

쾌 및 각성차원 기반 얼굴 표정인식

Facial expression recognition based on pleasure and arousal dimensions

신 영 숙* 최 광 남**
(Young-Suk Shin) (Kwang-Nam Choi)

요 약 본 논문은 내적상태의 차원모형을 기반으로 한 얼굴 표정인식을 위한 새로운 시스템을 제시한다. 얼굴표정 정보는 3단계로 추출된다. 1단계에서는 Gabor 웨이블릿 표상이 얼굴 요소들의 경계선을 추출한다. 2단계에서는 중립얼굴상에서 얼굴표정의 성긴 특징들이 FCM 군집화 알고리즘을 사용하여 추출된다. 3단계에서는 표정영상에서 동적인 모델을 사용하여 성긴 특징들이 추출된다. 마지막으로 다층 퍼셉트론을 사용하여 내적상태의 차원모델에 기반한 얼굴표정 인식을 보인다. 정서의 이차원 구조는 기본 정서와 관련된 얼굴표정의 인식 뿐만 아니라 다양한 정서의 표정들로 인식할 수 있음을 제시한다.

주제어 차원모형, Gabor 웨이블릿 변환, FCM(Fuzzy C-Means) 군집화 알고리즘, 동적연결모형, 표정 인식

Abstract This paper presents a new system for facial expression recognition based on dimension model of internal states. The information of facial expressions are extracted to the three steps. In the first step, Gabor wavelet representation extracts the edges of face components. In the second step, sparse features of facial expressions are extracted using fuzzy C-means(FCM) clustering algorithm on neutral faces, and in the third step, are extracted using the Dynamic Model(DM) on the expression images. Finally, we show the recognition of facial expressions based on the dimension model of internal states using a multi-layer perceptron. The two dimensional structure of emotion shows that it is possible to recognize not only facial expressions related to basic emotions but also expressions of various emotion.

1. 서론

얼굴은 인간의 의사소통에서 얻을 수 있는 비언어적

정보를 제공한다. 사람과 사람의 대면 상황에서 정보가 전달되는 다양한 채널 중에서 얼굴표정은 큰 부분을 차지한다. 얼굴표정이 큰 부분을 차지하는 주된 이유는 인간의 내적 정서 상태가 주로 얼굴표정을 통하여 표현되기 때문이다.

이러한 정서상태를 측정하는 방법으로 심장박동, 호흡율 등의 생리적 반응을 측정하는 방법이 시도되어 왔지만 이는 실제 상황에서 사용하기 어려울 뿐 아니라 생리적 반응만으로 정서상태의 구체적인 유형에 대한 정확한 정보를 얻기 어렵다는 점으로 인하여 주로 이론적 연구에 국한되어 사용되어 왔다. 얼굴표정은 정서상태의 주요 표현방법으로 서로 다른 문화권에서도 많은 공통점이 관찰된다는 점에서 문화적인 보편성을 지닌다고 할 수 있다.

* 조선대학교 정보통신과,
Department of Information and Telecommunication Engineering,
Chosun University
** 중앙대학교 컴퓨터공학부
School of Computer Science and Engineering
연구세부분야: 컴퓨터 비전, 패턴인식, 신경망
주소 : 광주광역시 동구 서석동 375번지
조선대학교 전자정보공과대학 정보통신과
우편번호: 501-759
전화 : 연구실: (062)230-7011
Cellular : 019-654-8473
FAX : (062)230-7011
E-mail : ysshin@mail.chosun.ac.kr



(그림 1) 얼굴표정 데이터베이스

초기의 얼굴표정 연구는 주로 연구자의 주관적 판단이나 일반 사람들의 평정자료에 근거를 두고 정서상태가 얼굴표정으로 표현되는 연구들이었다. 그러나 근래에는 체계적으로 얼굴근육의 움직임에 부호화하여 각 정서상태의 표현에 연결시키는 작업들이 진행되고 있다. 그 예로서 Ekman과 Friesen [1]이 개발한 Facial Action Coding System(FACS)을 들 수 있다. FACS는 얼굴근육의 움직임을 상세하게 기록한 후 이를 토대로 행복, 놀람, 공포, 분노, 혐오, 슬픔의 6가지 기본정서를 판별할 수 있도록 구성되어 있다.

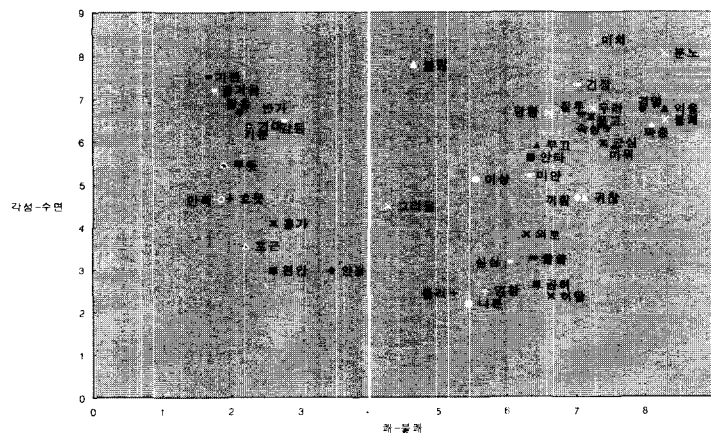
이러한 이론적 기반을 바탕으로 Ekman은 각 기본정서의 얼굴표정을 구성하는 얼굴근육 움직임을 분석하여 FACS를 구성함으로써 얼굴표정을 통한 정서모형을 주도

해왔다. 지금까지의 표정인식 대부분은 기본정서모형을 기반으로 한 표정연구들이다 [2, 3, 4, 5, 6, 7].

그러나, 기본정서모형의 주장에 많은 비판이 제기되었다. 특히, 기본정서모형의 주요한 이론적 근거가 되는 표정을 통한 정서 인식의 문화적 보편성은 방법론상의 문제로 과대 추정되었을 가능성이 있음을 지적하고 있다. 문화적 보편성을 보고한 대다수의 표정연구에서는 여섯 개의 기본정서 중에서 반응을 선택하게 하였는데, 기존의 다른 정서들을 추가하여 선택하게 하였을 경우에는 인식률이 훨씬 떨어졌다고 보고되었다 [8]. 또한 표정과 일치하지 않는 상황에 대한 정보를 줄 때는 표정보다 상황에 의존한 판단을 하는 경우도 많았으며, 기본정서 유형이 연구자마다 다르다는 점 등도 비판의 대상이 되고 있다 [9]. 이와같은 기본정서모형에 의한 범주적인 표정 분류는 표정을 통해서 연속적인 정서상태의 변화시 변화 강도가 반영되지 않는다는 제한점이 존재한다.

본 연구에서는 정서의 2차원 구조를 기반으로 한 표정인식을 통하여 기본정서를 바탕으로 한 표정인식의 한계성을 극복하며, 확장하는 새로운 시스템을 제안한다. 본 논문의 전체적인 구성을 살펴보면 다음과 같다. 1단계, 차원모형 기반 표정인식을 위한 얼굴표정 데이터베이스를 소개하며, 2단계, FCM군집화 알고리즘과 동적인 연결모형을 이용하여 특징점들의 자동추출을 제시한다. 3단계, 차원모형 기반 신경망을 이용한 얼굴표정인식 결과를 제시하며, 마지막으로 인간의 평정과 신경망을 이용한 표정인식 결과에 대한 결론 및 논의로 이루어진다.

44개 어휘목록



(그림 2) 44개 어휘목록의 차원분석

2. 차원모형 기반 얼굴표정 데이터베이스

제한한 표정인식 시스템을 구현하기 위하여 640x480 해상도를 가진 290개의 정면영상을 사용하였다. 6명의 남, 녀 각각의 영상들은 정서와 관련하여 83개의 내적 상태를 나타내는 어휘와 일치되는 표정영상으로 구성되어 있다[10]. (그림 1)은 연구에 사용된 표정영상의 일부를 보이고 있다. 83개 단어는 김영아, 김진관, 박수경, 오경자[11]가 연세대학교에 재학중인 학부생 200명과 30, 40대 피험자 150명을 대상으로 실험을 통하여 수집된 단어들로 구성되어있다. 본 연구에서는 83개의 내적 상태의 표정영상에서 표정영상과 내적 상태와의 일치성에 대하여 피험자들간의 높은 빈도수를 보인 44가지의 내적 상태 표정영상을 추출한 후, 실험을 위하여 6사람 각각의 44 내적 상태 표정영상에서 다시 11개의 표정영상을 추출하였다.

추출된 11개의 표정 영상들은 무표정을 포함한 기쁨, 놀람, 슬픔, 경멸, 공포, 만족, 편안함, 속상함, 나른함, 근심의 내적 상태 표정들로 이루어진다. 이 11개의 내적 상태 표정영상들은 두 차원 상에서 특정 공간에 편중되지 않고 전반적인 분포를 반영하도록 선택되었다.

(그림 2)는 두 차원(패-불패, 각성-수면)에서 피험자간에 내적 상태와 표정영상간의 일치율이 높은 빈도를 보인 44개 어휘목록에 대한 차원분석 결과이다. 두 차원들은 9점 척도상에서 표정데이터를 평정하도록 나타내진다.

3. 표정정보 추출

연구에 사용된 모든 표정 영상들은 얼굴의 특징 요소들만을 고려하여 인간과 유사한 기능을 수행하는 표정인식 시스템 구현을 위하여 다음과 같은 제한점을 제시한다. 모든 표정 영상들은 거의 정면 영상만을 대상으로 하며, 영상처리에 사용되기 전에 두 눈 안쪽사이의 거리를 일정한 화소값으로 보정하므로써 다양한 얼굴 크기를 정규화하였다. 또한 왼쪽 눈의 안쪽 끝점의 좌표 위치를 일정 위치로 고정시켰다.

3.1 Gabor 웨이브렛 변환을 이용한 경계선 추출

얼굴의 특징점 추출을 위한 전처리로 Gabor 웨이브렛 변환이 적용되었다. Gabor 웨이브렛 변환 표상은 다중해상도와 다중방향 정보를 이용하기 때문에 웨이브렛 변환된 계수의 크기(amplitude)는 단순한 배경에서 유효 정보인 경계선을 구별할 수 있게 된다. 영상에 적용된 Gabor 웨이브렛 함수는 Daugman[12]에 의하여 제시된 형식과

동일하다.

$$\psi_{\vec{k}}(\vec{x}) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) \left[\exp(i\vec{k}\vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (1)$$

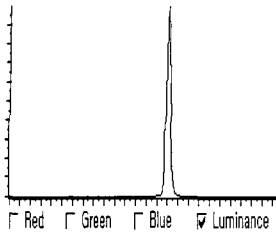
이차원 웨이브벡터 $\vec{k}=(k, \theta)$ 는 필터의 공간 주파수 k 와 방향 θ 를 통제하며, 벡터 $\vec{x}(x, y)$ 는 영상을 이루는 각각의 화소를 나타낸다. σ 는 k 에 관련하여 Gaussian 윈도우의 폭을 조절한다. $G_{\vec{k}}(\vec{x})$ 이 영상 $I(\vec{x})$ 에 적용되는 Gabor 함수라고 하면, $G_{\vec{k}}(\vec{x})$ 는 $\psi_{\vec{k}}(\vec{x})$ 과 $\varphi_{\vec{k}}(\vec{x})$ 의 두 필터가 존재한다. 이 필터들은 90도 위상 관계를 이루고 있다. 필터 쌍들이 위상차이에 의하여 영상에 미치는 영향을 적게 하기 위하여 각각의 필터를 영상에 적용한 후, 생성된 2개의 컨벌루션된 영상 v_1, v_2 는 다음과 같이 처리된다.

$$\begin{aligned} v_1 &= \sum \psi_{\vec{k}}(\vec{x})I(\vec{x}) \\ v_2 &= \sum \varphi_{\vec{k}}(\vec{x})I(\vec{x}) \\ S &= \sqrt{(v_1^2 + v_2^2)} \end{aligned} \quad (2)$$

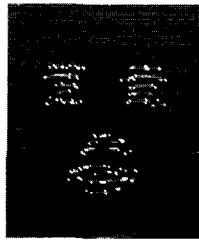
식(2)에 의하여 생성된 영상에 대한 이차원 Gabor 웨이브렛 계수 히스토그램은 (그림 3)과 같다. 이것은 히스토그램의 평균값을 이용하여 일정한 주파수와 방향에 대하여 얼굴의 주요 요소성분들의 경계선을 추출할 수 있음을 제시한다. (그림 4)는 Gabor 웨이브렛 계수 히스토그램의 평균값을 적용하여 얼굴의 주요 요소들의 경계선이 추출된 결과이다.

3.2 무표정영상에서 FCM 군집화 알고리즘을 이용한 특징점 추출

Gabor 웨이브렛 함수를 이용하여 컨벌루션된 영상들의 이웃하고 있는 화소들의 계수들은 높은 상관관계를 보이며 중복된 값들로 나타난다. 이것은 영상을 이루는 화소들이 서로 상관성을 갖지 않으며 중복되지 않는 독립적인 계수를 갖는 화소들로 다시 추출될 수 있는 가능성을 보인다. 이러한 독립적인 계수들을 갖는 화소들만을 추출하기 위해서는 화소들간의 거리간격이 필요하며, 또한 얼굴의 주요 요소성분들을 대표할 수 있는 기하학적인 특징점들을 보존할 수 있어야한다. 이러한 목적을 달성하기 위해서는 적절히 군집화되어 군집안에서 대표성 벡터들을 추출하는 것이 필요하다. 따라서 본 연구에서는 얼굴의 주요요소들의 경계선 정보들을 기반으로 얼굴의 기하학적인 특징들을 보유하면서 화소들간의 중복된 계수값들을 제거하고 대표성 벡터들만을 특징점으로 추출



(그림 3) Gabor웨이브렛 계수 히스토그램



(그림 4) 추출된 얼굴의 주요 요소 성분들의 경계선

하기 위하여 FCM 군집화 알고리즘을 적용하였다. FCM 군집화 알고리즘[13]은 각 데이터의 점과 각 군집의 중심과의 거리를 염두에 두고 유사도 측정을 기반으로 한 목적함수의 최적화 방식이다. 이것은 n차원의 유클리디안 공간상에서 N개의 특징 벡터들로 이루어지는 집합 X를 기반으로 한다.

$$X = \{x_1, x_2, \dots, x_N\}$$

사용된 목적함수는 다음과 같다.

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^N u_{ij}^m d_{ij}^2 \quad (3)$$

$$\sum_{i=1}^c u_{ij} = 1, \quad \forall j=1, \dots, N \quad (4)$$

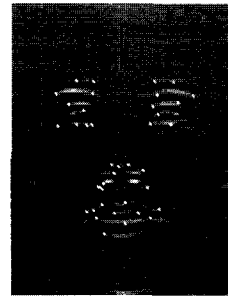
c_i 는 각 군집 i 의 중심을 의미하며, $d_{ij} = \|c_i - X_j\|$ 는 각 데이터의 점과 각 군집의 중심과의 유클리디안 거리를 나타낸다. 소속함수 u_{ij} 는 0과 1사이의 실수값으로 각 군집에서 데이터의 소속정도를 평가한다. 식(3)이 최소값에 도달하기 위한 필요 조건으로 식(5)와 식(6)이 있다.

$$c_i = \frac{\sum_{j=1}^N u_{ij}^m X_j}{\sum_{j=1}^N u_{ij}^m} \quad (5)$$

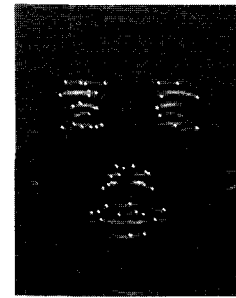
$$u_{ij} = \frac{1}{\sum_{k=1}^c \frac{d_{ij}^{2j}}{d_{kj}^{2j}}} \quad (6)$$

특징점 추출은 다음 단계를 이용하여 이루어졌다.

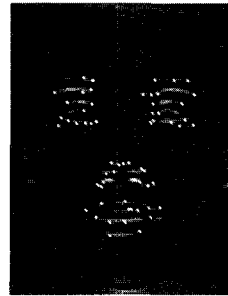
- 단계1: 데이터가 각 군집에 소속될 정도는 식(4)을 만족하는 0과 1사이의 임의의 값으로 초기화한다.
- 단계2: 각 군집의 중심 c_i 은 식(5)를 사용하여 계산된다.
- 단계3: 목적함수는 식(3)에 의하여 계산되며, 바로 전 분류결과와 현재의 분류결과와 에러차이를 비교하여 반복여부를 판단한다.



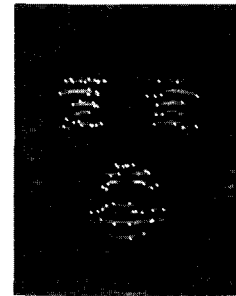
(a) 특징점 개수 40



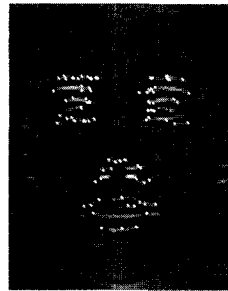
(b) 특징점 개수 50



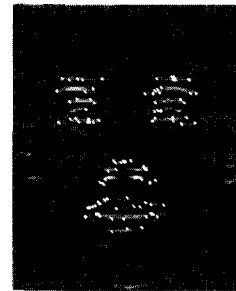
(c) 특징점 개수 60



(d) 특징점 개수 80



(e) 특징점 개수 90



(f) 특징점 개수 100

(그림 5) 군집수의 변화에 따라 추출된 특징점

단계4: 식(6)을 이용하여 데이터가 각 군집에 소속될 정도를 갱신하며, 단계2로 돌아간다.

군집의 개수는 인간의 시각으로 얼굴의 주요요소들과 동일한 위상관계를 반영하는 범위에서 실험을 통하여 결정하였다. (그림 5)은 FCM 군집화 알고리즘을 적용하여 군집수의 변화에 따라 추출된 특징점들을 보이고 있다. (그림 5)의 (a)는 군집수 40개를 적용한 경우로서 왼쪽 눈의 바깥쪽 끝점과 오른쪽 눈 안쪽 끝점, 그리고 아랫입술의 특징점들이 소실된 경우를 보여주며, (b)의 경우 왼쪽 눈의 안쪽 끝점과 아랫입술의 특징점이 소실된

경우를 보여준다. 군집수가 40, 50개를 적용할 경우에 얼굴의 주요 요소들의 기하학적인 특징점들이 소실되는 경우를 (그림 5)의 (a), (b)를 통하여 알 수 있다.

그러나 또한, 군집수가 너무 많으면 특징점들의 분포가 일부 영역에 중첩되어 나타나는 것을 볼 수 있는데, 군집수가 80인 (d)의 경우는 오른쪽 눈썹에 특징점들이 중첩된 경우를 볼 수 있으며, (e)의 경우는 왼쪽과 오른쪽 눈썹, 눈들의 특정영역에 중첩된 경우를 알 수 있다. 이러한 중첩의 정도는 군집수가 100개인 (f)의 경우에 더욱 두드러져 보이는 데, 눈썹, 눈, 코, 입 등 얼굴의 주요 요소 전반에 걸쳐 특징점들이 중첩되어 나타남을 볼 수 있다. 본 연구에서 사용된 군집수는 인간의 시각으로 얼굴의 주요 요소들과 동일한 위상관계를 반영하는 범위에서 실험을 통하여 결정하였다. 군집수는 60이 적용되었다 (그림 5(c)).

3.3 표정영상에서 동적 연결모형을 이용한 특징점 추출

FCM 군집화 알고리즘을 이용한 무표정 영상에서 추출된 특징점들은 사전에 추출된 표정영상들의 경계선 정보를 나타내는 점들로부터 특징점을 추출하는 템플릿으로 활용되었다. 템플릿을 이루는 무표정영상은 60개의 특징점을 갖는다. 표정영상에서 특징점 추출은 무표정 영상의 각각의 특징점에서 표정영상에서의 일대일 대응으로 이루어지는 동적 연결모형을 적용하였다. 본 연구에서는 Gabor 웨이브렛 변환을 사용하여 변환 계수를 각각의 특징점에서의 특징 파라미터로 이용하였다. 무표정 영상과 표정영상의 특징 파라미터간의 유사도 및 유클리디안 벡터를 비교하여 비용함수를 구하고, 비용함수 값이 가장 작은 점에서 표정영상으로의 최종적인 대응점으로 결정된다. 이러한 과정은 크게 두단계로 수행된다.

[개략적인 정합]

무표정 영상에서 표정영상으로의 특징점 연결은 무표정 영상의 사전 선택된 점들과 표정영상의 연결되는 점들과는 대략적으로 동일한 위상관계를 갖는다는 제한 아래 수행되어진다. 이는 영상처리에 사용되기 앞서 영상 정규화과정을 통해 얻어질 수 있다. "개략적인 정합"과정은 표정영상에서 얼굴의 주요요소들의 경계선 정보들을 기반으로 얼굴의 기하학적인 특징들을 보유하면서 대표성 벡터들만을 특징점으로 추출하는 과정이다. 표정영상들의 추출된 경계선 데이터들은 1차적으로 무표정영상과 동일한 군집수로 분류된다. 분류된 각 군집의 중심을 이루는 데이터들의 좌표점과 무표정 영상의 대응점간의 유클리디안 거리를 측정하여 가장 작은 거리를 갖는 군집의 중심이 표정영상으로의 1차적인 대응점이 된다. 식(7)은 무표정영역(N)과 표정영역(E)간의 유클리디안 거리 계산을 나타낸다.

$$\vec{\Delta}_{ij}^{NE} = \vec{x}_i^N - \vec{x}_j^E \quad (7)$$

[정밀한 정합]

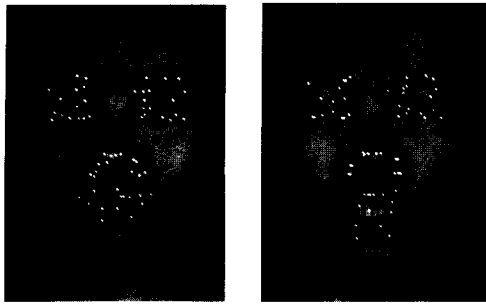
"정밀한 정합"과정은 개략적인 정합 단계에서 추출된 대표성 벡터를 포함하고 있는 각 군집 안의 점에서 Gabor 웨이브렛 변환을 통해 얻어진 젯(jet)들의 특징벡터를 이용한다. 즉, 1차적으로 무표정영상에서 표정영상의 대략적인 위치 결정 후, 다음으로는 무표정 영상의 각 특징점에서의 Gabor 웨이브렛 변환 후의 방향과 주파수로 이루어지는 Gabor 젯들과 표정 영상에서 1차적으로 선택된 군집안을 이루는 데이터들의 Gabor 젯간의 유사도 비교를 통하여 최종적인 대응점이 결정된다. 본 연구에서는 2개의 주파수, $\{\frac{\pi}{4}, \frac{\pi}{8}\}$ 와 30도 간격의 7개의 방향으로 이루어지는 14개의 Gabor 필터를 적용하였다. 대응점간의 유사도 측정함수를 식(8)에 보이고 있다.

$$S(\vec{J}_i^N, \vec{J}_i^E) = \frac{\vec{J}_i^N \cdot \vec{J}_i^E}{\|\vec{J}_i^N\| \|\vec{J}_i^E\|} \quad (8)$$

위의 두단계 과정의 동적연결 과정은 무표정 영상의 선택된 점들에서 표정 영상의 점들로의 최적화문제로 표현될 수 있다. 즉, 목적함수 H는 제한된 점들간의 연결 정도를 평가하게 되며, 식(9)와 같은 최소비용함수로 표현될 수 있다.

$$H = \vec{\Delta}_{ij}^{NE} - \sum S(\vec{J}_i^N, \vec{J}_j^E) \quad (9)$$

표정영상에서의 특징점들은 목적함수 H에서 다음 두 조건을 동시에 만족시킨 경우에 선택되었다. 첫째, 목적함수 H는 최소값에 도달해야 한다. 둘째, 표정영상에서 찾아지는 대응점들은 무표정 영상의 대응점과 대략적으로 동일한 위상관계를 형성하도록 하는 최대 거리를 초과하지 않아야 한다. (그림 6)은 동적 연결모형을 적용하여 표정영상에서 특징점이 추출된 결과를 보인다.



(그림 6) FCM 군집화 알고리즘과 동적연결모형이 적용된 표정영상에서 특징점 추출

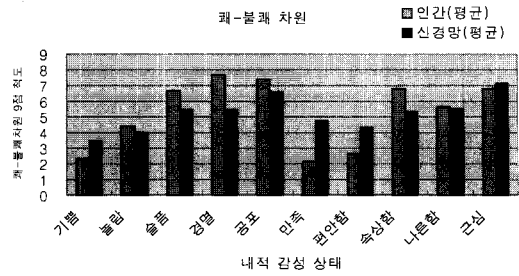
4. 표정인식 시스템

4.1 실험환경

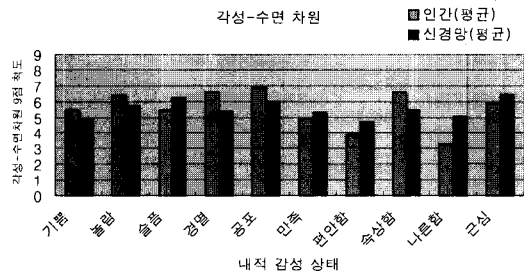
얼굴표정을 통한 표정인식 과정을 탐색하기 위해 인간 피험자를 대상으로 얻어진 데이터를 기반으로 신경망을 이용하여 표정인식을 수행하였다. 표정인식에 사용된 신경망구조는 세층으로 구성된다.

추출된 특징들은 표정을 인식하기 위하여 신경망의 입력층의 노드로 사용된다. 각 층에 있는 노드들의 출력은 0과 1사이에서 존재한다. 얼굴표정의 인식결과는 출력층으로 부터 얻어진다. 입력노드는 식(9)를 통해 표정영상에서 추출된 특징점들의 좌표값에서 무표정의 각각 대응되는 특징점들의 좌표값의 차이로 이루어진다. 따라서 입력노드는 60개의 특징점 각각에 대해 x, y 좌표값의 차이값으로 이루어진 120개의 입력노드로 형성되며, 0과 1사이의 값으로 정규화 되어있다. 은닉층은 240개의 노드로, 출력층은 2개의 노드로 이루어졌다. 출력노드의 수는 차원 모형을 기반으로 한 꽤-불쾌차원과 각성-수면차원을 반영한다. 출력층에서 얻어지는 2개의 값은 9점 척도의 값으로 변환되어 꽤-불쾌차원과 각성-수면차원상에 표현된다.

신경망의 학습방법은 오류 역전파 알고리즘을 사용하였다. 이 방법의 학습은 훈련되어 출력되는 값과 목표 값과의 오류를 은닉층에 역전파하여 수행하는 것으로 이 과정에서 연결 가중치를 조정한다. 이러한 연결 가중치 조절은 일반화된 델타 학습 규칙에 의해 수행된다. 은닉층의 활성화 함수(activation function)로 시그모이드 함수(sigmoid function)가 사용되었다. 은닉층과 출력층 사이의 평균자승오류가 전역적 최소값으로 수렴되었을 때, 더 이상의 가중치 갱신은 의미가 없으므로 학습이 완료된다. 따라서 훈련 데이터 집합의 1회전마다 그 전회전의 평균 자승 오류의 차가 0.00004 이하가 될 때까지 학습되었다.



(그림 7) 테스트데이터에 대한 꽤-불쾌 차원상의 비교



(그림 8) 테스트데이터에 대한 각성-수면차원상의 비교

<표 1> 학습에서 제외된 동일인의 다른 표정의 표정인식

내적감성 상태	쾌-불쾌차원		각성-수면차원	
	인간 (평균)	신경망 (평균)	인간 (평균)	신경망 (평균)
기쁨	2.35	3.48	5.47	4.88
놀람	4.38	3.97	6.4	5.72
슬픔	6.7	5.46	5.44	6.24
경멸	7.68	5.5	6.56	5.4
공포	7.38	6.61	6.83	5.96
만족	2.16	4.71	4.89	5.28
편안함	2.68	4.33	3.91	4.67
속상함	6.77	5.36	6.56	5.47
나른함	5.7	5.52	3.3	5.07
근심	6.77	7.14	5.87	6.47

4.2 표정인식결과 및 논의

신경망 학습을 위하여 250개의 얼굴표정이 사용되었으며, 학습에 사용되지 않은 40개의 얼굴표정이 인식을 위하여 사용되었다. 훈련 데이터 집합은 기쁨, 놀람, 슬픔, 경멸, 공포, 만족, 편안함, 속상함, 나른함, 근심으로 각각의 정서 어휘 범주에 25개씩 적용되며, 한 범주에 속하는

25개의 데이터 집합은 6사람 표정 데이터들로 구성된다. 신경망의 객관성을 검증하기 위하여 우선 신경망 학습에 사용되었던 250개의 표정데이터를 인식 실험 데이터로 사용하였다.

검증 실험결과에서 정서관련 어휘 분석을 통한 내적 감성 상태의 차원 모형 연구에서 얻어진 사람의 평정 결과와 신경망의 학습 결과는 100%에 가까운 학습률을 보였다. 검증을 위한 실험 후, 모든 피험자의 영상들로부터 학습에 사용되지 않은 표정 데이터 40개가 인식 실험에 사용되었다.

(그림 7)과 (그림 8)은 학습에서 제외된 동일한 사람의 다른 표정 데이터들의 각 차원 공간상에서 인식결과를 나타낸다. (그림 7)과 (그림 8)의 결과에서 패-불패 차원의 불패수준과 각성-수면차원에서 각성수준은 표정들을 변별하는 중요한 요인이 되었는데, 이는 얼굴에 나타나는 물리적인 변화를 탐지하는 데에 유리하기 때문이라고 해석된다. 또한 타인의 표정을 읽는 것은 두 차원들끼리의 조합에 의해 더욱 다양해지며, 정교화될 수 있다. 예를 들어, 높은 각성수준은 불패 차원 공간과의 조합에 의해 근심, 공포, 놀람과 같은 표정을 구별하는 중요한 요인이 되는 것으로 판단된다.

<표 1>은 테스트된 표정 데이터들의 인간 평정의 평균과 신경망 출력의 평균에 의한 각각의 두 차원상의 표정 인식 결과를 보이며, <표 2>는 Pearson 상관계수를 이용하여 패-불패 차원과 각성-수면 차원에서 인간평정과 신경망 출력의 상관정도를 나타낸다.

연구 결과에서 신경망의 인식결과는 인간의 평정결과와 정적상관을 보인다. 특히 패-불패 차원의 정서들은 각성-수면 차원의 정서들보다 안정적인 차원으로 설명할 수 있는 데, 이는 패-불패 차원 정서들이 유의도 .01수준에서 인간평정의 결과와 유의미한 상관 관계를 보이는 것으로 알 수 있다.

이러한 연구 결과는 김진관 등 [14]의 자신의 내적인 정서적 경험을 구분하고 해석하는 데 사용되었던 구조적 틀이, 표정을 통해 타인의 정서를 읽어내는 과정에서

<표 2> 인간평정의 평균과 신경망 출력의 평균의 Pearson 상관관계수 (**는 유의도 .01수준에서 유의미한 상관을 표시함)

dimension	correlation
패-불패차원	0.78**
각성-수면차원	0.47

도 유사하게 해석될 수 있음을 시사한다. 즉, 정서단어에 대한 인지적 평가를 이루는 과정에서는 패-불패 차원의 설명력이 높고 각성-수면차원의 설명력이 낮게 평가되었는데, 타인의 정서를 표정을 통해 추론하는 과정에서도 차원의 비중이 유사할 수 있다는 것이다.

다음은 신경망의 결과를 (그림 2)의 정서의 이차원 구조상의 감성어휘들과 유사도 비교를 수행하였다. 이는 정서의 이차원모형 공간에서 신경망의 표정인식이 어떻게 사상되는가를 살펴보기 위한 것이다. 유사도(S)는 다음과 같이 계산된다.

$$S(\vec{V}, \vec{M}) = \frac{\vec{V} \cdot \vec{M}}{\|\vec{V}\| \|\vec{M}\|} \min\left(\frac{\|\vec{V}\|}{\|\vec{M}\|}, \frac{\|\vec{M}\|}{\|\vec{V}\|}\right) \quad (10)$$

차원공간상에서 내적 감성 상태를 나타내는 어휘들 중 유사한 범주에 속하는 어휘들은 서로간에 매우 근접한 공간상에 배치되어 있다. 식(10)은 근접한 공간상의 내적 감성 상태를 동일한 범주로 인식 가능하게 한다. 각각의 두차원 상에서 인간피험자들에 의한 감성어휘 평정 값과 신경망의 인식결과에 대한 차원 값들이 \vec{V} 와 \vec{M} 벡터로 주어진다. <표 3>은 내적 감성 상태를 나타내는 어휘들의 평정 값과 신경망의 인식결과에 대한 유사도를 보여 준다.

<표 3>에서 신경망의 표정인식 결과는 내적 상태와 관련된 44 정서 어휘 중에서 가장 가까운 어휘로 대응될 수 있다. 인간 피험자들에 의한 내적상태 감성어휘 평정 값과 신경망에 의한 표정인식 결과는 매우 유사해 보인다. 특히 신경망의 불패 차원공간에서 높은 각성수준의 표정들은 높은 유사도를 보인다. 그러한 표정들로는 공포, 놀람, 속상함, 혐오 등이 있다. 이는 높은 각성의 불패 정서가 표정공간에서 상당히 보편화되고 안정된 정서로 판단된다. 또한 높은 각성수준은 패 차원 공간에서도 표정을 변별하는 중요한 요인이 되는 것으로 사료된다. 예를 들면, 만족, 나른함, 편안함 등이 패 정서에서 낮은 각성을 보이는 반면, 기쁨은 높은 각성 상태를 보인다.

편안함과 나른함과 같은 낮은 각성수준의 정서와 관련된 표정들은 눈과 입의 열린 정도가 작은 상태로 표현될 수 있다. 이러한 정서 표정들이 다른 표정들보다 인식되기 어려운 것은 무표정패턴과 표정패턴과의 변이가 작고 불확정적이기 때문이라고 판단되는데, 이는 얼굴의 주요 요소성분들의 물리적인 변화량을 감지하기 어렵게 할 수 있다. 이와는 대조적으로 경멸, 공포 등은 높은 각성수준을 보이는데, 이것은 얼굴의 주요 요소성분들의 무표정패턴과 표정 패턴과의 변이가 크기 때문에 물리적인 변

<표 3> 두 차원 공간상에서 인간의 정서어휘 평정값과 신경망의 표정인식 결과

감성 어휘	감성어휘 평정값		신경망		신경망 인식	유사도
	쾌-불쾌	각성-수면	쾌-불쾌	각성-수면		
기쁨	1.65	7.53	3.88	3.44	홀가분한	0.54
			4.92	4.6	심심한	0.71
			2.86	5.86	기분좋은	0.82
			1.31	5.69	뿌듯한	0.75
			4.43	4.8	그리운	0.73
만족	1.85	4.65	1.49	6.07	기분좋은	0.79
			2.14	4.96	흐뭇한	0.92
			7.08	3.43	울적한	0.47
			6.32	5.9	부끄러운	0.52
			6.53	6.06	부끄러운	0.50
편안함	2.61	2.98	5.0	5.7	이상한	0.52
			3.65	3.64	홀가분한	0.77
슬픔	7.22	6.57	7.07	5.23	부끄러운	0.89
			3.70	6.37	기대	0.72
			6.62	7.12	놀라운	0.91
나른함	5.44	2.2	7.94	6.29	긴장	0.55
			4.06	4.05	졸리운	0.90
			4.39	4.28	그리운	0.89
			4.8	5.09	이상한	0.76
			6.39	5.65	안타까운	0.65
근심	7.4	5.96	6.89	6.09	당황스런	0.97
			7.39	6.84	긴장	0.94
놀람	4.65	7.8	4.55	8.29	놀람	0.95
			4.61	7.67	놀람	0.98
			4.65	5.60	기대	0.79
경멸	7.93	6.74	6.35	3.42	쓸쓸한	0.68
			7.33	6.14	미워하는	0.91
			7.68	6.03	속상한	0.98
			3.89	2.63	안정	0.45
			6.05	6.72	놀람	0.86
공포	7.25	6.77	6.75	4.49	미안한	0.80
			6.43	5.21	답답한	0.83
			6.68	7.97	경멸	0.94
			7.30	7.96	억울한	0.91
			5.91	4.17	쓸쓸한	0.72
속상함	7.46	6.29	7.48	7.16	경멸	0.94
			4.28	5.81	희망	0.72
			4.77	4.97	심심한	0.70
			5.60	4.11	심심한	0.71
			5.81	5.05	이상한	0.79

화량을 감지하기가 쉽다고 할 수 있다.

그러나 이러한 정서 표정들에 대한 신경망 접근은 한 개의 차원만으로는 적절히 해석되어지지 않는다. 즉, 동일한 차원상에서 유사한 분포를 보이는 다른 정서들의 표정들 때문이다. 경멸과 공포는 유사한 각성수준의 정서 관련 표정이지만 쾌-불쾌 차원과의 조합을 통해 구별되어지며, 슬픔은 만족, 편안함과 같은 유사한 수면수준의 정서관련 표정이지만 쾌-불쾌 차원과의 조합을 통해 변별되어진다. 특히 높은 불쾌정서는 높은 각성수준과 더불어 얼굴의 물리적인 변화량에 많은 영향을 미칠 것으로 판단된다.

이는 인간의 평정과 비교적 유사한 결과를 보이는 공포, 놀람 등의 정서 관련 표정들은 대부분이 높은 불쾌정서를 동반하기 때문이다. 특히 이러한 불쾌 정서들은 쾌 정서들보다 무표정패턴과 표정패턴과의 변이가 더 확정적일 것으로 해석되는데, 이는 기쁨, 만족, 편안함 등이 공포나 놀람 보다 인간평정 결과와의 비교에서 더 큰 차이를 보이기 때문이다.

타인의 표정을 지각하고 추론, 해석하는 데 있어서 높은 불쾌 정서와 높은 각성 수준의 정서를 반영한 표정들인 근심, 경멸, 공포, 속상함, 놀람 등은 보편적으로 인식이 잘되는 것으로 평가할 수 있는데, 이는 얼굴의 주요 요소들의 무표정패턴과 표정패턴과의 변이가 크고 확정적이기 때문이라고 분석되어진다. 신경망을 이용하여 타인의 표정을 읽어내는 것은 내적 체계 공간에서의 차원들끼리의 조합에 의해 인간과 유사하게 인식함을 확인할 수 있었다. 따라서 내적 체계 공간에서의 차원들끼리의 조합에 의한 표정인식 방법은 기존의 기본정서모형에 의한 범주적인 표정 분류로 연속적인 정서상태의 변화시 변화강도가 반영되지 않는 제한점을 극복하며 확장시킬 수 있음을 알 수 있다.

5. 결론

본 논문에서는 정서의 이 차원 구조를 이용하여 기존의 기본 정서이론을 기반으로 한 표정인식의 한계성을 극복하고 확장할 수 있는 새로운 얼굴 표정인식 시스템을 제안하였다.

얼굴표정을 이루는 자동화된 특징점 추출은 인간의 시각 피질에 있는 수용장과 비슷한 모형으로 생물학적인 연관성을 갖는 Gabor 웨이브렛 변환을 통하여 이루어졌다. 특정 주파수와 방향을 갖는 Gabor 웨이브렛 벡터들의 이웃하고 있는 화소간의 높은 상관성과 중복성을 분석하여 히스토그램의 평균값을 적용함으로써 얼굴의 주요

요소성분들이 추출되었다. 이러한 Gabor 계수들의 특성은 추출된 얼굴요소성분을 이루는 경계선들로부터 대표성 벡터를 추출할 수 있는 FCM 군집화 알고리즘을 통하여 무표정영상에서 대표적인 특징점 추출을 가능하게 하였다. 무표정 영상의 특징점들은 표정의 변화 정도를 측정할 수 있는 기준으로 각각의 피험자들의 표정영상으로부터 특징점 추출을 위한 템플릿으로 사용되었다.

무표정영상의 특징점에서 표정영상의 특징점으로서의 일대일 대응은 두 영역에서 대응점간의 위상과 Gabor 젯(jet)들의 유사도 측정을 통하여 이루어졌다. 특히 대응점간의 정합은 FCM 군집화를 통한 대표 벡터의 특성을 적용하여 개략적인 정합과 정밀한 정합의 두단계를 통하여 빠르게 수행될 수 있었다. 추출된 특징점들은 신경망의 입력벡터로 사용되어 정서의 두 차원 공간상에서 하나의 표정으로 인식되어졌다.

동일인의 다른 표정들의 인식에서 신경망의 평균적인 인식결과를 통해, 쾌-불쾌 차원에서 각성수준의 정도는 만족, 나른함, 편안함, 기쁨, 근심, 공포, 슬픔, 놀람 등의 표정들을 변별할 수 있는 중요한 요인으로 해석되었다. 유사한 각성-수면차원에서는 쾌-불쾌 차원과의 적절한 조합을 통하여 더욱 다양하고 정교한 표정들로 구별되어질 수 있었다. 타인의 표정을 지각하고 추론, 해석하여 읽어내는 것은 얼굴의 물리적인 변화량에 더욱 민감하게 반응하며, 그에 따른 변화량을 측정하는 것으로 해석되며, 개인의 정서적 경험을 해석하는 것만큼 정교하지는 않을 것으로 추론된다.

본 연구의 결과는 두 차원 공간상에서 내적 상태와 관련된 대표성 정서어휘들을 이용한 표정인식을 통하여 기존의 기본 정서이론을 기반으로 한 제한된 표정인식의 한계를 극복하였으며, 자연스럽게 다양한 표정인식을 위한 토대를 제공하였다. 본 연구가 범용적인 표정인식 시스템이 되기 위해서는 타인에 대한 영상 및 기술어집이 있는 변환영상에서도 인식 가능한 시스템으로 확장되어야 할 것이다.

참고문헌

- [1] Ekman, P. & Friesen, W.(1978), *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press.
- [2] Mase, K.(1991), Recognition of facial expression from optical flow, *IEICE Transactions*, E 74, 10, 3473-3483
- [3] Pentland, A. P., Etocoff, N., Stamer, T.(1992), Expression recognition using eigenfeatures, MIT Media

Laboratory Vision and Modeling Group Technical Report No. 194.

- [4] Essa, I., Pentland, A.(1997), Coding, analysis, interpretation, and recognition of facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19** 757-763
- [5] Lien, J.(1998), Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity, Ph.D. Thesis, Carnegie Mellon University.
- [6] Oliver, N. Pentland, A., Berard, F.(2000), LAFTER-a real-time face and lips tracker with facial expression recognition, *Pattern Recognition* **33**, 1369-1382
- [7] Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S.(2003), Facial expression recognition from video sequence: temporal and static modeling, *Computer Vision and Image Understanding*, In Press
- [8] Carrol, J. M. & Russell, J. A.(1996), Do facial expressions signal specific emotions? Judging emotion from the face in context, *Journal of Personality and Social Psychology*, **70**(2), 205-218
- [9] Russell, J. A.(1994), Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, **115**(1), 112-141
- [10] 반세범, 한재현, 정찬섭(1997), 표정-내적상태 상관모형을 위한 얼굴 DB 구축, '97 한국 감성과학회 연차학술대회 논문집, 215-219
- [11] 김영아, 김진관, 박수경, 오경자(1997), 정서관련 어휘분석을 통한 내적상태의 차원연구, '97 한국감성과학회 연차학술대회 논문집, 209-214
- [12] Daugman, J.(1985), Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Opt. Soc. Amer.*, **2**, 1160-1169
- [13] Bezdek, J. C.(1973), Fuzzy mathematics in pattern classification. PhD thesis. Applied Math. Center, Cornell University, Ithaca.
- [14] 김진관, 문혜신, 오경자(1999), 감성개념 이차원 구조의 안정성, *감성과학논문지*, **2**(1), 43-52