

상황에 민감한 베이지안 분류기를 이용한 얼굴 표정 기반의 감정 인식

김진옥*

요약

사용자의 상황에 따라 적절한 서비스를 제공하는 컴퓨팅 환경을 구현하려는 유비쿼터스 컴퓨팅에서 사람과 기계간의 효과적인 상호작용과 사용자의 상황 인식을 위해 사용자의 얼굴 표정 기반의 감정 인식이 HCI의 중요한 수단으로 이용되고 있다.

본 연구는 새로운 베이지안 분류기를 이용하여 상황에 민감한 얼굴 표정에서 기본 감정을 강건하게 인식하는 문제를 다룬다. 표정에 기반한 감정 인식은 두 단계로 나뉘는데 본 연구에서는 얼굴 특징 추출 단계는 색상 히스토그램 방법을 기반으로 하고 표정을 이용한 감정 분류 단계에서는 학습과 테스트를 효과적으로 실행하는 새로운 베이지안 학습 알고리즘인 EADF(Extended Assumed-Density Filtering)을 이용한다. 상황에 민감한 베이지안 학습 알고리즘은 사용자 상황이 달라지면 복잡도가 다른 분류기를 적용할 수 있어 더 정확한 감정 인식이 가능하도록 제안되었다. 실험 결과는 표정 분류 정확도가 91% 이상이며 상황이 드러나지 않게 얼굴 표정 데이터를 모델링한 결과 10.8%의 실험 오류율을 보였다.

키워드 : 얼굴 표정 인식, 감정 인식, 베이지안 분류기, HCI

Emotion Recognition Based on Facial Expression by using Context-Sensitive Bayesian Classifier

Kim, Jin Ok*

ABSTRACT

In ubiquitous computing that is to build computing environments to provide proper services according to user's context, human being's emotion recognition based on facial expression is used as essential means of HCI in order to make man-machine interaction more efficient and to do user's context-awareness.

This paper addresses a problem of rigidly basic emotion recognition in context-sensitive facial expressions through a new Bayesian classifier. The task for emotion recognition of facial expressions consists of two steps, where the extraction step of facial feature is based on a color-histogram method and the classification step employs a new Bayesian learning algorithm in performing efficient training and test. New context-sensitive Bayesian learning algorithm of EADF(Extended Assumed-Density Filtering) is proposed to recognize more exact emotions as it utilizes different classifier complexities for different contexts. Experimental results show an expression classification accuracy of over 91% on the test database and achieve the error rate of 10.6% by modeling facial expression as hidden context.

Key Words : Facial Expression Recognition, Emotion Recognition, Bayesian Classifier, HCI

1. 서론

사람의 얼굴은 보고 듣고 맛보고 냄새를 맡는 센서로서의 생물학적 기능 외에 표정으로써 사람사이에 가장 기본적인 의사소통 신호를 전달하는 역할을 한다. 사람들은 얼굴 표정을 근간으로 하여 상대방의 감정 상태를 이해하기 때문에

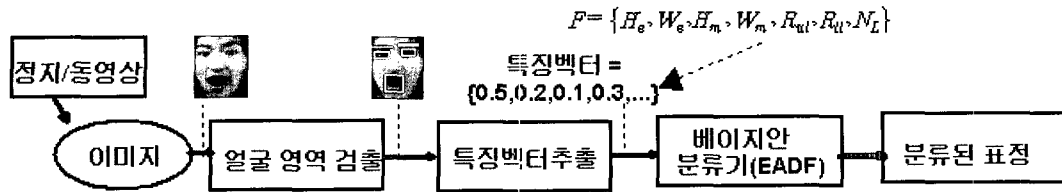
누군가의 행동과 감정을 해석하는 단서를 구할 때 일반적으로 얼굴 표정에 의존한다. 그래서 표정은 멀티 시그널 송수신기 역할이 가능한 유연성과 특이성을 가지고 있다.

얼굴 표정은 인간과 컴퓨터간의 자연스러운 인터페이스 구현에서 기계에 대한 요구사항 정보를 전달하는 기능을 한다. 따라서 얼굴 표정 연구는 사람 간에 다른 사람의 표정을 감지하여 그 의미를 해석하는 것과 유사하게 감정 기반의 의미 추론 체계를 구축하여 인간과 기계간에 자연스럽게 유연하게 상호작용이 이루어지는 방법을 다룬다.

얼굴 신호를 빠르게 분석하는 방법을 자동화하면 행동과

* 이 논문은 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2005-204-D00040).

† 정 회 원 : 대구한의대학교 정보경영대학 모바일콘텐츠학부 조교수
논문접수 : 2006년 8월 7일, 심사완료 : 2006년 10월 9일



(그림 1) 제안 얼굴 표정 기반 감정 추론 시스템

학, 심리학, 의학, 교육, 보안 등의 영역에 응용할 수 있다. 그래서 컴퓨터 분야에서는 얼굴에 나타나는 특징을 기계적으로 분석하려는 노력이 계속되고 있는데 특히 얼굴 표정 분석은 두려움, 슬픔, 혐오, 분노, 놀람, 행복과 같은 얼굴 표정이 드러내는 원형 감정의 단초를 인식하려는 연구[1-4]에 집중하고 있다. 이러한 시도는 기본 감정은 해당 원형 표정을 가진다고 제안한 Darwin[5]과 Ekman[6]의 연구에서 시발되었다. 특히 얼굴 근육의 움직임 관련 연구자들은 감정을 나타내는 얼굴 움직임 패턴이 특정 얼굴 근육의 움직임 요소에 의해 결정된다고 했다. 하지만 실제로 사람들은 원형의 표정은 덜 짓는 반면 꼭 다문 입 또는 살짝 들어 올린 눈썹과 같이 얼굴 근육의 움직임을 이용하여 감정 상태를 드러내는 변화만으로 다양한 감정을 표현한다. 따라서 이러한 표정의 미묘함을 감지하여 얼굴 표정이 전달하는 감정 정보를 만들어내기 위해 얼굴 신호를 빠르게 자동 인식하는 기술이 필요하다.

본 연구에서는 이와 같은 문제점에 대해 상황에 민감한 사용자의 얼굴 표정에서 원래의 감정을 정확하게 인식하기 위해 상황에 민감하게 표정 분류를 처리하는 확장된 베이지안 분류기(EADF)를 이용한 표정 분류 방법을 제시한다. 특히 본 연구의 확장된 베이지안 분류기(EADF)는 여러 상황에 대응하여 자연스럽게 표정 분류기를 변경할 수 있는 장점이 있다. (그림 1)은 본 연구에서 제안하는 얼굴 표정을 이용하여 사용자의 감정을 인식하는 시스템으로 제안 시스템은 얼굴 색상 정보를 이용하여 얼굴 영역을 검출한 후 얼굴 특징 벡터를 추출하고 상황에 민감한 확장된 베이지안 분류기를 통해 학습과 테스트를 거쳐 상황에 따라 다른 표정을 정확히 분류한다.

2. 관련 연구

얼굴 표정 인식 문제는 연속적 이미지를 입력 값으로 하여 표정의 역동적 특징을 분석해야 하는지 혹은 단일 이미지를 입력 값으로 한 정적 차이에 근간하여 표정을 구분해야 하는지에 따라 2 가지로 나뉜다[7]. 표정을 만들어 내는 얼굴 동작의 기록은 얼굴 움직임의 시간적 역동성을 연구하는데 필요한 반면 얼굴 동작의 정적 이미지는 얼굴 표정의 형태적 정보를 얻어내어 관련 의미를 추론하는데 필수적이다.

연속 이미지를 입력 값으로 한 연구에서 Tian[8]은 정면 얼굴 이미지 시퀀스에서 입술 추적, 템플릿 매칭과 신경망을 이용해 단독으로 또는 연결되어 발생하는 16 개 얼굴 근

육 움직임을 통해 표정을 인식하는 방법을 제시해서 87.9%의 표정 인식율을 제시했다.

Bartlett[9]은 가버 필터와 SVM(Support Vector Machine)을 이용해 전체 얼굴 비디오에서 18 개 얼굴 움직임을 자동적으로 정확하게 분석하여 95%에 달하는 표정 인식율을 보였으며 Valstar[10]은 시간적 템플릿과 조합된 K-최근접 방식과 규칙 기반의 분류기를 통해 얼굴 비디오에서 15 개 얼굴 움직임을 65%의 인식율로 인식하는 결과를 발표했다. 단일 이미지를 입력 값으로 한 Padget과 Cottrell의 연구[11]에서는 표정을 나타내는 특징 영역에서 7 개의 픽셀 블록을 사용했다. Cottrell과 Metcalfe[12]는 PCA와 feed-forward 신경망을 이용했다. Rahardja[13]는 신경망을 이용한 피라미드 구조를 이용했으며 Lanitis[14]는 얼굴 표정을 표현하기 위해 계층화된 비정형 템플릿을 이용했다. Lyons과 Zhang[15, 16]은 얼굴 표정을 코드화하기 위해 가버 웨이블릿 상관계수를 이용한 방법을 보였다. 최근 SVM과 Adaboost[17]와 같은 대규모 마진 분류기가 기계학습 분야에서 연구되면서 얼굴 표정을 분류하기 위한 새로운 시도가 계속되고 있다.

이와 같은 얼굴 표정 인식 연구는 최근 사용자 상황 인식의 주요 주제로서 유비쿼터스 컴퓨팅 분야에서 다뤄지고 있다[18]. 언제 어디서나 사용자의 상황에 따라 적절한 서비스를 제공하는 컴퓨팅 환경 제공을 목적으로 하는 유비쿼터스 컴퓨팅에서 인간과 컴퓨터간의 상호작용이 끊임없이 이루어져야 하며 이때 사용자 상황에 반응하며 필요한 서비스를 제공하기 위해서는 사용자의 얼굴 표정이 중요한 의사소통 신호로 이용되기 때문이다.

얼굴 표정 연구에서 표정을 통해 감정을 인식하는 것은 상황에 따라 동일한 감정에 대해 달라지는 얼굴 표정을 분석하여 상황에 민감하게 감정을 해석해야 함을 의미하기도 한다. 즉 사용자의 표정은 상황에 따라 달라지기 때문에 이를 잘 포착하여 표정 분류를 해야 하고 이를 통해 표정에 드러난 감정 추론이 이루어져야 한다. 하지만 현재 존재하는 얼굴 표정 분류기는 표정을 정확하게 드러낸 정지 장면으로 감정을 분석하기 때문에 실제 세계에서 사용자가 상황에 따라 조금씩 다르게 보이는 얼굴 표정 및 얼굴 표정이 보여지는 환경은 고려하지 않는 문제점을 보이고 있다.

3. 제안 연구

3.1 얼굴 특징 추출

본 연구에서는 선행 연구[19, 20]를 바탕으로 얼굴 특징을

추출한다. 얼굴 이미지에서 표정을 결정하는 눈썹, 입술, 입의 지속성 있는 얼굴 특징을 추출, 2D 기반의 위치 접근방식을 이용하여 이미지 시퀀스에서 얼굴 표정을 인식하는 방법이다.

얼굴 분석과 인식 과정에서 상황을 처리하기 위해 얼굴 특징 추출 전체 과정은 얼굴 영역 검출과 얼굴 특징 세그멘테이션 단계를 포함한다. 얼굴을 정면으로 하여 근접거리에서 포착한다고 가정하여 전체 과정이 이루어진다.

먼저 YCbCr 색상 공간을 이용하여 얼굴 이미지의 투영을 통해 서로 관련 있는 색조 영역 정의에 기여 역할을 하도록 한다. 즉 얼굴 영역의 이진영상에서 구멍이 없는 특정 영역의 넓이와 높이가 최고 수평, 수직 값이 됨을 이용, 이를 x, y 좌표로 하여 이미지에서 얼굴 피부 영역을 검출한다. 전체 이미지에서 피부 영역을 검출한 후 눈과 입 등의 부위별 피부 색상 조각을 얼굴 구조에 따라 공간적으로 배치하여 얼굴 후보 영역을 생성한다. 그리고 얼굴 후보 영역에서 얼굴 구조의 기하학적 특성에 따라 판단하여 (그림 2)의 (a)와 같이 경계선으로 얼굴 특징 영역을 추출한다. 얼굴 특징의 공간적 표현은 (그림 2)의 (b)와 같이 유클리디안 거리를 이용하여 특징간의 기하학적 혹은 지리적 관계에 근간하여 이루어진다. 식 (1)의 얼굴 특징 벡터 F 는 (그림 2)의 (c)와 같이 7개의 대표적 파라미터로 구성된다.

$$F = \{H_e, W_e, H_m, W_m, R_{ul}, R_{ll}, N_L\} \quad (1)$$

표정을 결정하는 얼굴 특징의 경우 행복한 표정은 위, 아래 입술과 비교할 때 입의 왼쪽, 오른쪽 모서리 사이가 더 많이 벌어지는 것으로 만들어지며 눈썹과 눈은 늘어지는 형태를 보인다. 반면 놀란 표정과 두려운 표정에서 입은 입술과 비교하여 왼쪽과 오른쪽 모서리가 벌어지고 눈은 더 열리고 크기가 더 커지는 형태를 보인다. 눈썹은 무표정한 위치보다 위쪽으로 당겨진다. 대표 파라미터 값으로 표정 분류가 제대로 이루어지지 않을 경우 각 파라미터를 세분화하여 처리할 수 있다.

이미지 시퀀스일 경우 특징 벡터의 모든 성분은 크기를 맞추기 위해 첫번째 프레임의 벡터크기에 맞춰 정규화한다. 위, 아래 입술 R_{ul}, R_{ll} 의 굴곡지름은 입술의 이진 마스크를

고려하여 계산한다. N_L 은 얼굴 이미지를 투영하여 분석했을 때 위, 아래 입술에 대해 감지한 입술간의 거리 값으로 입이 열린 상태인지 닫힌 상태인지를 나타낸다.

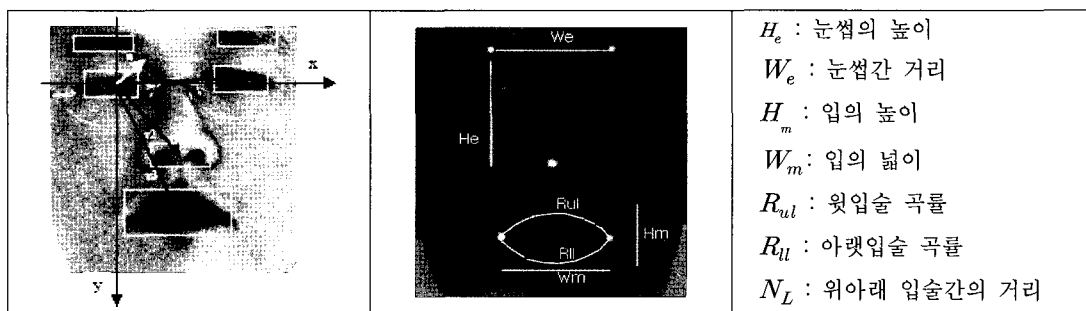
특징 벡터 F 에서의 변화는 무표정 상태에서 다른 표정 상태로 변화하는 것을 나타낸다. 특징 벡터의 능동적 성질은 모양과는 독립적이다. 식(2)의 ΔF 는 표정 분류기의 입력 값으로 이용된다.

$$\Delta F = \{\Delta H_e, \Delta W_e, \Delta H_m, \Delta W_m, \Delta R_{ul}, \Delta R_{ll}, \Delta N_L\} \quad (2)$$

3.2 상황에 민감한 베이지안 표정 분류기

사람들의 여러 가지 표정을 분류하여 감정을 인식하기 위해서는 표정 분류기가 학습 과정을 거쳐 표정 분류를 담당해야 한다. 사람들의 표정은 사무실과 같은 공적 장소와는 달리 집과 같은 사적 장소, 사적 상황에서는 다른 형태를 띈다. 따라서 표정 분류를 실세계에 응용하기 위해서는 표정 분류 대상자의 상황을 예측해야 하기 때문에 분류기가 복잡해 질 수밖에 없다. 하지만 이런 숨은 상황을 제대로 모델링해야만 더 나은 표정 분류기를 만들어 낼 수 있다. 본 연구에서는 다른 상황에 해당하는 분류 상태로 전환할 수 있는, 상황에 민감한 베이지안 표정 분류기를 제안한다. 베이지안 분류기는 각 클래스에 대해 확률 분포가 알려져 있을 때 최적이며 샘플의 양이 많지 않을 경우에 베이지안 분류기를 적용하기 위해서는 주어진 학습데이터의 밀집도 예측 문제를 단순화하는 가정을 거쳐야 한다. 하지만 베이지안 분류기는 베이지안 평균을 추정함으로써 좋은 분류 결과를 얻을 수 있으며 특히 베이지안 선형 분류기는 SVM처럼 커널 방법이나 특징 확장을 이용함으로써 비선형 분류기로 쉽게 변환될 수 있다는 장점이 있다[21,22].

제안 알고리즘은 현재의 데이터 집합을 학습시킨 후에 만약 다른 상황에 놓인 다른 데이터 집합이 있으면 이 새로운 데이터 집합에 새로운 상황의 분류기를 학습시키지만 하면 된다. 반면 SVM과 같은 전통적인 전역 학습 기술은 새로운 최적의 해를 얻어 내기 위해 데이터 전체에 걸쳐 전역 분류기를 재학습시켜야 한다. 그래서 테스트할 데이터 집합이 여러 개 있을 때는 제안된 베이지안 분류 방법이 전통적인 전역 학습 기술보다 컴퓨터 메모리를 덜 쓰면서 더 빨리



(a) 특징 추출 영역 설정

(b) 얼굴 특징 벡터

(c) 얼굴 특징 벡터 설명

(그림 2) 얼굴 특징 추출 영역 및 특징 벡터

학습시키는 결과를 보인다.

제안 알고리즘과 타 상황 학습 알고리즘간의 가장 큰 차이점 중의 하나는 학습 내용이 많아져도 다른 성분의 분류기에 데이터의 사후 확률을 적용하여 베이저안 형태에서 최적의 상황에 민감한 분류기를 유지할 수 있다는 점이다.

[베이저안 추론]

다른 추론 및 학습 기술과 비교하여 베이저안 접근 방식은 몇 가지 특징이 있다

베이저안 방식은 끊임없이 데이터와 사전 지식을 결합하여 예측과 추측에 오류정도와 신뢰정도를 조절할 수 있고 초계수를 튜닝하는 방법뿐만 아니라 모델링과 특징 선택에 필요한 메카니즘을 제공하는 장점이 있다.

(그림 3)은 베이저안 방식을 설명한 그래픽 모델로서 그래프의 회색 노드는 관찰된 변수이고 무색의 노드는 숨어 있는 변수이다. 왼쪽 그래프는 베이저안 네트워크의 방향성이 있는 모델이 만들어지는 추론과정으로 모델 샘플링 과정은 관측치가 처리되는 추이를 나타낸다. 오른쪽 그래프는 조건적 분류를 이용한 학습 방식으로 관측치에서 조건에 따라 w 모델 파라미터의 사후 분포를 추론한다. 조건적 모델의 데이터 레벨은 각각 다른 주어진 입력값과 모델 파라미터에 독립적이다.

베이저안 추론의 계산은 수치를 다룬 적분이므로 적분을 빠르고 정확하게 함으로써 좋은 결과를 얻을 수 있다. 베이저안 추론의 일반적 시나리오는 학습 데이터와 테스트할 데이터가 일정양 있다고 가정한다.

미지수 x 의 추론은 식 (3)과 같이 학습 데이터 D 가 주어진 사후 분포에 기반한다.

$$p(x|D) = \frac{p(x, D)}{p(D)} = \frac{p(x, D)}{\int_x p(x, D) dx} \tag{3}$$

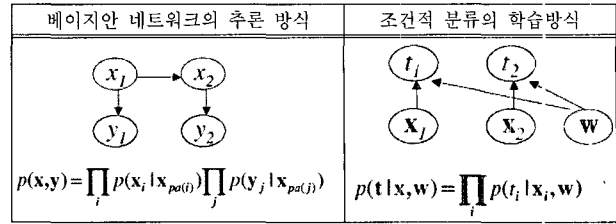
미지수의 사후 평균과 분산이 필요하면 이것 역시 식 (4, 5, 6)과 같이 사후 분포 전체를 적분함으로써 구한다.

$$E[x|D] = \int_x xp(x|D)dx = \frac{\int_x xp(x, D)dx}{\int_x p(x, D)dx} \tag{4}$$

$$E[x^2|D] = \int_x x^2p(x|D)dx = \frac{\int_x x^2p(x, D)dx}{\int_x p(x, D)dx} \tag{5}$$

$$var(x|D) = E[x^2|D] - E^2[x|D] \tag{6}$$

이 과정을 통해 수치 적분은 점차 실제 베이저안 추론이 접근해 간다. 추론에 베이저안 방식을 적용할 때 계산이 복잡하다는 것이 문제점인데 이는 베이저안 추론과 학습에서



(그림 3) 베이저안 네트워크의 추론과 학습 방식

폭넓게 쓰이는 Monte Carlo 방법이 정확한 결과를 끌어내기 위해 많은 양의 샘플을 필요로 하므로 계산상 비용이 많이 들기 때문이다. 또한 Monte Carlo 방법은 무작위적 알고리즘이어서 낮은 확률의 영역에서 많은 샘플들을 이용해야 좋은 결과치를 보이기 때문이기도 하다. 본 연구에서는 표정 데이터를 분류하여 정확한 감정을 추론하기 위해 베이저안 방식의 장점은 수용하면서 Monte Carlo 등 다른 결정적 근사화 기술보다 복잡도는 낮으면서도 정확한 예측치를 얻을 수 있는 베이저안 접근 방식의 데이터 분류 기술인 상황에 민감한 베이저안 분류기를 제안한다.

식(7, 8)과 같이 L 개의 하위집합을 가진 학습 데이터 집합 $D = D_1, \dots, D_L$ 이 주어졌을 때 테스트 데이터인 벡터 \tilde{x} 는 식 (9)와 같이 분류할 수 있다

$$D_1 = \{(x_1^1, t_1^1), \dots, (x_{N_1}^1, t_{N_1}^1)\} \tag{7}$$

⋮

$$D_L = \{(x_1^L, t_1^L), \dots, (x_{N_L}^L, t_{N_L}^L)\} \tag{8}$$

$$p(\tilde{x}|D) = \sum_i p(\tilde{x}, D_i) P(i|\tilde{x}, D) \tag{9}$$

i 는 하위집합 D_i 를 선택함을 의미하고 이는 하위 집합 D_i 는 각각 다른 상황에 해당함을 뜻한다. 그래서 먼저 기대치 확산에 근거한 분류신뢰도 $p(\tilde{x}, D_i)$ 를 계산하는 새로운 방법을 제시하고 $P(\tilde{x}|D_i)$ 을 계산하는 데 가우시안 혼합모형을 적용한 방법을 설명한다.

선형 분류기는 파라미터 벡터 w 를 이용하여 $t = \text{sign}(w \cdot \phi(x))$ 에 의해 포인트 x 를 분류한다.(두 클래스 $t = \pm 1$ 이다). 기저함수 $\phi(x_i)$ 는 원래의 특징치에서 비선형이 되는 분류경계를 만들어 내는데 이용한다. 학습 집합 $D = \{(x_1, t_1), \dots, (x_N, t_N)\}$ 이 주어지면 w 에 대한 유사도는 식 (10)과 같이 쓸 수 있다.

$$p(t|w, X) = \prod_i p(t_i | x_i, W) = \prod_i \Psi(t_i, w^T \phi(x_i)) \tag{10}$$

여기서 $t = \{t_i\}_{i=1}^N, X = \{x_i\}_{i=1}^N$ 이며 $\Psi(a)$ 는 스텝 함수로 만약 $a > 0$ 이면 $\Psi(a) = 1$ 이고 $a \leq 0$ 이면 $\Psi(a) = -1$ 이다. 설명을 간략하게 하기 위해 식 (7)에서 식 (9)의 데이터 집합 지수 i 를 제거한다. 새로운 입력값 x_{N+1} 이 주어지면 예측 분포를 식 (11), (12)로 근사할 수 있다.

$$p(t_{N+1}|x_{N+1}, t) = \int p(t_{N+1}|x_{N+1}, w)p(w|t)dw \quad (11)$$

$$\approx P(t_{N+1}|x_{N+1}, \langle w \rangle) \quad (12)$$

여기서 $\langle w \rangle$ 는 베이지안 포인트라고 불리는 가중치의 사후평균을 의미한다.

일반적으로 베이지안 네트워크에서 사용되는 Monte Carlo 방법을 사용하는 대신 본 연구에서는 $\langle w \rangle$ 와 $p(t_{N+1}|x_{N+1}, t)$ 를 계산하기 위해 ADF(Assumed-Density Filtering)[23]에 기반한 방법인 EADF(Extended Assumed-Density Filtering)를 적용했다

[EADF(Extended Assumed-Density Filtering)]

EADF는 유사도가 단순한 조건의 산물이라는 사실을 이용한다. 만약 조건 각각을 제대로 근사한다면 더 정확한 사후 근사 값을 구할 수 있다.

EADF는 조건을 정확하게 이용하여 사후 확률과 같은 근사치를 선택하게 된다. 사후 분포를 근사하기 위해 항을 나열하는 ADF의 강력한 확장방법인 EADF는 근사치를 이용하기 때문에 KL(Karhunen-Loeve) 확산 방법과 유사하다. 하지만 ADF 알고리즘은 (그림 4)의 (a)와 같이 데이터가 처리되는 순서에 종속되기 때문에 산포 확률이 현재의 w 예측치 값에 따라 달라진다. EADF는 이러한 반복적인 배치 버전을 개선한 방법으로 (그림 4)의 (b)와 같이 연속 데이터 전체 정보를 이용하면서 근사 값을 더 정확하게 구한다. (그림 4)는 ADF 알고리즘과 EADF를 설명하는데 $q(w)$ 는 사후 근사치이고 t_i 는 인수 분해한 $n+1$ 개 항이며 \tilde{t}_i 는 근사 항을 의미하는데 이 값은 새로운 사후 예측치를 이전 사후 예측치로 나눈 값에 상수인 정규화 계수 값을 곱해서 얻는다.

먼저 EADF의 w 에 대한 가우시안 사전 분포는 식 (13)과 같다.

$$p(w|\alpha) = \prod_i N(w_i|0, \alpha_i^{-1}) \quad (13)$$

여기서 $\alpha = \{\alpha_i\}$ 는 초계수 벡터이다 나중에 모든 i 에 대해 $\alpha_i = 1$ 로 할당한다.

정확한 사전 항을 $g_i(w)$ 로 하고 $\tilde{g}_i(w)$ 는 근사치 항을 의미하도록 정리하면 식 (14), (15)와 같다.

$$p(w|t, \alpha) \propto p(w|\alpha) \prod_i p(t_i|w) = p(w|\alpha) \prod_i g_i(w) \quad (14)$$

$$\approx p(w|\alpha) \prod_i \tilde{g}_i(w) \quad (15)$$

베이지안 선형 분류기에서 가우시안 형태가 되기 위해 근사화 조건은 (m_i, v_i, s_i) 파라미터로 표현한다.

$$\tilde{g}_i = s_i \exp(-\frac{1}{2v_i}(t_i \phi^T(x_i)w - m_i)^2) \quad (16)$$

파라미터는 근사화 사후 분포를 가우시안 형태로 만든다.

$$p(w|t, \alpha) \approx q(w) = N(m_w, V_w) \quad (17)$$

이상과 같이 EADF는 최적의 조건 근사항을 찾기 위해 (그림 5)와 같은 알고리즘으로 이루어진다.(기호를 줄이기 위해 $t_i \phi(x_i)$ 은 ϕ_i 로 나타낸다)

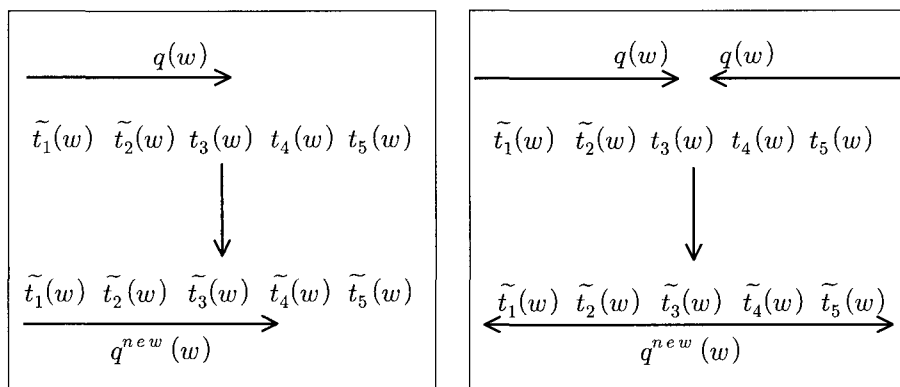
EADF 알고리즘에 의해 데이터 집합 D_i 에 대해 학습된 결과가 나오면 식 (18)에 의해 데이터 집합의 분류를 테스트한다. $p(\tilde{t}|\tilde{x}, D_i)$ 는 테스트 데이터 집합을 의미한다.

$$p(\tilde{t}|\tilde{x}, D_i) = \int p(\tilde{t}|\tilde{x}, W)p(W|D_i)dW = \Psi(\tilde{z}) \quad (18)$$

여기에서 $\tilde{z} = \frac{(t M_w)^T \phi(\tilde{x})}{\sqrt{\phi^T(\tilde{x}) V_w \phi(\tilde{x})}}$ $x^i = \{x_1^i, \dots, x_{N_i}^i\}$, $t^i = \{t_1^i, \dots, t_{N_i}^i\}$

그리고 $D_i = \{x^i, t^i\}$ 라 하고 식 (19)와 같이 x^i 를 가우시안 혼합모델 H_i 로 모델링하는데

$$p(x^i) = \sum_{j=1}^J \beta_j N(\mu_j, \Sigma_j) \quad (19)$$



(a) 각 항을 근사한 다음 정확하게 사후 예측치 $q^{new}(w)$ 를 계산하는 ADF 방식

(b) 모든 다른 항의 상황에서 EADF는 각 항을 고려하여 사후 예측치를 구함

(그림 4) ADF 방식과 EADF 방식

1. 초기 단계
 $\tilde{g}_i=1$ (근사항 값), $v_i = \infty$ (이전 항), $s_i = 1$ (크기 계수)로 설정한다.
 사전 조건으로 $m_w = 0, V_m = \text{diag}(\alpha)$, 그리고 모든 i 에 대해 $\alpha_i = 1$ 로 설정한다.

2. 모든 (m_i, v_i, s_i) 가 수렴할 때까지 루프를 돈다.
 Loop $i = 1, \dots, N$:
 (a) $N(m_w^i, V_w^i)$ 의 가우시안 형태인 $q(w)$ 로부터 사후 'leave-one-out' $q^i(w)$ 를 구하기 위해 근사치 \tilde{g}_i 를 제거한다. $q^i(w) \propto q(w)/\tilde{g}_i$ 은 다음 내용을 내포하고 있다.

$$V_w^i = V_m + \frac{(V_w \phi_i)(V_w \phi_i)^T}{v_i - \phi_i^T V_w \phi_i}$$

$$m_w^i = m_w + (V_w^i \phi_i) v_i^{-1} (\phi_i^T m_w - m_i)$$

(b) 조건과 함께 i 없이 사전확률을 넣으면 $\hat{p}(w) \propto g_i(w) q^i(w)$ 가 된다.
 $KL(\hat{p}(w) \parallel q(w))$ 를 최소화하기 위해 $q(w)$ 를 선택한다.
 Z_i 를 정규화 계수라 하면

$$m_w = m_w^i + V_w^i \rho_i \phi_i$$

$$V_m = V_w^i - (V_w^i \phi_i) \left(\frac{\rho_i \phi_i^T m_w}{\phi_i^T V_w^i \phi_i} \right) (V_w^i \phi_i)^T$$

$$Z_i = \int_W g_i(W) q^i(W) dW = \Psi(z_i)$$

여기서 $z_i = \frac{(M_w^i)^T \phi_i}{\sqrt{\phi_i^T V_w^i \phi_i}}$

$$\rho_i = \frac{1}{\sqrt{\phi_i^T V_w^i \phi_i}} \frac{N(z_i; 0, 1)}{\Psi(z_i)}$$

(c) $\tilde{g}_i = Z_i \frac{q(W)}{q^i(W)}$ 를 이용하여 조건 근사치를 갱신한다.

$$v_i = \phi_i^T V_w^i \phi_i \left(\frac{1}{\rho_i \phi_i^T M_w} - 1 \right)$$

$$m_i = \phi_i^T M_w^i + (v_i + \phi_i^T V_w^i \phi_i) \rho_i$$

$$s_i = Z_i \sqrt{1 + v_i^{-1} \phi_i^T V_w^i \phi_i} \exp\left(\frac{\rho_i \phi_i^T V_w^i \phi_i}{2 \phi_i^T M_w} \right)$$

(그림 5) EADF 알고리즘

여기서 혼합모델 H_i 는 $\{\beta_j, \mu_j, \Sigma_j\}_{j=1}^J$ 로 계수화 된다. 계수 $\{\beta_j, \mu_j, \Sigma_j\}_{j=1}^J$ 를 예측하는데 전통적인 알고리즘인 최고 기대치(EM) 알고리즘을 적용할 수 있다.

확률 $P(\hat{x}, D)$ 이 학습 집단의 분류에 의존하지 않는다고 가정하고 $p(\hat{x}|\hat{H}_i)$ 를 이용하여 $p(\hat{x}|D_i)$ 를 근사화 한다. \hat{H}_i 는 최고 기대치(EM)가 예측한 파라미터라고 정의한다.

하위집합 i 에 대해 단일한 사전 확률을 적용하면 다음과 같은 식 (20)을 산출한다.

$$P(\hat{x}, D) \approx \frac{p(\hat{x}|\hat{H}_i)}{\Sigma_i p(\hat{x}|\hat{H}_i)} \tag{20}$$

식 (9)에 식 (18)와 식 (20)을 적용하면 $p(\hat{x}, D_1, \dots, D_L)$ 을 구할 수 있다.

만약 $p(\hat{x} = 1\hat{x}, D_1, \dots, D_L) > p(\hat{x} = -1\hat{x}, D_1, \dots, D_L)$ 이면 $\hat{i} = 1$ 이고 아니면 $\hat{i} = -1$ 이다.

4. 실험

표정 분류기를 실험하기 위해 동양인의 얼굴 데이터베이스인 JAFFE[24]와 상황에 따라 바뀌는 얼굴 표정으로 남자 7인, 여자 3인이 각각 다른 상황에서 보이는 표정을 포착하여 이용했다. 사람의 얼굴 표정은 처한 상황에 따라 달라지므로 먼저 학교 연구실과 같은 공적인 장소에서 실험 주관자 입회하에 보인 표정과 집과 같이 사적인 장소에서 실험 대상자가 직접 자신의 표정을 잡아낸 표정 클래스를 대상으로 하였다. JAFFE는 정확한 실험 환경 설정 하에서 감정을 드러낸 표정이기 때문에 가장 정확한 감정을 드러낸 표정 데이터로 이용했다. 본 연구는 얼굴 표정의 형태적 정보를 이용한 감정 추론을 다루기 때문에 단일 이미지로부터의 얼굴 표정 인식에 중점을 둔다. 중간 표정, 기쁨, 슬픔, 놀람, 화냄, 혐오, 두려움 등 7개 표정 이미지는 256×256 크기로 해당 표정별로 4-6개의 이미지를 사용했다. 주어진 데이터 표정이 잘 분류되면 표정을 통해 감정을 인식할 수 있으므로 표정 분류를 위해 각 표정의 특징 벡터에 대한 평균 값과 분산 값을 추출한다.

표정 데이터베이스에서 *i*번째 대상자의 데이터 하위집합의 특성 벡터 포인트 전체에 대해 $\phi(x) = [K(x, x_1), \dots, K(x, x_i), \dots, K(x, x_N)]^T$ 와 같이 식 (11)의 비선형 기저 전개 함수 $\phi(x)$ 를 선택한다. 여기서 $K(x, x_i)$ 는 가우시안 형태의 기저함수이다.

각 표정 데이터들은 무작위로 60%와 40%씩 학습 하위집합과 테스트 집합으로 나뉜다. 먼저 학습 하위집합을 베이지안 성분 분류기로 학습한 다음 두 개의 가우시안 혼합을 통해 하위 집합을 모델링한다. 이때 파라미터는 최대 기대치(EM) 알고리즘으로 예측한다. 첫번째 두 개의 성분 분류기에 대해 가우시안 커널 넓이를 0.1로 하고 다른 쪽은 0.5로 한다. 테스트 단계에서 테스트 데이터 집합이 어느 데이터 범주로부터 온 것인지 알지 못한다고 가정하고 알고리즘이 테스트 데이터에 자동적으로 적용되도록 한다. 각 테스트 집합에 대한 5 번의 테스트를 거친 분류 오류를 기록한 결과 평균 오류율은 <표 1>과 같다. 커널 넓이에 따라 다른

오류율을 보이지만 제안 방법은 서로 다른 커널 넓이에서 더 나은 결과를 선택하므로 10.6%로 오류율이 낮아졌다. 다른 분류기와 비교하기 위해 가우시안 커널 넓이가 0.1과 0.5로 된 두 개의 SVM(Support Vector Machine)를 이용하여 데이터 집합을 학습한 테스트 결과 역시 <표 1>과 같다. SVM의 오류율은 0.5의 커널에서 11.4%를 보임으로써 제안 방법의 오류율이 더 낮음을 알 수 있다.

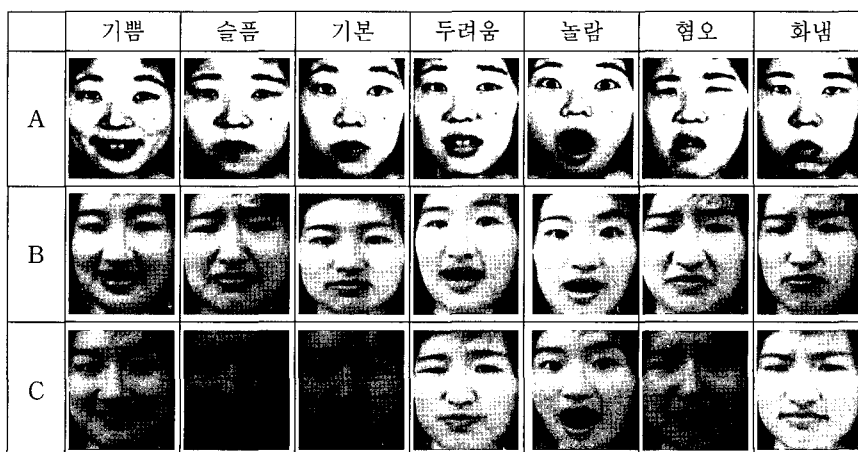
<표 1> EADF로 학습한 베이지안 분류기와 SVM를 비교한 오류율(%)

테스트 오류 분류기 (커널넓이)	테스트1	테스트2	테스트3	테스트4	테스트5	평균 오류
EADF(0.1)	7.3	8.2	12.2	11.6	27.7	13.4
EADF(0.5)	8.1	9.8	9.1	5.7	23.7	11.3
EADF(0.1&0.5)	7.3	8.2	9.1	5.7	23.7	10.8
SVM(0.1)	7.8	8.5	12.8	13.3	29.4	14.7
SVM(0.5)	8.2	8.8	10.3	6.2	23.8	11.5

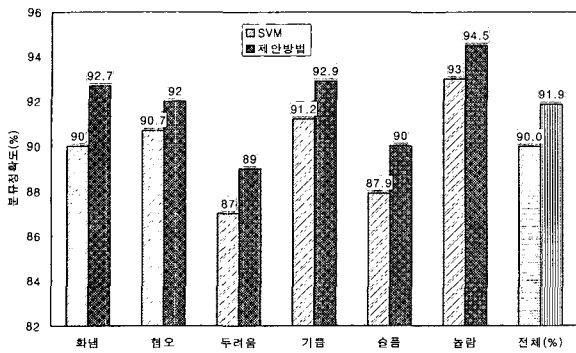
제안된 베이지안 분류기와 비교했을 때 SVM은 평균 분류 오류율은 더 높았다. 또한 더 넓은 커널 매트릭스를 계산하고 더 많은 메모리 크기와 긴 학습 시간을 필요로 한다. 데이터 집합을 적절하게 선택하면 SVM에 필요한 메모리 크기는 줄일 수 있지만 데이터 수렴을 위해 대상 집단을 나누는데 많은 학습단계의 반복 과정이 필요하다.

한편 제안된 분류기는 다른 학습 집단에 대해서는 반복할 필요가 없다. 또한 제안 알고리즘은 처리 결과를 최적화하기 위해 다른 성분 분류기에는 다른 커널 크기를 적용할 수 있다.

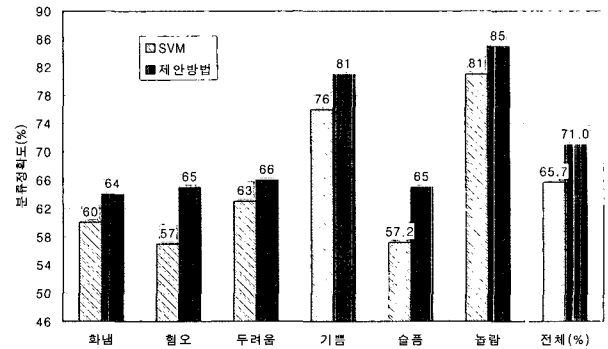
<표 2>는 상황에 따라 표정이 어떻게 다른지를 설명한다. 상황에 따라 달라지는 표정 분류를 위해 훈련 받지 않은 자연스러운 표정 데이터를 훈련된 상태와 비교한 결과는 <표 2>와 같다. A의 경우 실험주도자가 입회한 상태에서 실험대상자가 지시에 따라 표정을 연출한 JAFFE의 얼굴 표정이다. B와 C는 동일인의 표정으로 B는 실험대상자가 표정



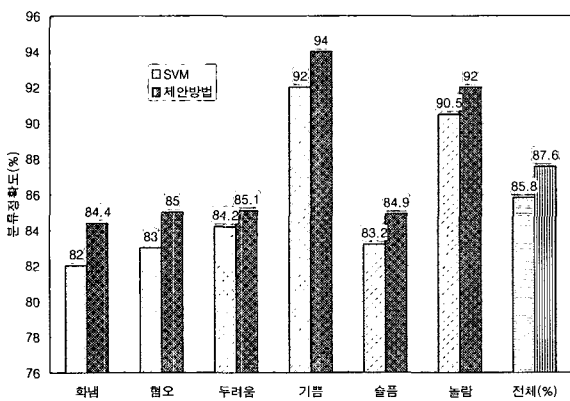
<표 2> 표정에 드러난 감정의 분류



(그림 6) A 상황의 표정 분류



(그림 8) C 상황의 표정 분류



(그림 7) B 상황의 표정 분류

짓기에 대한 실험 내용을 인지하고 조명 등 실험 환경을 갖춘 다음 연구실에서 찍은 표정 결과이다. JAFEE에 비해서는 표정에 감정이 덜 나타났다. C의 경우 실험대상자가 표정 짓기에 대한 사전 지식 및 조명 등 주변 환경 설정 없이 혼자 찍은 표정이다. 환경 및 상황에 따른 표정 데이터 분류 실험의 비교를 위해 세 가지 상황의 표정에 본 연구에서 제안한 방법과 SVM을 적용하였다.

A의 JAFEE 표정 데이터는 실험 지시자가 입회한 상태에서 실험 대상자에게 기본 표정을 짓도록 연습시킨 후 구축된 데이터베이스로써 사용자 표정이 감정을 드러내기에 가장 좋은 상황이다. (그림 6)과 같이 SVM 방법은 90%의 분류 정확도를, 제안된 방법은 91.9%의 높은 표정 분류 정확도를 보였다.

B 상황의 데이터에 대해서는 (그림 7)과 같이 SVM과 제안 방법이 85.8%와 87.6%의 분류 정확도를 보인다. 감정을 표정으로 드러내는데 대해 실험 대상자가 사전에 인지한 상태이기 때문에 사용자의 표정이 자연스러운 상태에서 더 강하고 정확하게 나타났지만 표정에 감정 충분히 나타나지 않아 A보다 분류 정확도가 낮다.

C의 표정 분류는 (그림 8)과 같이 SVM의 경우 평균 65.7%, 제안된 방법은 71%의 분류 정확도를 보이는데 C 표정 분류율이 타 상황의 표정에 비해 낮은 주요 원인은 실험 대상자가 기본 감정을 어떻게 표정에 담아 표현하는지에 대

한 사전 지식이 없는 상황이고 감정을 나타내는 표정의 형태가 정확하지 않았기 때문이다.

하지만 제안 방법의 결과를 SVM에 비교해 볼 때 제안된 방법이 훨씬 나은 분류 정확도를 보였다.

상황에 따라 같은 감정을 드러내는 표정이 달리 나타날 수 있기 때문에 HCI 기반의 사람과 컴퓨터간의 상호작용을 위해서는 어떤 상황이라도 표정에 따른 감정을 적절히 인식할 수 있어야 한다. 이 때문에 상황에 민감하게 반응하는 표정 분류 방법이 필요하다.

실험 결과를 통해 상황이 달라진 상태에서도 본 연구에서 제안한 EADF를 이용한 베이지안 분류기가 SVM보다 오류율은 더 낮으면서 더 정확하게 표정을 분류함을 알 수 있다.

5. 결 론

사용자 상황에 따라 다른 서비스 환경을 제공해야 하는 유비쿼터스 컴퓨팅에서 필요한 기술 중의 하나가 사용자 상황을 인식하는 기술이다. 얼굴 표정을 이용하여 사용자 감정을 인식하는 것은 그 중의 한 방법으로 사람과 컴퓨터간의 인터페이스에서 감정 정보를 포함하면 더 자연스럽게 효과적인 상호작용이 가능해진다. 얼굴 표정을 이용한 감정 인식 방법은 별도의 장비를 이용한 전처리 과정 없이 바로 카메라 센서를 통해 실시간으로 사용자 상황을 인식할 수 있어 유비쿼터스 컴퓨팅 환경에서 더 중요한 역할을 할 것이다.

본 연구는 상황에 따라 달리 나타나는 사용자의 얼굴 표정에서 정확하게 원래의 사용자 감정을 인식하는 방법을 다룬다. 이를 위해 먼저 얼굴 색상 정보를 이용하여 얼굴 특징을 추출하고 새로운 베이지안 학습 알고리즘으로 EADF (Extended Assumed-Density Filtering)을 제안하여 표정 분류 및 감정을 인식한다. 제안 알고리즘은 주어진 표정 데이터 집합을 효과적으로 학습하여 새로운 지식을 쉽게 추론, 생성할 수 있게 한다. 실험 결과 제안 알고리즘은 불안정한 데이터 분포를 정확하게 분류하면서 데이터의 다른 상황에 대해 다른 복잡도의 커널들을 다양하게 사용함으로써 10.8%의 낮은 오류율을 나타냈다. 반면 유사한 분류 알고리즘이지만 전역 학습 알고리즘인 SVM는 지역성과 처리 용량 간

에 균형을 취하기 어려워 최적의 결과를 보여 주지 못했다. 본 연구에서 제안된 베이지안 분류기를 서로 다른 상황의 얼굴 표정으로부터 사용자의 감정을 인식하는데 적용한 결과 다른 상황의 사용자 표정에 대해 SVM 보다 더 높은 감정 분류 정확도를 보임으로써 제안 알고리즘이 사용자의 다른 상황에 대응하여 더 나은 응답 결과를 목표로 하는 HCI 인식 시스템의 한 부분으로 사용할 수 있음을 보여 준다. 향후 연구에서는 상황의 모델링 방법을 개선하고 학습과정에서 알지 못한 숨겨진 상황을 감지하도록 하여 분류 오류율을 개선하도록 한다. 또한 서로 다른 데이터 성분 분류기의 학습 과정에서 분류기간에 아무런 정보를 공유하지 못하므로 이를 개선하는 방법을 연구하여 멀티모달 인터페이스를 이용한 사용자 감정 추론에 적용할 수 있도록 할 계획이다.

참 고 문 헌

- [1] P. Ekman and W. Friesen, "Facial Action Coding System," Consulting Psychologist Press, 1978.
- [2] K. Mase, "Recognition of Facial Expression from Optical Flow," IEICE Transactions, Vol.E74, No.10, pp.3474-3483, 1991.
- [3] M. Pantic, L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The state of the art," IEEE Trans. Pattern Anal. Mach. Intel., Vol.22, No.12, pp.1424-1445, 2000.
- [4] I. Essa, A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions," IEEE Trans. PAMI, Vol.19, No.7, pp.757-763, 1997.
- [5] C. Darwin, "The Expression of the Emotions in Man and Animals," Univ. of Chicago Press, 1965.
- [6] D. Keltner and P. Ekman, "Facial Expression of Emotion," Handbook of Emotions, pp.236-249, 2000.
- [7] Y. Tian, T. Kanade, J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," IEEE Trans. PAMI, Vol.23, No.2, pp.97-115, 2001.
- [8] T. Kanade, J. Cohn, Y. Tian, "Comprehensive Database for Facial Expression Analysis," Proc. IEEE Conf. FGR, pp.46-53, 2000.
- [9] M. S. Bartlett, "Machine Learning Methods for Fully Automatic Recognition of Facial Expressions and Actions," Proc. IEEE Conf. SMC, pp.592-597, 2004.
- [10] M. F. Valstar, M. Pantic and I. Patras, "Motion History for Facial Action Detection from Face Video," Proc. IEEE Conf. SMC, pp.635-640, 2004.
- [11] C. Padget, G. Cottrell, "Identifying Emotion in Static Images," Proc. 2nd Joint Symp. Neural Computation, Vol.5, pp.91-101, 1997.
- [12] G. Gottrell, J. Metcalfe, "Face, Gender and Emotion Recognition using Holons," Adv. Neural Inform. Process. Syst. Vol. 3, pp. 564-571, 1991.
- [13] A. Rahardja, A. Sowmya, W. Wilson, "A Neural Network Approach to Component vs. Holistic Recognition of Facial Expressions in Images," SPIE Proc. Intelligent Robots and Computer Vision X : Algorithms and Techniques, Vol.1607, pp.62-70, 1991.
- [14] A. Lanitis, C. Taylor, T. Cootes, "Automatic Interpretation and Coding of ace Images using Flexible Models," IEEE Trans. Pattern Anal. Mach. Intel., Vol. 19, No.7, pp.743-756, 1997.
- [15] M. J. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intel., Vol.21, No.12, pp.1357-1362, 1999.
- [16] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, "Feature Comparison between Geometry Based and Gabor-Wavelets Based Facial Expression Recognition using Multi-layer Perceptron," Proc. 3rd IEEE Int. Conf. Automatic Face and Gesture Recognition, pp.454-459, 1998.
- [17] G. Guo, C. R. Dyer, "Learning from Examples in the Small Sample Case: Face Expression Recognition," IEEE Trans. on Systems, Man and Cybernetics, part B: Cybernetics, Vol.35, No.3, pp.477-489, 2005.
- [18] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," IEEE Trans. on PAMI, Vol. 22, No.1, January. 2000.
- [19] J. O. Kim, K. H. Jung, C. H. Chung, "On a Face Recognition by the Modified Nonsingular Discriminant Analysis for a Ubiquitous Computing," Lecture Notes in Computer Science 3980, Vol.1, pp.929-937, May., 2006.
- [20] 김진욱, "빛보상과 외형기반의 특징을 이용한 얼굴 특징 검출," 한국인터넷정보학회 논문지, 제7권 제4호, pp. 143-153, 2006.
- [21] T. Minka, "A Family of Algorithms for Approximate Bayesian Inference," Ph.D. Thesis, MIT, 2001.
- [22] M. Harries, K. Horn, "Learning Static Concepts in Domains with Hidden Changes in Context," 13th ICML, workshop on Learning in Context Sensitive Domains, Italy, 1996.
- [23] H. J. Kushner, A. S. Budhiraja, "A Nonlinear Filtering Algorithm based on an Approximation of the Conditional Distribution," IEEE Trans. Automatic Control, Vol.45, pp.580-585, 2000.
- [24] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, "Coding Facial Expressions with Gabor Wavelets,"

Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, pp.200-205, 1998.



김진옥

e-mail : bit@dhu.ac.kr

1989년 성균관대학교 졸업(학사)

1998년 성균관대학교 대학원 정보통신
공학과 졸업(석사)

2002년 성균관대학교 대학원 전기전자 및
컴퓨터공학과 졸업(박사)

1992년~1999년 현대정보기술 인터넷사업본부 선임

2004년~현재 대구한의대학교 정보경영대학 모바일콘텐츠학부
조교수

관심분야: 멀티미디어공학, 패턴인식, 영상처리, 유비쿼터스
컴퓨팅 등