C., Torralba, A., Murphy, K. P. and Freeman, W. T., LabelMe: a database and web-based tool for image annotation, *International Journal of computer vision*, 77(1), 157-173, 2008.

Salvucci, D. D., A model of eye movements and visual attention, *Proceedings of the International Conference on Cognitive Modeling*, (pp. 252-259). Veenendaal, TheNetherlands: Universal Press, 2000.

Sanocki, T. and Epstein, W., Priming Spatial Layout of Scenes, Psychological Science, 8(5), 374-378, 1997.

Torralba, A., Olivia, A., Castelhano, M. S. and Henderson, J. M., Contextual guidance of eye movements and attention in real-world scenes: the

role of global features in object search, Psychological Review, 113(4), 766-786, 2006.

Torralba, A., Choi, M. J., Lim, J. J. and Willsky, A. S., A Tree-Based Context Model for Object Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2012.

Treisman, A. M. and Gelade, G. A., A Feature-integration Theory of Attention, *Cognitive Psychology*, 12(1), 97-136, 1980.

Walther, D., Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics. *PhD thesis*, California Institute of Technology, Pasadena, CA, 23th February 2006.

Walther, D. B. and Koch, C., Attention in Hierarchical Models of Object Recognition, *Progress in Brain Research*, 165, 57-78, 2007.

Silfverberg, M., MacKenzie, I. S. & Korhonen, P., Predicting Text Entry Speed on Mobile Phones. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 9-16, The Hague, The Netherlands, 2000.

St. Amant, R., Horton, T. E. & Ritter, F. E., Model-Based Evaluation of Cell Phone Menu Interaction. *In Proceedings of the Conference on Human Factors in Computing Systems*, 343-350, 2004.

St. Amant, R., Horton, T. E. & Ritter, F. E., Model-Based Evaluation of Expert Cell Phone Menu Interaction. *ACM Transactions on Computer-Human Interaction*, Vol. 14, No. 1, Article 1, 2007.

TextwareSolution., The FITALY one-finger keyboards, http://fitaly.com/fitaly, 1998.

Zhai, S., Hunter, M., & Smith, B. A., The Metropolis Keyboard - An Exploration of Quantitative Techniques for Virtual Keyboard Design, *In Proceedings of the UIST 2000*, CHI Letters 2(2), ACM Press, 119-128, 2000.

Author listings

Daecheol Park: zodiac17@korea.ac.kr

Highest degree: BE in School of Media, Soongsil University
Areas of interest: HCI, Cognitive Modeling

Rohae Myung: rmyung@korea.ac.kr

Highest degree: Ph.D., Industrial Engineering, Texas Tech University
Position title: Professor in School of Industrial Management Engineering, Korea University
Areas of interest: HCI, Cognitive Modeling

Sang-Hyeob Kim: shk1028@etri.re.kr

Highest degree: Ph.D., Department of Apply Physics, Tohoku University
Position title: Principle Member of Engineering Staff, BT Convergence Technology Research Department, Electronics and Telecommunications Research Institute
Areas of interest: Cognition Convergence, Emotion Recognition

Eun-Hye Jang: cleta4u@ etri.re.kr

Highest degree: Ph.D., Department of Psychology, Chungnam National University
Position title: Researcher, BT Convergence Technology Research Department, Electronics and Telecommunications Research Institute
Areas of interest: Cognition Convergence, Emotion Recognition

Byoung-Jun Park: bj_park@etri.re.kr

Highest degree: Ph.D., Department of Electrical Engineering, Wonkwang University
Position title: Senior Member of Engineering Staff, BT Convergence Technology Research Department, Electronics and Telecommunications Research Institute
Areas of interest: Computational Intelligence, Pattern Recognition

Date Received : 2012-05-11

Date Revised : 2012-10-04

Date Accepted : 2012-10-22