

실제 이미지에서 현저성과 맥락 정보의 영향을 고려한 시각 탐색 모델

최윤형 · 오형석 · 명노해[†]

고려대학교 산업경영공학과

Visual Search Model based on Saliency and Scene-Context in Real-World Images

Yoonhyung Choi · Hyungseok Oh · Rohae Myung

Department of Industrial Management Engineering, Korea University

According to much research on cognitive science, the impact of the scene-context on human visual search in real-world images could be as important as the saliency. Therefore, this study proposed a method of Adaptive Control of Thought-Rational (ACT-R) modeling of visual search in real-world images, based on saliency and scene-context. The modeling method was developed by using the utility system of ACT-R to describe influences of saliency and scene-context in real-world images. Then, the validation of the model was performed, by comparing the data of the model and eye-tracking data from experiments in simple task in which subjects search some targets in indoor bedroom images. Results show that model data was quite well fit with eye-tracking data. In conclusion, the method of modeling human visual search proposed in this study should be used, in order to provide an accurate model of human performance in visual search tasks in real-world images.

Keywords: Visual search model, Saliency, Scene-context, Real-world images, ACT-R

1. 서론

인간공학 및 Human-Computer Interaction(HCI) 분야에서는 사용자와 컴퓨터간의 활동에서 사람의 시각 탐색이 어떻게 진행되는지에 대한 처리 방식에 관심이 많기 때문에 사람의 시각 정보처리 과정에 대한 여러 연구(Halverson and Hornof, 2011; Kujala and Saariluoma, 2011; Ling and Van, 2004; Tuch *et al.*, 2009; Jung, 2015)가 진행되고 있다.

사람의 시각 탐색은 크게 두 가지 시각 정보 처리 방식이 존재한다. 첫 번째 처리 방식은 상향 처리(bottom-up)방식으로써 목표 대상을 찾기 위해 각각의 객체들에게 시각 주의를 이동시키며 탐색하는 방법이다. 두 번째 처리 방식은 하향 처리

(top-down)방식으로써 목표 대상의 위치를 예상하고, 예상한 영역에서 객체를 탐색하는 방법이다. 그리고 상향 처리 방식은 주로 현저성에 영향을, 하향 처리 방식은 맥락 정보에 영향을 받으며, 인지과학과 인지심리학분야에서는 이미지의 특성에 따라 각각의 영향이 고려되어야 함을 주장하였다(De Graef *et al.*, 1990; Henderson, 2003; Koch and Ullman, 1985; Niebur, 2007).

하지만 눈의 움직임과 주의에 대한 Bayesian 모델을 제안한 Torralba *et al.*(2006) 연구에 따르면, 실제 이미지는 다양하고 복잡한 이미지들로 이루어져 있기 때문에 현저성과 맥락 정보의 영향이 동등하게 고려되어야 함을 주장하였고, 이는 계산 모형 개발을 통해 입증하였다.

또 다른 시각 탐색 연구에서는 사람이 언제, 어떤 시각 정보를

제10회 석사논문경진대회 수상논문.

이 논문은 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2012R1A2A2A03047330).

[†] 연락저자 : 명노해 교수, 136-713 서울특별시 성북구 안암동 5가 고려대학교 산업경영공학과, Tel : 02-3290-3392, Fax : 02-929-5888,

E-mail : rmyung@korea.ac.kr

2015년 2월 9일 접수; 2015년 5월 11일 게재 확정.

지각하고 그 정보를 어떻게 처리하는지에 대한 깊은 이해를 얻고, 시각 탐색 작업의 수행도도 예측하기 위하여 인지모델을 이용한 많은 연구(Anderson *et al.*, 1997; Hornof, 2004; Halverson and Hornof, 2012; John and Kieras, 1996; Plumlee and Ware, 2006; Ritter *et al.*, 2006; Oh *et al.*, 2011; Oh *et al.*, 2014; Oh and Myung, 2012)가 진행되고 있다. 실제로, 많은 연구들에서는 UCIE(Understanding Cognitive Information Engineering; Lohse, 1993), VAM(Visual Attention Model; Schneider, 1995), EPIC(Executive Process-Interactive Control; Hornof, 2004), 그리고 VisMap(Visual Manipulation; St.Amant and O.Riedl, 2001)과 같은 여러 인지모델링을 사용하였고, 시각 탐색에 대한 수행 시간 등의 정량적 값을 예측할 수 있었다. 그 중, ACT-R(Adaptive Control of Thought-Rational; Anderson, 2007)인지 아키텍처는 여러 인지 아키텍처 중에서도 인간의 인지과정을 하위수준에서 가장 세부적이고 정량적으로 묘사할 수 있는 아키텍처로 인정받고 있다. 그리고 많은 연구들은 ACT-R/PM(ACT-R/Perceptual and Motor; Byrne, 2001), EMMA(Eye Movements and the Movement of Attention; Salvucci, 2001), 그리고 PAAV module(Pre-Attentive and Attentive Vision module; Nyamsuren and Taatgen, 2013)과 같이 여러 시각 모듈의 개발을 통해 컴퓨터 인터페이스에서의 메뉴나 아이콘과 같은 가상 이미지에서 사람의 시각 탐색을 좀 더 세부적으로 묘사하기 위한 연구가 진행되었다. 그렇기 때문에, 본 연구에서는 실제 이미지의 ACT-R 시각 탐색 인지모델을 개발하여 사람이 목표 대상을 찾고자 할 때 어떤 인지과정으로 시각 정보를 처리하는지에 대한 세부적인 차원에서의 설명이 가능하고, 더 정확한 사람의 시각 탐색 모델을 만들기 위한 한 단계 높은 수준의 방법론을 제시할 수 있을 것이다.

따라서 본 연구의 목적은 Torralba *et al.*(2006)이 주장하는 실제 이미지에서의 시각 탐색 패턴인 현저성과 맥락 정보를 모두 고려하여 사람의 시각 탐색에 대한 ACT-R 인지모델을 개발하는 것이다.

본 연구에서는 맥락 정보의 영향과 현저성의 영향을 설명하기 전에 먼저 시각 탐색 모델의 이해를 돕기 위해 ACT-R 인지 아키텍처와 인지모델이 대한 프레임워크에 대해 설명한 뒤, 현저성과 맥락 정보에 의한 시각 탐색 과정에 대해 설명하고자 한다.

1.1 ACT-R 인지 아키텍처

ACT-R은 목적에 대한 행동을 저장하는 목적 모듈(goal module), 지식을 저장하는 선언적 모듈(declarative module), 시각과 청각에 대한 자극을 수용하고 처리하는 시각 모듈(visual module), 청각 모듈(aural module), 행동과 관련된 운동 모듈(motor module) 그리고 음성 모듈(vocal module)이 있으며, IF-THEN의 형태로 사람의 인지적 행동의 규칙을 표현한 절차적인 지식(procedural memory)을 통해 순차적으로 모듈들을 처리하고 통합하는 생

산 시스템(production system)으로 이루어져 있다.

본 연구는 실제 이미지를 대상으로 사람의 시각 탐색 모델을 수립하는 것으로, 생산 시스템, 선언적 모듈 그리고 무엇을 보는지 결정하는데 사용되는 시각 모듈(Byrne, 2001)이 사용된다. 하지만 ACT-R 시각 모듈은 버튼 및 텍스트와 같은 시각 객체를 자동적으로 인식할 수 없기 때문에, 연구자가 직접 객체들의 특징들을 선언적 지식의 처리 단위인 군집(chunk)으로 표현해야 한다. 따라서 다음 장에서는 실제 이미지에서 객체의 특징을 추출하는 방법에 대해 먼저 설명하고, 이어서 이번 연구에서 개발하고자 하는 시각 탐색 모델의 프레임워크를 설명한 뒤 시각 탐색에 있어서 동등하게 고려한 두 가지 영향(현저성과 맥락 정보)에 대해 설명하고자 한다.

2. 실제 이미지에서 사람의 시각 탐색 모델

2.1 실제 이미지에서 객체의 특징 추출

현재 ACT-R 모델에서 처리 될 수 있는 특징들은 종류(텍스트, 줄, 그리고 버튼), 색, 크기, 높이, 너비, 값, 그리고 위치와 같이 매우 간단하다. 그러나 실제 이미지에서 객체의 특징들은 모델에서 표현할 수 있는 요소보다 복잡하고 다양하다. 하지만 실제 사람은 모든 특징들의 조합을 지각함으로써 객체를 구별하는 것이 아닌, 특정 특징을 지각함으로써 객체를 구별할 수 있다(Treisman, 2006).

따라서 실제 이미지에서 사람의 시각 탐색 모델을 개발할 때에는 이미지안의 객체들을 구별할 수 있는 정도로 충분히 세부적인 특징들로 나누어져야 한다. 예를 들어, <Figure 1> 안의 의자를 다른 객체들로부터 구별하기 위한 특징들로는 안장, 다리, 그리고 의자의 등받이를 들 수 있다.



Figure 1. The example of real-world image

이 외에도 <Figure 1>에서 관측할 수 있는 다른 객체들을 <Table 1>처럼 여러 특징들로 나눌 수 있다.

Table 1. Distinctive features of the categorized objects

Categorized Object	Distinctive Features
Painting(frame)	Frame, painting, picture
Bed	Mattress, legs, pillow, blanket, fabric
Chair	Legs, wood, back of chair
Chest of drawers	Drawer, knob, handle

2.2 모델 프레임워크

앞에서 언급했던 대로, 실제 이미지에서 사람의 시각 탐색은 지식 기반의 맥락 정보, 또는 현저성-기반의 구체적인 영역에 관한 탐색이 수행된다. 따라서 실제 이미지에서 사람의 시각 탐색을 모델링 할 때는 맥락 정보와 현저성의 영향이 모두 고려되어야 한다. 그러므로 현저성, 맥락 정보 그리고 임의선택(randomness)을 기반으로 한 서로 다른 세 가지의 탐색 전략이 있어야 한다. 임의선택 기반의 전략이 고려된 이유는 맥락-기반(context-based) 영역과 현저성-기반(saliency-based) 영역 외의 존재하는 객체들을 탐색하는 전략을 고려하기 위함이다. 따라서 본 연구에서 제안하는 모델에는 현저성-기반 전략, 맥락-기반 전략, 그리고 임의선택 기반 전략이 있으며, 시각 탐색 모델의 흐름은 <Figure 2>와 같이 세 가지의 전략들 중 한 가지의 전략이 선택됨으로써 수행 될 것이다. 위 전략들은 ACT-R에서의 각 생산규칙(production rule)에 해당되고 변수 유용성(utility) 값이 바뀌면서 한 가지의 생산규칙이 선택될 것이다. ACT-R 인지 아키텍처의 생산규칙은 유용성으로부터 선택될 확률이 정해지며, 높은 유용성을 가진 생산규칙은 선택될 확률이 높아 빈번하게 실행된다(Anderson et al., 2004; Anderson, 2007). 생산규칙의 유용성 값과 선택될 확률은 아래의 두 식으로 결정된다.

$$U_i(n) = U_i(n-1) + \alpha [R_i(n) - U_i(n-1)] \quad (1)$$

$$probability(i) = \frac{e^{U_i/\sqrt{2}s}}{\sum_j e^{U_j/\sqrt{2}s}} \quad (2)$$

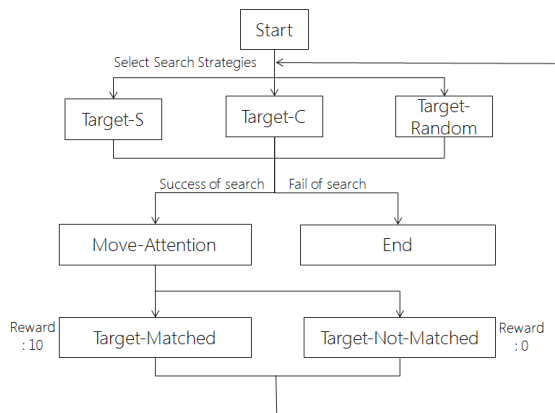


Figure 2. Framework of the model

먼저 생산규칙 i 의 유용성 값은 (1)에 의거하여 결정되며, $U_i(n)$ 와 $U_i(n-1)$ 는 생산규칙 i 의 n 번째와 $n-1$ 번째의 유용성 값, $R_i(n)$ 은 생산규칙 i 가 실행되었을 때 보상(reward) 값, 그리고 α 는 학습 비율(learning rate)로써 기본 값은 0.2이다. 그리고 식 (2)에 의거하여 생산규칙 i 의 선택될 확률은 생산규칙 i 의 유용성 값을 모든 생산규칙 유용성 값의 합으로 나누어 계산된다. 여기서 s 는 유용성의 노이즈(noise) 파라미터(parameter)이다.

탐색 전략이 선택된 후에는 주의(attention)를 주지 않았던 객체가 있는지 확인하는 과정이 진행된다. 만약 주의를 주지 않았던 객체가 있다면 모델은 그 객체로 주의를 이동하고, 주의를 주지 않았던 객체가 없다면 모델은 종료된다. 주의를 이동한 이후에는 현재 주의를 준 객체가 찾고 있던 대상인지 아닌지에 따라 보상을 받는다. 그리고 보상을 받게 되면 식 (1)에 의거하여 각 탐색 전략의 유용성이 바뀌게 된다. 보상을 주는 과정까지 끝났다면 주어진 이미지에서 찾을 수 있는 모든 대상이 찾아질 때까지 모델은 모든 단계를 다시 반복한다.

앞에서 언급한 것과 관련하여, 세 가지 전략에 해당하는 생산규칙은 각 전략의 영향에 비례하여 초기 유용성 값이 결정되어야 한다. 그래서 본 연구에서는 맥락 정보와 현저성을 동등하게 고려해야 함을 강조한 Torralba et al.(2006)의 주장에 의거하여, 처음 단계에서의 맥락 정보-기반과 현저성-기반 전략의 선택될 확률은 동등하다고 가정했다. 이에, 각 두 가지 전략 맥락-기반 전략과 현저성-기반 전략은 같은 유용성 값으로 설정하였고 임의선택-기반 전략의 유용성 값은 맥락-기반 전략과 현저성-기반 전략보다 낮은 값으로 설정하였다. 그리고 각 전략의 영향은 찾고자 하는 대상의 위치에 따라 변화한다. 예를 들어, 만약 모델이 처음 단계에서 현저성-기반 전략을 선택했는데 현저성-기반 영역에 찾고자 하는 대상이 없다면 현저성-기반 전략의 선택될 확률이 ACT-R의 유용성 시스템에 따라 줄어들 것이다.

다음 절에서는 실제 이미지의 현저성과 맥락 정보가 어떻게 구분되어 졌는지에 대해서 설명할 것이다.

2.3 현저성

사람의 시선에 대한 상향식 처리(bottom-up)를 설명하기 위해 사용된 ‘현저성 지도(saliency map)’는 가장 두드러진 곳을 지형적으로 배열한 지도이다(Niebur, 2007). 어떤 위치에서의 현저성은 주로 그 위치에서의 색(color), 방향(orientation), 동작(motion), 밝기(intensity) 등이 주위와 얼마나 다르냐에 따라 결정되며, 현저성 영역에 의해 사람의 시각 주의를 이끌려 진다고 설명하고 있다(Koch and Ullman, 1985). 그리고 Walther(2006)가 개발한 MATLAB의 saliency toolbox을 이용하면 색, 밝기, 방향 대비가 다른 위치보다 두드러진 영역을 정량적으로 계산한 현저성 지도를 만들 수 있다. 그 예로 <Figure 3>은 <Figure 1>에 대한 saliency toolbox의 결과로써 원으로 표시된 세 영역이 <Figure 3>에서의 현저성 영역에 해당한다.



Figure 3. High salient regions of <Figure 1>

2.4 맥락 정보

Torralba *et al.*(2006)에 따르면, 광범위의 영상에서 목표 대상을 찾을 때 맥락 정보를 통한 탐색은 수직상의 위치는 예측할 수 있지만 수평상의 위치는 예측하기 어렵다. 따라서 본 연구에서의 맥락 정보-기반 영역의 모양은 수평의 층(horizontal layer)이다. 예를 들어, <Figure 1>과 같은 실내 장면에서 액자들을 찾는 시각 탐색 과제를 수행할 때 맥락 정보에 대한 영역은 <Figure 4>와 같이 이미지 상단의 수평 부분이 선택될 것이다.



Figure 4. Scene-context region of <Figure 1>

3. 실험

3.1 실험 방법

본 연구에서는 모델에 대한 입증을 위해 5개의 실내 이미지를 가지고 실험이 수행되었다. 실험의 진행은 10명의 피험자를 대상으로 3초 동안 컴퓨터 스크린에 띄운 이미지 속에서 액자들의 수를 세도록 하였다. 3초 후에는 다음 이미지로 바뀌도록 하였고 총 5개의 이미지가 실험에 사용되었다. 그리고 액자를 찾는 작업동안 eye-tracking 장치를 통해서 이미지를 응시하는 시간을 측정하였다. 5개의 이미지는 <Figure 5>와 같다. 실험의 결과는 각 이미지에 대해서 3초 동안 맥락-기반 영역에 응시하는 시간과 현저성-기반 영역에 응시하는 시간의 비율, 그리고 모델과 비교함으로써 도출하였다.



Figure 5. Experimental images for the indoor bedroom

3.2 장비

eye-tracking 장치는 FaceLab Eye-Tracking(version 4.2.2)이 사용되었고 분석 프로그램은 GazeTracker software가 사용되었다.

3.3 모델

모델은 본 연구에서 앞서 제안한 프레임워크를 기반으로 만들어졌다. 이 모델에서는 대기 시간 요인(latency factor) 값과 유용성의 초기 값을 제외하고 기존 ACT-R 모듈에서 사용되는 파라미터들의 기초 값을 사용하였다. 각 탐색 전략에 대한 생산규칙의 초기 값은 앞서 언급한 내용과 같이 Torralba *et al.*(2006)의 주장에 의거하여 설정하였다(현저성, 맥락, 임의선택) = (10, 10, 6). 대기 시간 요인 값은 액자의 개수를 세는 작업 동안 단기 기억(short-term memory)만 요구하는 모델이라는 가정에 입각하여 0.14로 설정하였다(숙련된 기억을 인출하는 것에 대한 모델링을 연구한 Lewis(2005)의 연구에서도 대기 시간 요인의 값을 0.14로 설정). 모델은 <Figure 2>와 같이 7개의 생산규칙으로 구성되었고, 각 생산규칙에 대한 설명은 다음과 같이 설명될 수 있다.

먼저, 2번째 줄에 해당하는 세 가지의 생산규칙들에 대해 설명하자면, TARGET-S는 현저성-기반 영역, TARGET-C는 맥락 정보-기반 영역 그리고 TARGET-RANDOM은 주어진 이미지 내에서 임의의 영역으로 시각 주의를 이동시키기 위한 전략에 해당하는 생산규칙이다. 그리고 시각 탐색 모델이 시작되면, 각 영역 내에 있는 객체의 시각적인 위치를 탐색하기 위해서 세 가지의 생산규칙(TARGET-S, TARGET-C, 그리고 TARGET-RANDOM)들 중에 한 가지의 생산규칙이 선택되어 실행된다.

세 가지의 생산규칙 중 한 가지가 선택 되고 해당 영역에서 객체의 위치를 찾을 수 있다면, MOVE-ATTENTION 생산규칙이 실행된다. 이 규칙은 객체의 시각적인 위치로 주의를 옮기는 역할을 한다. 그 다음, TARGET-MATCHED 생산규칙과 TARGET-NOT-MATCHED 생산규칙 중 한 가지의 생산규칙이 실행되는데, 만약 현재 주의를 주고 있는 객체가 액자(목표 대상)라면 TARGET-MATCHED 생산규칙이 선택될 것이고, 만약 그렇지 않으면 TARGET-NOT-MATCHED 생산규칙이 선택될 것이다. 이 두 개의 규칙은 각각에 해당되는 보상을 가지고 있으며(TARGET-MATCHED 생산규칙이 실행되면 10의 보상을 받고, TARGET-NOT-MATCHED 생산규칙이 실행되면 0의 보상을 받는다), 어느 생산규칙을 선택할지 결정되어 실행되면 해당 생산규칙의 보상으로부터 실행되었던 생산규칙들의 유용성 값이 갱신된다.

3.4 결과

<Figure 6>은 피험자가 3초 동안 각 이미지의 맥락 정보-기반 영역을 응시하는 시간의 비율과 모델의 비율을 보여주고 있다. 회귀분석을 통한 0.99의 r-square 값으로부터 모델과 사람의 맥락-기반 영역에 응시하는 시간 사이에는 밀접한 선형 관계를 가지고 있음을 보여주고 있다.

<Figure 7>은 피험자가 3초 동안 각 이미지의 현저성-기반의 영역을 응시하는 시간의 비율과 모델의 비율을 보여주고 있다. 본 결과 역시 회귀분석을 통한 0.99의 r-square 값으로부터 모

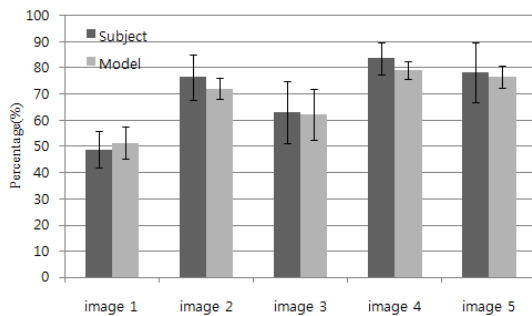


Figure 6. Ratio of gaze duration of the model and subjects in the context region of each image for three seconds

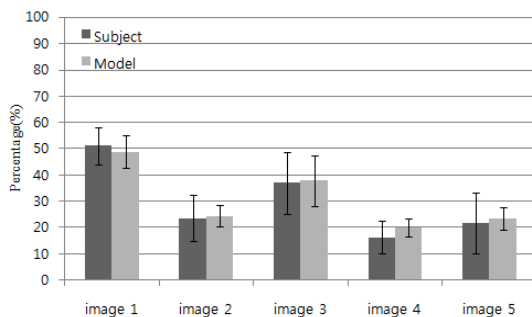


Figure 7. Ratio of gaze duration of the model and subjects in the salient region of each image for three seconds

델과 사람의 현저성-기반 영역에 응시하는 시간이 밀접한 선형관계를 가지고 있음을 보여주고 있다. 위 두 결과를 통해 본 연구에서 개발한 시각 탐색 모델의 결과는 사람의 데이터와 일치함을 볼 수 있다.

하지만 실험에 사용한 이미지들의 모든 객체들은 모두 맥락 정보-기반 영역과 현저성-기반 영역에 포함되기 때문에 모델 결과 값과 실제 실험 결과 값을 검토했을 때 그 외 영역(임의선택-기반 영역)에 응시하는 시간을 관찰할 수 없었다.

4. 토 의

결과를 통해 본 연구에서 개발된 시각 탐색 모델이 실제 이미지에서 사람의 시각 탐색을 정확하게 예측할 수 있음을 볼 수 있었다. 서론에서 언급했듯이 실제 이미지에서는 현저성 뿐만 아니라 맥락 정보를 고려하는 것이 중요하다. 따라서 우리는 현저성과 맥락 정보 모두 동등하게 고려한 시각 탐색 모델을 제시하였다. 이번 논문에서는 실제 이미지에서 현저성과 맥락 정보의 영향을 ACT-R의 유용성시스템을 사용함으로써 보여 줄 수 있었다.

<Figure 8>은 각 이미지에서 모델이 종료했을 때 맥락 정보-기반 전략의 최종 유용성 값과 현저성-기반 전략의 최종 유용성 값의 차이를 보여주고 있다. 1번 이미지의 경우는 목표 대상이 현저성-기반 영역뿐 만 아니라 맥락 정보-기반 영역에도 위치하고 있기 때문에 각 영역으로 시각적 주의를 이동시키는 전략이 유사한 비율로 선택되었을 것이다. 그에 따라 두 시각 탐지 전략의 최종 유용성 값이 유사해 짐을 볼 수 있다. 3번 이미지의 경우는 현저성 지도에 속하는 영역이 맥락 정보 영역과 유사했기 때문에 다른 2, 4, 5번 이미지보다 현저성-기반 영역에 응시하는 시간이 더 많았다. 그 결과 3번 이미지에서의 모델에서는 현저성-기반 규칙의 유용성 값과 맥락 정보-기반 규칙의 유용성 값의 차이가 이미지 2, 4, 5번보다 작은 것을 볼 수 있다. 이 결과를 통해 사람은 시각 탐색 작업을 수행할 때 현저성-기반 전략과 맥락 정보-기반 전략을 모두 이용하여 시각 탐색을 진행 하는데, 목표 대상이 위치한 영역에 따라 전략을 바꾸면서 탐색 하는 것을 볼 수 있다.

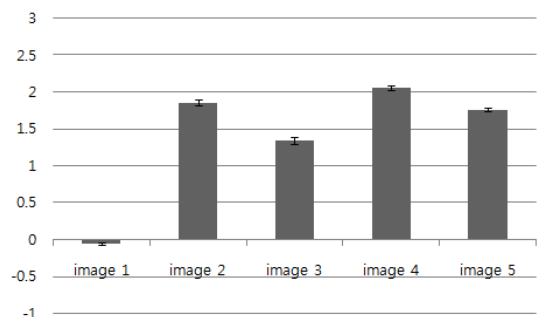


Figure 8. The differences of final utilities between context-based rule and saliency-based rule of the model of each image

또한 검증 실험을 통해 모델의 각 생산규칙에 대한 유용성의 변화 양상과 피험자의 시각 탐색 전략의 변화 양상이 일치함을 확인할 수 있었다. 특히, 이미지 2, 4, 그리고 5번에서의 시각 탐색 모델은 현저성-기반 영역을 응시하는 비율 보다 맥락-기반 영역을 응시하는 비율이 높음을 볼 수 있었다. 그 이유는 목표 대상이 이미지 2, 4, 그리고 5번에서 오직 맥락-기반 영역에만 있었기 때문이다. 이 와 같은 결과는 실제 실험을 통한 eye-tracking 결과에서도 볼 수 있었다.

<Figure 9>는 2번 이미지에서 한 피험자의 시각 탐지 패턴을 보여주는 결과이고 <Figure 10>은 2번 이미지에서 모델의 시각 탐지 패턴을 보여주는 결과이다. 실험 결과에서 사람의 시각 탐지 패턴은 현저성-기반 영역에 해당하는 좌측의 협탁으로 먼저 주의를 준 다음 맥락 정보 영역인 이미지의 상단으로 주의를 이동시키는 것을 볼 수 있었다. 그리고 피험자의 주의를 맥락 정보 영역 안을 이동하면서 액자를 탐색하였다. 모델 결과에서의 사람의 시각 탐지 패턴 역시 현저성-기반 영역에 해당하는 좌측의 협탁으로 먼저 주의를 준 다음 맥락 정보 영역인 이미지의 상단으로 주의를 이동시키는 것을 볼 수 있었다. 위 두 결과를 통해 맥락 정보 영역을 응시하는 비율이 현저성-기반 영역을 응시하는 비율보다 큰 것을 볼 수 있었고, 본 연구의 모델 역시 이러한 현상을 잘 표현하는 것을 보임에 따라 모델이 사람의 시각 탐지 패턴을 잘 묘사했음을 알 수 있다.

반면 실험 결과에 의하면, 이미지 1번과 3번은 맥락-기반 영역의 응시하는 시간이 나머지 이미지 2, 4, 그리고 5번에 비해 적은 것을 볼 수 있었다. 이를 더 명확하게 하기 위해 SPSS 통계

프로그램(IBM SPSS Statistics 21)을 사용하여 ANOVA의 Tukey 사후분석을 실시한 결과 2, 4, 5번의 그룹과 1번 그룹, 3번 그룹이 각각 형성되는 것을 볼 수 있었다. 그 이유로, 1번 이미지는 목표 대상들이 맥락-기반 영역뿐만 아니라 현저성 지도 영역(오른쪽 램프 부분)에도 위치하고 있기 때문이다. 이에, 두 영역의 응시하는 시간의 비율이 다르지 않게 나타났다.

이와 같은 결과는 한 피험자에 대한 1번 이미지의 eye-tracking 실험 결과인 <Figure 11>과 모델 결과인 <Figure 12>를 통해 확인할 수 있다. 실험 결과에서 피험자의 탐지 패턴은 현저성-기반 영역인 램프와 창가 주위에 많은 시간을 할애하는 것을 볼 수 있다. 모델 결과 역시 현저성-기반 영역인 램프 주위에 오랫동안 주의를 주고 그 이후에 맥락 정보-기반 영역으로 주의를 이동시키는 것을 볼 수 있다. 따라서 현저성-기반 영역과 맥락-기반 영역에서의 응시하는 비율의 차이가 크지 않음을 실험 결과와 모델 결과를 통해 확인할 수 있었고, 이번 연구에서 제안한 시각 탐지 모델은 사람의 시각 탐지 패턴을 충분히 잘 묘사하고 있음을 보여주고 있다.

본 논문에서는 실제 이미지에서 사람의 시각 탐색을 이해하기 위해 현저성의 영향뿐만 아니라 맥락 정보의 영향을 고려하여 인지모델을 개발하였다. 특히, 맥락 정보-기반 탐색 전략은 실제 이미지에서 사람의 시각 탐색을 정확하게 예측할 수 있었다. 따라서 본 연구에서 제안한 사람의 시각 탐색 모델링 방법은 실제 이미지에서 시각 탐색 작업을 수행할 때 사람의 수행도(performance)에 대한 정확한 모델을 제공하기 위하여 고려되어야 한다.



Figure 9. The eye tracking result of #2 image



Figure 11. The eye tracking result of #1 image



Figure 10. The model result of #2 image

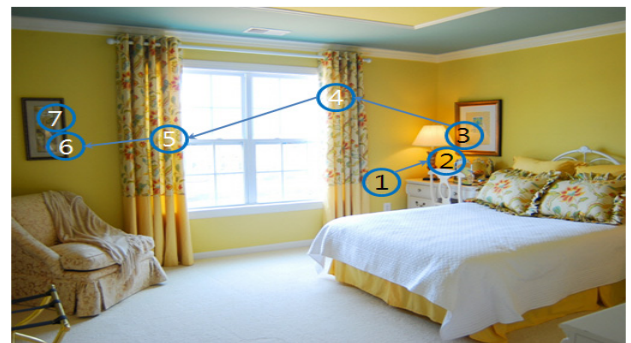


Figure 12. The model result of #1 image

추후 연구에서는 모델의 타당성을 확인하기 위해 실내 이미지를 사용한 것처럼 실제 실외 이미지를 사용한 검증을 통해 더욱 개선된 시각 탐지 인지모델링이 가능할 것이다.

5. 결론

본 연구에서는 컴퓨터 인터페이스상의 이미지에서 국한된 기존의 시각 탐색 인지모델과는 달리 실제 이미지에서 사람의 시각 탐색에 대한 인지모델을 개발하기 위해 현저성의 영향뿐만 아니라 맥락 정보의 영향을 고려한 시각 탐색 인지모델을 개발하였다. 그리고 ACT-R의 유용성 시스템을 이용하여 실제 이미지에서 현저성 영역으로 주의를 이동시키는 전략과 맥락 정보의 영역으로 주의를 이동시키는 전략에 따른 시각 탐색 과정에 대한 인지모델을 개발하였다. 또한, eye-tracking을 이용한 검증 실험을 통해 현저성-기반 영역과 맥락 정보 기반 영역에서 피험자와 모델의 응시하는 시간의 비율이 통계적으로 밀접한 선형관계를 가지고 있음을 보였고, 모델의 각 생산규칙에 대한 유용성 변화 양상을 피험자의 시각 탐색 전략의 변화 양상과 비교함으로써 모델이 실제 사람의 시각 탐색 패턴을 잘 묘사하고 있음을 확인하였다. 이에, 결론적으로 실제 이미지에서 사람의 시각 탐색을 모델링하기 위해 현저성-기반 전략과 맥락 정보-기반 전략이 고려된 방법을 사용해야 한다.

참고문헌

- Anderson, J. R. (2007), *How can the human mind occur in the physical universe?*, Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004), An integrated theory of mind, *Psychological Review*, **111**, 1036-1060.
- Anderson, J. R., Matessa, M., and Lebiere, C. (1997), ACT-R : A theory of higher level cognition and its relation to visual attention, *Human-Computer Interaction*, **12**(4), 439-462.
- Byrne, M. D. (2001), ACT-R/PM and menu selection : Applying a cognitive architecture to HCI, *International Journal of Human-Computer Studies*, **55**(1), 41-84.
- De Graef, P., Christiaens, D., and d'Ydewalle, G. (1990), Perceptual effects of scene context on object identification, *Psychological research*, **52**(4), 317-329.
- Halverson, T. and Hornof, A. J. (2011), A computational model of "active vision" for visual search in human-computer interaction, *Human-Computer Interaction*, **26**(4), 285-314.
- Henderson, J. M. (2003), Human gaze control during real-world scene perception, *Trends in cognitive sciences*, **7**(11), 498-504.
- Hornof, A. J. (2004), Cognitive strategies for the visual search of hierarchical computer displays, *Human-Computer Interaction*, **19**(3), 183-223.
- John, B. E. and Kieras, D. E. (1996), Using GOMS for user interface design and evaluation: Which technique?, *ACM Transactions on Computer-Human Interaction(TOCHI)*, **3**(4), 287-319.
- Jung, K. (2015), Legible and preferred korean sizes for various colors and fonts, *Journal of the Korean Institute of Industrial Engineers*, **41**(1), 59-63.
- Koch, C. and Ullman, S. (1985), Shifts in selective visual attention : towards the underlying neural circuitry, *Human Neurobiology*, **4**(4), 115-141.
- Kujala, T. and Saariluoma, P. (2011), Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks, *Ergonomics*, **54**(8), 716-732.
- Lewis, R. L. and Vasishth, S. (2005), An activation-based model of sentence processing as skilled memory retrieval, *Cognitive science*, **29**(3), 375-419.
- Ling, J. and Van Schaik, P. (2004), The effects of link format and screen location on visual search of web pages, *Ergonomics*, **47**(8), 907-921.
- Lohse, G. L. (1993), A cognitive model for understanding graphical perception, *Human-Computer Interaction*, **8**(4), 353-388.
- Niebur, E. (2007), Saliency map, *Scholarpedia*, **2**(8), 2675.
- Nyamsuren, E. and Taatgen, N. A. (2013), Pre-attentive and attentive vision module, *Cognitive Systems Research*, **24**, 62-71.
- Oh, H., Jo, S., and Myung, R. (2011), Head Movement Module in ACT-R for Multi-display Environment, *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **55**(1), 1836-1839.
- Oh, H., Jo, S., and Myung, R. (2014), Computational modeling of human performance in multiple monitor environments with ACT-R cognitive architecture, *International Journal of Industrial Ergonomics*, <http://dx.doi.org/10.1016/j.ergon.2014.09.004>
- Oh, H. and Myung, R. (2012), Modeling Human Visual Processing Within and Beyond the Oculomotor Range Using ACT-R Cognitive Architecture, *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **56**(1), 995-999.
- Plumlee, M. D. and Ware, C. (2006), Zooming versus multiple window interfaces : Cognitive costs of visual comparisons, *ACM Transactions on Computer-Human Interaction (TOCHI)*, **13**(2), 179-209.
- Ritter, F. E., Van Rooy, D., Amant, R. S., and Simpson, K. (2006), Providing user models direct access to interfaces : An exploratory study of a simple interface with implications for HRI and HCI, *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, **36**(3), 592-601.
- Salvucci, D. D. (2001), An integrated model of eye movements and visual encoding, *Cognitive Systems Research*, **1**(4), 201-220.
- Schneider, W. X. (1995), VAM : A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action, *Visual Cognition*, **2**(2-3), 331-376.
- St Amant, R. and Riedl, M. O. (2001), A perception/action substrate for cognitive modeling in HCI, *International Journal of Human-Computer Studies*, **55**(1), 15-39.
- Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006), Contextual guidance of eye movements and attention in real-world scenes : the role of global features in object search, *Psychological review*, **113**(4), 766.
- Treisman, A. (2006), How the deployment of attention determines what we see, *Visual cognition*, **14**(4-8), 411-443.
- Tuch, A. N., Bargas-Avila, J. A., Opwis, K., and Wilhelm, F. H. (2009), Visual complexity of websites : Effects on users' experience, physiology, performance, and memory, *International Journal of Human-Computer Studies*, **67**(9), 703-715.
- Walther, D. (2006), Interactions of visual attention and object recognition : computational modeling, algorithms, and psychophysics, *Phd thesis*, California Institute of Technology.