

얼굴 특징점의 지각적 위계구조에 기초한 표정인식 신경망 모형

A neural network model for recognizing facial expressions based on perceptual hierarchy of facial feature points

반 세 범* 정 찬 섭**
(Sei-Bum Ban) (Chan-Sup Chung)

요 약 얼굴 특징점의 지각적 위계구조를 반영한 표정인식 신경망 모형을 설계하였다. 입력자료는 MPEG-4 SNHC (Synthetic/Natural Hybrid Coding)의 얼굴 정의 파라미터(FDP) 중 39개 특징점 각각에 대해, 150장의 표정연기 사진을 5개의 크기와 8개의 방위를 갖는 Gabor 필터로 분석한 값이었다. 표정영상에 대한 감정상태 평정 값과 39개 특징점의 필터 반응 값을 중다 회귀분석한 결과, 감정상태의 쾌-불쾌 차원은 주로 입과 눈썹 주변의 특징점과 밀접한 관련이 있었고, 각성-수면 차원은 주로 눈 주변의 특징점과 밀접한 관련이 있었다. 필터의 크기는 주로 저역 공간 주파수 필터와 감정상태가 관련이 있었고, 필터의 방위는 주로 비스듬한 사선 방위와 감정상태가 관련이 있었다. 이를 기초로 표정인식 신경망을 최적화한 결과, 원래 1560개(39x5x8) 입력요소를 400개(25x2x8) 입력요소로 줄일 수 있었다. 표정인식 신경망의 최적화 결과를 사람의 감정상태 평정과 비교하여 볼 때, 쾌-불쾌 차원에서는 0.886의 상관관계가 있었고, 각성-수면 차원에서는 0.631의 상관관계가 있었다. 표정인식 신경망의 최적화 모형을 기쁨, 슬픔, 놀람, 공포, 분노, 혐오 등의 6가지 기본 정서 범주에 대응한 결과 74%의 인식률을 얻었다. 이러한 결과는 사람의 표정인식 원리를 이용하면 작은 양의 정보로도 최적화된 표정인식 시스템을 구현할 수 있다는 점을 시사한다.

주제어 표정인식, 특징점, 감정상태, 신경망, Gabor 필터, MPEG-4

Abstract Applying perceptual hierarchy of facial feature points, a neural network model for recognizing facial expressions was designed. Input data were convolution values of 150 facial expression pictures by Gabor-filters of 5 different sizes and 8 different orientations for each of 39 mesh points defined by MPEG-4 SNHC (Synthetic/Natural Hybrid Coding). A set of multiple regression analyses was performed with the rating value of the affective states for each facial expression and the Gabor-filtered values of 39 feature points. The results show that the pleasure-displeasure dimension of affective states is mainly related to the feature points around the mouth and the eyebrows, while arousal-sleep dimension is closely related to the feature points around eyes. For the filter sizes, the affective states were found to be mostly related to the low spatial frequency, and for the filter orientations, the oblique orientations. An optimized neural network model was designed on the basis of these results by reducing original 1560(39x5x8) input elements to 400(25x2x8). The optimized model could predict human affective rating values, upto the correlation value of 0.886 for the pleasure-displeasure, and 0.631 for the arousal-sleep. Mapping the results of the optimized model to the six basic emotional categories (happy, sad, fear, angry, surprised, disgusted) fit 74% of human responses. Results of this study imply that, using human principles of recognizing facial expressions, a system for recognizing facial expressions can be optimized even with a relatively little amount of information.

* 연세대학교 심리학과, 현재 성균관대학교 심리학과 BK21 박사후 과정.
E-mail: stuntman@psylab.yonsei.ac.kr.
** 연세대학교 심리학과

E-mail: cschung@yonsei.ac.kr.
이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2000-005-C00011).

1. 서론

표정으로부터 내적 감정상태를 추정하는 시스템을 개발하기 위해서는 영상영역과 심리영역을 연결하는 해석체계가 필요하다. 이러한 해석체계를 만들기 위해서는 먼저 다양한 표정자료를 수집하여, 사람들이 각 표정에 대해 어떤 심리적 해석을 내리며 그러한 해석의 원리가 되는 영상의 구성적 특징들이 무엇인가를 조사하여야 한다. 최근, 얼굴 및 표정인식에 대한 연구관심이 높아지면서 표정인식 기법과 모형의 개발에 커다란 발전이 있었지만 이와 같은 종합적인 접근을 시도한 예는 드물다. 대부분의 연구들이 임의로 자체 촬영한 소수의 표정 자료와 네 개 내지 여섯 개의 감정상태 기술 형용사들을 연결짓는 시스템을 개발하여 선보이고 있다. 이러한 제한점을 극복하기 위하여 정찬섭, 오경자, 이일병, 변혜란 (1998)은 영화 및 텔레비전 연속극에 나오는 배우들의 표정연기 사진과 연극 배우들이 연기한 대규모의 사진자료를 수집하여 이를 내적 감정상태에 대응시킨 자료집(DB)을 개발하였다. 이들은 사람들의 평정작업과 군집분석, 의미변별분석, 중다차원 분석 등을 통하여 내적 감정상태를 기술하는 범주모형과 차원모형을 만들고 그에 따라 자료집의 모든 표정을 분류 또는 좌표화 함으로써 경험적 연구에 기초하여 표정과 감정상태를 잇는 해석체계를 개발하였다. 따라서, 우리는 이 자료집을 표정인식 시스템 개발의 출발점으로 설정하고 그것에 포함된 표정과 내적 감정상태 자료를 이용하여 영상의 구성적 특징을 찾아내기 위한 영상처리 작업과 그 결과를 해석하는 모형의 개발을 시도하였다.

얼굴 또는 표정 인식에 현재 다양한 영상처리 기법이 사용되고 있다. 각 기법들이 나름대로의 장점을 지니고 있지만 이 연구에서는 인간의 시각피질에 있는 단순세포(simple cell)와 유사한 연산특성을 지니고 있는 여러 규모의 Gabor wavelet이 국소 영상의 특징을 위계적 및 효율적으로 포착할 수 있다는 가정 하에 그 연산 결과를 영상처리의 기본 단위로 채택하였다. 영상처리결과를 내적 감정상태로 잇는 모형은 표정인식이 비선형적인 해를 요한다는 가정 하에 신경망 모형을 사용하기로 하였다. 이러한 접근법의 특성과 장점을 이해하기 위해서는 다른 기법과 모형을 사용하는 접근법과 비교를 통하여 표정인식의 문제를 얼마나 잘 해결할 수 있는지를 살펴 볼 필요가 있다.

표정은 주어진 얼굴에 포함된 특징요소들의 변형을 통하여 발현된다. 얼굴의 특징요소들은 표정을 결정하

는 데에 서로 다른 비중을 지니며, 하나의 특징요소도 표정에 따라 상대적인 중요도가 달라질 수 있다. 예를 들어, 유쾌한 표정을 나타낼 때에는 입의 양끝 부분이 결정적이고, 불쾌한 표정에는 양 눈썹 사이의 주름이 결정적이다. 따라서 표정인식 시스템을 설계하는 데에 있어서 상대적으로 중요한 특징요소를 선별하여 분석하고, 각 표정별로 특징요소들의 상대적 기여도를 반영하는 것이 시스템의 효율성과 민감성을 높일 수 있는 관건이 된다. 기존의 표정인식 시스템은 지형적 특징요소 기반 시스템, 형판정합 기반 시스템, 주성분 분석(PCA) 기반 시스템, Gabor wavelet 기반 시스템의 네 가지로 크게 나누어 볼 수 있는데 이들은 각기 특징요소의 선별과 표정별 특징요소의 기여도를 반영하는 비중이나 방법에서 차이를 보인다.

지형적 특징요소 기반 시스템에서는 입력영상으로부터 눈, 코, 입 등 얼굴의 주요 요소들을 구성하고 있는 소수의 특징점들(feature point)을 탐지한 후, 중성 얼굴로부터의 각 요소 내 및 요소 간 특징점 사이의 거리와 각도 변화를 통하여 표정을 인식한다 (예를 들어, Zhang, Lyons, Schuster와 Akamatsu, 1998). 특징요소에 기초한 접근방법은 복잡한 입력영상을 소수의 특징점 집합으로 압축하여 정보량과 연산 부하를 획기적으로 감소시켜주는 장점이 있지만 표정 인식에 유효한 많은 단서들이 누락되는 문제점도 아울러 지니게 된다. 즉 소수의 특징점 집합만으로는 각 특징점 주변의 얼굴구조 변화를 충분히 포착할 수 없다는 것이다. 이에 덧붙여 어느 특징점들이 표정인식에 더 중요한지에 대한 객관적인 근거가 아직 확립되어 있지 않아서 연구자들마다 주관적으로 선정한 서로 다른 특징점 집합을 사용한다는 문제점도 있다. 대부분의 연구자들은 이러한 이유 때문에 지형적 특징요소 기반 시스템의 문제들을 보완하는 방법의 하나로서 광류(optic flow) 정보나 주성분 분석을 함께 적용하는 복합적인 접근방법을 사용하고 있다(Yacoob과 Davis, 1994; Bartlett, Sejnowski, Larsen, Hager와 Ekman, 1996; Lien, Kanade, Cohn과 Li, 1998).

형판정합 기반 시스템에서는 여러 사람의 표정을 표준화하여 몇 가지 표정의 프로토타입 형판을 만든 후, 입력영상을 그것과 대조하여 표정을 인식한다. 일반적으로 형판 기반 시스템은 특징요소가 변형된 지역적 정보가 반영되므로 얼굴 특징요소를 찾는 과정이 필요하지 않으며, 얼굴 특징요소 정보를 이용하기 위한 규칙의 최적화를 고려하지 않아도 된다는 장점을 지닌다. 그러나 이 방법은 형판모형이 갖는 제약점을 그대

로 지너 제한된 수의 형판으로 다양한 얼굴의 표정에 일반화시켜 적용하기 힘들다는 점, 얼굴의 기울기나 회전에 취약하다는 점등의 이유 때문에 보안장치와 같이 제한된 사람의 얼굴을 인식할 수 있는 시스템에는 사용되지만 표정인식시스템에는 거의 사용되지 않는다. 따라서, 표정인식의 경우에는 얼굴표정 영상 자체를 형판화하기보다는 표정변화에 수반된 운동에너지의 분포와 같은 동적정보를 추상화 한 형판을 사용하기도 한다 (Essa와 Pentland, 1995).

주성분 분석(Principal Component Analysis) 기반 시스템은 원래 KL(Karhunen-Loeve) 변환이라는 영상 압축 기법을 이용한 것으로 입력영상을 2차원 벡터로 변환하여 직교적인 공분산 행렬을 구한 다음, 주성분 분석을 하여 고유값을 토대로 그 영상을 대표하는 소수의 고유벡터(또는 고유 영상)를 구한다. 입력영상으로부터 고유값이 비교적 큰 소수의 영상벡터가 구해지면 이들의 가중조합에 의해 원래에 근접한 영상을 복원할 수 있다. 고유값이 큰 영상벡터는 일반적으로 얼굴의 기본적인 구조나 공통요소와 관련된 정보를 포함하고 있기 때문에 성별이나 인종과 같은 얼굴범주화에 유용한 정보를 제공하며 고유치가 작은 고유벡터는 개별 얼굴에 특징적인 정보를 전달하기 때문에 서로 다른 얼굴을 구별하는 데에 유용한 정보를 제공한다(O'toole, Abdi, Deffenbacher와 Valentin, 1993). 주성분 기반 시스템은 모든 얼굴의 보편성과 개별 얼굴의 독특성을 위계적으로 포착한 고유 얼굴을 추출해내기 때문에 형판 정합 기반 시스템에서보다 작은 양의 정보로 자료 압축을 할 수 있어 얼굴인식에 널리 사용되고 있다. 그러나 표정인식의 경우에는 개별 얼굴에 특징적인 고역 공간 주파수 정보가 결정적인 역할을 하기 때문에, 주로 고유치가 큰 고유얼굴을 사용하는 주성분 분석으로는 표정정보를 효율적으로 포착할 수가 없다(Donato, Bartlett, Hager, Ekman과 Sejnowski, 1998). 또한 이 시스템은 영상의 방향과 조명변화에는 영향을 덜 받지만 크기나 해상도 변화에는 영향을 많이 받기 때문에 입력자료가 표준화되고 최적화되어 있지 않을 때는 수행력이 저하되는 결점을 지닌다(Chellapa, Sirohey, Wilson과 Barnes, 1995).

Gabor wavelet 기반 시스템은 지금까지 논의된 기존 시스템들이 지니는 많은 문제점을 수용하여 해결할 수 있다. 시각 대상의 방위와 크기에 선별적인 인간의 시각피질 세포와 비슷한 연산특성을 지닌 Gabor 필터는, 지역적인 특징요소의 변화를 여러 방위에 걸쳐 포착할 수 있으며, 여러 규모의 필터 크기를 사용

함으로써 한 영역내의 영상정보를 위계적으로 포착할 수 있다. Gabor wavelet 접근법에서는 하나의 모함수 (mother wavelet)에 포함된 두 개의 변수, 즉 필터 방위와 크기값만을 조정함으로써 다중해상 다중방위의 필터를 구성할 수 있기 때문에, 저역 공간 주파수 필터로는 형판정합과 같은 전역적인 분석이 가능하고, 고역 공간 주파수 필터로는 지형적 특징요소 추출과 같은 지역적인 분석이 가능하다. 이러한 유연성과 효율성 때문에 많은 연구자들은 Gabor 필터를 입력 영상의 분석에 가장 효율적인 도구로 간주하고 있다 (Field, 1987).

Gabor 필터를 사용하는 경우, 몇 가지의 필터를 몇 개의 점에 적용해야 하는가의 문제가 생긴다. 영상의 모든 픽셀에 대해 필터를 적용하면 완벽한 정보가 얻어지겠지만 연산부하와 정보중복(redundancy)이 매우 커지게 되기 때문에 보통 적정 간격을 지닌 사각형 격자(grid)의 교점이나(Lades, Vorbruggen, Buhman, Lange, Malsburg, Wurtz와 Konen, 1993), 얼굴의 특징점에 대해서만 여러 규모 및 방위의 필터를 적용하는 것이 보통이다. 후자의 경우, 연구자에 따라 선정된 특징점의 위치 및 개수와 필터의 종류 및 개수가 다를 수 있다. 예를 들어, Wiskott, Fellous, Kruger와 Mlasburg(1997)는 45개의 특징점에 대해 40가지 (크기 5 X 방위 8)의 필터를 적용한 반면, Zhang 등 (1998)과 Lyons, Akamatsu, Kamachi와 Gyoba (1998)는 34개 특징점에 대해 18가지(크기 3 X 방위 6)의 필터를 적용하였다. 따라서 얼굴 특징점에 필터를 적용하는 경우에는 특징점과 필터 가지 수 선정에 타당한 근거가 있어야 한다.

Gabor wavelet기반 시스템은 필터가 적용되는 점 주위의 얼굴 구조를 여러 규모에 걸쳐 포괄적으로 포착하기 때문에 국소적으로 전역적인(locally global) 해법을 제공한다. 따라서, 다른 접근법들의 문제점으로 지적됐던 소수의 특징점이나 형판으로 미묘한 표정 변화를 포착하는데 따르는 어려움, 형판정합의 결과가 얼굴의 회전이나 기울기에 민감하게 변화되는데 따르는 어려움을 완화해 줄 수 있다. 엄밀한 의미에서 Gabor wavelet도 하나의 형판이지만, 지형적 특징요소 기반 시스템에서 보듯이 너무 단순화된 점의 분포에 의존하거나 형판정합 또는 PCA 기반 시스템에서 보듯이 너무 복잡한 필터에 의존하는 대신 그 중간 수준에 속하는 영상의 기초특징(primitive features)을 사용하기 때문에 이러한 바람직한 속성들을 보장받는다 할 수 있다. 그러나, Gabor wavelet 기반 시스템에는

지형적 특징요소 기반 시스템과 같이 어느 위치에, 몇 개의 특징점(이 경우에는 convolution 적용점)을 사용해야 하는가를 결정해야 하는 어려움이 수반된다. 이에 추가하여, 시스템의 성능을 최적화하기 위해서는 몇 종류 및 몇 개의 필터를 사용해야 하는지를 결정해야 하는 것도 이 접근법에서 선결되어야 하는 문제들 중 하나이다. Gabor wavelet을 이용한 기존 연구들의 대부분이 이와 같은 문제점들에 대한 뚜렷한 해결책을 제시하고 있지 못하기 때문에 이 접근법이 보편적인 지지 기반을 얻기 위해서는 이에 대한 체계적인 공략이 필요하다.

2. 표정인식을 위한 특징점의 선정 및 Gabor wavelet 변환

표정인식 시스템을 최적화하기 위해서는 감정상태와 밀접한 얼굴 특징점을 선정해야 한다. 지금까지의 연구들에서는 일반적으로 얼굴 및 표정인식을 위해서는 눈, 눈썹, 코, 입 등에서 추출한 20개에서 40개의 특징점이 사용되고 있다. 그러나 이들 특징점들은 대개 연구자의 주관적 판단에 따라 임의 선정된 것으로 표정인식과 관련된 그것들의 상대적 중요성이 분명하지 않다. 상대적으로 중요한 특징점만을 선별하려면 그것을 합리화할 수 있는 선정 표준이 필요하다. 이와 관련하여, 현재 화상통신 분야에서 표준규약으로 제안되고 있는 MPEG-4 SNHC(Synthetic/Natural Hybrid Coding)의 얼굴 정의 파라미터(FDP) 및 얼굴 애니메이션 파라미터(FAP)는 얼굴표정을 결정하는 특징점의 선정에 적합한 요건을 갖추고 있다. 따라서, 이들 얼굴 특징점 중에서, 표정인식에 상대적으로 중요한 점들을 선정하여 위계적으로 Gabor 필터들을 구성하는 것이 신뢰롭고 타당한 표정정보의 수집 가능성을 높이는 하나의 방법이 될 수 있다.

2.1 MPEG-4 FDP 기반 특징점 선정

이 연구에서는 얼굴에 필터를 적용하기 위한 특징점들을 정찬섭 등(1998)의 DB를 참조하여 선정하였다. 이 DB에 있는 표정자료는 FDP/FAP 특징점 84개중 그림 1에 나와 있는 것과 같이 얼굴외곽선, 귀, 머리상부 및 이면 등 표정과 비교적 관계가 없는 점들을 제외한 39개의 점에 대한 2차원 좌표 값을 포함하고 있다. 그림에서 볼 수 있듯이 얼굴 윤곽을 나타내기 위한 소수의 점들을 제외하면 대부분의 점들이 눈, 눈썹, 코, 입의 모양을 규정하도록 되어 있다.

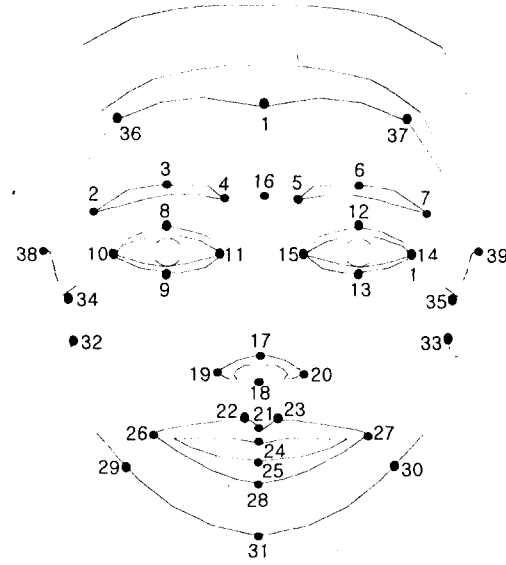


그림 1. FDP 특징점 중 본 연구에서 선정된 39개의 특징점.

2.2 Gabor wavelet 변환

표정 영상의 39개 특징점 각각에 대해 40개의 필터(크기 5 X 방위 8)를 적용하여 하나의 영상을 Gabor 제트(jet)로 나타내도록 하였다. 즉, 영상 $I(\vec{x})$ 에서 주어진 특징점 $\vec{x} = (x, y)$ 의 위치에 다중해상 다중방위의 Gabor 필터를 convolution 시켜서, 모든 필터 반응값으로 구성된 Gabor 제트 $J(\vec{x})$ 를 얻었다. 공식 1은 j번째 필터 반응값을 나타낸다.

$$J_j(\vec{x}) = \int I(\vec{x}') G_j(\vec{x} - \vec{x}') d^2\vec{x}' \quad (\text{공식 1})$$

이 공식의 Gabor 웨이블릿 G_j 는 공식 2에 있는 Daugman(1985)의 convolution kernel 이다.

$$G_j(x, y) = g_j(x, y) \exp\{i\omega_j[x \cos \theta_j + y \sin \theta_j]\} \quad (\text{공식 2})$$

여기서, ω 와 θ 는 변조 주파수와 방위이며 가우시안 윈도우 g 는 공식 3과 같다.

$$g_j(x, y) = \exp\left\{-\frac{1}{4\pi\sigma^2}[x^2 + y^2/a^2]\right\},$$

$$(x', y') = (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta) \quad (\text{공식 3})$$

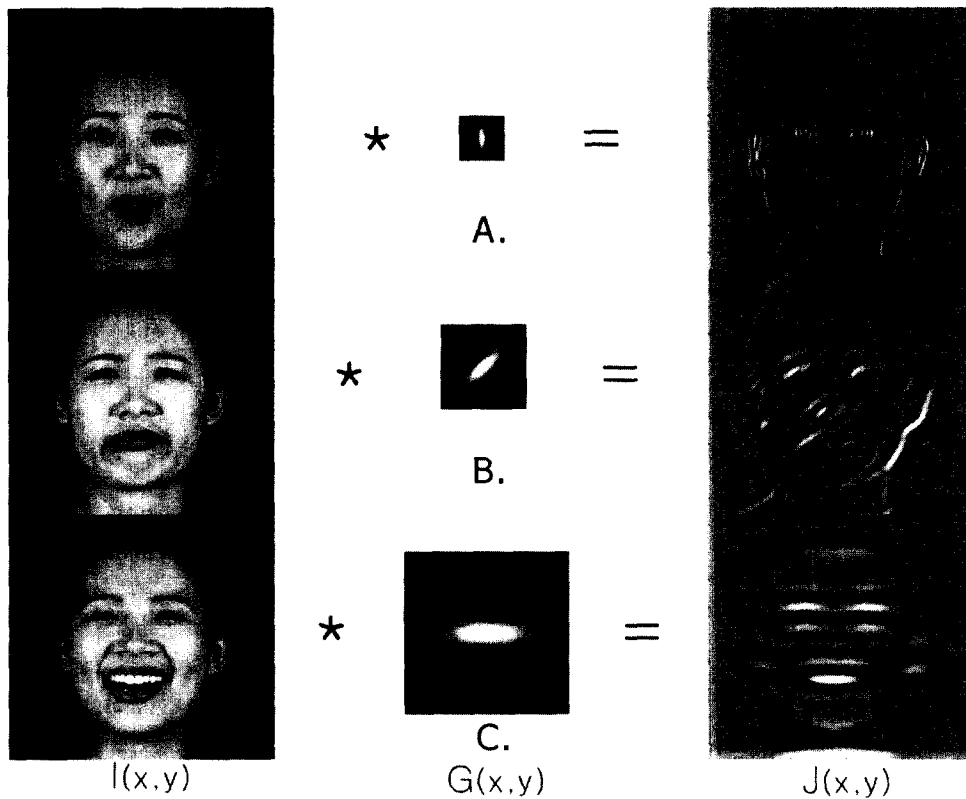


그림 2. Gabor wavelet 변환. 표정영상 I에 대해 다중해상 다중방위의 Gabor 필터 G를 적용하면 필터 반응 값 J를 얻어낸다.
 A. 필터크기 20화소, 필터방위 0도 B. 필터크기 40화소, 필터방위 45도
 C. 필터크기 80화소, 필터방위 90도

공식 3에서 x' 와 y' 는 θ 만큼 회전된 방위의 영상과 표이며, σ 와 a 는 가우시안 윈도우의 크기와 이심률(eccentricity)이다. 그림 2는 공식 2의 Gabor wavelet 중 우함수 부분만을 이용하여 얻은 하나의 표정영상에 대한 다중해상 다중방위의 필터 적용 결과이다. Gabor wavelet은 공식 2에 있는 것과 같이 분리 가능한 우함수와 기함수로 이루어져 있는데 이 연구에서는 우함수 값(G+)과 기함수 값(G-)의 합인 Gmag을 사용하였다. 이 두 우함수와 기함수의 값은 공식 4를 이용하여 계산하였다.

$$\begin{aligned}
 G_+ &= g(x, y) \cos[\omega_x(x \cos \theta_j + y \sin \theta_j)], \\
 G_- &= g(x, y) \sin[\omega_x(x \cos \theta_j + y \sin \theta_j)], \\
 g(x, y) &= \exp\left\{-\frac{\lambda^2 \omega_j}{4\pi\sigma^2} [x^2 + y^2]\right\} \quad (\text{공식 4})
 \end{aligned}$$

그림 3은 우함수와 기함수의 필터 쌍들이 지닌 위상 차이를 최소화하기 위하여, 그 두함수로 얻은 2개의 영상을 합쳐 하나의 Gabor 강도(magnitude)로 나타낸 것이다. 이 그림은 Gmag 내 각 영역의 밝기를 256단계의 흑백도(gray level) 값으로 변환한 것이다.

2.3 Gabor Jet의 최적화

Gabor wavelet을 사용하여 표정 영상을 처리하면 대상의 특징구조를 효율적으로 표상할 수 있다. 그러나 Gabor wavelet을 사용한 영상처리에는 복잡한 계산과정이 포함되어 있기 때문에, 모든 얼굴 영역을 처리하기 위해서는 많은 시간과 대용량의 정보 저장이 필요하다. 이러한 처리시간과 용량한계를 줄이는 방법 중의 하나는 표정인식에 중요한 얼굴 특징점과 필터만을 선정하여 영상처리를 하는 것이다. Gabor

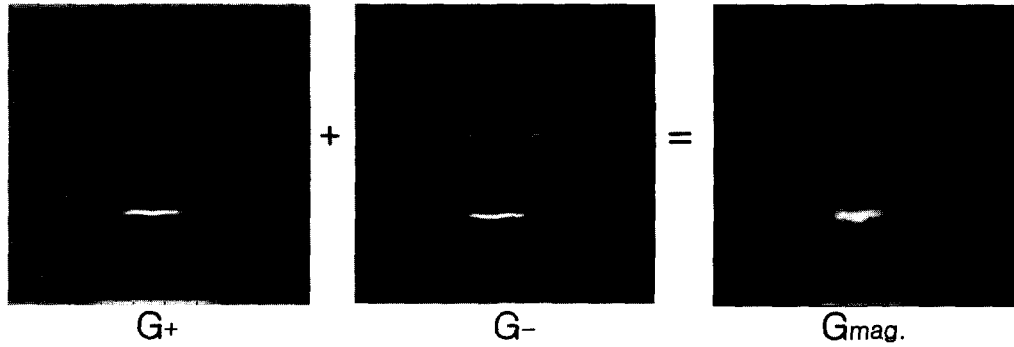


그림 3. Gabor magnitude 계산.

복소 Gabor wavelet 중 cosine함수 G+ 필터와 sine함수 G- 필터를 영상에 컨벌루션시킨후 두 개를 합산하여 Gabor magnitude 영상을 생성한다.

wavelet을 사용하면 한 특징점의 주변 영역을 서로 다른 필터를 적용하여 포괄적으로 표상할 수 있어서 39개의 특징점과 40개의 필터보다 적은 연산으로도 얼굴을 효율적으로 표상하는 것이 가능할 수 있다. 이와 같은 가능성, 즉 39개의 특징점과 40개의 필터 중 표정인식에 상대적으로 중요한 것이 무엇인지를 회귀분석을 통해 알아보았다. 회귀분석의 독립변수는 1560개(얼굴 특징점 39 X 필터 40)의 필터 연산 값이었고, 종속변수는 정찬섭 등(1998)의 DB에 수록되어

있는 각 표정 영상에 대한 사람들의 쾌-불쾌, 각성-수면 평균 평정 값이었다.

2.3.1 얼굴 특징점에 대한 Garbo 필터 연산 값과 감정상태의 중다 회귀분석

표정인식에는 어떠한 얼굴 특징점이 상대적으로 중요한가를 알아보기 위해 회귀분석을 하였다. 그림 4는 회귀분석 결과를 토대로 39개 특징점의 상대적 중요성을 내적 감정상태의 쾌-불쾌 차원과 각성-수면 차

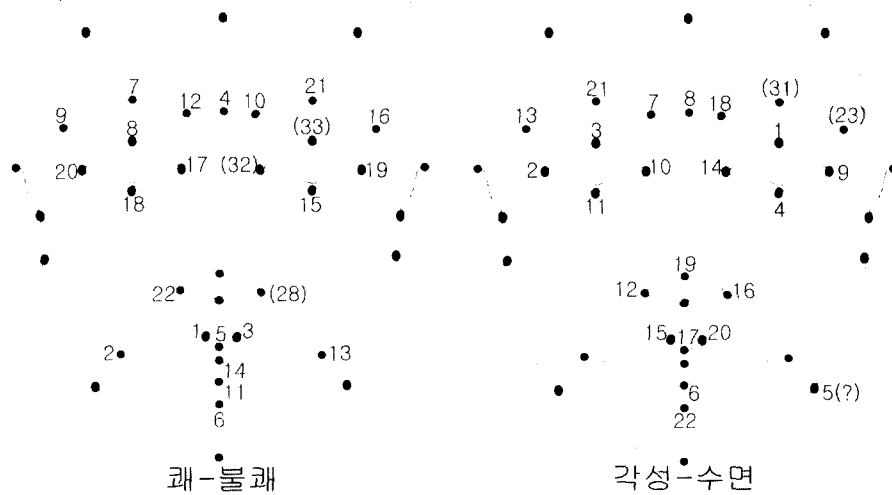


그림 4. 얼굴 특징점 39개의 상대적 중요성.

특징점의 숫자는 회귀분석 결과 감정 상태를 예측하는 설명량(R²)이 큰 순서를 가리킨다.

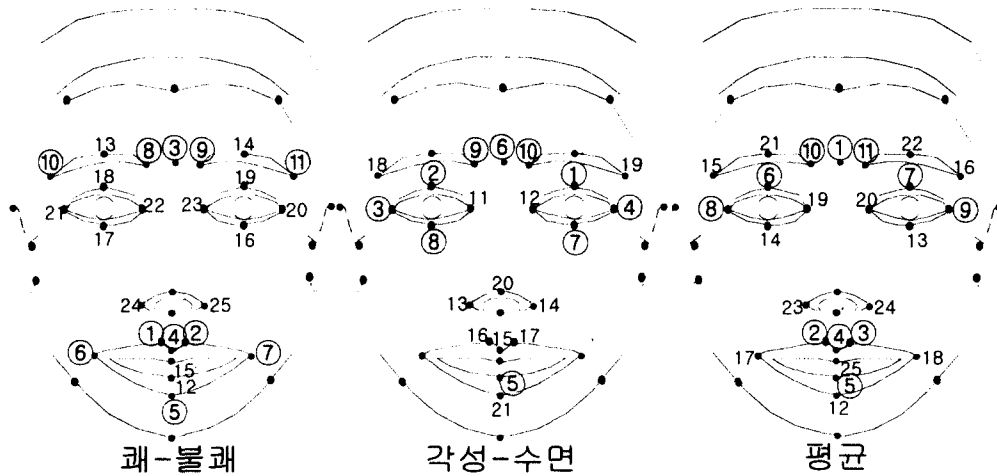


그림 5. 특징점 회귀분석의 좌우대칭 보정모형.

감정상태의 쾌-불쾌 차원은 주로 입과 눈썹 주변의 특징점과 밀접한 관련이 있고,

각성-수면 차원은 주로 눈 주변의 특징점과 밀접한 관련이 있다.

원별 순위로 나타난 것이다. 쾌-불쾌 차원은 주로 입과 눈썹, 그리고 눈과 관련된 특징점과 밀접한 관계가 있는 것으로 나타났다. 회귀분석결과(R2)를 특징점 순위로 보면 1번에서 6번 점까지는 쾌-불쾌 감정의 30% 이상을 설명하는 것으로 나타났고, 7번에서 14번 점까지는 20%, 그리고 15번에서 22번 점까지는 10% 이상을 설명하는 것으로 나타났다. 각성-수면 차원은 주로 눈과 관련된 특징점들과 밀접한 관계가 있었다. 회귀분석 결과, 순위가 1번에서 11번 점까지는 20% 정도, 12번에서 22번 점까지는 10% 정도의 설명량을 지닌 것으로 나타났다. 그림에서 10% 미만의 설명량을 지닌 점들은 따로 표시하지 않았다. 그림에서 볼 수 있듯이 좌우 대칭되는 특징점들의 설명량이 유사하지 않은 것들도 있었다 (순위가 괄호안에 표시된 점들은 설명량이 10% 미만인 것 가운데 그것의 대칭 짝의 설명량이 큰 것을 나타낸다).

특징점의 회귀분석 결과, 얼굴의 좌우 대칭점끼리 설명량이 동일하지 않은 점들(순위가 쾌-불쾌 차원에서 2번과 13번 점, 각성-수면 차원에서 2번과 9번 점 등)의 좌우 비대칭성을 보정하기 위해 좌우 대칭 짝이 있는 점들은 두 점의 설명량을 평균함으로써 전체 특징점들의 순위를 재조정하였다. 그림 5에서 볼 수 있듯이 재조정 결과 쾌-불쾌 차원에서는 입과 눈썹 주변의 특징점이, 각성-수면 차원에서는 눈 주변과 양미간 사이의 특징점이 좀 더 부각되었다. 쾌-불쾌 및 각성-수면 차원을 합쳐서 보았을 때에는 눈썹의 양미간 사

이와 눈의 윗 부분, 그리고 입의 위아래 부분이 가장 중요한 것으로 나타났다(그림 5의 평균 순위를 볼 것).

얼굴 특징점에 적용된 Gabor 필터 값에 근거한 회귀분석 결과에 대해 정리해보면 첫째, 분석에 사용된 39개의 특징점 가운데 표정인식에 중요한 것과 중요하지 않은 것이 확연히 구분되는 것으로 나타났다. 예를 들어 눈, 눈썹, 입과 관련된 특징점들은 중요한 것으로, 턱이나 귀, 머리와 관련된 점들은 중요하지 않은 것으로 나타났다. 둘째, 감정상태의 쾌-불쾌 차원과 밀접한 특징점은 각성-수면 차원과 밀접한 특징점과 비교적 독립적인 것으로 나타났다. 즉 그림 5에서 볼 수 있듯이, 쾌-불쾌 차원에서 가장 중요한 특징점은 입술 주변의 특징점이고, 각성-수면 차원에서 가장 중요한 특징점은 눈 주변의 특징점이다. 이러한 결과는 Gabor 필터 연산값 대신 특징점들의 위치변화를 토대로 한 기존연구(Smith와 Scott, 1997; 한재현과 정찬섭, 1998)들의 결과와 일치한다. 즉, Smith 등은 입술의 양끝점이 '쾌'와 관련이 있으며, 눈썹 미간사이의 점이 '불쾌'와 관련이 있다고 보고하였는데, 본 회귀분석에서도 그 점들의 순위가 높은 것으로 나타났다. 또한 한재현 등은 얼굴 특징요소 중 '눈 열림'이나 '입 열림' 등의 세로 요소가 각성상태와 밀접한 관련이 있다고 하였는데, 이것도 본 회귀분석의 결과와 유사하다. 셋째, 특징점들 중 좌우 대칭 짝을 이루는 것들이 표정인식에서 동일한 정도의 중요성을 지니지는 않는 것으로 나타났다. 기존 연구에 의하면 이러한 좌

우 비대칭성은 주로 얼굴로 나이를 판단하거나 표정 및 매력에 대한 감성적인 판단을 할 때 나타나며, 관찰자의 시점으로 보아서 좌측 시야의 특징요소가 상대적으로 더욱 중요한 것으로 알려져 있다(Burt와 Perrett, 1997). 이와 같은 경향이 나타나는 것은 표정과 매력 등의 판단에 감성적인 대뇌활동이 개입되며, 이 기능을 주로 맡고있는 대뇌 우측 반구의 활동이 우세하기 때문인 것으로 설명되고 있다. 본 연구의 회귀분석에서도 왼쪽의 특징점들이 상대적으로 더 중요한 것으로 나타났는데, 이는 정찬섭 등(1998)의 연구에 참가한 사람들이 표정을 보고 내적 감정상태를 평정할 때 주로 좌측 얼굴에 더욱 주목하였다는 것을 암시한다. 넷째, 얼굴 특징점의 Gabor 필터 값에 의한 쾌-불쾌 차원의 설명량은 각성-수면 차원의 설명량보다 큰 것으로 나타났다. 두차원 모두에서 설명량이 10% 이상인 점들의 개수가 22개로 동일하지만, 30% 이상인 점들을 보면 쾌-불쾌 차원에서는 6개가 있는 반면 각성-수면 차원에서는 하나도 없었다. 따라서 얼굴 특징점에 대한 Gabor 필터 값으로 각성-수면 상태보다 쾌-불쾌 상태를 더 정확하게 예측할 수 있으며, 쾌-불쾌 상태는 각성-수면 상태와 달리 소수의 예측력이 높은 특징점들이 있다는 것을 알 수 있다.

2.3.2 Gabor 필터의 종류와 감정상태의 회귀분석

표정인식에서 어떠한 필터 크기와 필터 방위가 상대적으로 중요한가를 알아보기 위해 회귀분석을 하였다. 얼굴 특징점 39개에 대해 40개의 필터(크기 5 X 방위 8)를 적용하였는데, 5, 10, 20, 40, 80 화소의 필터 크기와 0도(수직), 22.5도, 45도, 67.5도, 90도(수

평), 112.5도, 135도, 157.5도의 필터 방위를 사용하였다. 여기서 사용된 표정 영상은 480x480 화소 크기의 24bit 칼라 영상이었다.

감정상태와 필터구성의 회귀분석 결과를 보면, 그림 6에서 볼 수 있듯이 필터의 크기가 커질수록 상대적으로 내적 감정상태와 밀접한 것으로 나타났다. 즉, 감정상태와 밀접한 것으로 나타난 필터는 5개의 필터 크기 중에서 상대적으로 저역 공간 주파수에 해당되는 40화소와 80화소의 필터 크기였다. Nasanen(1999)은 얼굴 정보처리에 8~13 cycle/face의 필터 크기가 가장 최적이라고 보고하였는데, 이를 본 연구의 결과와 비교해보면 480 X 480 화소의 표정 영상에 대해 약 36화소에서 60화소의 필터 크기가 최적이며, 이는 본 연구의 회귀분석 결과와 일치하는 범위 안에 있다고 볼 수 있다. 감정상태와 밀접한 필터의 방위는 비스듬한 필터인 것으로 나타났다. 그림 6에서 볼 수 있듯이, 8개의 필터 방위 중에서 감정상태와 밀접한 것으로 나타난 것은 0도(수직)나 90도(수평) 방위를 제외한 비스듬한 방위이다. 이는 쾌-불쾌의 감정상태와 밀접한 요소가 기울어짐(slantedness) 요소라는 선행 연구(Yamada와 Shibui, 1998)에서도 확인된 바 있다. 따라서 감정상태의 쾌-불쾌 차원을 예측하기 위해서는 가로나 세로 방위보다는 비스듬한 방위의 필터를 사용하는 것이 효과적이라는 것을 알 수 있다.

3. 표정인식 신경망의 최적화

얼굴 특징점에 대한 Gabor 필터 연산 값과 필터 구성의 회귀분석 결과를 참조하여 위계적으로 표정인식 신경망을 최적화하였다. 신경망의 기본모형(full model)

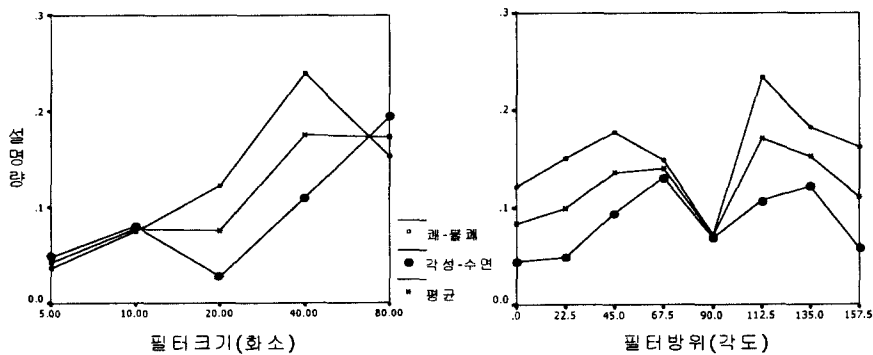


그림 6. 감정상태와 필터구성에 대한 회귀분석 결과
 감정상태와 밀접한 필터 크기는 상대적으로 저역 공간 주파수에 해당되는 것이고,
 필터 방위는 주로 비스듬한 방위의 필터였다.

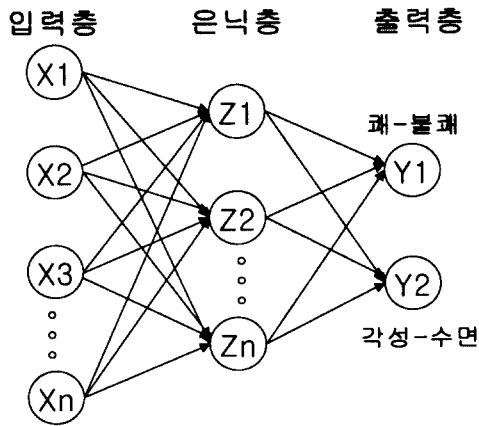


그림 8. 표정인식 신경망 최적화 모형의 구성

3.3 신경망의 학습

정찬섭 등(1998)의 DB에서 선정한 200개의 표정영상 가운데 150개를 무선 표집하여 신경망의 학습자료로 사용하였다. 신경망의 학습에는 오류 역전파(error backpropagation) 알고리즘을 사용하였고 연결 가중치의 조절에는 일반화된 델타 학습규칙을 사용하였다. 신경망의 목표 값으로는 한 표정영상에 해당되는 정찬섭 등의 DB에 수록된 패-불패 및 각성-수면의 감정상태 2차원 공간 값을 사용하였다. 신경망의 출력 기대 값과 실제 출력 값의 차이로서 평균 자승오류를 최소화하는데에는 traingdx 함수를 사용하였다. 신경망의 활성화 함수로서는 입력층에서 은닉층으로의 전달 함수로 log sigmoid 함수를, 은닉층에서 출력층으로의 전달 함수로는 positive linear 함수를 사용하였다. 이들 두 함수는 각각 $\text{logsig}(n) = 1 / (1 + \exp(-n))$ 과 $\text{poslin}(n) = n, \text{ if } n \geq 0; = 0, \text{ if } n < 0$ 이었다. 은닉층과 출력층의 평균 자승오류가 전역적인 최소값으로 수렴되었을 때 학습이 완료되는데, 주어진 입력자료 값에 대한 오차가 0.00001 이하로 떨어졌을 때 학습을 종료하도록 하였다.

신경망 학습에 사용된 개인용 컴퓨터는 Pentium III 500MHz CPU와 256M RAM이 장착된 IBM 컴퓨터 호환기종이었다. 신경망의 입력 영상 하나에 1560개의 모든 입력요소를 사용하는 기본모형의 경우, 5320번의 회기(epoch) 후에 학습이 종료되었으며 약 12시간의 학습시간이 필요하였다. 표 1에서 보듯이 회귀분석결과 설명량이 큰 특징점들이나 필터들만을 포함시켜 단계적으로 입력 요소의 수를 줄여 나가면

신경망의 학습시간이 5시간까지 점진적으로 감소한다는 결과를 얻었다. 모든 모형에서 학습은 사람의 수행과 거의 100% 일치하는 수준에서 수렴되었다.

표 1. 표정인식 신경망의 최적화 단계별 학습회기, 학습시간 및 감정차원 평정값과의 상관관계. 기본 입력요소는 1560 개 요소이며, 39개의 특징점과 5개의 필터 크기, 8개의 필터 방위로 이루어져 있다. 입력요소란의 괄호안에 있는 숫자들은 자기 신경망에 포함된 특징점×필터크기×필터방위의 수를 나타낸다.

| 입력요소 | 학습회기 | 학습시간 | 패-불패 | 각성-수면 |
|------------------|-------|------|-------|-------|
| 1560 (39x5x8) | 5320 | 12시간 | 0.708 | 0.495 |
| 1000 (25x5x8) | 7944 | 8시간 | 0.765 | 0.496 |
| 760 (19x5x8) | 8025 | 6시간 | 0.738 | 0.505 |
| 400 (25x2x8) | 23399 | 5시간 | 0.886 | 0.631 |

3.4 신경망의 수행력 검사

신경망의 검사에는 정찬섭 등(1998)의 DB에서 선정한 200장의 사진중 학습에 사용하지 않은 50장의 사진을 사용하였다. 신경망의 수행력은 신경망의 출력과 정찬섭 등의 DB에 수록된 사람들의 감정차원 평정값과의 상관관을 나타내는 Pearson 적률(r)을 이용하였다.

입력요소(노드)의 수가 1560개인 기본모형의 수행결과와 사람의 평정결과의 상관계수는 패-불패 차원이 0.708 이었고, 각성-수면의 차원은 0.495 이었다. 표 1에서 보듯이, 회귀분석 결과를 반영하여 설명량이 큰 특징점과 필터들의 수를 단계적으로 줄이면 상관계수가 대체로 증가하여 수행력이 향상되는 것으로 나타났다. 각 모형에서 특징점과 필터 크기의 수는 회귀분석 결과를 토대로 줄어나갔지만 필터의 8개 방위는 그대로 모두 포함시켰다. 설명량이 10% 이상인 특징점들만을 포함시킨 1000개 입력노드의 신경망은 패-불패 및 각성-수면의 상관계수가 각각 0.765와 0.496, 설명량이 20% 이상인 특징점들만으로 된 760개 노드 모형은 0.738과 0.505, 감정상태 설명량이 10% 이상인 특징점과 17% 이상인 필터 크기만으로 된 400개 노드 모형은 0.886과 0.631로서 1560의 노드로 된 기본모형보다 모두 사람과 더 유사한 수행력을 보였다. 특히, 기본모형의 약 26% 정도인 400개 노드 모형에서 매우 높은 상관계수를 얻었다는 것은 회귀분석의

결과를 토대로 신경망의 학습과정과 수행력을 현격하게 최적화시킬 수 있다는 것을 보여 준다. 지금까지 개발된 표정인식모형은 대부분은 감정상태의 차원모형보다 범주모형을 사용하고 있어 이 모형의 수행력을 기존 모형과 직접 비교하기 어렵다. Lyons 등(1998)은 그들이 설계한 신경망과 사람의 평정자료간 등위 상관계수로서 0.568을 얻었는데, 이에 비해 회귀분석 결과를 이용하여 최적화된 모형의 수행력이 월등하게 높다는 것을 알 수 있다.

표 2. 표정인식 분야의 선행연구와 본 연구의 수행결과 비교

| 연구자 | 인식 방법 | 표정 자료 | 표정 종류 | 수행결과 | 참고 |
|-------------|----------|-------|-------|-------------|--------------|
| Zhang(98) | 지형 요소 | 213장 | 6가지 | 73% | 정확 인식률 |
| Essa(95) | 형판 정합 | 48장 | 4가지 | 98% | 슬픔, 공포 제외 |
| Padgett(97) | 주성분 분석 | 97장 | 6가지 | 86% | 정확 인식률 |
| Lyons(98) | Gabor 필터 | 213장 | 6가지 | 0.568 | 등위 상관계수 |
| Zhang(98) | Gabor 필터 | 213장 | 6가지 | 92% | 정확 인식률 |
| 본 연구 | Gabor 필터 | 200장 | 6가지 | 74% | 대표 범주 정확 인식률 |
| 본 연구 | Gabor 필터 | 200장 | 60가지 | 0.886/0.631 | 쾌/각성 차원 상관계수 |

3.5 범주화에 의한 모형의 수행력 검증

내적 감정상태의 범주를 사용한 기존 모형과 수행력을 비교하기 위하여 신경망이 출력한 두 감정상태의 차원 값을 이용하여 검사에 사용된 각 표정을 그림 7의 감정상태 공간의 기쁨, 슬픔, 분노, 공포, 경멸, 놀람 등 6가지 기본정서의 좌표 값 중 가장 가까운 것으로 분류하여 범주화하였다. 그 결과, 신경망의 출력값에의 범주화 결과가 사람의 평정결과와 74% 일치하였다. 이러한 결과는 표정인식 분야의 여러 가지 선행 연구 결과를 비교하여 볼 때, 대체적으로 Zhang 등(1998)의 지형적 특징요소를 이용한 모형의 인식률(73%)와 비슷하였으나 Zhang 등의 Gabor 필터 기반 모형(92%), Essa(1995) 등의 형판정합 기반 모형(98%), Padgett과 Cottrell(1997)의 주성분분석 기반 모형(86%)보다는 낮은 것으로 나타났다(표 2). 그러나 이들 기존 연구들은 처음부터 소수 범주에 분명히

속하는 얼굴표정을 입력자료로 사용한 것에 비해, 이 연구에서는 감정상태 공간에 널리 분포하면서 6개의 전형적인 표정범주가 아닌 다양한 표정을 이용했다는 점을 고려하면 이 연구에서 사용된 신경망의 수행력이 기존모형에 비해 상대적으로 낮은 것으로 평가할 수 없다.

4. 종합 논의

감정상태의 구조화 측면에서 볼 때, 얼굴 표정으로 인식할 수 있는 감정상태는 쾌-불쾌와 각성-수면의 차원모형이 기본 정서 범주모형보다 포괄적이고 유용하다. 이 연구에서 시도된 Gabor 필터 연산 값과 감정상태의 회귀분석 결과에서 보듯이, 얼굴 표정은 사람의 감정상태를 안정적으로 예측하는 지표가 될 수 있다. 그러나 기쁨, 슬픔, 놀람, 공포, 혐오, 분노만을 인식하는 기본 정서 범주모형은 일반적인 감정상태 공간의 전역을 모두 포괄하지 못하며, 불쾌한 감정상태와 높은 각성상태 쪽으로 편파되어 있다. 따라서 기본정서 모형의 표정 범주들은 자연스럽게 다양하게 발생하는 미묘한 표정변화를 포착할 수 없다(Russell, 1980). 사람의 의사소통 과정에서는 미묘한 표정변화도 중요한 단서가 될 수 있는데 이 연구의 결과는 Gabor Jet을 입력값으로 한 신경망이 이러한 미묘한 변화를 포착하여 비교적 정확하게 표정을 인식할 수 있다는 것을 보여준다.

회귀분석을 토대로 입력 단위의 종류와 수를 최적화하는 것이 Gabor wavelet 기반 표정인식 시스템의 수행력을 크게 향상시킬 수 있다는 결론을 얻었다. 이는 Gabor wavelet이 한 영상 영역에 대한 여러 규모의 영상정보를 포함하고 있어 지리적으로 인접해 있고 상대적으로 표정인식에 덜 중요한 특징점들을 제외시킬 수 있는 연산특성을 지니고 있기 때문인 것으로 보인다. 회귀분석은 이에 덧붙여 모든 종류의 필터를 무차별적으로 이용하는 것보다 각 특징점별로 설명력이 높은 것만을 추려 모형에 반영할 수 있는 길을 열어주어 시스템 성능을 최적화하는데 큰 도움을 줄 수 있다는 것도 확인되었다. 얼굴표정과 관련된 모든 특징요소들과 이에 대한 영상처리를 위한 모든 유형의 필터들을 모형 내에 모두 반영하는 것은 무모하고 비효율적인 시도이다. 이러한 점 때문에, 기존 연구들에서는 연구자의 직관에 따라 연산부하를 줄일 수 있도록 특정 특징점과 필터를 선별하여 사용해 왔다(예를 들어 Lyons 등, 1998). 그러나, 회귀분석 등 사전 분석에 의한 시스템 최적화 전략을 사용하면 직관보다

는 경험적 검증을 통해 시스템의 성능을 향상시킬 수 있으며 그 결과에 따라 시스템 설계의 개선방향을 제시받을 수 있다.

선형모형에서는 이론적으로 독립변수가 많을수록 종속변수에 대한 예측력이 증가하며 절대 감소될 수가 없다. 따라서, 표정인식 과정이 선형모형에 적합한 것이라면 1560개의 입력단위를 갖는 기본 모형의 예측력이 그보다 적은 입력 단위를 갖는 축소모형보다 높을 수가 없다. 이러한 예측과 달리 기본모형보다 적은 수의 입력단위를 갖는 모든 최적화 단계의 모형들에서 수행력이 기본모형보다 우수했다는 것은, 그리고 입력단위 수가 가장 적은 400개 노드 모형에서 수행력이 가장 높았다는 것은 표정인식 과정이 비선형적 과정이라는 것을 강하게 암시한다. 이와 같은 결과는 결국 입력 노드의 수가 증가하는 것이 반드시 신경망의 수행력을 향상시키지 않으며 오히려 저해할 수 있다는 것을 시사하는데 구체적으로 이러한 저해효과가 초래되는 정확한 원인에 대한 연역적 및 귀납적 설명이 추후 연구에서 시도될 필요가 있다.

참고문헌

- 김영아, 김진관, 박수경, 오경자, 정찬섭. (1998). 정서 관련어휘분석을 통한 내적상태의 차원연구. **한국감성과학회지**, 제 1권 제 1호, 145-152.
- 정찬섭, 오경자, 이일병, 변혜란. (1998). **표정/제스처에 의한 감성측정기술 및 DB개발**. 과학기술부 감성공학기술개발사업 최종보고서, G-17-01-09.
- 한재현, 정찬섭. (1998). **얼굴표정에 의한 내적상태 추정**. **한국감성과학회지**, 제 1권 제 1호, 41-58.
- Bartlett, M. Viola, P. Sejnowski, T. Larsen, J. Hager, J. and Ekman, P. (1996). Classifying facial action. In D. Touretski, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems, volume 8*, 823-29. Morgan Kaufmann, San Mateo, CA.
- Burt, D. M. & Perrett, D. I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, speech and expression. *Neuropsychologia*, Vol. 35, No. 5, 685-693.
- Chellappa, R. (1995). Human and Machine Recognition of Faces: A Survey. *Proceedings of the IEEE* 83(5), 704-740.
- Daugman, J. G. (1985) Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by 2D Visual Cortical Filters. *J. Opt. Soc. Amer. (A)* Vol. 2, pp. 1160-1169.
- Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999) Classifying facial action. in *IEEE Transactions on PAMI* 21(10), 974-989.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA : Consulting Psychologists Press.
- Essa, I. A. & Pentland, A. P. (1995). Facial Expression Recognition using a Dynamic Model and Motion Energy. *IEEE ICCV95*, 360-367.
- Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, (4), 2379-2394.
- Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3), 300-311.
- Lien, J. J., Kanade, T., Cohn, J. F. & Li, C. (1998). Automated facial expression recognition based on FACS action units. *Proceedings of the third international conference on automatic Face and Gesture Recognition*, 390-395.
- Lyons, M., Akamatsu, S., Kamachi, M. & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proceedings of the third international conference on automatic Face and Gesture Recognition*, 200-205.
- Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Reserch* 39, 3824-3833.
- O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America A*, 10(3), 405-411.
- Padgett, C. and Cottrell, G. (1997). Representing

- face images for emotion classification. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems, volume 9*, Cambridge, MA, MIT Press.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Smith, C. and Scott, H. (1997). A componential approach to the meaning of facial expression. In Russell, J. and Fernandez-Dols, J. *The Psychology of Facial Expression*. Cambridge University Press, New York.
- Wiskott, L., Fellous, J. M., Kruger, N., and Von Der Malsburg, C. (1997) Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. on Patt. Anal. and Machine Intell. Vol. 19*, 775-779.
- Yacoub, Y. and Davis, L. S. (1994). Computing Spatio-Temporal Representations of Human Faces. in Proceedings, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Yamada, H. and Shibui, S. (1998) *A psychological model of recognizing facial expressions of emotion. ATR Symposium on Face and Object Recognition '98*, 23-24.
- Zhang, Z., Lyons, M., Schuster, M. & Akamatsu, S. (1998). Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. *Proceedings of the third international conference on automatic Face and Gesture Recognition*, 454-459