
모션 기반 선택적 주의 시스템

A Motion-driven Selective Visual Attention System

박민철*, 최경주**

한국과학기술연구원 시스템 연구부*, 충북대학교 전기전자컴퓨터공학부**

Min-Chul Park(minchul@kist.re.kr)*, Kyung-Joo Cheoi(kjcheoi@chungbuk.ac.kr)**

요약

본 논문은 동영상에서 ROI(region of interest)나 FOA(focus of attention)를 탐지하기 위한 방법으로써 여러 시각자극(stimuli) 중 모션(motion) 정보를 사용한 선택적 주의 시스템을 소개한다. 2개 이상의 연속적인 프레임으로 구성되는 모션은 시간정보가 포함된 현실세계의 시각장면에서 시간적으로 현저함(temporal saliency) 정도를 나타내는 특징이라 할 수 있으며, 본 연구에서는 이러한 모션을 인간 뇌의 대뇌 시각피질에 존재하는 MT(middle temporal cortex) 영역에 존재하는 “이중 대립 수용야(double opponent receptive fields)”의 특성과 “노이즈 여과(noise filtration)” 기능에 대한 기존의 심리학적 연구 결과를 바탕으로 분석하고, 이렇게 분석된 결과를 MT에서의 “모션통합” 이론을 기반하여 통합한다. 이러한 과정을 통해 동영상에서 다른 주변의 영역과 현저히 다른 돌출 영역으로 정의되는 ROI 또는 FOA를 탐지할 수 있다. 제안하는 시스템의 성능분석을 위해 인간을 대상으로 한 심리학적 실험을 추가적으로 수행하여 제안된 시스템을 통해 얻어진 결과와 비교함으로써, 본 시스템이 얼마나 타당성이 있는지 검증하였다.

■ 중심어 : | ROI | FOA | 모션 | 선택적 주의 | MT | 이중 대립 수용야 | 노이즈 여과 | 통합 |

Abstract

In this paper, a selective visual attention module based on motion stimuli is introduced for the purpose of detecting ROI(region of interest) or FOA(focus of attention) in motion pictures. Analysis of motion fields in our approach is in direct contrast to some of the previous studies of selective visual attention module. Motion that presents temporal visual saliency in an aspect between two successive frames is analyzed based on psychological studies in “DORF(double opponent receptive fields)” and “NF(noise filtration)” in MT(middle temporal cortex). Analyzed results are integrated based on the theory of “motion integration” in MT to obtain a single conspicuous region. Experiments through a human subjective evaluation showed generally accepted results.

■ keyword : | ROI | FOA | Motion | Selective Visual Attention | Double Opponent Receptive Fields | Noise Filtration | Integration |

* 본 연구는 한국전산원의 광개토 컴소사업 : HFC/FTTH 기반의 통신 방송융합서비스 연구과제로 수행되었습니다.

I. 서론

인간을 포함한 영장류는 신경세포라는 하드웨어로 구성되어 있어 정보 전달에 있어 속도제한이 있을 것이라는 기대에도 불구하고, 입력되는 모든 복잡한 시각장면을 실시간으로 해석하고 이해하는 뛰어난 능력을 가지고 있다. 인간은 시각장면을 분석할 때 이에 대한 정보 처리의 복잡도를 낮추기 위해 가장 유용한 감각정보를 미리 선택하여 그 부분에 대하여 더욱 자세한 처리를 수행하도록 한다[1]. 다시 말해 ‘우리는 우리 주변의 모든 것을 본다’라는 보편적인 믿음에도 불구하고, 순간순간 입력되는 수많은 정보 중 일부분만이 시각 시스템에 입력되고 이 일부분에 해당하는 정보들에 대해 추가적인 자세한 처리가 수행되어 이 일부 정보에 의해 우리의 행동이 직접적인 영향을 받게 된다는 것이다. 또한 선택적 주의는 인간이 살아가는 데 있어 가장 중요한 기능 중 하나라 볼 수 있는데 한 예로, 진화적인 입장에서 보았을 때 동물은 선택적 주의 능력을 통해 먹이나 적의 존재 유무를 재빨리 인지하도록 하여 생태계에 쉽게 적응하여 살아갈 수 있도록 해 준다[2]. 시각장면에서의 특정한 위치는 행동적 관련성 또는 국부적인 영상의 단서에 기반하여 선택되어진다. 인간을 포함한 영장류는 한곳에 주의를 집중하는 시각적 주의와 도약 안구운동을 사용하여 복잡한 시각정보를 실시간으로 분석하며, 어떤 부분에 집중하고 안구를 움직이는가는 이들의 행동습성에 크게 좌우된다[3].

기존의 전형적인 특징기반 선택적 주의 시스템은 인간의 선택적 주의의 상향식 방식에 기반을 둔 것으로, 정지영상에서 색상, 방위, 명암도 등의 특징을 사용하여 ROI를 찾는다[4]. 이러한 선택적 주의 시스템은 시간특징 중 하나인 모션정보를 하향식 단서로 사용하여 동영상에서도 사용되어질 수 있도록 확장되어지곤 하였다[5]. 기존의 공학적인 모션정보의 분석을 위한 방법으로는 국소적 움직임, 고차운동, 색 운동 등을 이용하는 방법 등이 사용된다. 먼저 국소적 움직임을 검출하는 방법에는 모션에너지나 광류(optical flow)의 구속식을 이용하는 구배법 등이 존재하며, 고차운동을 이용하는 방법에서는 영상의 2차이상의 특징을 사용한다. 또한 색

운동을 이용하는 방법에서는 등휘도에서 색변화를 이용하여 영상의 움직임을 검출한다[6]. 이러한 모션분석방법을 선택적 주의 시스템에 적용한 연구들은 모션의 크기나 방향을 고려하지 않고 단지 입력된 영상에서 움직임이 있는 영역만을 탐지하는데 그치고 있으며, 또한 생물학적인 기반을 가지지 않는다.

본 논문에서는 생물학적인 기반을 갖는 선택적 주의 시스템으로써 동영상에서 ROI를 탐지하는 시스템을 제안한다. 제안된 시스템에 모션자극이 입력되면 블록매칭 기법을 이용하여 모션을 탐지하고 탐지된 시간특징 중 하나로 분석한다. 그리고 이를 하나로 통합[7][8]하여 그 결과로 ROI를 탐지하게 된다. 탐지된 모션의 분석은 대뇌 시각피질 영역의 MT 영역에서의 “이중 대립 수용야”의 특성과 “노이즈 여과” 기능에 관한 심리학적 연구결과를, 그리고 분석된 결과의 통합 역시 MT에서의 “모션 통합” 이론에 관한 심리학적 연구결과를 토대로 수행되어졌다. 또한 모션 분석 시 모션벡터에 대해 중심-주변 연산자를 사용하여 시간특징에 대한 현저도를 산출하였다. 이렇게 본 논문에서 제안하는 시스템은 모션분석 및 통합에 있어 기존의 심리학적이 연구결과를 사용하므로 생물학적인 기반을 가지며, 공간특징만을 대상으로 한 기존의 선택적 주의 연구들과 달리 시간특징인 모션을 분석하고 이를 통해 얻어진 결과를 사용한다. 또한 인공적으로 간단히 만들어진 심리학적 실험데이터가 아닌 외부환경에서 흔히 접할 수 있는 실영상에 적용할 수 있다.

다음 2장에서는 모션분석 시 토대가 되었던 심리학적 연구들에 대해 소개하고, 3장에서는 본 논문에서 제안하는 모션에 기반한 선택적 주의 시스템에 대해 기술한다. 이어 4장에서 제안된 시스템의 실험결과에 대해 설명하고, 5장에서 결론을 맺는다.

II. 심리학적 연구를 토대로 한 모션정보의 분석 및 통합

영장류의 대뇌 시각피질 영역 중 MT(middle temporal area) 영역이 모션분석에 있어 주요한 역할을

하고 있다는 사실은 많은 해부학적, 생리학적 연구들로부터 밝혀져 왔다.

MT의 뉴런들은 방위와 속도에 모두 강하게 선택적으로 반응한다. 즉 MT 세포들은 각각 특정 속도와 방향으로 움직이는 시각자극에 최대로 반응하도록 조율되어 있기 때문에 이 최적속도와 방향으로 움직이는 시각자극이 주어졌을 때만 최대로 반응하도록 되어있다. 대뇌 1차 시각피질인 V1과 2차 시각피질인 V2는 모두 MT로 투영된다[9]. 따라서 MT 세포는 V1과 V2 세포가 가지고 있는 특성을 기본으로 그 위에 MT 세포만의 고유 특성을 더한 위계적으로 구성되어 있다[10].

본 논문에서는 시각장면에서 탐지된 모션을 분석하기 위해 대뇌 시각피질 영역 중 MT 영역의 “이중 대립 수용야”의 성질과 “노이즈 여과” 기능에 대한 심리학적 연구결과를 사용하였고, MT 영역의 “모션 통합”에 대한 심리학적 연구결과를 참고하여 모션 분석을 통해 분석된 결과를 통합하였다.

1. 이중 대립 수용야

MT 세포가 어떻게 반응하게 되느냐는 MT 세포의 수용야 내부 영역과 주변 영역에 모션자극(움직이는 시각자극)이 어떻게 주어졌는가에 따라 다르다. 수용야 내부 영역에 최적방향으로 모션자극이 주어졌을 때 MT 세포는 최대로 반응한다([그림 1](b) 참조). 그런데 위와 같은 모션자극이 있는 상태에서 수용야의 주변영역에 모션자극이 최적방향의 반대방향으로 주어지면 MT 세포의 반응은 더욱 증대된다([그림 1](c) 참조). 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([그림 1](a) 참조). 이러한 “이중 대립 수용야”의 특성은 모션의 경계를 탐지하고, 탐지된 모션의 경계를 기반으로 하여 모션의 범위를 정의하는데 유용하게 사용할 수 있다[10].

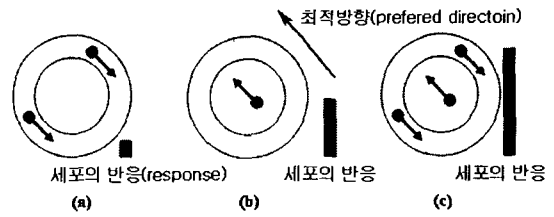


그림 1. 최적방향과 수용야 위치에 따른 세포의 반응

2. 노이즈 여과

MT는 모호한 모션자극을 걸러내기도 한다. 만일 실험자가 MT 세포의 수용야에 하나는 세포의 최적방향으로 움직이고 다른 하나는 최적방향의 반대방향으로 움직이는 2개의 점을 제시하면 세포의 반응은 매우 약해진다([그림 2] 참조). 사실상, 이러한 자극은 일반적으로 시각장면에 있어서의 “노이즈” 때문에 생긴 결과이며, MT는 이것을 노이즈로 해석한다.

V1과 MT 영역의 세포들은 모두 특정 방향에 선별적인 반응을 보이는데, MT 세포의 경우 투명한 모션자극이 주어졌을 때는 세포의 반응이 억제되는 것에 비해 V1 세포들은 투명한 모션자극일지라도 최적방향으로 모션자극이 주어졌다면 이 자극에 최대로 반응한다. 이렇듯 MT 세포는 보다 단순한 처리를 수행하는 V1 세포의 발화 패턴을 기본 바탕으로 좀 더 정제된 모션처리를 수행한다.

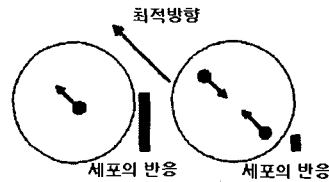


그림 2. MT 세포의 노이즈 여과 기능

3. 모션통합

영장류에 있어 복잡한 모션 처리는 2단계에 걸쳐 일어난다. 첫 번째 단계에서는 1차원의 국부적인 모션이 분석되고, 두 번째 단계에서 국부적인 모션이 전역적인 모션으로 통합된다. 이러한 프로세스를 “모션통합”이라

하며, 본 연구에서는 이러한 기본지식을 바탕으로 모션을 통합한다.

III. 모션자극에 기반한 선택적 주의 시스템

동물의 뇌에서 중앙 신경 시스템 전반에 걸쳐 위치한 수많은 신경 세포들은 모션자극에 대해 큰 반응을 보인다. 이러한 모션자극에 대한 선별성으로 인해 감각정보의 입력만으로도 모션을 인지할 수 있다. 또한 이러한 모션분석을 통해 몸의 균형과 자세를 잡는 것뿐만 아니라 관찰자의 움직임 방향과 주변의 물체나 표면과의 충돌 시간의 추측도 가능하다[11]. 이번 장에서는 모션자극에 기반을 둔 선택적 주의 시스템을 소개하고자 한다. 이 장 맨 마지막에 제안하는 주의 시스템을 통해 얻어지는 결과를 ROI 또는 FOA로써 제시하였다.

[그림 3]은 제안하는 선택적 주의 시스템의 일반적인 동작과정을 보여준다. 그림에서도 보이는 바와 같이 제안하는 시스템은 모션분석과 모션통합이라는 2개의 주요 모듈로 구성되어 있다. 모션분석 모듈은 블록매칭 기법으로부터 탐지된 모션벡터(모션맵)에 대한 분석을 수행하는 부분으로써, 모션벡터의 분석은 상향식 방법으로 분석되어지고, 병렬적으로 처리된다. 모션분석은 모션벡터를 나타내는 모션맵의 각 픽셀마다 모션의 방향과 크기를 계산함으로써 모션맵에 시간특징이라는 하나의 측정치를 할당하는 방식으로 분석이 수행된다. 모션분석 모듈은 'DORF(이중대립수용야)'와 'NF(노이즈여과)', 그리고 'LMC(local motion conspicuity)'로 구성되어 있다. DORF와 NF는 인간 대뇌 시각피질 영역 중 MT 영역의 이중 대립 수용야의 성질과 노이즈 여과 기능에 대한 기존의 심리학적 연구결과를 토대로 만들어진 것이고, LMC는 중심-주변 연산자를 통하여 주변의 국부적인 모션과는 현저히 다른 모션을 얻기 위해 계산되어진다. 또 다른 모듈인 모션통합 모듈은 위 단계에서 분석된 결과를 통합하여 하나의 현저도 맵으로 만들어낸다.

1. 모션분석 프로세스

시각장면에서 모션정보를 얻기 위하여 현재 프레임(t) 및 이전 프레임(t-1)로 구성되어 있는 연속적인 명암도 영상 시퀀스를 사용하였으며[6], 이 영상 시퀀스에 블록매칭 기법을 사용하여 모션맵 $M_{x,y}$ 을 얻는다. 모션맵을 얻기 위한 방법으로 블록 매칭 방법을 사용함으로써 화소 단위가 아닌 블록 단위의 계산이 이루어져 보다 안정적인 모션정보를 추출할 수 있다. 모션맵은 모션에 대한 수평, 수직 요소인 V_x 과 V_y 에 의해 2-D영상으로 그려진다. 제안하는 시스템은 이 모션맵에 대하여 다음과 같은 3개의 모듈을 통해 모션분석을 수행한다.

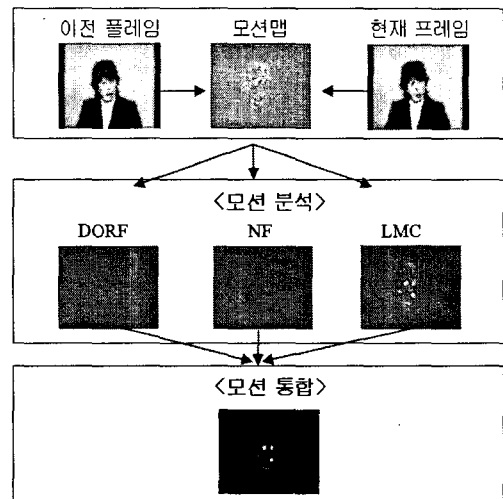


그림 3. 제안하는 시스템의 전반적인 동작과정

1.1 DORF 모듈 : 이중 대립 수용야

DORF 모듈은 [그림 4]와 같이 주변의 모션자극쌍의 방향이 최적방향과 반대방향이고, 중심의 모션자극이 최적방향으로 주어졌을 때 최대 반응하도록 설계되었다. 향이 최적방향과 반대방향이고, 중심의 모션자극이 최적방향으로 주어졌을 때 최대 반응하도록 설계되었다.

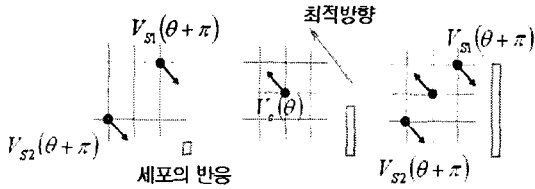


그림 4. "DORF"를 기반으로 설계된 DORF 모듈

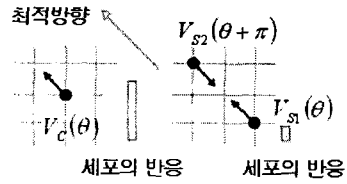


그림 5. "NF"를 기반으로 설계된 NF 모듈

각 픽셀의 반응값은 8개의 방위 ($\theta \in (0, \pi/4, \dots, 7\pi/4)$)마다 계산된다. [표 1]은 8개 방위에 따른 최적방향과 관련된 중심과 주변영역의 모션벡터 위치를 보여준다. 픽셀의 반응값은 식(1)에 의해 계산된다.

1.2 NF 모듈 : 노이즈 여과

NF 모듈은 [그림 5]와 같이 2개의 주변 모션자극의 방

표 1. 방향과 위치에 대한 DORF 모듈의 룩업테이블

방위	주변1 (S1)	중심 (C)	주변2 (S2)
0, 4π/4	(x+0, y-1)	(x, y)	(x+0, y+1)
π/4, 5π/4	(x-1, y+1)	(x, y)	(x+1, y-1)
2π/4, 6π/4	(x-1, y+0)	(x, y)	(x+1, y+0)
3π/4, 7π/4	(x-1, y-1)	(x, y)	(x+1, y+1)

$$DORF_{x,y} = \sum_{\theta} \left(\sum_{m,n} |R_{m,n}(\theta)| \right)$$

만일 $|V_c(\theta)| \neq 0$ 이면,

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

그렇지 않을 경우,

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

대 방향으로 주어졌을 때 최대로 반응하도록 설계되었다. NF 모듈은 4개의 방위($2\pi/4, \dots, 5\pi/4$)를 가질 수 있는 주변의 모션벡터 쌍의 방향이 서로 반대방향이 되었을 때 반응한다.

[표 2]는 4개 방향에 따른 최적방향과 관련된 중심과 주변의 모션벡터의 위치를 보여준다. 픽셀의 반응값은 식(2)에 의해 계산된다.

1.3 LMC 모듈 : 국부적으로 현저한 모션 탐지

LMC(Local Motion Conspicuity) 모듈은 모션맵 $M_{x,y}$ 에 특정 연산자를 사용하여 국부적인 영역마다 주변의 모션들과 현저히 다른 모션을 탐지한다.

LMC를 탐지하기 위해 본 모듈에서는 모션맵에 8개의

표 2. 방향과 위치에 대한 NF 모듈의 룩업테이블

방위	주변1 (S1)	중심 (C)	주변2 (S2)
2π/4	(x+0, y-1)	(x, y)	(x+0, y+1)
3π/4	(x+1, y-1)	(x, y)	(x-1, y-1)
4π/4	(x+1, y+0)	(x, y)	(x-1, y+0)
5π/4	(x+1, y+1)	(x, y)	(x-1, y+1)

$$NF_{x,y} = \sum_{\theta} \left(\sum_{m,n} |R_{m,n}(\theta)| \right)$$

만일 $|V_c(\theta)| \neq 0$ 이면, $|R(\theta)| = |V_c(\theta)| \quad (2)$

그렇지 않을 경우,

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

항이 하나는 최적방향이고, 다른 하나는 최적방향의 반방위에 조율된 ON-중심과 OFF-주변 연산자인 다중 스케일의 필터뱅크인 DOOrG 필터를 회선(convolution)시킴으로써(식 3 참조) 모션맵의 각 픽셀 위치마다 이들의 현저도를 나타내는 현저도 측정치를

할당하였다.

DOOrG 필터에서 ON과 OFF의 폭간 비율은 $r_1 = \sigma_{off} / \sigma_{on}$ 에 의해 정의되며 또한 고정된다. 2개의 가우시안은 동심원 형태(원형의 좌우대칭 형태)보다는 타원 형태를 가지며, 이때 2개의 가우시안의 이심률(eccentricity)은 $r_2 = \alpha_x / \alpha_y$ 이다. [그림 6]은 동심원 형태의 필터 예를, [그림 7]은 타원 형태 필터 예를 보여준다. [그림 6]과 같은 동심원 형태의 ON-중심, OFF-주변 모형에서는 빛의 밝기가 일어나는 영역의 위치나 크기에 따라 세포가 서로 다른 반응을 보인다. [그림 6]에서 점선으로 된 박스는 빛의 밝기가 일어나는 영역을, 회색빛의 막대바는 빛의 밝기 변화가 일어나는 영역에 따른 수용야 별 세포의 반응도를 나타낸다. 빛의 밝기 변화가 수용야의 "ON" 영역에 일어나면 이 세포의 반응도는 매우 커지는 반면([그림 6](b) 참조), "OFF" 영역에 일어나면 반응은 일어나지 않는다([그림 6](c) 참조). 또한 빛의 밝기 변화가 수용야 전반에 걸쳐 일어나도 세포는 이에 거의 반응하지 않는다([그림 6](d) 참조). [그림 6]과 [그림 7]에서 보이는 DOOrG 필터의 특성을 사용하면 모션맵에서 형태가 뚜렷하지 않은 모양의 영역을 탐지할 수 있다.

각 DOOrG 필터는 식 3과 같이 8개의 방위별로 계산된다. 특정 방위를 가진 DOOrG 필터 $DOOrG_{x,y}(r_1, r_2, \theta, \sigma)$ 는 $D^1 G^1 - D^2 G^2$ 로 정의될 수 있다. 여기서 $G_{x,y}(\alpha_x, \alpha_y, \theta)$ 는 식 4와 같은 2-D 가우시안이고, 상수 D^1 과

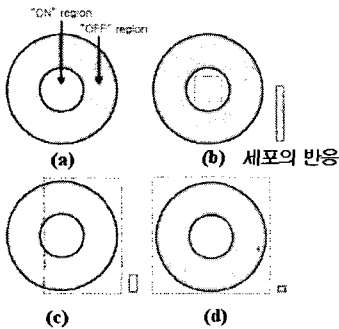


그림 6. 동심원 형태의 ON-중심, OFF-주변 모형

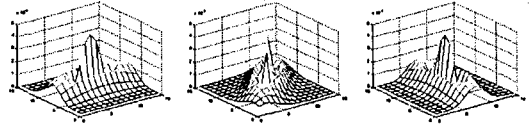


그림 7. 서로 다른 방위에 조율된 몇몇 ON-중심, OFF-주변 연산자

D^2 의 값은 각각의 요소들에 대한 계수의 합이 1로 정규화되도록 설정된다. 여기서 스케일 파라미터는 실험적으로 결정하여 사용하였다. 얻어진 결과값의 부호를 없애고, 콘트라스트를 강조하기 위하여 최선한 결과를 제공하는 후 각 방위별 만들어진 맵을 하나의 유일한 현저한 맵으로 만들기 위해 방위별로 나뉜 맵을 극부적인 최대값을 사용하여 하나로 통합하였다.

$$F_{x,y} = \sum_{\theta} \left(\sum_{m,n} M_{m,n} \cdot |D^1 \cdot G^1 - D^2 \cdot G^2| \right)^2 \quad (3)$$

단, $G^1 = G_{x-m, y-n}(\sigma, \gamma_1, \sigma, \theta)$,

$G^2 = G_{x-m, y-n}(\gamma_2, \sigma, \gamma_1, \gamma_2, \sigma, \theta)$

$$G_{x,y}(\sigma_x, \sigma_y, \theta) = e^{-\frac{(x \cos \theta + y \sin \theta)^2}{2\sigma_x^2}} \cdot e^{-\frac{(-x \sin \theta + y \cos \theta)^2}{2\sigma_y^2}} \quad (4)$$

2. 모션통합 프로세스

기존 연구들에 의해 인간의 선택적 주의에 영향을 준다고 알려진 많은 특징들이 확인되어져 왔다. 그러나 이러한 특징들에 대해서 어떠한 특징이 다른 특징에 비해 더 중요한지, 이들 특징간의 관련성은 어떠한지에 대해서는 거의 알려진 바가 없으며, 이를 알아내기도 매우 어렵다. 예를 들어 어떤 특징의 경우 특정 영상에서는 다른 특징에 비해 중요도가 높을 수 있을지 모르지만 다른 영상에서는 서로 반대 입장이 될 수도 있는 것이기 때문이다.

이렇게 특징의 중요도 및 이들에 관련성에 대한 정보가 미비하므로 선택적 주의를 시스템화 할 때에는 수많은 요소들을 고려해야만 한다. 주의 시스템이 얻고자 하는 목표물(타겟)에 대한 템플릿에 대한 지식을 미리 알

고 있다면 보다 높은 정확도를 가진 주의 시스템이 만 들어질 수 있을 것이다. 그러나, 일반적으로 순간순간 바라보고 있는 시각장면에 내용이나 정황에 대해서 미리 알기란 매우 어려우므로 고수준의 정보는 사용되어 질 수 없다[12]. 따라서 본 논문에서는 고수준의 선지식을 사용하지 않는다. 서로 다른 3개의 모션분석 모듈로부터 얻어진 3개의 특징맵 $DORF_{x,y}$, $NF_{x,y}$, $F_{x,y}$ 은 [그림 3]에서 보인 바와 같이 단순히 정규화하고 합해져서 하나의 현저도 맵 $C_{x,y}$ 로 통합되어진다.

$$C_{x,y} = \frac{1}{3} \sum_{x,y} (DORF_{x,y} + NF_{x,y} + F_{x,y}) \quad (5)$$

IV. 실험결과 및 분석

제한된 선택적 주의 시스템을 평가하기 위한 테스트 영상으로써 비디오 압축에 주로 쓰이는 표준 이미지 시퀀스 중 2개의 얼굴 영상 시퀀스와 도로 영상 시퀀스를 사용하였다. 테스트 영상은 CIF(Common Intermediate Format) 포맷의 352X288 해상도를 가진 “클레어(Claire)”, “파리(Paris)”, “도로(Highway)”의 3가지 종류의 영상 시퀀스로, 각 영상 시퀀스는 150 프레임으로 구성되어 있으며, 5분간 지속된다(1초당 30 프레임).

제한된 시스템은 입력 영상 시퀀스에 대한 어떠한 선지식을 가지고 있지 않다는 가정을 전제로 하고 있기 때문에, 입력되는 영상 시퀀스에서 탐지하고 하는 ROI는 영상 시퀀스에 대한 이해과정보다는 움직임에 대한 모션 자극을 통해서 우선적으로 탐지된다.

선택적 주의 시스템에 대한 평가는 시스템이 주의가 가는 곳(ROI)을 얼마나 잘 찾아내는 것인가이다. 하지만 현실적으로 영상의 어떤 부분에 대해 주의가 가는지는 각 영상마다 다 다르다. 즉 선택적 주의 시스템에 대한 객관적인 실험 데이터 및 이에 대한 평가기준이 없다. 이에 본 연구에서는 우선적으로 동일한 테스트 영상에 대해 인간을 대상으로 한 심리학적 실험을 우선적으로 실행하여 그 결과를 평가의 기준으로 삼았으며, 인간을 대상으로 한 실험 결과에 비해 시스템에서 나온 결

과가 얼마나 매칭이 되는지 분석하여 시스템의 성능을 평가하였다.

인간을 대상으로 한 심리학적 실험에는 남녀 40명이 참여하였다. 임의적으로 선택된 3개의 영상 시퀀스를 20~30세 사이의 20명의 남성과 20명의 여성에게 보여준 후, 제시된 영상 시퀀스에서 가장 주의가 가는 영역이나 물체에 마크 표시를 하라 지시하였다. 만일 주의가 가는 곳이 여러 곳이라면 주의가 가는 우선순위로 체크하도록 하였다.

본 시스템으로부터 탐지된 ROI는 그레이 스케일 영상의 현저도 맵으로 최종 출력된다. 최종 출력된 현저도 맵안의 명암도는 현저함 정도를 나타내는데, 이 때 더 밝은 명암도를 가진 부분일수록 더욱 현저하다는 것을 나타낸다. [표 3]~[표 5]는 인간을 대상으로 한 심리학적 실험결과를 나타내고, [그림 8]~[그림 10]은 이에 대응되는 본 논문에서 제안하는 주의 시스템을 통해 얻은 ROI를 보여준다.

“클레어(Claire)” 영상 시퀀스의 경우, 인간을 대상으로 한 심리학적 실험에서는 입과 눈 영역이 가장 현저한 영역이라고 마크되었다(62.5%, [표 3] 참조). 만일 입과 눈 영역을 얼굴 영역으로 포함한다면 이의 비율은 70%까지 올라간다. [그림 8]은 본 시스템을 통해 얻어진 결과 중 하나를 보여주는데, 본 시스템 역시 입과 눈 영역을 ROI로 식별함을 알 수 있다.

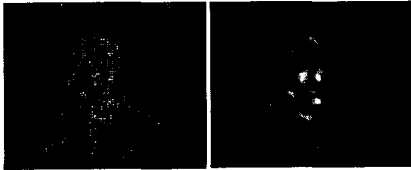
표 3. “클레어” 영상 시퀀스에 대한 인간의 심리학적 실험결과

성별	1순위	2순위	3순위
남자:20	입:7	입:6	입:1
	옷:5	헤어:4	헤어:1
	눈:4	눈:3	
	헤어:2	옷:3	
	얼굴:1	귀걸이:1	
	귀걸이:1	고개:1	
여자:20		머플러	
	입:8	입:6	헤어:1
	눈:6	헤어:6	
	옷:3	눈:3	
	얼굴:2	볼터치:1	
귀걸이:1	귀걸이:1		

전체:40	입:15	입:12	입:1
	눈:10	헤어:10	헤어:1
	옷:8	눈:6	
	얼굴:3	옷:3	
	헤어:2	귀걸이:2	
	귀걸이:2	볼터치:1	
		고개:1	



(a) 전 프레임 (b) 현재 프레임



(c) 모션맵 (d) 현저도맵

그림 8. 제안된 시스템에 입력되는 “클레어” 영상 시퀀스와 이에 대한 시스템의 결과 예

“파리(Paris)” 영상 시퀀스의 경우, 인간을 대상으로 한 심리학적 실험에서는 공 영역이 가장 현저한 영역이라고 마크되었다(60%, [표 4] 참조). 공은 [그림 9]에서도

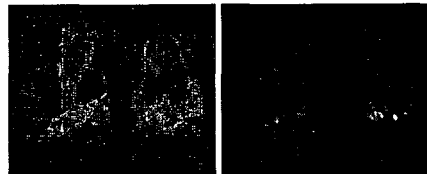
표 4. “파리” 영상 시퀀스에 대한 인간의 심리학적 실험결과

성별	1순위	2순위	3순위
남자:20	공:12	팬:6	팬:3
	팬:12	공:6	남자얼굴:2
	노란책:2	여자얼굴:3	공:1
	남자헤어:1	파란인형:1	
	남자얼굴:1	팔찌:1	
	여자얼굴:1	서류:1	
	서류용치:1		
여자:20	공:12	팬:7	팬:2
	책:2	공:5	넥타이:1
	여자손:2	넥타이:2	
	남자얼굴:1	종이컵:2	
	여자얼굴:1	책:2	
	테이블:1	서류:1	
	남자입:1		

전체:40	공:24	팬:15	팬:5
	책:4	공:11	남자얼굴:2
	남자얼굴:2	여자얼굴:3	공:1
	여자얼굴:2	넥타이:2	넥타이:1
	여자손:2	종이컵:2	
	팬:2	책:2	
	서류:1	서류:2	
	남자헤어:1	팔찌:1	
	테이블:1	인형:1	
	남자입:1		



(a) 전 프레임 (b) 현재 프레임



(c) 모션맵 (d) 현저도맵

그림 9. 제안된 시스템에 입력된 “파리” 영상 시퀀스와 이에 대한 시스템의 결과 예

표 5. “도로” 영상 시퀀스에 대한 인간의 심리학적 실험결과

성별	1순위	2순위	3순위
남자:20	표지판:9	도로선:6	표지판:1
	도로선:8	표지판:6	검은점:1
	구름:3	구름:4	
		검은점:1	
		왼쪽막대:1	
여자:20	도로선:13	표지판:8	구름:3
	표지판:5	도로선:6	
	구름:1	구름:4	
	길(전체):1	아스팔트:1	
전체:40	도로선:21	표지판:14	구름:3
	표지판:14	도로선:12	검은점:1
	구름:4	구름:8	표지판:1
	길:1	검은점:1	
		아스팔트:1	
		왼쪽막대:1	

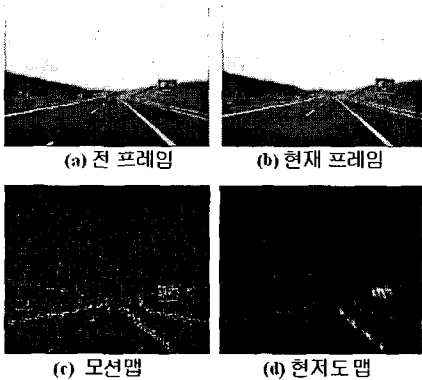


그림 10. 제안된 시스템에 입력된 “도로” 영상 시퀀스와 이에 대한 시스템의 결과 에

보이는 바와 같이 제안된 주의 시스템에서도 가장 현저한 영역이라고 탐지되었다.

“도로(Highway)” 영상 시퀀스의 경우, 인간을 대상으로 한 심리학적 실험에서는 도로와 교통표지가 가장 현저한 영역이라고 마크되었다(87.5%, [표 5] 참조). 역시 [그림 10]에서 보이는 바와 같이 제안된 주의 시스템에서도 같은 영역이 탐지되었다.

비록 우리가 인간을 대상으로 한 심리학적 실험에 사용한 3가지 영상 시퀀스가 모두 색상이 함유된 컬러 영상이었지만, 사람들은 일반적으로 움직이는 객체 또는 영역에 보다 더 주의를 기울이고 있다. 이와 같이 모션은 선택적 주의를 일으키는데 가장 강한 영향을 주는 특징 중 하나이다[6][13].

V. 결론

본 논문에서는 동영상을 위한 모션정보에 기반한 선택적 주의 시스템이 제안되었다.

제안된 시스템은 모션자극이 주어지면 이 자극을 시간특징 중 하나로 분석한다. 시스템에서 사용된 모션자극의 분석과 통합은 기존의 심리학적 연구결과를 바탕으로 하고 있다. 또한 모션분석을 통해 얻어진 결과를 통합하기 위한 방법으로써 기존의 전형적인 특징통합 모형에서 제시된 통합 방법을 사용하였다. 비록 제안된

시스템이 인간의 선택적 주의의 특정 부분을 모방한 것이기는 하지만, 아직은 보완해야 할 점이 많다. 그 이유에는 무엇보다도 우리에게 알려진 인간의 인지 메커니즘에 대한 지식의 양이 너무나도 적다는 것이고, 또 다른 이유로는 입력된 정보를 처리하는 과정에서 추출된 특징맵들을 하나의 현재도 맵으로 통합되는 과정이 결코 수월한 작업이 아니라는 것이다. 통합에 있어 단순히 모든 특징맵을 합하는 것은 신뢰성 있는 ROI를 탐지하지 못하는 경우가 많다고 이미 많은 연구들로부터 보고가 되고 있기는 하지만 이러한 연구들은 단순히 공간특정만을 통합하는 것을 목적으로 한 연구들이다.

제안된 시스템을 통해 얻은 결과는 인간을 대상으로 한 심리학적 실험을 통해 나온 결과와 대부분 일치한다. 본 시스템의 좀 더 정확한 ROI 탐지능력은 어떤 방법으로든 향후 더욱 개선될 것으로 기대하고 있다. 한 예로 기존의 심리학적 연구결과에 기반을 두어 시공간특징을 통합하는 방법[13]을 사용하여 성능을 개선할 수 있을 것이다. 향후 더욱 다양한 영상 시퀀스를 대상으로 실험을 수행하여 제안된 시스템의 성능을 평가해야 할 것이다.

참고 문헌

- [1] L. Itti, C. Koch, and E. Niebur, "A Model of saliency-based visual attention for rapid scene analysis," *IEEE Trans.on PAMI*, Vol.20, No.11, pp.1254-1259, 1998.
- [2] L. Itti, "Models of Bottom-Up Attention and Saliency," *Neurobiology of Attention*, San Diego, CA:Elsevier, pp.576-582, 2005.
- [3] V. Navalpakkam, M. Arbib, and L. Itti, "Attention and Scene Understanding," *Neurobiology of Attention*, San Diego, CA:Elsevier, pp.197-203, 2005.
- [4] M. Park, K. Cheoi, and T. Hamamoto, "A Smart Image Sensor with Attention Modules," *Proc. of Computer Architecture for Machine Perception*, pp.46-51, 2005.

[5] S. Olivier, K. Yasuo, and C. Gordon, "Development of a Biologically Inspired Real-Time Visual Attention System," Lecture Notes in Computer Science, Vol.1811, pp.150-159, 2000.

[8] A. Treisman and G. Gelade, "A Feature-integration Theory of Attention," Cognitive Psychology, Vol.12, No.1, pp.97-136, 1980.

[9] S. Shioiri, "Visual Psychophysics (11) : Attention and Visual Search," Journal of Image Information and Television Engineers, Vol.58, No.11, pp.1619-1625, 2004.

[10] S. Yano, "Human Sensations from Physiological and Psychological Aspects: Human Visual Perception and System," Journal of Image Information and Television Engineers, Vol.57, No.11, pp.1408-1413, 2003.

[11] <http://web.psych.uaberta.ca/~iwinship/visio n/visualpathways.html>

[12] R. Anderson, L. Snyder, D. Bradley, and J. Xing, "Multimodal Representation of Space in the Posterior Parietal Cortex and Its Use in Planning Movements," Annual Review Neuroscience, Vol.20, pp.303-330, 1997.

[13] 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.

[15] W. Osberger and A. Maeder, "Automatic identification of Perceptually important regions in an image," Proceedings of Fourteenth Intl. Conf. On Pattern Recognition, No.1, pp.701-704, 1998.

[16] E. Adelson and J. Bergen, "Spatiotemporal Energy Models for the Perception of Motion," Journal of Optical Society of America, Vol.2, No.2, pp.284-299, 1985.

저 자 소개

박 민 철(Min-Chul Park)

정회원



- 1993년 2월 : 홍익대학교 전자공학과(공학사)
- 1997년 3월 : 일본 동경대학 전자정보공학과(공학석사)
- 2000년 4월 : 일본 동경대학 전자정보공학과(공학박사)

• 2001년 6월~현재 : 한국과학기술연구원 시스템 연구부 선임연구원

<관심분야> : 내용기반 영상검색, 멀티미디어, 3차원 영상 디스플레이, 컴퓨터비전

최 경 주(Kyung-Joo Cheoi)

정회원



- 1996년 2월 : 충북대학교 컴퓨터과학과(공학사)
- 1999년 2월 : 연세대학교 컴퓨터과학과(공학석사)
- 2002년 8월 : 연세대학교 컴퓨터과학·산업시스템공학과(공학박사)

• 2002년 7월~2005년 2월 : LG CNS 연구개발센터

• 2005년 3월~현재 : 충북대학교 전기전자컴퓨터공학부 전임강사

<관심분야> : 컴퓨터비전, 영상처리, 뇌과학, 바이오컴퓨팅, 유비쿼터스컴퓨팅