

사용자 행동 자세를 이용한 시각계 기반의 감정 인식 연구

김진옥*

요약

사람의 의도를 인지하기 위해 감정을 시각적으로 인식하는 연구는 전통적으로 감정을 드러내는 얼굴 표정을 인식하는 데 집중해 왔다. 최근에는 감정을 드러내는 신체 언어 즉 신체 행동과 자세를 통해 감정을 나타내는 방법에서 감정 인식의 새로운 가능성을 찾고 있다. 본 연구는 신경생리학의 시각계 처리 방법을 적용한 신경모델을 구축하여 행동에서 기본 감정 의도를 인식하는 방법을 제안한다. 이를 위해 시각 피질의 정보 처리 모델에 따라 생물학적 체계의 신경모델 검출기를 구축하여 신체 행동의 정적 자세에서 6가지 주요 기본 감정을 판별한다. 파라미터 변화에 강건한 제안 모델의 성능은 신체행동 자세 집합을 대상으로 사람 관측자와의 평가 결과를 비교 평가하여 가능성을 제시한다.

키워드: 행동자세 인식, 감정 인식, 시각계

A Study on Visual Perception based Emotion Recognition using Body-Activity Posture

Kim, Jin Ok*

ABSTRACT

Research into the visual perception of human emotion to recognize an intention has traditionally focused on emotions of facial expression. Recently researchers have turned to the more challenging field of emotional expressions through body posture or activity. Proposed work approaches recognition of basic emotional categories from body postures using neural model applied visual perception of neurophysiology. In keeping with information processing models of the visual cortex, this work constructs a biologically plausible hierarchy of neural detectors, which can discriminate 6 basic emotional states from static views of associated body postures of activity. The proposed model, which is tolerant to parameter variations, presents its possibility by evaluating against human test subjects on a set of body postures of activities.

Keywords: Activity Posture Recognition, Emotion Recognition, Visual Perception

1. 서론

최근 컴퓨팅 기술 발전 방향은 사람의 행동을 통해 의도를 인식하여 이에 적절한 서비스를 제공하는 인간중심의 컴퓨팅환경 제공을 향하고 있다. 의도는 사람이 하고자 하는 것에 대한 정신적 상태 또는 생각으로 사람의 행동을 먼저 관찰하고 여기에 그 사람이 처한 상황을 반영하여 대상자의 목적을 인지한다. 암묵적으로 표현되는 의도 인지의 기본인 행동 카테고리는 아주 다양하기 때문에 행동이 드러내는 상태를 감정으로 인식하는 것이 일반적이다.

감정 인식은 여러 다른 문화에서도 동일한 표정 근육이

감정을 표현할 때 동일하게 사용된다는 개념에서 시작했고 감정 의도 인식은 심리학과 뇌과학, 신경생리학에서도 광범위하게 연구[1, 2, 3]되어 왔으며 이 분야에 컴퓨터와 공학기술을 접목한 연구자들은 표정을 이용한 감정 인식을 위해 신경학적 구조를 생물학적 시각으로 분석하는 방법을 제안하고 있다[4,5,6].

1970년대 이후 가장 광범위한 기본 감정은 Ekman의 분류를 따라 분노, 혐오, 공포, 기쁨, 슬픔, 놀람과 같은 여섯가지 카테고리로 나뉜다[7]. 얼굴 표정을 기본 감정 카테고리로 하나로 분류하기 위해서 여러 감정 카테고리에 미리 정확한 인지적 경계를 가정하여 처리해야 하므로 여러 감정 카테고리간의 시각적 유사성을 어떻게 정확히 판별하는가 하는 것이 감정 인식의 가장 중요한 기술이다.

수년간 감정 인식 연구 대부분은 얼굴 표정의 감정 인식에 집중하였고 최근에는 신체의 자세와 움직임을 통해 감정 의도를 인식하는 연구 즉 감정적 신체언어로 선회하기 시작

* 이 논문은 2010년도 대구한의대학교 기린연구비 지원에 의한 것임.
† 종신회원: 대구한의대학교 국제문화정보대학 모바일콘텐츠학부 부교수
논문접수: 2011년 4월 28일
수정일: 1차 2011년 6월 13일
심사완료: 2011년 6월 13일

했다[8, 9, 10]. 감정 신체 언어 연구에서 설정한 암묵적이고 일반적 가정은 신체 언어가 기본 감정을 표정과 동일하게 나타낼 수 있는 수단이라는 점이다. 신체는 얼굴보다 훨씬 많은 자유도가 있어서 신체의 전체 모양은 관절 동작에 따라 다양하게 변하기 때문에 신체 표정의 인식 자체가 어렵다. 그런데 컴퓨터 비전과 기계 학습 관련 연구[11, 12]에서 행동에 나타나는 낮은 수준의 시각적 단서를 통해서도 감정 카테고리 결정이 신뢰도 높게 이루어짐을 보여줌으로써 대상자의 움직임으로 그 사람의 의도를 추론하여 컴퓨팅 환경은 대상자에 적절한 서비스를 제공에 대한 가능성을 제시하고 있다.

다만 감정의 카테고리 분류를 사람 관찰자와 함께 모델링한 연구에서 경험치를 쌓아 결정했는데[13, 14, 15, 16] 그 과정에서 감정 카테고리 결정을 신속하고 신뢰성 있게 처리하기 위해 서로 구분하기 힘든 표정 카테고리들을 억지로 선택하여 문제를 축소하고 오로지 서로 다른 감정 카테고리를 판별하는 기능에만 중점을 두는 인지적 측면을 고려하기 때문에 논란의 여지가 있다.

본 연구는 시각피질을 포함한 시각계를 통해 대상의 감정을 인지하는 신경생리학적 매카니즘 형태의 신경 모델을 구축하여 행동 자세에서 감정 의도를 인식하는 방법을 제시한다. 제안 연구의 실험은 먼저 감정을 나타내는 행동 자세로

구성된 실험 집단을 설정하고 이 집단의 감정 상태를 사람 관찰자가 분류하도록 한 다음 제안 모델이 실험 집단을 분류하도록 테스트하여 그 결과를 비교한다. 이 방법은 동물 비전의 능력에 도달하기에는 아직 갈 길이 먼 기계 학습에 신경과학적 접근 방식을 도입하면서 전산생체공학 분야 융합연구에 새로운 담론을 형성하는데 기여한다.

본 연구에서는 행동자세의 형태 분석에 한정하고 신체의 능동성(광류)은 다루지 않는다. 본 연구가 시각적 자극의 인지와 인식, 분류에 관여하는 배쪽 경로(ventral system)의 신경계 흐름[17]으로 시각계 처리 과정을 모델링하여 동작정보 없이 단일자세의 어떤 카테고리가 신체에서 감정을 추출하는데 기여하는 지에 집중하기 때문이다. 함축된 동작 그 자체를 민감하게 추출하고 인코딩하는 연구가 제시되고 있으나 그 매카니즘이 아직 확실하기 않기 때문이기도 하다.

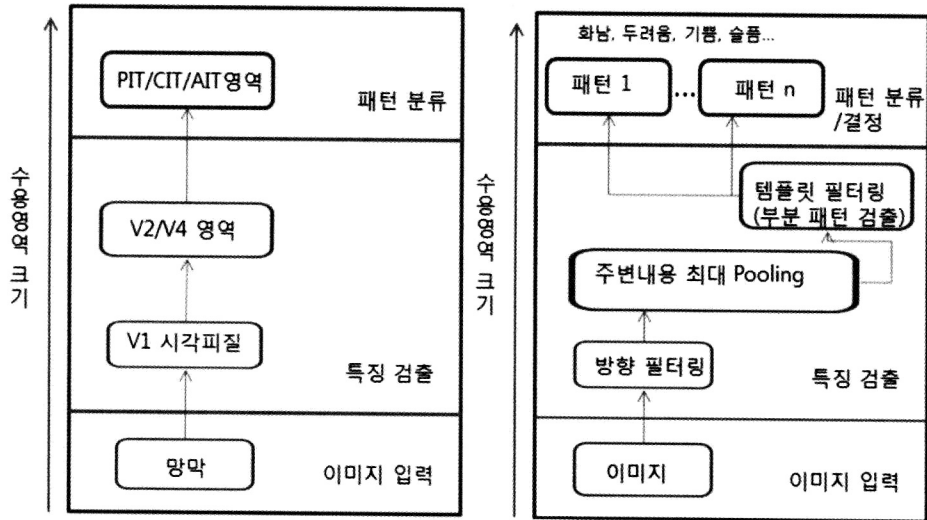
2. 시각적 인지

2.1 행동 자세의 감정 표현

본 연구는 사람들이 나타내는 다양한 감정을 Ekman에 의해 정립된 감정 카테고리인 분노, 혐오, 공포, 기쁨, 슬픔, 놀람

	분노	혐오	공포	기쁨	슬픔	놀람
정인식 자세						
						
						
오인식 자세	분노와 혐오	혐오와 공포	공포와 놀람	기쁨과 놀람	슬픔과 분노	놀람과 분노
						

(그림 1) 행동 자세의 감정 표현



(a) 시각계 (b) 신경 모델
(그림 2) 시각계와 제안 신경모델

놀람으로 한정하고 해당 감정을 나타낸 행동자세를 통해 대상의 감정 카테고리를 인식하는 과정은 시각계 처리 형태를 따른다. 사람이 감정을 드러낸 행동자세의 예는 (그림 1)과 같다. (그림 1)의 정인식 자세는 사람 관찰자와 제안 모델이 제대로 인식한 감정 자세이며 오인식 자세는 해당 카테고리의 감정을 다른 감정으로 잘못 인식한 사례이다.

2.2 시각계

본 연구에서 제안하는, 사용자 행동에서 감정을 인식하기 위한 신경 모델은 (그림 2)의 (a)와 같은 신경생리학의 시각 피질을 통해 정보를 인지하는 시각계[12, 18]를 참조한다. 시각계는 시각적으로 입력된 정보에서 단서를 추출하여 이를 순차적으로 더 고위의 인지 영역에서 감정을 결정하는 데 이용하는데 특히 얼굴과 신체 이미지의 감정적 표현에 민감하다. 시각계의 인식 과정은 시각 정보가 시각 피질 내에 전달되면 V1 신경은 색과 동작과 2차원적인 정보에 선택적으로 반응하며 순서대로 V2/V4/V5 영역을 경유한다. 시각계의 정보 전달 경로는 등쪽 경로(dorsal stream)와 배쪽 경로(ventral stream)로 나뉘는데 등쪽 경로는 눈과 손의 움직임을 담당하는 영역과 소통하고 배쪽 경로는 시각적 자극의 인식, 확인과 분류 등을 담당한다.

제안 모델의 구조는 시각적 자극의 인식과 분류를 처리하므로 시각계의 배쪽 경로를 따른 정보 처리 흐름 개념으로 행동자세를 인지하기 위해 (그림 2)의 (b)와 같이 설정한다. 인지 과정은 먼저 계층적 신경망 형태의 특징 검출기를 구성하고 실험 데이터에 이를 적용하여 이루어진다.

신경생리학적 시각계와 본 연구에서 제안한 신경모델의 처리를 비교하면 먼저 주어진 이미지에서 움직임 자세를 검출하고 신체의 부분방향을 필터링한 다음 (시각계의 V1 영역) 공간적인 주변내용을 최대한 풀링하여(시각계의 V2/V4 영역), 학습한 복합 특징인 템플릿으로 필터링을 한다(시각

계의 V4/I). 템플릿으로 필터링 한 결과는 패턴 분류기의 입력값(시각계의 IT 출력)이 되고 분류기를 통해 움직임 자세를 감정 패턴의 하나로 분류하게 된다(PIT/CIT/AIT 영역). 시각 정보처리에 응답하는 자극 영역인 수용영역의 크기 결정은 행동자세 이미지에서 감정을 결정하는 영역의 범위를 결정하는 것과 유사하다.

3. 제안 신경 모델

3.1 저수준 특징 검출

제안하는 신경 모델의 첫 번째 계층에서는 입력된 이미지 데이터에 로그-가버 필터 집합을 적용하여 여러 가지 크기의 방향을 추출하는 저수준 특징 검출 단계이다. 이를 위해 먼저 감정을 나타내는 이미지에서 각 신체방향에 민감한 셀의 반응을 로그-가버필터 뱅크로 계산한다. 움직임 자세 이미지에 가버 방식과 같은 필터링을 이용하는 방법은 시각 피질의 첫 번째 V1영역에서 단순 셀의 반응을 근사화하기 위한 표준 방법과 같은 것이다. 필터 선택은 주파수 크기에 비례하며 자연 이미지의 주파수 스펙트럼을 보상하고 모든 크기를 같은 중요도로 처리해야 한다. 로그-가버 필터를 이용하면 선호 주파수가 적은 선형 필터보다 더 나은 스펙트럼의 수렴 결과를 보이며 특징 추출이 잘 이루어진다[19].

첫 번째 계층에서 방향에 민감한 셀의 반응은 식 (1)과 같이 가버필터 뱅크로 계산하며 반응 g 는 공간주파수 w 에서 필터의 선호 주파수 μ 와 상수 σ 를 이용, 스펙트럼의 범위를 포괄하도록 한다.

$$g(w) = \frac{1}{\mu} \left\| e^{-\frac{\log(w/\mu)}{2\log\sigma}} \right\| \quad (1)$$

셀 반응은 22.5° 씩 8개 방향과 4개의 선호주파수로 계산한다[20]. 로그-가버 필터 반응 크기는 이미지의 모든 픽셀을 계산하여 결정하며 수용영역(receptive field)이 잘 중복되도록 처리한다.

그 다음 단계에서는 시각 피질 모델의 V2 영역에서의 셀 관측과 같이[21] 공간적으로 인접한 셀로부터 필터 반응을 최대 풀링한다. 시각계 정보처리 단계의 기본 동작인 최대 풀링은 센서 잡음에 대한 위치불변과 강건함을 증대시키며 시각정보 처리를 담당하는 시각피질의 V2/V4영역에서 관측된다[22, 23]. 본 연구의 풀링 과정은 승자독식 방식인 최대 연산자(Max operator)를 통해 처리하는데 이는 가장 강한 반응을 보이는 풀링 뉴런을 출력으로 결정하는 것으로 풀링은 각 방향과 크기에 대해 개별적으로 이루어진다.

두 번째 단계 뉴런의 수용영역 크기는 첫 번째 단계보다 5배 크게 선택하여 식(2)와 같이 각 뉴런 응답은 5×5 크기의 입력 값 전체에서 최대치가 되도록 한다.

$$r_i = \max_{(x,y) \in g_i} g_i(x,y) \quad (2)$$

(x,y)는 나타나는 뉴런의 공간적 위치이고 g_i 는 i^{th} 뉴런의 수용영역을 의미한다. 이렇게 하면 나타나는 뉴런이 중복되어 해당 수용영역의 크기는 반으로 줄면서 이 단계의 데이터 양은 축소한다. 즉, 5×5 픽셀로 정한 풀이 2.5 픽터 정도로 공간 해상도가 낮아져 다음 단계 뉴런의 양은 84% 까지 줄어 든다.

3.2 고수준 특징 검출

세 번째 계층은 복잡한 구조에 민감한 특정 특징 검출기로 구성되며 신경생리학의 V4/IT에 해당하는, 성분 조절 유닛과 같다. PCA는 선형 기저를 찾아 데이터양을 줄이며 정해진 상관계수로 데이터에서 변동량의 대부분을 제거하므로 본 연구에서 PCA의 이런 특성을 고려하여 특징 검출기를 구성한다. 본 연구에서 제안한 신경 모델은 이전 계층의 결과 뉴런 집합의 PCA를 통해 감정을 표현한 복잡한 신체 특징 구조를 학습한다. 한번의 학습을 통해 축소된 PCA 기저벡터가 검출기 템플릿이 된다. 이 검출기 전체를 이전 계

층의 입력 신호에 적용하여 PCA 기저위에 투영한 결과의 상관계수가 이 계층의 출력을 형성한다.

이를 위해 전체 방향과 주파수를 달리한 각 훈련 데이터에 대한 부분 반응이 반응벡터를 형성하도록 연결하고 이 벡터집합에 PCA를 수행하여 검출기 템플릿 $\{b_i, i = 1 \dots N\}$ 를 학습하도록 한다. 따라서 세 번째 계층에서는 새로운 감정 데이터가 주어지면 식 (3)과 같이 이전 계층의 응답 벡터 r 를 정규화하여 각 검출기 템플릿과 비교한 후 그 결과로 출력 신호 v_i 를 만든다.

$$v_i = \left\langle \frac{r}{|r|}, b_i \right\rangle \quad (3)$$

(그림 3)은 본 실험에서 사용된 32 채널의 템플릿 필터백크(4 스케일 × 8 방향)에 대한 첫 번째, 두 번째 기저벡터가 서로 다른 다리의 움직임을 어떻게 포착하는지 보여준다. 양의 값은 붉은 색으로, 음의 값은 초록색으로 나타나며 더 밝은 색이 큰 값을 의미한다. 템플릿의 왼쪽과 위쪽의 표시는 로그 가버 필터의 스케일과 방향을 나타낸다.

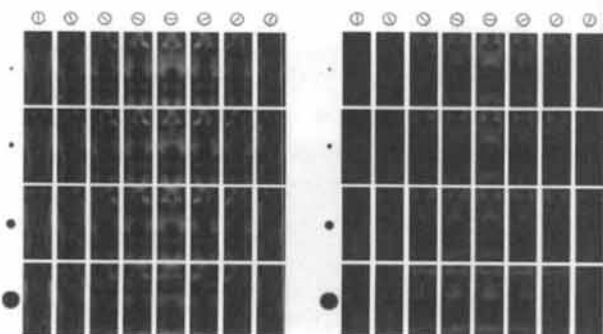
3.3 분류결정 계층

신경모델의 최고 계층에서는 감정 카테고리를 분류 결정한다. 감정 카테고리는 신체 행동 자세의 크기로 결정하므로 행동 자세 인식과 밀접한 관련이 있고 기능적으로는 신경심리학 실험에서 IT 영역[24]에 해당한다. 결정 과정을 단순화하기 위해 본 연구에서는 먼저 감정 카테고리를 정하고 이전 계층의 출력을 이 분류결정 계층에서 SVM(Support Vector Machine) 분류기를 통해 감정 카테고리 분류한다. SVM은 단힌 형태의 최적화 문제 해를 제시하는 방법으로 써 통계적 학습 이론에 뿌리를 두고 두 개 카테고리간의 초평면을 예측한다. SVM은 훈련 데이터 양이 정해진 상태에서 동일한 분류기를 이용하여 여러 다른 상태를 신경 매카니즘으로 학습할 수 있기 때문에 본 연구에 적합하다[25].

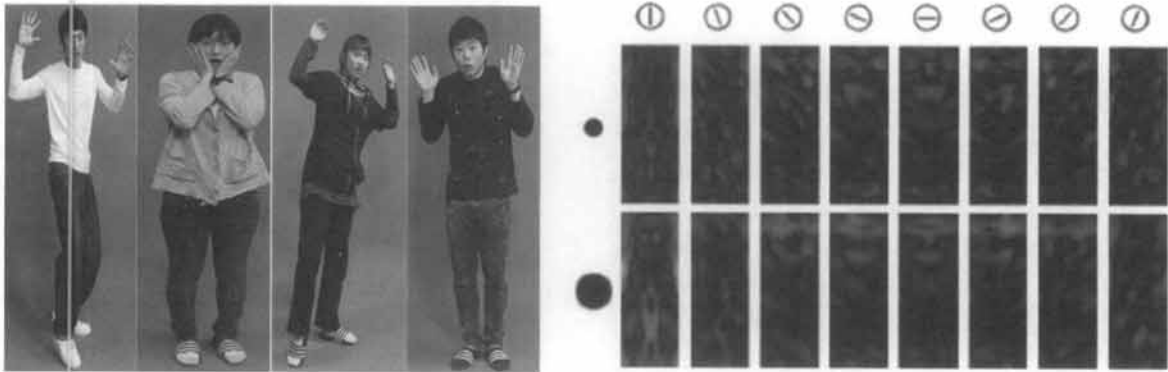
분류할 감정이 여러 가지이므로 이전 SVM 개념을 $N > 2$ 으로 확장하기 위해 일대다(one-vs-all)방법을 이용, 이전 분류기가 학습한 각 카테고리를 다른 카테고리들에서도 판별하도록 한다. 입력 신호는 그 다음 모든 N 개의 분류기에 입력하여 그 중 가장 큰 출력값을 보이는 카테고리를 선택한다. N 개의 각 분류기가 한 개 유닛을 모델링하여 특정 카테고리에 대해서만 활성화되도록 하고 가장 활성성이 강한 유닛이 카테고리를 결정하는 과정은 정보해석에 대한 생물학적 방법과 같다.

해당 유닛들은 [26]의 스냅샷뉴런, [27]의 키프레임과 유사하며, 동일한 감정이지만 다른 자세로 나타나는 경우를 제외하고는 [28]에서 관측된 것처럼 신체 구성을 결정하는 뉴런으로 간주한다. (그림 4)는 감정을 드러낸 행동자세의 해당 뉴런 특징을 보여준다.

감정 분류결정 과정은 선형 SVM이 각 감정 카테고리를



(그림 3) PCA 기저벡터 템플릿



(그림 4) 표현된 감정과 학습한 뉴런 특징

훈련하여 분류과정을 학습한 다음 뉴런을 템플릿(분리 초평면의 표준 벡터)에 투영하여 다른 감정 카테고리로부터 해당 감정 카테고리를 분류하는 최적의 분리 초평면 $\{n_i, d_i\}$ 을 찾아서 이루어진다. 식 (4)와 같이 이전 단계의 출력 값 벡터 v 를 각 분류 유니트에 입력하여 가장 강한 반응의 유니트가 카테고리를 결정하도록 하는 것이다.

$$c_i = \langle v, n_i \rangle - d_i \quad (4)$$

(그림 4)는 놀람 감정을 분류하기 위해 여러 사람의 놀람 자세와 3단계 출력에 할당된 가중치를 이용한다. 양의 가중치(붉은색)가 감정 카테고리의 전형적 구성 부분을 나타내고 음의 가중치(초록색)는 특이한 구성을 나타낸다. 분류기는 놀람 자세를 특징짓는 팔과 손의 형태를 강조하고 있다.

4. 실험

4.1 행동 자세 데이터

본 연구에 적용한 데이터는 정적 신체 움직임으로 사람의 감정을 드러낸 것을 수집한 것으로 남녀(34명의 여자와 16명의 남자)가 다른 감정을 표현하는 행동의 자세 이미지로 이루어져 있다. 모든 이미지는 카메라를 정면에서 바라보는 자세로 찍었으며 뒷 배경은 크로마키로 동일하다. 이미지 크기는 대상에 따라 25% 범위 편차로 다양하며 대상은 사진 중심에 위치해 있고 10% 범위 편차로 넓이가 다르다. 대상의 얼굴은 제거하는 대신 신체 자세의 일부로 간주한다. 이미지를 관찰한 결과 얼굴이 보이건 안보이건 대상이 드러내는 표정은 감정의도와 관련이 있어 슬프거나 두려울 경우 얼굴을 머리카락이나 손으로 가리고 놀란 상태에서는 종종 입을 가린다. 얼굴이 드러난 몇 개 이미지만을 일방적으로 걸러내면 실험 결과를 왜곡시킬 염려가 있고 감정 의도 분류에 영향을 미치므로 얼굴 상태를 그대로 둔다. 또한 실험 이미지 크기로는 대략의 표정만 드러나므로 얼굴 자체가 사용자의 감정적 의도 포착에 크게 영향을 미치지 않는다고 판단했다.

4.2 사람에 의한 행동 자세 분류

실험데이터의 실효성 확인을 위해 실험 데이터를 대상으로 사람이 인식하고 감정을 결정하는 과정을 먼저 테스트한다. 여자 3명, 남자 11명의 실험 관찰자는 데이터 집합에서 각각 25개 이미지를 보고 6가지 감정과 무표정 카테고리중 하나로 선택하여 분류한다. 25개 감정 이미지를 임의로 선택하여 이미지를 주의 깊게 보게 함으로써 사전 학습효과를 배제한다. 실제 상황에서 미확인 대상의 감정을 인식하는 것이 필요한데 같은 사람이 취한 신체 움직임 상태를 미리 보면 감정 상태 선택이 어렵기도 하고 사전에 학습한 자세와 현재 상태의 자세를 직접 비교할 수도 있기 때문이다. 이와 같은 조건은 제안 모델에도 적용하여 같은 대상이 학습과 테스트 집합에 동시에 나타나지 않도록 한다.

사람 관찰자가 결정한 6가지 감정과 무표정 카테고리에 대한 감정 인식율의 정확도는 <표 1>과 같이 87%에 달한다. 표의 각 열은 실험 대상이 나타낸 자신의 자세 표정을 의미하고 각 행은 실험 대상의 자세 표정을 보고 사람 관찰자가 선택한 카테고리이다. 특정 감정을 나타내는 자세는 특징이 있기 때문에 대부분의 감정을 사람 관찰자들이 정확하게 분류한다. 하지만 혐오와 공포, 무표정과 슬픔 자세는 서로 쉽게 혼동함을 알 수 있다. <표 1>에 의하면 분노 상태의 50개 이미지 중 40개는 정확하게 분류하지만 4개는 실험관찰자가 공포로 잘못 분류했으며 공포를 나타내는 2개 이미지는 분노로 잘못 분류한 것이다.

<표 1> 사람 관찰자의 감정 분류 결과

	분노	혐오	공포	기쁨	슬픔	놀람	무표정
분노	40	2	2	0	0	2	0
혐오	1	39	3	0	0	3	0
공포	4	8	44	0	1	2	0
기쁨	1	0	0	49	0	0	0
슬픔	1	0	1	0	42	1	2
놀람	2	0	0	0	0	42	0
무표정	1	1	0	1	5	0	48

4.3 제안 모델 테스트

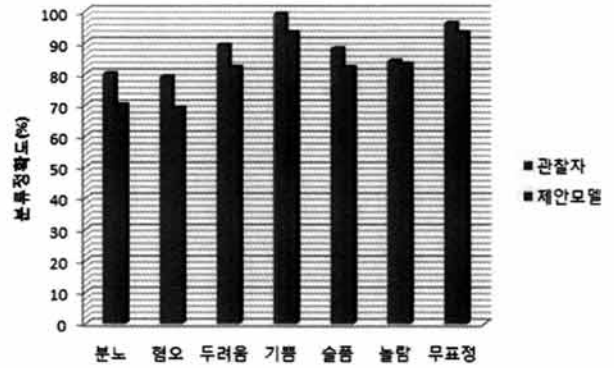
제안 모델을 설정한 이미지 집합으로 테스트한다. 2장에서 설명한 바와 같이 동일한 배경을 통해 시선이 암묵적으로 사람에게 향하도록 조절하고 계산 편이를 위해 이미지는 200× 75 픽셀로 잘라 같은 크기로 조정한다.

모델 테스트는 10중 상호 교차(10-fold cross-validation)를 통해 얻은 평균값으로 결과를 나타낸다. 10회의 테스트를 수행하는 동안 매회 데이터는 임의로 다른 감정 이미지를 10개의 집단으로 묶어 나누어서 10개의 집단 이미지 중 9개의 집단으로는 훈련을 하고 남은 이미지 한 집단으로는 테스트를 한다. 테스트 집단의 이미지를 분리하는 것은 시스템이 모든 감정을 동일하게 학습한 후 훈련동안 못 본 대상의 감정을 인식하게 하기 위함이다. 전체 훈련은 각 테스트 과정에서 반복하여 모델이 훈련 이미지를 통해 SVM 분류기는 물론 PCA기저를 학습한 다음, 새로운 이미지를 통해 테스트한다.

전체적으로 <표 2>와 같이 제안 모델의 인식 정확도는 82%에 달한다. <표 2>의 각 열은 실험 대상이 나타낸 자신의 자세 표정을 의미하고 각 행은 실험 대상의 자세를 보고 제안 모델이 감정을 선택한 결과이다. 제안 모델의 평균 오인식율은 5%로 사람 관찰자가 테스트한 것보다는 더 높은

<표 2> 제안 모델의 감정 분류 결과 (10회 반복)

	분노	혐오	공포	기쁨	슬픔	놀람	무표정
분노	35	4	1	1	2	4	06
혐오	6	34	6	0	1	2	0
공포	2	5	41	0	0	0	0
기쁨	2	3	1	47	0	0	0
슬픔	3	1	0	0	40	0	3
놀람	2	3	2	2	0	42	1
무표정	1	0	0	0	5	1	46



(그림 5) 사람관찰자와 제안 모델의 인식 분류 결과 비교

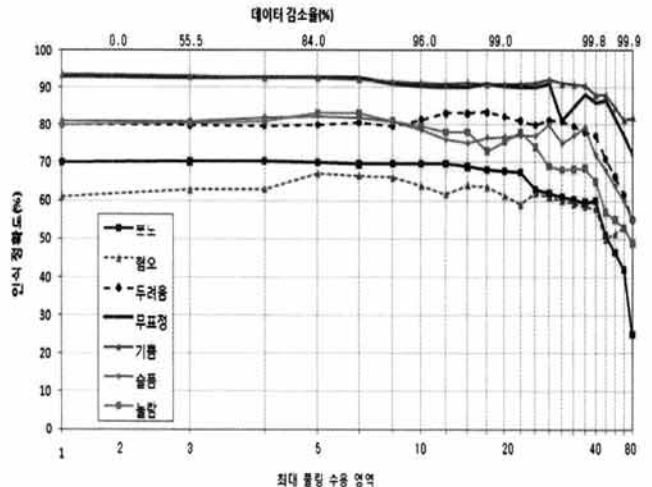
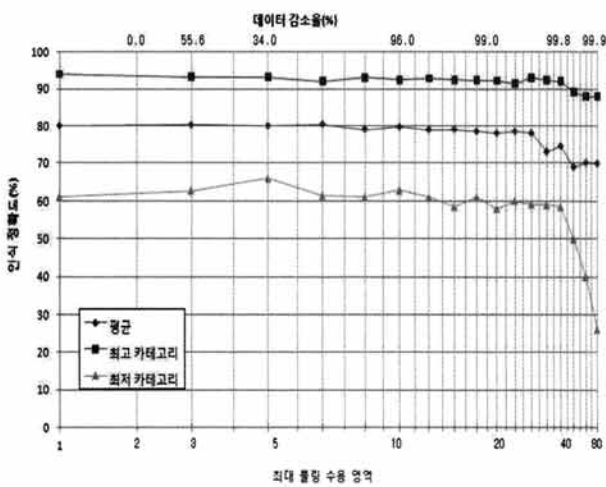
오류율을 보여 사람 관찰자의 결과를 100%로 볼 때 94%에 달하는 인식율을 보인다. (그림 5)는 사람 관찰자와 제안 모델의 수행결과를 감정 카테고리별로 비교한 결과이다. 관찰자 한 사람당 무작위로 25개의 이미지를 분류한 결과와 10회의 테스트 횟수동안 제안 모델이 분류한 결과의 평균과 표준편차를 고려한 결과를 비교하였다. 사람관찰자와 제안 모델이 보인 두 결과 간에는 높은 상관도가 존재하는데 사람관찰자가 나타낸 인식하기 쉬운 감정과 인식하기 어려운 감정 결과는 제안 모델도 유사한 결과를 나타내며 혐오와 두려움 같은 감정들에 대해서는 역시 혼란스러워 해 분류 결과가 낮다. 정인식 결과와 오인식 결과의 사례 이미지는 2장의 (그림 1)에 해당한다.

4.4 제안 모델의 파라미터 설정

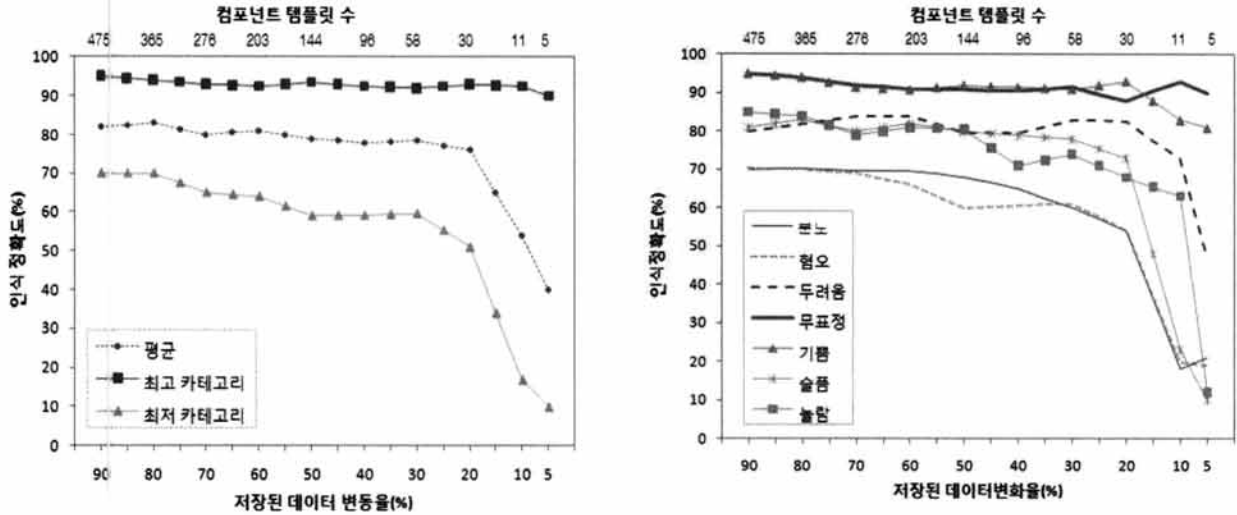
4.4.1 수용영역의 크기

제안 모델의 구성 시 V1 뉴런의 방향과 크기, 뉴런을 공유하는 수용영역, 컴포넌트 조절 유닛 수와 같은 기본 파라미터에 대한 설정 값 범위를 전체적으로 비교하여 파라미터 변화에 안정적으로 대응하도록 한다.

먼저 수용영역의 크기 비교를 한 결과 (그림 6)과 같이 수



(그림 6) 풀링 뉴런의 수용영역 크기에 따른 성능



(그림 7) 컴포넌트 템플릿 수(PCA 기저 벡터)에 따른 분류판별 결과

용영역의 크기가 커지면 인식 결과는 완만하게 하향곡선을 그린다. 이미지에서 표정을 결정하는 각 부위별 영역의 크기를 어느 정도까지 할 것인가를 결정할 때 시각계의 뉴런이 영향을 미치는 범위인 수용영역 크기를 고려하여 처리하는데 각 감정별로 수용영역의 크기가 커지고 고려할 데이터가 줄어들수록 인식율이 낮아짐을 알 수 있다. 이 결과를 반영하여 행동 자세 이미지에서 감정 부분 범위를 설정한다.

4.4.2 템플릿의 수

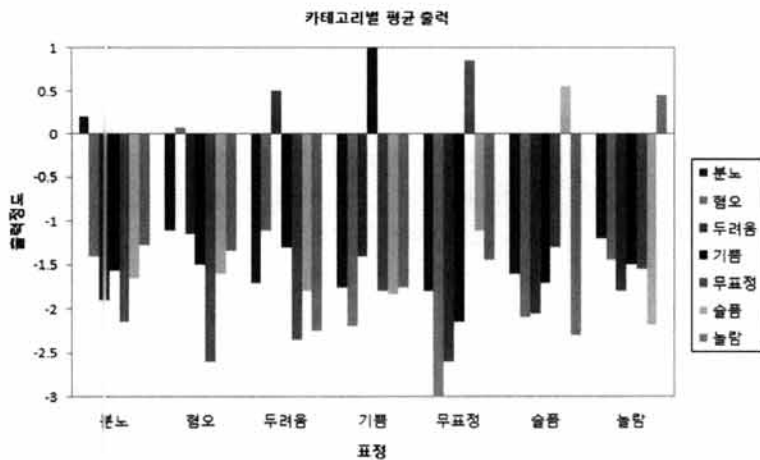
제안 모델이 행동 자세를 통해 감정을 분류 결정하는 과정에서 높은 인식율을 보이는 감정 카테고리는 몇 개의 템플릿만으로도 높은 인식 결과를 나타내지만 인식이 어려운 감정 카테고리는 많은 템플릿을 적용해도 인식율이 더 낮다. 이는 기쁨과 같은 감정 클래스의 판별 가능 특성은 혐오와 같은 다른 감정 특성보다 강하게 인식됨을 의미한다.

그래서 잘 판별되는 카테고리는 몇 개의 템플릿만으로도 명확하게 이루어지고 판별이 어려운 카테고리는 특징 템플릿을 더 많이 이용해야 함을 감안해야 한다. 단순히 모든 특징 템플릿 수를 줄이게 되면 인식율이 나쁜 감정 카테고리의 경우 판별 결과가 더 나빠질 수 있기 때문이다.

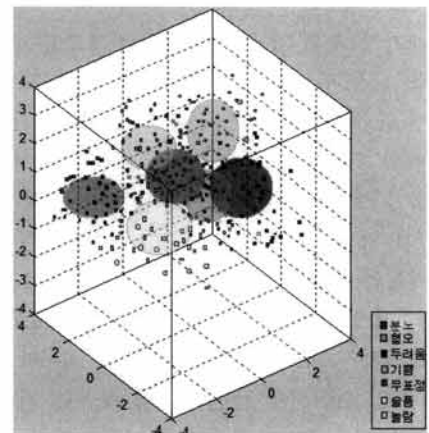
(그림 7)은 이 내용을 설명하고 있다. 특징 성분 템플릿 수를 다양화해서 분류를 수행한 결과 분노, 혐오 감정에서는 낮고 기쁨, 무표정 템플릿에 대해서는 더 높은 인식율을 보인다. 또한 특징 검출기가 많으면 많을수록 인식율이 높지만 그에 대한 비용과 시간이 만만치 않으므로 템플릿을 58개 정도 적용했을 때 적절한 검출 결과가 도출됨을 확인하여 이를 실제 감정 인식 실험에 반영한다.

4.4.3 뉴런의 평균 출력 신호

(그림 8)은 여러 카테고리의 데이터가 같이 제시되었을



(a) 감정 카테고리별 출력 정도



(b) 감정별 다차원 분포도

(그림 8) 감정 출력

때 각 감정 카테고리 뉴런의 평균 출력 신호를 보여준다. (그림 8)의 (a)는 카테고리에서 민감한 감정의 평균 출력값(카테고리에서 '분노' 이미지는 평균 0.2 값을 보이고 '혐오'는 -1.1의 값을 보인다)을 나타낸다. 출력 값은 최대치를 1로 하여 관련 크기로 나타낸다. 음의 출력값은 결정 경계의 모든 다른 카테고리를 대상으로 한다. (그림 8)의 (b)는 카테고리의 출력 값을 3D 다차원으로 나타낸다(스트레스 = 0.196). 다차원의 각 축은 감정 카테고리 뉴런의 출력 신호 크기를 의미한다. 점은 단일 표정을 나타내고 색깔은 실제 카테고리를 의미하며 타원은 클러스터의 중심과 표준편차를 보여준다.

실험에 적용한 행동 자세 이미지 데이터에서 카테고리의 구조를 설명하기 위해 분류기의 응답 공간(사각형 스트레스 카테고리에서의 non-metric MDS)상에 무표정을 포함한 7개 감정차원의 거리를 3차원적 크기로 보여준다. '기쁨'과 같은 인식이 쉬운 카테고리는 평균 출력 신호가 크고 결정 마진이 크게 밀집해 있으므로 다른 카테고리들과 잘 분리되어 있음을 알 수 있다. 반대로 '혐오', '두려움'은 다른 카테고리들과 섞여서 나타난다.

결과적으로 얼굴 표정의 감정 인식과 비교해 움직임 자세를 이용해 감정을 인식하는 문제는 내재적으로 높은 복잡성을 보인다. 유사한 신경구조를 이용할 경우 2차원이면 충분한 얼굴 이미지와는 달리[29] 움직임 자세를 이용해 다차원 크기의 잘 분리된 클러스터를 구하려면 5차원이 필요하다. 하지만 얼굴 표정의 감정 인식에서 모호하게 여겨지는 감정을 정확하게 인식하는데 본 연구를 적용하면 감정 인식율을 개선하는데 기여할 수 있다.

5. 결 과

본 연구에서는 신경생리학에서 적용한 방법을 도입하여 신체 자세를 이용한 감정 인지에 대한 신경 모델을 제시한다. 감정을 드러낸 신체자세 이미지가 제시되었을 때 모델은 6가지 주요 감정과 무표정 카테고리 중 하나로 해당 감정을 분류 판별한다. 실험 결과 제안 모델은 사람이 관찰하여 분류한 결과인 87%에 비해 82%의 인식율을 보였다.

제안 연구와 비교하여 최근 10년간 얼굴 표정과 신체 자세를 대상으로 감정 인식을 한 연구 결과를 분석한 결과 6-7개 카테고리로 얼굴 표정을 분류 인식하는데 85-98%의 인식율을 보였다[30, 31]. 신체 자세 인식은 8-12가지 도식적 자세에 대해 85-91%의 인식율을 보였고 서고 앉고 누워 있는 3가지 주요 자세에 대해서만 94%의 인식결과를 나타냈다[32, 33]. 신체 감정의 자동 분류는 제안 연구와 유사한 방법으로는 시도된 적이 없다.

본 연구는 인지과학측면에서 두가지 기여를 한다. 먼저 신체 자세를 이용한 감정 인식은 꽤 복잡한 현상이지만 카

테고리화를 위한 중요한 부분은 3D 동작 재현과 움직임 없이 저차원 형태의 처리과정으로 해결할 수 있다. 이는 신체의 감정을 인식함에 있어 2D 자세 인식과 2D 처리가 제 역할을 함을 의미한다. 이는 신체 자세를 이용한 감정 인식을 3D 방식으로만 처리하고 2D 처리는 3D 자세와 동작 재현 과정에서 중간 단계로만 이용하는 시각을 개선한 것이다.

두 번째, 본 연구는 [17]의 연구처럼 신경생리학의 시각계에서 배쪽 경로를 따른 정보 처리 흐름 개념으로 행동자세를 인지함으로써 감정 인식에 대한 새로운 방법론을 제시하고 있다. 특히 제안 구조는 정확한 인식에 필요한 다양한 파라미터의 불확실성과 주어진 많은 데이터 특징, 파라미터의 변화량을 고려하여 파라미터 변화에 강건하게 인식 과정을 처리한다.

본 연구의 한계점은 연구 대상을 단일 이미지의 형태 처리에 한정시켜서 최적의 모델이 되기에는 시각적 경로의 한 부분만을 다루고 있으며 적용한 행동 자세를 통한 감정 표현 데이터가 제한적인 양이라는 점이다. 또한 제안 신경 모델은 순전히 전향적이어서 어떤 정보도 높은 계층에서 낮은 계층으로 피드백되지 않고 있다. 시각계를 이용하는 사람은 판별이 어려운 상황에 봉착할 시 어려운 상황을 타결하기 위해 다양한 피드백 정보를 이용하기 때문에 현재 사람이 직접 관찰한 것과 본 연구에서 제안하는 모델간의 감정 인식 수행 결과 차이는 존재할 수 밖에 없다.

향후에는 이러한 제안 모델의 한계를 극복하도록 풍부한 신경망 회로가 높은 계층에서 낮은 계층으로 피드백 되도록 구성하고 오랜 관찰시간동안의 정보를 활용함으로써 사람이 시각계로 인지하는 것과 같은 방식으로 개선할 계획이다. 동시에 전체적 신경생리학적 인지 모델을 구축하기 위해 다양한 신경 전달 경로와 고수준의 인지 기능 그리고 여러 가지 신체기관의 동조를 모델링하는 연구를 계속할 예정이다.

참 고 문 헌

- [1] P. Ekman, Universal facial expressions of emotion, California Mental Health Research Digest, Vol.8, pp.151-158, 1970.
- [2] N. H. Frijda, The Emotions, Cambridge University Press, 1986.
- [3] P. Ekman, Facial expression and emotion, American Psychologist, Vol.48, 1993.
- [4] C. Padgett, G. W. Cottrell, Representing face images for emotion classification, Proc. Advances in Neural Information Processing Systems, Vol.9, MIT Press, 1996.
- [5] M. N. Dailey, G. W. Cottrell, C. Padgett, R. Adolphs, EMPATH: a neural network that categorizes facial expressions, Journal of Cognitive Neuroscience, Vol.14, No.8, pp.1158-1173, 2002.
- [6] N. Fragopanagos, J. G. Taylor, Emotion recognition in

- human-computer interaction, *Neural Networks*, Vol.18, pp.389-405, 2005.
- [7] P. Ekman, An argument for basic emotions, *Cognition and Emotion*, Vol.6, 169, pp.384-392, 1992.
- [8] M. V. Peelen, P. E. Downing, The neural basis of visual body perception, *Nature Reviews Neuroscience*, Vol.8, No.8, pp.636-648, 2007.
- [9] J. J. Stekelenburg, B. de Gelder, The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect, *Neuroreport*, Vol.15, No.5, pp.777-780, 2004.
- [10] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.886-893, 2005.
- [11] Z. Zeng, M. Pantic, G. Roisman, T. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.31, No.1, pp.39-48, 2009.
- [12] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio, Robust object recognition with cortex-like mechanisms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.29, No.3, pp.411-426, 2006.
- [13] A. J. Calder, A. W. Young, D. I. Perrett, N. L. Etco, D. Rowland, Categorical perception of morphed facial expressions, *Visual Cognition*, Vol.3, pp.81-117, 1996.
- [14] B. de Gelder, J. P. Teunisse, P. J. Benson, Categorical perception of facial expressions: categories and their internal structure, *Cognition and Emotion*, Vol.11, No.1, pp.1-23, 1997.
- [15] B. de Gelder, Towards the neurobiology of emotional body language, *Nature Reviews Neuroscience*, Vol.7, No.3, pp.242-249, 2006.
- [16] H. K. M. Meeren, C. van Heijnsbergen, B. de Gelder, Rapid perceptual integration of facial expression and emotional body language, *Proc. National Academy of Sciences of the USA*, Vol.102, No.45, pp.16518-16523, 2005.
- [17] D. J. Felleman, D. C. van Essen, Distributed hierarchical processing in the primate visual cortex, *Cerebral Cortex*, Vol.1, pp.1-47, 1991.
- [18] M. Riesenhuber, T. Poggio, Hierarchical model of object recognition in cortex, *Nature Neuroscience*, Vol.2, pp.1019-1025, 1999.
- [19] D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells, *Journal of The Optical Society of America A*, Vol.4, No.12, pp.2379-2394, 1987.
- [20] J. Hedge, D. C. van Essen, Selectivity for complex shapes in primate visual area V2, *Journal of Neuroscience*, Vol. 20(RC61), pp.1-6, 2000.
- [21] K. Fukushima, Neocognitron: a self-organizing neural network model for mechanisms of pattern recognition unacted by shift in position, *Biological Cybernetics*, Vol.36, pp.193-202, 1980.
- [22] T. J. Gawne, J. Martin, Response of primate visual cortical V4 neurons to two simultaneously presented stimuli, *Journal of Neurophysiology*, Vol.88, No.17, pp.1128-1135, 2002.
- [23] I. Lampl, D. Ferster, T. Poggio, M. Riesenhuber, Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex, *Journal of Neurophysiology*, Vol.92, pp.2704-2713, 2004.
- [24] N. K. Logothetis, J. Pauls, T. Poggio, Shape representation in the inferior temporal cortex of monkeys, *Current Biology*, Vol.5, pp.552-563, 1995.
- [25] D. Anguita, A. Boni, Improved neural network for SVM learning, *IEEE Transactions on Neural Networks*, Vol.13, pp.1243-1244, 2002.
- [26] M. A. Giese, T. Poggio, Neural mechanisms for the recognition of biological movements, *Nature Neuroscience*, Vol.4, pp.179-192, 2003.
- [27] J. A. Beintema, M. Lappe, Perception of biological motion without local image motion, *Proc. National Academy of Sciences of the USA*, Vol.99, pp. 5661-5663, 2002.
- [28] P. Downing, Y. Jiang, M. Shuman, N. A. Kanwisher, A cortical area selective for visual processing of the human body, *Science*, Vol.293, pp.2470-2473, 2001.
- [29] N. Sebe, Y. Sun, E. Bakker, M. Lew, I. Cohen, T. Huang, Towards authentic emotion recognition, *Proc. International Conference on Systems, Man, and Cybernetics*, 2004.
- [30] 김진옥, 상황에 민감한 베이즈안분류기를 이용한 얼굴표정 기반의 감정인식, *한국정보처리학회논문지*, 13-B권, pp. 653-662, 2006.
- [31] B. Boulay, F. Bremond, M. Thonnat, Applying 3D human model in a posture recognition system, *Pattern Recognition Letters*, Vol.27, No.15, pp.1788-1796, 2006.
- [32] M. W. Lee, I. Cohen, A model-based approach for estimating human 3D poses in static images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.6, pp.905-916, 2006.
- [33] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.32, No.2, pp.288-303, 2010.



김진옥

e-mail : bit@dhu.ac.kr

1989년 성균관대학교(학사)

1998년 성균관대학교(석사)

2002년 성균관대학교 전기전자및컴퓨터
공학과(박사)

1992년~1994년 현대전자산업(주)

정보통신사업부 재직

1994년~1999년 현대정보기술(주) 인터넷사업부 재직

2004년~현 재 대구한의대학교 국제문화정보대학

모바일콘텐츠학부 부교수

관심분야: 멀티미디어공학, 패턴인식, 영상처리, 유비쿼터스
컴퓨팅, 웹서비스 등