

선택적 주의집중 인공시각 시스템

이민호 (경북대학교)

I. 서론

인간 시각 시스템은 무의식적 작용이나 의식적 작용을 통해 주변에 비해 상대적으로 두드러짐이 큰 영역이나 선호하거나 목적하는 영역 또는 대상으로 시선을 이동하는 선택적 주의 집중 기능을 갖는다. 이러한 선택적 주의 집중 기능은 복잡도가 크고 용량이 큰 정보의 선택적 처리를 가능하게 하여, 대용량 정보처리가 필요한 영상기반 감시시스템 및 인공비서 로봇의 인공시각 시스템 등 다양한 시각정보처리 시스템에 적용되어 효율적 정보처리 기능을 제공해 준다.

인간 시각 시스템의 선택적 주의 집중을 유발하는 요소는 다양하다. 하나의 입력 영상에서 시각 정보의 분포 특성에 의해 결정될 수도 있고, 연속 입력 영상에서 시각정보의 동적 특성의 변화량에 의해 결정될 수도 있다. 한편, 원근의 차이도 시선을 결정하는데 영향을 줄 수 있다. 이와 같이 인간시각시스템이 시각정보의 특성에만 기반하여 무의식적으로 시선을 결정하는 것을 상향식 시선결정이라고 한다. 반면, 목표하는 대상이나 선호하는 대상이 있거나 감정상태에 따라 시선을 결정하는 하향식 시선결정기능도 있다.

시선을 선택적으로 결정하는 인간 시각의 선택적주의집중 기능을 갖는 인공 시각 시스템 개발과 관련된 기존 연구는 다음과 같다. Itti 등은 인간 시각시스

템을 모방하여 시각 주의집중을 위한 돌출맵(Saliency Map: SM) 모델을 개발했으며^[1], Ramstrom과 Christensen은 다중 해상도의 피라미드와 다중 특징을 사용하여 처리함으로써 돌출맵을 계산하였고^[2], Koike 등은 확률적인 승자독식(winner take all: WTA) 방법을 기반으로 주의집중의 전환 모델을 개발했다^[3]. Kadir 등은 여러 돌출 요소들을 통합한 주의집중 모델을 제안하였으며 주의집중 된 영역내의 객체 크기를 추정하기 위해 엔트로피 정보를 사용하였다^[4]. Perona 등은 상향식 돌출맵 및 객체 이해 기반의 하향식 바이어스를 통해 객체중심의 주의집중 모델을 개발하였다^[5]. Saiz-Valverde 등은 영상내의 움직임 및 깊이정보를 통합하여 동적 양안기반의 시각 주의집중 모델을 개발하였으며^[6], Pirri 등은 깊이 정보를 고려하여 생물학 기반의 로봇 주의집중 모델을 개발하였다^[7]. 또한, 특정 영역 및 객체의 움직임을 해석하기 위한 모델도 Fukushima에 의해 개발 되었다^[8].

지금까지 제안된 대부분의 주의집중 모델에서는 특징 통합 이론(feature integration theory)을 기반으로 여러 특징들의 통합을 통한 시선결정 기능을 갖는 모델을 제안하였지만^[9], 특징간의 통합에 대한 명확한 메커니즘은 아직까지 불확실한 상황이다. 이와 관련해 인간 대뇌의 시각 피질에서는 시각 정보의 중복도 감소에 의한 시각 특징 추출 및 인지 작용이 일어난다는 Barlow의 가설^[10]에 기반하여 독립성분분석(Independent Component Analysis: ICA)기법을 이용하여 다양한 시각 특징간의 중복도를 줄이고 또한, 효과적으로 통합하는 과정을 통해 시선을 결정하는 연구도 진행되었다^{[11],[12]}.

이 논문에서는 무의식적 작용에 의한 상향식 시선결정과 의식적 작용에 의한 하향식 시선결정 기능을 통합 구현한 양안 시선결정 모델을 소개한다. 한편, 개발한 시선결정모델의 선택적 주의 집중 기능을 기반으로 하는 얼굴인식, 얼굴표정인식, 일반객체 인식, 시각 환경인식 등을 위한 다양한 시각 인지 모델들을 개발하였다. 개발한 양안기반의 시각자극을 통한 상향식 모델과 사람의 선호도를 기반으로 한 하향식 주의집중 통합 모델 및 시각 인지 모델들은 생물학적 시각 정보처리 메커니즘에 대한 이해 및 공학적 모방을 통해 구현되었다. 감각정보처리에 중요한 역할을 담당하는 인간 대뇌 피질중 시각 피질은 크게 'what-path'(객체 인지 등 '무엇'에 해당하는 정보를 해석하는 시각영역)와 'where-path'(공간 인지 등 '어디'에 해당하는 정보를 해석하는 시각영역)로 구분된다^[13]. 본 논문에서는 이러한 시각피질의 'what-path'와 'where-path'의 일부 기능의 모방을 통해 개발된 선택적주의 집중 모델 및 시각 인지 모델들을 기반으로 사용자에 따른 대응적 서비스를 제공해 줄 수 있는, 환

경 적응형 능동시각기능을 갖는 로봇두뇌 시스템을 개발하였다. 개발 시스템은 실시간 시스템으로 구현되어 업무도우미(Office-mate) 시스템으로 응용되었으며, 다양한 실험을 통해 개발 시스템의 여러 기능에 대한 성능 검증을 수행하였다.

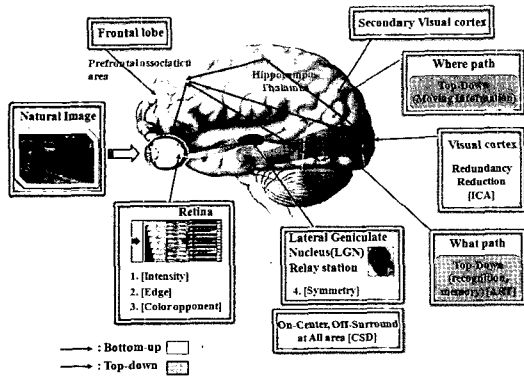
본 논문의 2장에서는 제안한 시각 주의집중 모델을 설명하고, 3장에서는 하드웨어 구현에 대해 설명한다. 4장에서는 실제 자연영상을 통한 실험 및 하드웨어 시스템의 응용 예를 통해 알고리즘의 성능 및 응용 가능성에 대해 나타내고 5장에서 결론을 맺는다.

II. 생물학 기반 선택적 주의집중 모델

선택적 주의집중 모델은 인간의 시각 구조 및 처리의 일부 과정을 모델링한 것으로 입력된 영상에 대하여 즉각적으로 반응하는 자료 주도적 처리 과정과 사전에 학습된 정보를 이용하여 개념 주도적 처리 과정으로 나누어진다.

인간시각시스템은 이러한 자료 주도적 처리과정과 개념 주도적 처리과정의 병행적 수행을 통한 선택적 주의 집중 기능을 수행한다. <그림 1>은 인간의 생물학적인 시각 처리 메커니즘을 나타낸다. 생물학적 시각 처리에서는 망막으로부터 입력된 시각 정보가 외측슬상핵(Lateral Geniculate Nucleus: LGN)을 거쳐 시각피질의 'what-path'와 'where-path'를 통해 다양한 시각인지처리가 이루어지게 된다. 보다 고차원적인 시각 인지 추론 기능 등은 전두엽(frontal cortex)에서 시각 피질과의 상호작용을 통해 이루어진다.

<그림 2>는 인간의 생물학적인 시각 처리 메



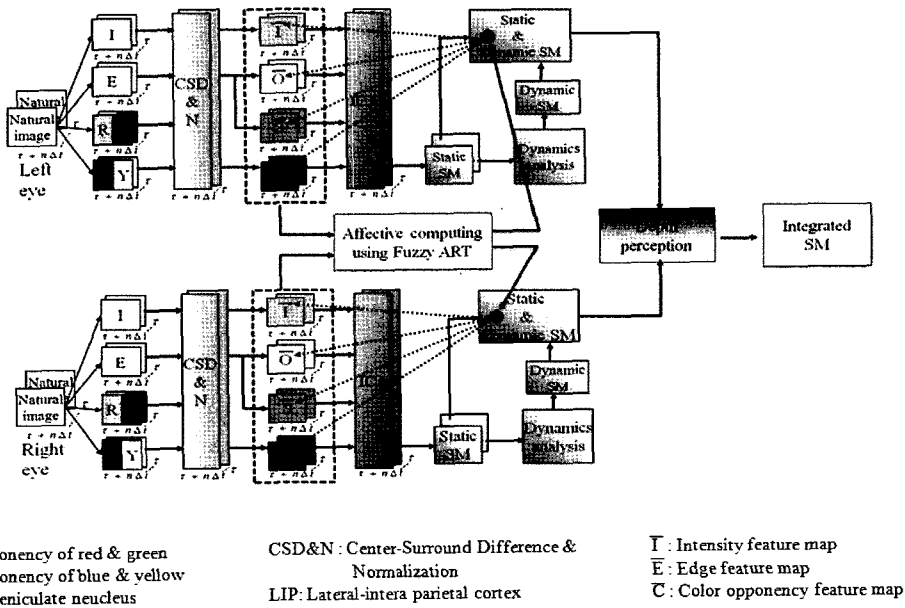
〈그림 1〉 생물학적 시각 처리 경로

커니즘을 모델링하여, 눈의 망막(retina)에서부터 외측슬상핵을 지나 시각피질(visual cortex)까지의 상향식 시각 메커니즘 및 하향식 메커니즘이 일어나는 시상배계(pulvinar) 영역까지를 표현한 그림이다^[14]. 양안 기반의 선택적 주의집중 모델은 각각의 눈에 서로 다른 영상을 입력으로 각 영상에서의 시각 자극을 통해 4가지 특징 정보가 추출되며 독립성분해석을 통해 각 특징

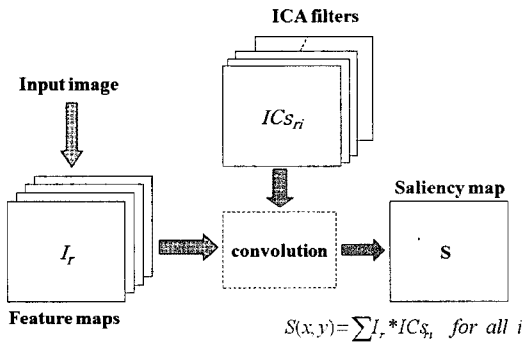
정보가 통합되어 정적 돌출맵이 생성된다. 특정 정보로는 명암, 윤곽선, 색상 정보뿐만 아니라 대칭성 정보를 적용한다. 생성된 정적 돌출맵을 기반으로 시간 축에서의 엔트로피 해석을 통해 동적 돌출맵을 생성하고 사람의 감정 및 선호도, 양안 기반의 깊이정보를 토대로 정적/동적 돌출맵과의 통합을 통해 최종 돌출맵을 생성한다.

1. 정적 돌출맵 기반 선택적 주의 집중 모델

Treisman의 특징통합 이론을 기반으로 Itti 등은 3가지 기저 특징맵으로 명암, 윤곽선, 색상 정보를 이용하여 시선 결정을 위한 상향식 돌출맵을 구현하였다^{[11],[9]}. 하지만 이 논문에서 제안한 모델에서는 인간 시각시스템에서 객체에 대한 중요한 특징정보 중 하나인 대칭성정보를 추가하였으며, 또한 효과적인 특징맵 통합을 위해



〈그림 2〉 생물학 기반 시선결정 선택적 주의집중 모델



<그림 3> ICA 필터를 이용한 통합 돌출맵

독립성분분석(ICA) 필터를 추가하였다 [11],[12].

대칭성 정보는 객체의 검출 및 분할을 위해 효율적인 정보로 알려져 있으며 [15-17], 이에 초기 시각피질 영역으로부터 획득되는 대칭축을 기반으로 대칭성정보를 추출하는 방법과 주의집중 모델과의 통합을 통해 효과적인 돌출맵을 구현하였다 [18]. 따라서 <그림 2>에서 보는 것과 같이 4개의 특징맵과 ICA 필터를 사용하여 효율적으로 특징맵들을 통합하였다.

<그림 3>은 ICA 필터 적용을 통한 특징맵 통합 및 이에 따른 돌출맵 생성방법을 보이고 있다. 입력 영상으로부터 획득한 4개의 특징맵과 다양한 자연영상들로부터 획득한 ICA 필터들의 컨벌루션에 의해 돌출맵이 형성된다.

2. 동적 돌출맵 기반 선택적 주의 집중 모델

동적 돌출맵은 연속된 정적 돌출맵을 기반으로 엔트로피 정보를 이용하여 구현하였다 [19]. 이는 Kadir 등이 제안한 방법과 유사한 접근 방법으로 공간상에서의 엔트로피 정보가 아닌 정적 돌출맵의 시간축 해석을 통해 엔트로피가 최대가 되는 영역을 주의집중 하도록 구현한 것으로 다음

식(1)을 통해 시간축 엔트로피를 구할 수 있다.

$$X_s = \operatorname{argmax} \{ H_D(s, X, \tau) \times W \} \quad (1)$$

식 (1)에서 τ 는 현재 시점, s 는 영역의 크기, X_s 는 시간축으로 엔트로피가 최대가 되는 공간상의 영역이다. H 는 엔트로피를 나타내며, W 는 공간상에서 서로 다른 크기 영역간의 특징 분포의 변화 정도를 나타내고 있으며 식 (2)와 (3)에 의해 구해진다.

$$H_D(s, X, \tau) \equiv - \sum_{d \in D} P_{d,s,X,\tau} \log_2 P_{d,s,X,\tau} \quad (2)$$

$$W_D(s, X, \tau) \equiv \frac{s^2}{2s-1} \sum_{d \in D} |P_{d,s,X,\tau} - P_{d,s-1,X,\tau}| \quad (3)$$

식 (2)의 P 는 확률밀도함수를 나타내며, D 는 X_s 영역내의 명암 히스토그램의 데이터 값들의 집합을 나타낸다. 확률밀도함수는 히스토그램의 엔트로피값으로 구해진다. 식 (3)의 s 는 설정된 영역의 크기를 나타내고 있다. 최종적으로 동적 돌출맵의 정보는 식 (4)에 의해 구해진다.

$$T_D(X_s, X, \tau + n\Delta t) \equiv - \sum_{d \in D} PH_{D(X_s, X, \tau + n\Delta t)} \log_2 PH_{D(X_s, X, \tau + n\Delta t)} \quad (4)$$

이와 같이 구해진 동적 돌출맵과 정적 돌출맵의 통합으로 시선결정을 위한 최종 돌출맵이 구해진다.

3. 감정 선호도를 반영한 돌출맵 모델

선택적 주의 집중 모델은 상향식 시선 결정 기능뿐만 아니라 사람의 감정 및 선호도를 고려한

하향식 시선 결정기능도 사람의 주의집중을 위한 중요한 요소로 작용된다. 이에 상향식 돌출맵에 의해 결정된 주의집중 영역이 사람의 감정 및 선호도에 따라 주의집중이 더욱 강화되거나 또는 억제되는 현상을 모델링하여 알고리즘을 개발하였다. Lee 등은 Hebbian 학습법에 따라 사람의 선호도를 고려하여 주의집중 영역의 강화 및 억제 기능을 갖고 주의집중을 할 수 있는 모델을 개발하였다^[20]. 즉, 제안한 모델은 사람이 감독자 학습을 통해 선호하거나 선호하지 않는 클래스를 분류하고 기억할 수 있도록 하였으며, 이에 따라 입력 영상에 대해 상향식 돌출맵 및 선호도에 따라 주의집중 영역을 결정할 수 있는 것이다. 이와 같은 모델을 구현하기 위해 점진적 추가 학습이 가능한 Fuzzy adaptive resonance theory (ART) 모델을 사용하였으며^[21], 이 모델을 통해 특정 영역에 대한 선호도를 기억 학습하였다.

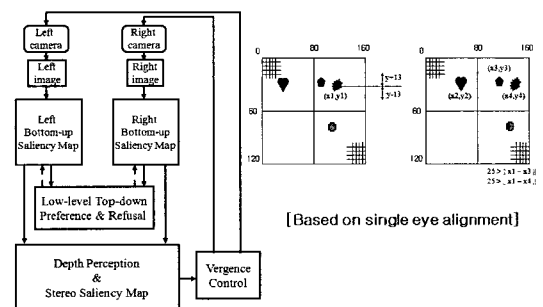
4. 양안 기반 돌출맵 모델

인간의 시각 시스템은 보는 영상에 대하여 양안 수렴을 거친 후 깊이 정보 또는 거리정보를 추출하는 시각 메커니즘을 가지고 있다. 이에 제안하는 모델에서는 정확한 양안 수렴 모델을 구성하기 위해서 단안 정렬을 우선 고려하였다^[12]. 제안하는 모델은 인간의 single eye dominant 가설을 기반으로 양안 수렴을 위한 주된 눈을 찾기 위해 각 카메라 영상에 돌출맵을 적용하고 주된 눈과 돌출영역을 구한 후 단안 정렬을 한다^[12]. 또한 부 카메라를 주 카메라에 대해 단안 정렬함으로써 양안 정렬을 이루게 된다. 제안된 모델에서는 돌출맵을 이용하여 얻어진 각각의 돌출영역 에너지 값을 비교하여 큰 값을 가지는 영역을 포함한 카메라를 주 카

메라로 결정하고 나머지 카메라를 주 카메라의 돌출점에 정렬되기 위한 부 카메라로 결정한다.

주/부 카메라의 정렬을 위해 각 카메라의 돌출영역 간의 유사도를 비교하는 연산량을 최소화하기 위해서 주 카메라의 돌출점과 부 카메라의 연속적인 돌출점들의 각 x, y축 좌표차를 이용한다. 이는 양안 카메라에서의 돌출점들이 특정 영역에 유사하게 나타나는 특성을 기반으로 돌출점들의 발생가능 영역을 한정하는 것이다. 즉, 두 카메라사이의 돌출점 위치가 일정 거리를 가지는 범위 안에 존재할 경우 후보 돌출영역으로 결정되는 것을 뜻한다.

최종적으로 양안의 돌출 점들이 같은 영역임을 확인하기 위해 각 돌출 영역의 에너지 값을 비교한다. 이 때 주 카메라의 돌출영역과 부 카메라의 돌출영역이 가지는 각각의 에너지 값들의 유사도를 보고 가장 유사도가 높은 돌출점을 선택하여 양안 수렴을 이루는 최종 특이점으로 간주한다. 또한 이러한 영역 비교 과정에서 양안 카메라의 폐색 영역으로 인해 나타나지 않는 부분은 후보 돌출영역의 조건에 의해 배제되며, 계속적으로 다음 돌출영역의 값을 비교함으로써 특이점을 결정할 수 있다. <그림 4>는 양안 수렴 모델에 대한 전체 구조를 보여준다.



<그림 4> 양안 수렴 모델

III. 선택적 주의집중 기반 업무도우미 시스템을 위한 응용 알고리즘 모듈

선택적 주의 집중 모델에 의한 시선 결정 기능은 인간의 시각 시스템처럼 ‘어디’를 볼 지를 스스로 결정할 수 있게 해 준다. 하지만, 좀 더 고차적으로 ‘무엇’을 보고 있는지를 인지할 수 있는 시각 시스템 개발을 위해서는 얼굴 인식, 물체인식, 환경 인식 등의 시각 인지 기능이 중요하다. 이에 본 논문에서는 제안하는 시각 시스템이 고차적인 실세계 응용시스템으로 활용이 가능할 수 있도록 하기 위해 다양한 시각 인지 모듈을 개발하였다.

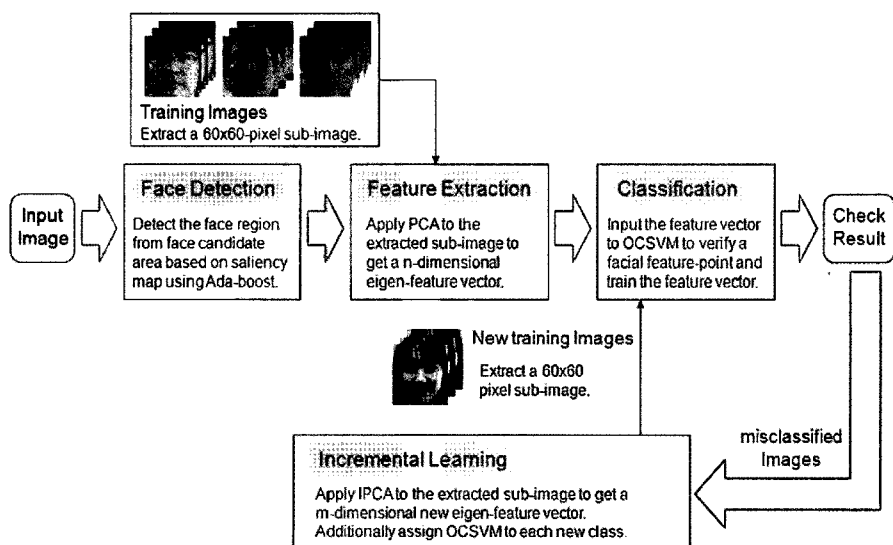
1. 점진적 얼굴 인식 모듈

본 논문에서 제안하는 얼굴인식 시스템은 선택적 주의집중 기반의 얼굴검출 및 얼굴 인식 모듈과 점진적 학습 모듈로 구성되어 있다. 이러한 모듈은 감독자의 참여 없이 실시간으로 온라인 학습을 통

해 스스로 기억/학습이 가능하다.

검출된 얼굴 영역은 추축성분분석(Principal Component Analysis: PCA) 기법 및 점진적 주축성분분석(Incremental PCA: IPCA) 기법으로 추출된 고유 얼굴(Eigen-faces)를 이용하여 얼굴 영역을 표현하는 특징벡터를 구하고^[22], 최적의 클래스 경계면을 구성할 수 있는 알고리즘인 Support Vector Machine (SVM) 중 단일 클래스에 대한 최적의 임계값 결정이 가능한 One class SVM을 이용하여 얼굴데이터 모델을 구한다^{[23],[24]}.

PCA 및 IPCA에 의해 추출된 영상은 얼굴 영상이 가지는 고유벡터(Eigen-vector)들로 구성되어 있으며 이러한 고유벡터들에 의해 데이터 양은 줄어든 반면 얼굴패턴의 특징은 그대로 보존되기 때문에 얼굴인식에 자주 사용되고 있는 방법이다. 하지만 PCA는 오프라인 학습법으로 모든 데이터를 획득한 후 고유벡터를 구한다는 단점이 있지만, IPCA의 경우 모든 데이터를 가지고 있지 않더라도 새롭게 입력되어진 데이터를 가장 잘 표현할 수 있는 고유벡터를 추가하거나 기존의 고유벡터를



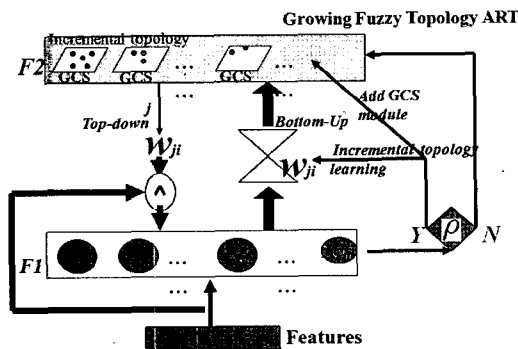
〈그림 5〉 점진적 얼굴 인식 모듈

회전시킴으로써 점진적인 학습이 가능하다. 이와 같이 추출된 고유벡터는 점진적으로 데이터베이스로 저장되거나 또는 기존에 학습된 데이터의 고유벡터와 유사성을 비교해 분류 인식하였다. <그림 5>는 제안된 시스템의 전반적인 구조를 보이고 있다.

2. 선택적 주의 집중 기반 점진적 객체 인지 모델

인간의 시각 시스템은 시각 피질과 IT (infero-temporal), 전두엽 등에서 선택적 주의 집중과 물체 인식, 지식 획득, 추론 처리 등을 체계적 협동을 통해 복잡한 시각 정보를 처리한다. 제안된 점진적 객체 인지 모델은 인간의 두뇌 기능 중 'what path'의 이해를 기반으로 한다.

우리는 점진적인 객체 검출 및 인지를 위하여 새로운 형태의 Growing Fuzzy Topology ART (GFTART) 네트워크를 설계 하였다^[25]. GFTART는 기존의 fuzzy ART 모델의 F2 층의 노드들이 각각 Growing Cell Structures(GCS) 신경망 모듈로 대체되어 기존의 fuzzy ART가 가지는 과다 노드 생성등 노드 생성에 관한 문제점을 보완하는 모델이다^[25].



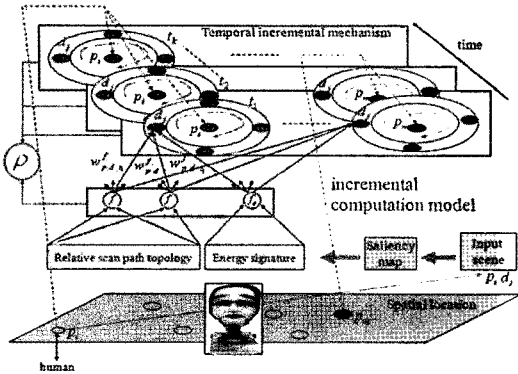
<그림 6> 점진적 객체 인지를 위한 GFTAT

<그림 6>은 점진적 객체 검출 및 인지 기능의 GFTART 구조를 보이고 있다. GFTART는 물체를 표현하기 위한 색상과 형태를 입력으로 이를 클래스화하는 맵을 형성한다. 클래스화된 정보는 특정한 물체의 묘사에 적절할 뿐만 아니라 이 정보를 이용하여 배우지 못한 물체에 대해 스스로 추정하여 새로운 클래스를 만들 수 있다.

3. 적응자율지능 성장을 위한 환경인지 모듈

생물학적 메커니즘을 기반으로 입력영상에 대한 돌출맵을 생성하는 시각모델을 개발했지만 돌출맵의 각 지역적인 특성을 기억해 주어진 환경을 인지하는 것은 효율성이 떨어진다. 따라서 제안한 모델에서는 구해진 돌출맵 정보로부터 얻어지는 돌출영역들의 topology를 기반으로 주어진 환경을 기억할 수 있도록 하였다. 더욱이 돌출맵으로부터 구한 에너지 정보는 장면 변화의 유무를 판단하는 추가적인 요소가 된다. 획득한 돌출영역의 topology와 에너지 정보를 기반으로 환경변화에 대한 전반적인 지식을 점진적으로 기억하게 된다^[26].

<그림 7>은 시공간차원에서의 영상 장면분석을 통해 환경인지를 하는 모델을 보여준다. 먼저 기억된 장면과 새롭게 획득된 입력영상과의 비교를 통해 장면 변화를 감지하게 되며 이때 사용되는 정보는 기억된 장면 정보와 새로운 입력영상과의 유사도이다. 유사도를 측정 한 후 유사도에 의해서 새로운 장면변화에 대한 결정을 하게 된다. 그리하여 모델은 기억되어 있는 장면정보와 새로 생성된 장면정보와의 비교를 통해서 가장 유사한 정보를 통해 해당 장면의 특정 공간에서의 위치 정보를 획득할 수 있게 된다. 즉, 기억된 장면 정보를 이용하여 시각 환경의 변화를 인지할 뿐 만 아니라 장면



〈그림 7〉 적응형 환경인지 모델

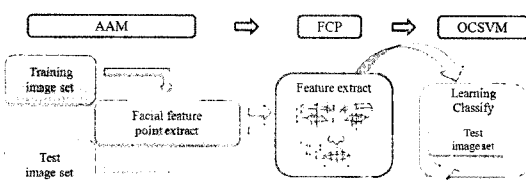
정보를 바탕으로 공간상의 위치 정보를 스스로 인지할 수 있는 기능을 가지게 되는 것이다.

4. 얼굴 표정 기반 감정 인식 모듈

얼굴 표정 특징 정보는 인간의 감정 상태에 대한 정보를 제공한다. 이러한 얼굴 특징 정보를 이용하여 사람의 감정을 분석하기 위해 Active Appearance Model(AAM)을 이용하여, 얼굴의 모양 및 특징에 관한 정보를 추출한다^[27].

검출된 특징점들을 얼굴 표정의 변화에 따라 가장 변화 작은 점들을 기준으로 가장 변화가 큰 점들과의 관계를 벡터로 나타내어 특징벡터를 생성하고 OCSVM을 통해 최적의 분류기를 설계해 5 가지 얼굴 표정을 인식한다.

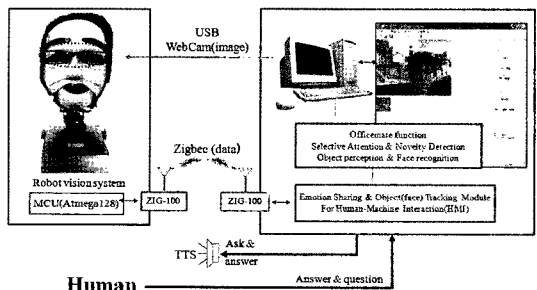
〈그림 8〉은 얼굴 표정 기반 감정인식 모듈의 처리 과정을 보이고 있다.



〈그림 8〉 얼굴 표정 기반 감정 인식 모듈

IV. 선택적 주의집중 및 사물 인지 기능을 포함한 업무도우미 (Office-mate)

〈그림 9〉는 실시간 선택적 주의 집중 모델 및 사물 인식 등 시각 인지 기능을 포함한 업무 도우미 시스템인 SMART-v2.0(Self Motivated Artificial Robot with a Trainable selective attention model version 2.0)의 구조를 보여주고 있다^[28]. 구현된 SMART-v2.0 시스템은 실시간 선택적 주의 집중 모델을 통해 현재 주어진 각도의 입력 장면에 대하여 주변보다 돌출적인 영역에 대해 스스로 주의 집중을 결정할 수 있을 뿐만 아니라 선택적으로 얼굴이나 특정객체에 시선을 결정할 수 있는 인식 모듈이 내장되어 있다. 이러한 알고리즘을 수행하기 위해 320x240 해상도의 영상을 초당 30프레임 이상 획득 가능한 USB 카메라 2개를 장착하였다. 또한 카메라 위치 제어 및 로봇감정 표정을 제어하기 위해서 4개의 RC-서보 모터와 3개의 DC 모터를 사용하였다. SMART-v2.0은 7개의 모터 제어기로 ATmega 128을 사용하였으며, PC와 통신을 하기 위하여, zigbee 무선통신 모듈을 사용하였다. 자세한 기능 및 사양은 실험결과에 나타내었다.



〈그림 9〉 SMART-v2.0 하드웨어 구조

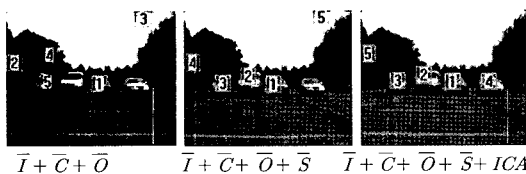
V. 실험 결과

<표 1>은 주어진 영상 내 특징맵을 이용하여 다양한 통합 방법에 따른 주의집중 실험 결과를 나타내고 있다. <표 1>에서와 같이 대칭성 정보가 객체를 검출하는 데 좋은 요소임을 보여주고 있으며, 구해진 특징맵들 간의 통합을 위해 중복도를 제거하여 각 특징맵들을 효율적으로 통합할 수 있는 독립성분분석 기법을 사용하는 것이 가장 높은 성능을 보임을 알 수 있다.

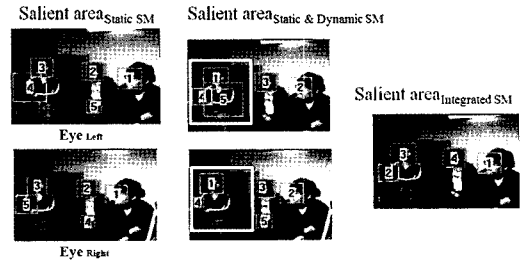
<그림 10>은 <표 1>에 나타난 것과 같이 3가지 특징맵 통합에 따른 주의집중 영역 선택결과를 나타내고 있다. <그림 10>에서와 같이 3가지 특징맵이 통합하는 경우 원하는 객체에 주의집중이 잘 일어나지 않는 것을 볼 수 있다. 하지만 대칭성 정보를 추가했을 경우 객체에 조금 더 주의집중 하는 것을 볼 수 있으며 독립성분분석을 통해 중복도를 제거했을 경우 모두 객체에 주의

<표 1> 특징맵 통합 방법의 주의 집중 결과

객체포함 돌출영역	$\bar{I} + \bar{C} + \bar{O}^{(1)}$	$\bar{I} + \bar{C} + \bar{O} + \bar{S}$	$\bar{I} + \bar{C} + \bar{O} + \bar{S} + ICA$
1번째	143	150	165
2번째	103	104	106
3번째	64	77	68
4번째	47	57	66
5번째	28	32	46
총 영역	385	420	451
객체포함 영역검출 성능	77 %	84 %	90 %



<그림 10> 특징맵 통합방법의 주의집중 결과

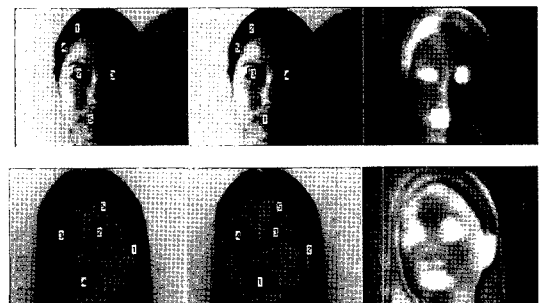


<그림 11> 통합 돌출맵 모델의 주의집중결과

집중이 일어나는 것을 볼 수 있다.

<그림 11>은 정적인 돌출맵과 동적인 돌출맵의 통합을 통한 실험 결과를 보여주고 있다. 영상에서 가장 왼쪽 사람이 연속된 영상프레임에서 왼쪽에서 오른쪽으로 조금 움직이고 있으며 다른 두 명은 움직임이 없었고 가장 오른쪽에 있는 사람이 깊이정보가 가장 짧게 나타났다. 이에 정적인 돌출맵에서는 가장 오른쪽 사람이 첫 번째로 주의집중이 일어났으며, 동적 돌출맵과 통합 돌출맵에서는 가장 왼쪽 사람이 첫 번째로 주의집중이 일어났다. 최종적으로 주의 집중된 위치에서의 깊이정보를 반영하여 가장 오른쪽에 있는 사람이 첫 번째로 주의집중이 일어났으며 가장 왼쪽에 앉아 있는 사람에게 2, 3번째 주의 집중이, 가운데 사람에게 4번째 주의집중이 일어났다.

<그림 12>는 감정 및 선호도를 고려한 주의집중 실험 결과를 나타내고 있다. 가장 왼쪽 그림에

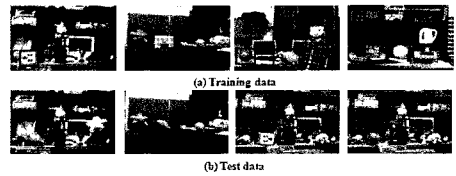


<그림 12> 감정, 선호도에 따른 입술검출 결과

서는 선호도를 고려하기 전의 돌출맵에 의해 생성된 주의집중 순서를 나타내고 있다. 얼굴 중 입술에 주의집중이 되도록 학습된 모델에 의한 주의 집중 결과는 가운데 위치한 그림과 같다. 가운데 위치한 그림에서는 입술의 주의집중 순서가 1 번으로 바뀐 것을 확인 할 수 있으며, 가장 오른쪽에 위치한 그림은 주의 강화 모델에 의해 생성된 돌출맵으로 입술부근에서 더욱 밝게 강화 학습이 되고 있는 것을 볼 수 있다. 이와 같이 사용자의 감정이나 선호도에 따라 돌출맵 영역에 강화/억제를 통해 상향식 및 하향식 통합 주의 집중에 의한 시선 결정 기능을 갖는다.

<표 2>는 감정 및 선호도를 기반으로 입술 검출 성능을 나타내는 것으로 총 96개의 학습 영상을 이용해 하향식 주의집중 모델을 학습하였으며, 90개의 시험 영상을 통해 모델의 성능을 검증하였다. <표 2>와 같이 상향식 주의집중 모델을 사용했을 경우 첫 번째 주의집중 순서로 입술이 선택된 것이 39%로 나타난 것에 비해 하향식 주의집중을 했을 경우 98%의 성능을 보여 하향식 모델에 의한 선호도에 따른 선택적 주의 집중이 성공적으로 수행됨을 확인할 수 있다.

<그림 13>은 점진적 객체인식의 성능을 보이고 있다. 5개 색상과 형태가 다른 객체의 정보를 40개의 영상을 통하여 학습 하였으며, 80장의 영상에서 120개의 객체에 대한 인식 성능을 나



<그림 13> GFTART를 이용한 객체인식 결과

타내고 있다. 5개 객체의 색상 인식을 평균은 85% 성능을 보이며, 형태 인식률은 평균 79%의 성능을 보이고 있다.

또한 <그림 14>는 상황변화에 따른 적응적 환경인지 모듈인 Novelty 검출 실험 결과로 91.8%의 변화감지 성능을 보이고 있다.

SMART-v2.0 시스템은 인간과의 상호작용 및 친밀감을 위하여 <그림 15>와 같이 다양한 상황에 대응해 감정 표현을 할 수 있도록 설계 하였다.

<그림 16>은 SMART-v2.0 시스템에 구현된 감정인식 모듈을 이용한 4가지 감정 인식 실험 결과를 보이고 있다.



	Novelty scenes (44 scenes)	Non-novelty scenes (17 scenes)	Total (61 scenes)
Novelty perception	TP 41	FP 2	F(P+N) 5
Non-novelty perception	FN 3	TN 15	T(P+N) 56
Correct Recognition Rate	95.3%	88.2%	91.8%

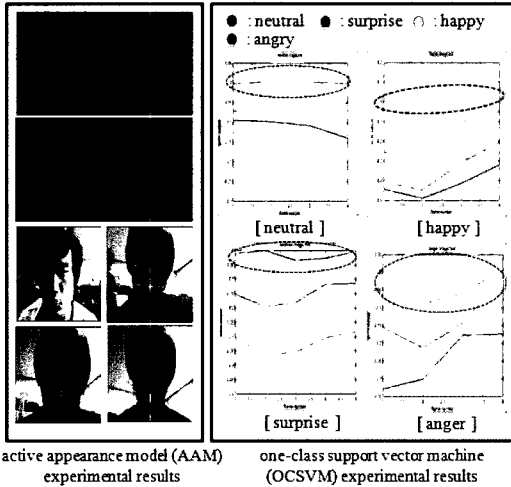
<그림 14> 선택적 주의집중 기반 환경인지 모듈 실험

<표 2> 감정 선호도 기반 입술 검출 성능

	학습 영상	90 Test images	
		상향식 돌출맵	감정 선호도 기반 돌출맵
1번째 돌출영역내 입술 포함 (개)	96	35	88
1번째 돌출영역내 입술 미포함 (개)	0	55	2
입술 검출 성능 (%)	100	39	98



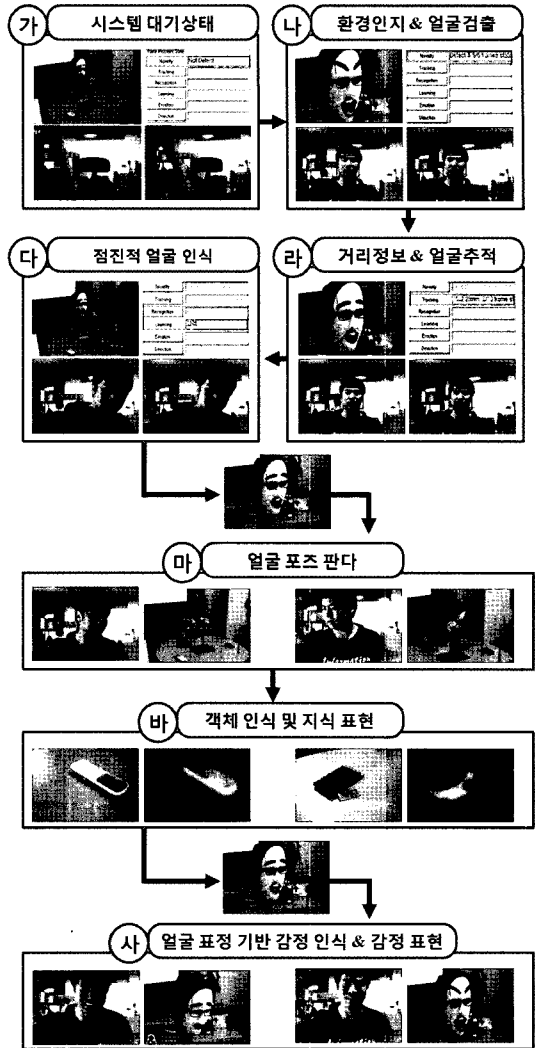
〈그림 15〉 SMART-v2.0 감정 표현



〈그림 16〉 4가지 감정 인식 실험 결과

〈그림 17〉은 방문자 서비스를 위한 업무도우미(Office-mate) 시스템인 SMART-v2.0에 내장 되어 있는, 인간-기계 상호작용 기반 업무도우미 시나리오의 흐름도를 나타내고 있다.

- 가) SMART-v2.0은 핸드폰 가게의 휴대전화 판매 도우미로, 현재 장소의 환경을 인지하고 있으며, 환경 변화를 체크 하고 있다.
- 나) SMART-v2.0은 가게에 손님이 들어오는 것을 스스로 인지한다. 이후, 손님 얼굴을 검출하며, SMART-v2.0은 손님에게 “안녕하세요” 라 말하며 손님을 반긴다.
- 다) 제안한 업무 도우미(Office-mate)는 손님 얼굴을 추적하며, 시스템에 내장되어 있는 양안을 이용하여, 손님과 떨어져 있



〈그림 17〉 업무도우미 데모 흐름도

- 는 거리 정보를 구한다.
- 라) SMART-v2.0은 앞서 검출된 얼굴 정보를 기반으로 얼굴 인식을 하여, 기존 방문한 손님인지, 새로운 손님인지 판단하여, 손님에 따른 적응적 서비스를 제공한다.
- 마) 손님의 얼굴 포즈 정보를 판단하는 모듈은, 손님이 관심 있어 하는 물건이 어떤 것인지, 손님의 의도를 판단 할 수 있다. 이후, SMART-v2.0은 손님의 얼굴 방향과 같

은 방향을 볼 수 있도록 개발 되었다.

- 바) SMART-v2.0은 손님이 관심 있어 하는 물건에 대해서, 객체에 대한 점진적 인식을 하여, 객체 정보를 손님에게 알려 준다.
- 사) SMART-v2.0은 손님의 얼굴표정 정보를 분석하여, 소개해준 휴대 전화에 대해 손님의 감정에 대해 좋아하는지/싫어하는지 판단한다.

VI. 결론

본 논문에서는 생물학적 시각정보처리 메커니즘을 모방한 시선결정 모델과 시각인지 모델을 기반으로 업무도우미 시스템에 응용되는 인공시각시스템을 개발하였다. 인간 시각의 선택적 주의 집중 처리를 통한 효율적 시각 정보처리 메커니즘에 대한 모방을 통해 구현한 상향식 선택적 주의집중모델과 하향식 선택 집중 모델의 통합 모델을 제안하였다. 제안하는 시선결정모델은 시각정보뿐 만 아니라, 시각 정보에 대한 사람의 의도 및 선호도를 반영하는 기능을 가진다. 한편, 선택적 주의 집중 기반 점진적 얼굴/객체 검출 및 인식 모델, 사용자 및 주변 환경과 상호작용을 위한 환경변화 인지를 위한 Novelty detection 모델과, 얼굴표정 기반 감정인식 모델 등 다양한 시각 인지 모델을 개발하였다. 개발한 시각 인지 모델은 시선결정모델과의 통합 구현을 통해 사무실 환경에서 업무 도우미 역할을 수행하는 실시간 인공시각플랫폼 SMART-v2.0을 개발하였다.

인공시각 플랫폼 SMART-v2.0은 시각 인지 측면에서 좀 더 자율적이며 점진적 지식 성장 기능 보완을 통해 앞으로 다양한 지능형시스템에

중요한 역할을 수행할 수 있을 것으로 기대된다.

VII. 감사의 글

이 논문은 산업자원부의 중장기산업기술개발 사업의 “뇌정보처리에 기반한 인공두뇌 기본모형 및 인공비서 개발(과제번호 10023913)” 및 교육과학기술부의 재원으로 한국연구재단-“신기술융합형 성장동력사업”의 지원을 받아 수행된 연구임(No. 2009-0082262).

참고문헌

- [1] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20(11), pp. 1254-1259, 1998.
- [2] O. Ramström, and H. I. Christensen. Visual attention using game theory. *Lecture Notes in Computer Science*, 2525, pp. 462-471, 2002.
- [3] T. Koike, and J. Saiki, Stochastic guided search model for search asymmetries in visual search tasks. *Lecture Notes in Computer Science*, 2525, pp.408-417, 2002.
- [4] T. Kadir, and M. Brady, Scale, saliency and image description. *International Journal of Computer Vision*, pp. 83-105, 2001.
- [5] D. Walther, U. Rutishauser, C. Koch, and P. Perona, Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Computer*

- Vision and Image Understanding, 100(1-2), pp. 41-63, 2005.
- [6] A. Fernández-Caballero, M. T. López, and S. Saiz-Valverde, Dynamic stereoscopic selective attention (DSSVA): Integrating motion and shape with depth in video segmentation. *Expert Systems with Applications*, 34, pp. 1394-1402, 2008.
- [7] A. Belardinelli, and F. Pirri, A biologically plausible robot attention model, based on space and time. *Cognitive Processing*, 7, pp. S11-S14, 2006.
- [8] K. Fukushima, Extraction of visual motion and optic flow. *Neural Networks*, 21(5), pp. 774-785, 2008.
- [9] A. M. Treisman, and G. Gelade, A feature integration theory of attention. *Cognitive Psychology*, 12(1), pp.97-136, 1980.
- [10] H. B. Barlow and D. J. Tolhurst, Why do you have edge detectors?, *Optical society of America Technical Digest*, 23, p. 172, 1992.
- [11] S.-J. Park, K.-H. An, and M. Lee, Saliency map model with adaptive masking based on independent component analysis. *Neurocomputing*, 49, pp. 417- 422, 2002.
- [12] S.-B. Choi, B.-S. Jung, S.-W. Ban, H. Niitsuma, and M. Lee, Biologically motivated vergence control system using human-like selective attention model. *Neurocomputing*, 69, pp. 537-558, 2006.
- [13] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the brain* (2nd ed.). Baltimore: Lippincott Williams & Wilkins, 2001
- [14] S. Jeong, S.-W. Ban, and M. Lee, Stereo Saliency Map Considering Affective Factors and Selective Motion Analysis in a Dynamic Environment, *Neural Networks*, 21, pp. 1420-1430, 2008.
- [15] A. M. Norcia, T. R. Candy, M. W. Pettet, V. Y. Vildavski, and C. W. Tyler, Temporal dynamics of the human response to symmetry. *Journal of Vision*, 2, pp.132-139, 2002.
- [16] D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context-free attentional operators: The generalized symmetry transform. *International Journal of Computer Vision*, 14, pp.119 -130, 1995.
- [17] J. Duncan, Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, pp.501-517, 1984.
- [18] K. Fukushima, Use of non-uniform spatial blur for image comparison: symmetry axis extraction. *Neural Networks*, 18, pp. 23-32, 2005.
- [19] S.-W. Ban, I. Lee, and M. Lee, Dynamic visual selective attention. *Neurocomputing*, 71(4-6), pp. 853-856, 2008.
- [20] S.-B. Choi, S.-W. Ban, and M. Lee, Human-like selective attention model with reinforcement and inhibition mechanism, *LNCS*, 3316, pp. 694-699, 2004.
- [21] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, *Fuzzy ARTMAP: A neural network*

architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks, 3(5), pp. 698-713, 1992.

- [22] S. Ozawa, S. L. Toh, S. Abe, S. Pang, and N. Kasabov, Incremental learning for online face redognition, Proceedings of International Joint Conference on Neural Networks, 18, pp. 575-584, 2005.
- [23] B.-H. Asa, H. David, T. S. Hava, and V. Vladimir, "Support Vector Clustering," Journal of Machine Learning Research, 2, pp. 125-137, 2001.
- [24] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a High-Dimensional Distribution," Microsoft Research Technical Report, MSR-TR-99-87, 1999.
- [25] B. Hwang, S.-W. Ban, and M. Lee, Top-down Object Preferable Attention Using Growing Fuzzy Topology ART. Proceedings of International Conference on Neural Information Processing, pp. 284-285, 2008.
- [26] S. W. Ban and M. Lee, Selective attention-based novelty scene detection in dynamic environments, Neurocomputing, Vol.69, No.13-15, pp.1723-1727, 2006.
- [27] A. U. Batur and M. H. Hayes, Adaptive active appearance models, IEEE Trans. Image Processing, 14, pp.1707-1721, 2005.
- [28] M. Lee, Y.-M. Jang, and S.-W. Ban, Office-mate: Selective attention and

incremental object perception, Proceedings of 2009 IEEE International Conference on Systems, Man, and Cybernetics, pp. 2212-2217, 2009.

저자소개



이 민 호

1988년 2월 경북대학교 전자공학과(공학사)
 1992년 2월 한국과학기술원 전기및전자공학과(공학석사)
 1995년 8월 한국과학기술원 전기및전자공학과(공학박사)
 1995년 9월~1998년 2월 한국해양대학교 전임강사/
 조교수
 1998년 3월~현재 경북대학교 전자전기컴퓨터학부 전임강
 사/조교수/부교수/교수
 2004년 10월~2005년 8월 대구 테크노파크 모바일단말상
 용화센터 실장(겸임)
 2006년 7월~2007년 7월 Dept. of Brain and
 Cognitive Science, MIT (방문교수)

주관심 분야 : 뇌신경정보학, 인공지능 및 주의집중, 인
 지 로봇 및 지능시스템