

# 정지영상과 동영상의 융합모델에 의한 얼굴 감정인식

## Face Emotion Recognition by Fusion Model based on Static and Dynamic Image

이대종\*\* · 이경아\*\* · 고현주\* · 전명근\*

Dae-Jong Lee\*\*, Kyong-Ah Lee\*\*, Hyoun-Joo Go\*, Myung-Geun Chun\*

\*충북대학교 전기전자컴퓨터공학부

\*\*다산네트웍스

\*Chungbuk National University School of Electrical and Computer Engineering

\*\*Dasan Networks

### 요 약

본 논문에서는 인간과 컴퓨터의 인터페이스를 좀더 자연스럽게 쉬운 형태의 능동적인 휴먼 인터페이스로 구현하기 위해 정지영상 및 동영상에서의 감정인식기법을 제안하고자 한다. 제안된 얼굴의 감정인식 기법은 Hidden Markov Model(HMM), 주성분분석기법(PCA)와 웨이블릿 변환을 기반으로 구성하였다. 얼굴의 감정인식을 위하여 심리학자인 Ekman과 Friesen의 연구에 의해 문화에 영향을 받지 않고 공통으로 인식하는 6개의 기본 감정인 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 바탕으로 실험하였다. 감정인식에서 입력영상은 이산 웨이블릿을 기반으로 한 다해상도 분석기법을 사용하여 데이터 수를 압축한 후, 각각의 영상에서 PCA 특징벡터를 추출한 후 이를 사용하여 HMM의 모델을 생성한다. 인식단계에서는 정지영상에서의 인식값과 동영상에서의 인식값을 정규화 과정을 통하여 상호보완 함으로써 인식률을 높일 수 있었다.

### Abstract

In this paper, we propose an emotion recognition using static and dynamic facial images to effectively design human interface. The proposed method is constructed by HMM(Hidden Markov Model), PCA(Principal Component) and wavelet transform. Facial database consists of six basic human emotions including happiness, sadness, anger, surprise, fear and dislike which have been known as common emotions regardless of nation and culture. Emotion recognition in the static images is performed by using the discrete wavelet. Here, the feature vectors are extracted by using PCA. Emotion recognition in the dynamic images is performed by using the wavelet transform and PCA. And then, those are modeled by the HMM. Finally, we obtained better performance result from merging the recognition results for the static images and dynamic images.

**Key Words** : 감정인식(Emotion Recognition), 웨이블릿 변환(Wavelet Transform), HMM(hidden markov model), 주성분 분석기법(Principal Component Analysis)

## 1. 서 론

정보화 시대를 맞이하여 컴퓨터를 비롯한 많은 시스템들이 급속도로 대중화되어 가고 있다. 고기능의 컴퓨터들이 개인용으로 확산되어 보급되면서 사람이 지시하는 업무만을 일정한 규칙에 따라 처리하는 것에서 그치지 않고 인간의 내면에 접근하는 방법에 대한 연구가 활발히 이루어지고 있다. 예를 들어 사용자의 기분이 우울한 날이면 오디오가 자신의 기분에 어울리는 음악을 들려주거나, 공항의 설치된 모니터를 통해 감정 인식하는 컴퓨터가 무언가 불안해하는 사람들

을 인지하여 수배중인 범죄자나 테러범을 찾아 범죄를 예방할 수도 있을 것이다. 이와 같은 사람과 기계와의 인터페이스의 궁극적 목적은 사람과 사람이 대화하듯이 컴퓨터와의 인터페이스가 이루어지는 것이다. 이러한 연구에 있어서 인간의 감정을 어떻게 측정하고 인식하느냐가 가장 어려운 문제로 부각되고 있다.

심리학에서는 얼굴을 마주 보고 하는 대화에서 표정이 매우 중요하다고 알려지고 있다. Mehrabian은 메시지의 단지 7%만 순수한 언어에 의해 전달되며 신체 움직임이나 음성과 같은 준언어에 의해서는 38%, 얼굴표정에 의해서는 55%의 메시지가 전달된다고 보고한 바 있다. 이는 얼굴 표정은 사람과 컴퓨터가 자연스런 메시지의 전달을 위해서는 향후 더욱 더 그 중요성이 더해 가는 인터페이스임을 뜻한다. 이러한 사람의 감정 표현에 대해서 심리학자인 Ekman과 Friesen에 따르면 사람의 여섯 가지 기본 표정인 놀람, 공포, 혐오, 행복감, 두려움, 슬픔 등은 그 보편성 가지고 있다고 보고하고 있다[1]. 얼굴표정 인식 방법은 정적영상을 기반으

접수일자 : 2004년 10월 4일

완료일자 : 2005년 5월 3일

감사의 글 : 이 논문은 2004년도 전반기 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음 (KRF-2004-214-D00138)

로 하는 방법과 동적영상에 기반한 모션을 사용하는 방법의 두 부류로 크게 구분할 수 있다. 특징치를 기반으로 하는 방법은 얼굴의 주요 특징인 눈, 코, 입의 위치를 찾고, 모양과 기하학적 관계를 파악함으로써 인식기를 구성하는 방법이다. 그러나, 이 방법은 피부색이나 조명에 관계없이, 정확한 위치와 모양을 실제 영상으로부터 찾기가 어렵다는 단점이 있어 사람이 직접적으로 윈도우나 마크를 사용한 수동적 방법에 의존한다[2]. 홀리스틱 분석은 얼굴전체에 대한 분석으로 얼굴 영상들의 통계학으로부터 학습된 데이터 구동 커널을 사용한다. 대표적인 방법으로는 PCA(Principal Component Analysis)[3], LFA(Local Feature Analysis)[4], LDA(Linear Discriminant Analysis)[5], ICA(Independent Component Analysis)등이 연구되어지고 있다. 국부적인 표현방법으로는 얼굴영상의 세부적인 영역을 다루는 Local PCA[6], Gabor 웨이블릿[7] 표현방법 등이 있다.

표정의 동적인 특성은 광 플로우 등을 이용한 접근을 시도하게 되었다. Yacoob와 Davis는 몇 가지 얼굴의 특정 부위를 지정하고 이 부분에서 광 플로우를 이용하여 움직임을 추적하여 모션 특징과의 관계를 간단한 규칙을 표현함으로써 인식을 시도하였고 실험결과, 여섯 가지 기본 표정에 대해 32명에 대해 80%~94%의 높은 인식 성공률을 얻었다. Essa는 광 플로우와 인공 신경망을 이용하여 얼굴 표정의 인식을 시도하였다. 얼굴에 대한 컴퓨터 그래픽 모델을 기초로 얼굴의 모션을 추정하고 이로부터 얼굴 표정의 인식을 시도하였다. 인식을 위해 모션 에너지(motion energy) 개념을 소개하고 행복감, 분노, 놀람, 혐오에 대해 실험하였다. Fumio Hara는 역전파 학습 알고리즘에 의한 feed forward 인공 신경망을 이용하여 여섯 가지 얼굴 표정으로 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 인식할 수 있는 얼굴 로봇을 구현한 바 있다. 이 실험에서는 특징점 사이의 거리를 사용한 특징을 사용하고 있으므로, 특징점의 신뢰성 있는 추출문제와 표정 이외의 다른 다양한 얼굴 모양에 대해 신뢰성 있는 시스템을 구축하기 어려운 단점이 있다[8]. Takagi 와 Ushida는 개념적 퍼지 집합이라는 개념을 도입하여 거리를 기반으로 한 특징치를 사용하여 표정인식을 시도하여 78.7%의 인식률을 얻었다[9].

최근에는 여러 개의 개별 특징들을 효과적으로 결합시키는 융합법칙에 대한 연구가 활발히 이루어지고 있다[10-12]. 본 논문에서도 정지영상과 연속적인 감정표현을 갖는 동영상에서의 감정 추출의 장점을 살리기 위하여 두 가지 기법을 모두 사용하여 높은 인식률을 얻고자 한다. 얼굴의 감정인식을 위하여 심리학자인 Ekman과 Friesen의 연구에 의해 문화에 영향을 받지 않고 공통으로 인식하는 6개의 기본 감정인 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 바탕으로 실험하였다[1]. 동영상에서의 감정 데이터는 중성의 감정표현에서 6개의 기본감정 중의 하나로 친이 되어지는 단계를 연사촬영하여 총 5개의 영상으로 구분하였다. 감정인식에서 입력영상은 이산 웨이블릿을 기반으로 한 다해상도 분석기법을 사용하여 데이터 수를 압축한 후, 각각의 영상에서 PCA 특징벡터를 추출한 후 이를 사용하여 HMM의 모델을 생성한다. 인식단계에서는 정지영상에서의 인식값과 동영상에서의 인식값을 정규화 과정을 통하여 상호보완 함으로써 인식률을 높일 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 얼굴 감정에서 특징을 추출하는 방법과 모델 형성 및 인식을 위한 HMM 모델링 기법에 대해서 설명한다. 3장에서는 본 논문에서 제안된 정지영상 및 동영상에서의 얼굴 감정인식 기법에 대해

서 설명한다. 4장에서는 얼굴 감정인식 실험을 통하여 제안한 방법의 타당성을 검토하고, 마지막 5장에서 결론을 맺는다.

## 2. 얼굴 감정 특징벡터 추출 및 HMM 모델링

### 2.1 웨이블릿과 PCA(principle component analysis)를 이용한 특징벡터 추출

2차원의 경우 웨이블릿은 아주 작은 비트 율로 정보를 표현함에도 불구하고 영상의 전체적인 정보뿐만 아니라 에지와 같은 미세한 정보도 스케일 계수로 모두 유지시킬 수 있다. 따라서, 계수들이 변환 전 영상의 위치정보를 포함하기 때문에 사용자가 원하는 영상정보를 변환 후에도 유지시킬 수 있다. 그리고, DWT를 영상신호에 적용하는 것은 영상을 공간상의 x축과 y축 방향으로 저대역 통과필터(LPF)와 고대역 통과 필터(HPF)를 사용하여 신호를 추출하는 것을 의미하는 것으로, DWT를 거친 신호는 총 네 개로 분리될 수 있다. 이때, 분리된 영역은 그림 1과 같이 영상에 적용된 필터에 따라 구분된다.

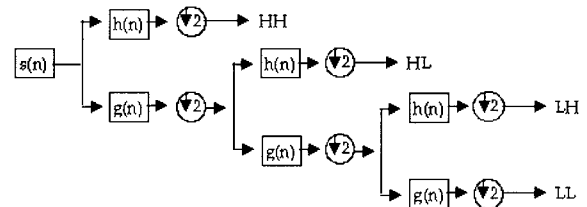


그림 1. 웨이블릿 구조  
Fig. 1. Wavelet structure

그림 1의 LL영역은 영상에 저대역 통과 필터를 적용한 것으로, 영상에서 고주파 성분이 배제된 계수들로 이루어져 있다. 또한, HH영역은 영상에 고대역 통과 필터를 사용한 것으로 LL과는 반대의 고주파 성분만 나타나게 된다. 따라서 LL 영역의 주파수 대역은  $0 < \omega < \pi/2$ 이고, HH영역의 주파수 대역은 반대로  $\pi/2 < \omega < \pi$ 이다. 또한, HL영역과 LH영역의 주파수 대역은 공간상에서 x축과 y축 방향으로 서로 다른 필터를 사용했기 때문에 공간적인 방향에 따라 달라질 수 있다. 그림 2는 원 영상에 대해 한번의 웨이블릿 변환을 거친 후 LL1, LH1, HL1, HH1으로 분리된 영상을 보여준 것이며, LL1 영상에 대해 세번 웨이블릿 변환을 거친 LL3, LH3, HL3, HH3으로 분리된 영상을 보여주고 있다[13].

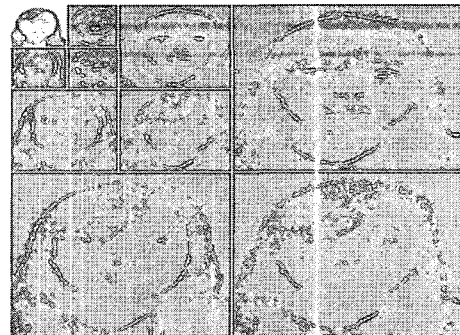


그림 2. 웨이블릿 변환후 4개의 밴드로 분리된 영상  
Fig. 2. Four subband images after wavelet transform

주성분 분석 기법은 고차원의 데이터를 저차원의 데이터로 선형 변환하는 것으로 데이터의 차원을 줄이는데 널리 쓰인다. 일반적으로 얼굴 영상은 매우 고차원의 데이터로 표현되기 때문에 특징 추출과 분류를 위해서는 저차원의 데이터로 표현되는 것이 요구된다. 얼굴 인식에서 주성분 분석 기법은 학습영상의 2차 통계적 특성을 이용하여 학습영상의 전체적인 특성을 표현하는 직교기저영상인 고유얼굴로 분해할 수 있으며 이 고유얼굴의 선형조합으로 임의의 얼굴 영상을 표현하는 방법으로 입력 데이터를 저차원의 데이터로 표현하는 효과적인 방법이다. 여기서 정규 직교하는 열로 구성된 고유벡터를 고유얼굴(Eigenfaces)라고 한다. 그림 3은 PCA를 이용한 얼굴인식으로 앞에서 얻어진 고유얼굴과 특징벡터의 선형적인 결합에 의해 얼굴영상들을 표현할 수 있다. 본 논문에서는 세 번의 웨이블릿 변환을 거친 후의 영상인 LL3, LH3, HL3, HH3를 특징벡터와 고유얼굴로 표현하였다.

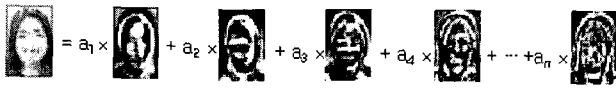


그림 3. PCA를 이용한 고유얼굴  
Fig. 3. Eigenfaces obtained by PCA

## 2.2 HMM(Hidden Markov Mode) 알고리즘에 의한 모델링

은닉 마코브 모델(Hidden Markov Model, HMM)은 신호에 포함된 특성들의 발생과정을 확률적으로 모델화한 것으로 학습을 통하여 모델을 형성한다[14]. 이러한 HMM은 세가지 요소로 구성되어 있는데 상태(state)의 개수, 시간에 따른 상태의 변화를 결정하는 상태전이 확률분포, 그리고 각 상태에서의 출력심벌의 확률분포이다. 이러한 정의를 이용하여 HMM은  $\lambda=(A, B, \pi)$ 로 나타낼 수 있다. 여기서,  $\Pi=\pi_{ij}$  : 초기 상태전이 확률분포,  $A=a_{ij}$  : 상태 전이확률분포,  $B=b_{ij}$  : 관측심벌 확률분포를 나타낸다.

HMM을 이용하기 위해서는 다음 두 단계를 거치게 된다. 즉, 모델형성 과정과 형성된 모델을 이용하여 관측 심벌의 확률값을 구하는 두 과정을 거치게 된다.

- 단계 1 : 모델 형성과정 ( $P(O|\lambda)$ 를 최대로 하는 모델 파라미터  $\lambda=(A, B, \pi)$ 를 구하는 문제)
- 단계 2 : 모델 인식과정 (관찰된 심벌의 시퀀스  $O=O_1O_2\cdots O_T$ 와 모델  $\lambda=(A, B, \pi)$ 가 주어졌을 때 모델에 대한 심벌의 likelihood  $P(O|\lambda)$ 를 구하는 문제)

본 논문에서 단계 1의 모델형성과정을 살펴보면 다음과 같다. 관측된 심벌의 시퀀스는 PCA를 이용하여 사용된 모든 얼굴 영상의 특징을 검출한 후 이를 이용하여 얼굴 감정을 대표하는 코드북을 생성한 후 각각의 얼굴감정별 시퀀스는 입력 얼굴 감정 영상과 대표 코드북과의 특징벡터를 비교함으로써 얻어진다. 이렇게 얻어진 관측 심벌의 시퀀스를 이용한 HMM 모델링은 Baum-Welch 알고리즘을 모델 파라미터  $\lambda=(A, B, \pi)$ 를 설정하였다. 단계 2의 인식단계에서는 각각 설정된 얼굴 감정별 HMM 모델과 관측된 심벌 시퀀스와 의 부합확률은 forward 알고리즘을 이용하여 산출하였다.

## 3. 정지영상 및 동영상에서의 얼굴감정인식

### 3.1 PCA를 이용한 정지영상에서의 감정인식

정지영상에서의 감정인식 시스템은 주로 2차원 영상으로 획득되는 얼굴의 감정데이터에서 서로 다른 감정의 미세한 차이를 구별하고, 동시에 같은 감정의 얼굴감정영상으로부터는 작은 변화에도 불구하고 동일한 특징들을 찾아내는 기술이 우선적으로 필요하다. 본 논문에서는 이러한 문제를 해결하기 위해 웨이블릿을 통해 영상을 다 해상도로 분석을 하였으며, 이를 통해 얻어진 영상으로부터 PCA를 이용하여 특징 벡터를 추출하였다. 앞의 관련연구에서 설명하였듯이 PCA는 여러 변수가 서로 상관관계를 가지고 있을 때 이들을 설명할 수 있는 구조를 찾아내는 방법이다. 이 분석에서는 서로 상관관계를 가지고 있는 고유벡터들을 이용해서 상관관계를 갖지 않는 새로운 특징벡터들을 찾아낸다. 각각의 얼굴영상은 이 고유얼굴과 특징벡터의 선형적인 결합에 의해 표현된다.

본 논문에서는 다해상도 분석을 위해 학습이미지에 대해 이산 웨이블릿 변환을 두 번 적용하였으며, 원 이미지(640x480)에 대해 사이즈 160x120의 4개의 해상도 영상을 얻을 수 있다. 여기서, 4개의 해상도 영역은 다 해상도 분할 방법의 마지막 단계로 이전 대역의 LL(저주파 혹은 스케일링 함수에 의해 사영된 영역)영역의 정보만을 연속으로 분해하여 얻어진 영역들이다. 각각의 영역은 스케일링함수에 의해 원 영상의 모습을 그대로 표현하고 있는 한 개의 영역과, 웨이블릿 함수에 의해 사영되어진, 수직, 수평, 대각선의 방향성을 가진 세 개의 영역으로 구성되어 있으며, 각각의 영역에 대한 정보들을 해상도별로 저장한다. 그리고 앞에서와 같은 웨이블릿 결과로 얻어진 LL3 영상(160x120)에 대해, PCA 방법을 적용하여 특징벡터를 추출하였다. 이때, 고유벡터에 대해 남자 60개, 여자 80개를 사용하였을 때, 인식 결과가 가장 우수하였다. 위와 같은 방법에서 얻어진 학습영상에 대한 특징벡터와 검증영상에 대한 특징벡터의 유사도는 유클리디안 거리에 근거한 KNN 알고리즘을 이용하여 얻어진다. 그림 4에 정지영상에서의 얼굴 감정인식 과정을 정리하여 나타냈다.

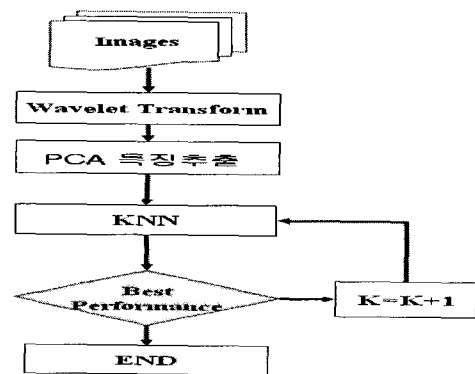


그림 4. 정지영상에서의 얼굴감정인식 순서도  
Fig. 4. The flowchart for emotion recognition of static facial images

### 3.2 동영상에서의 얼굴 감정인식

본 논문에서 제안한 동영상에서의 얼굴인식 시스템은 그림 5에서 보인 바와 같이 특징 추출부, 코드북 생성부, HMM 생성부 및 인식부로 구성되어 있다. 사용된 동영상 데이터는 중성의 감정표현에서 6개의 기본 감정(화남, 혐오, 행

복, 슬픔, 공포, 놀람) 중의 하나로 처리 되어지는 과정을 연사 촬영하여 5개의 영상으로 구성하였다.

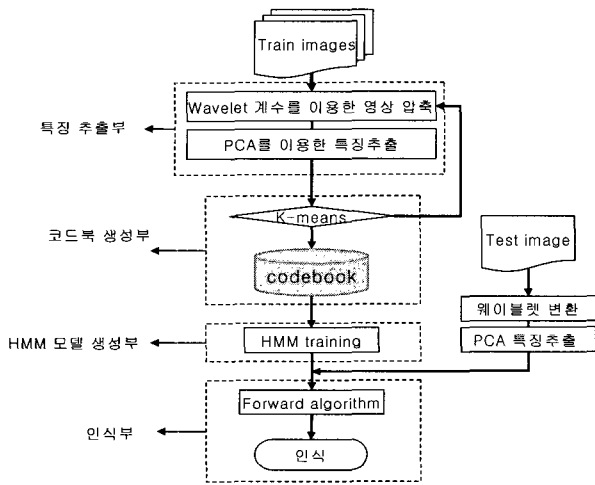


그림 5. 동영상에서의 감정인식 순서도

Fig. 5. The flowchart for emotion recognition of dynamic facial images

640×480사이즈의 입력영상은 nearest neighbor interpolation에 의하여 160×120의 영상으로 재조정하였다. 재조정된 영상은 3번의 웨이블렛 변환 과정을 통하여 20×15크기의 4개의 대역별 영상을 얻었는데 이는 다해상도 분석을 위하여 이전대역의 LL영역(저주파 함수에 의해 사영된 영역)을 연속 분해하여 얻어진 영상이다. 웨이블렛 변환 후 대역 별로 저장된 영상은 각 시퀀스 별로 PCA(주성분 분석 기법) 과정을 거친다. 이 과정 후 생성된 PCA 특징벡터는 각 영상을 대표하는 특징값으로 저장된다. 저장된 특징벡터는 그림 6에서 보는 바와 같이 K-means 알고리즘을 거쳐 일정한 코드북 크기만큼 독립적인 코드북을 대역별로 만들어 놓는다. 이때, 고유벡터에 대해 남자 60개, 여자 80개를 사용하였을 때, 인식 결과가 가장 우수하였다. 위와 같은 방법에서 얻어진 학습영상에 대한 특징벡터와 검증영상에 대한 특징벡터의 유사도는 유클리디안 거리에 근거한 KNN 알고리즘을 이용하여 얻어진다.

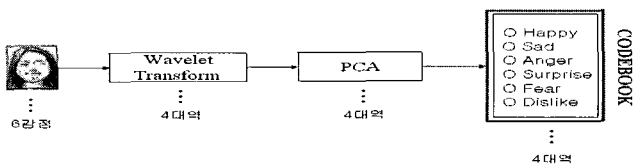


그림 6. 코드북 생성과정

Fig. 6. Formation process of codebook

PCA를 거쳐 얻어진 각 영상에 대한 특징벡터는 코드북과 비교하였다. 여기에서, 특징벡터와 코드북과의 유사도는 유클리디안 거리에 근거한 kNN 알고리즘을 이용하여 특징벡터와 코드북 사이의 공간상의 거리가 가장 가까운 코드북의 인덱스를 그림 7에 나타난 과정을 거쳐 산출한다. 입력 영상은 시간의 흐름에 따라서 5개의 시퀀스가 생성되었기 때문에, 이에 따라 인덱스 값은 1 세트당 5개가 산출되고 이는 HMM의 관측벡터가 된다. 입력벡터와 기준 코드북과의 비교에 의해 예러가 최소인 코드북내 Index는 BAUM-WELCH 알고리즘에 의하여 각각의 감정별 HMM 변수인

A, B,  $\pi$ 를 결정한다. 앞에서 설명하였듯이 A는 상태전이 확률 분포, B는 관찰 심볼 확률분포,  $\pi$ 는 초기 상태전이 확률 분포이다. 이는  $\lambda=(A, B, \pi)$ 로 표현되며 이 과정을 마치면 6개의 각 감정별 HMM 모델이 생성된다. 본 논문에서는 HMM의 상태모델을 좌우(left to right) 모델을 사용하였다.

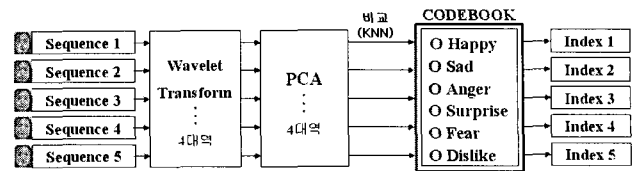


그림 7. 인덱스 산출 과정

Fig. 7. Index calculation process

인식부에서는 모델 생성과정과 동일하게 검증 시퀀스에 대해서 웨이블렛 변환을 통하여 데이터를 압축 한 후 PCA를 통하여 각 대역별 특징벡터를 추출한다. 이렇게 추출된 특징벡터는 학습영상에 의해 만들어진 코드북과의 비교를 통하여 관측벡터인 인덱스를 산출한다. 즉, KNN을 이용하여 검증영상의 특징벡터와 코드북 사이의 거리가 가장 가까운 코드북의 인덱스를 산출한다. 이렇게 생성된 검증영상의 인덱스는 각각의 영상별 HMM 모델에 적용한다. 이때, Forward 알고리즘을 이용하여 각각의 모델에 해당하는 관측벡터의 확률값을 구하게 된다. Forward 알고리즘은 모델  $\lambda=(A, B, \pi)$ 가 주어질 때 시간 t에서 상태 j이고 시간 t까지 관측열  $O_1 O_2 \dots O_t$  일 확률을 구하는 알고리즘이다. 아래 그림 8은 Forward 알고리즘의 과정을 나타내었다.

Forward 알고리즘을 통하여 구해진 확률이 가장 큰 감정 모델이 검증영상의 감정으로 인식하게 된다. 더 나은 인식을 위해 본 논문에서는 그림 9에 나타난 웨이블렛 변환을 이용하여 다해상도 분석을 하였다. 위의 과정을 웨이블렛 대역 별로 거친 후 얻어진 각 대역별 확률값의 합으로 인해 최종 확률값을 구하게 된다. 이렇게 산출된 확률값이 가장 높은 모델을 선택하여 인식한다.

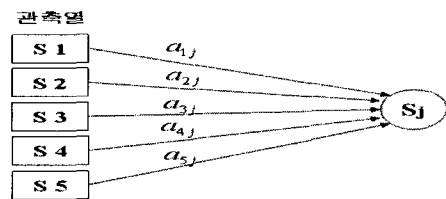


그림 8. Forward 알고리즘의 흐름도

Fig. 8. Forward algorithm of HMM

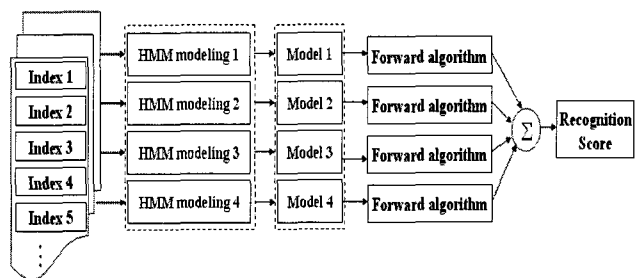


그림 9. 웨이블렛 4대역에 대한 감정인식 인식과정

Fig. 9. Facial emotion recognition using wavelet

3.3 정지영상과 동영상을 이용한 감정인식 융합

얼굴영상에 있어서 얼굴표정의 시간적 변화에 따라 감정을 인식하려는 시도가 있었다. 이것은 정지영상에서의 감정 표현이 가지는 애매성을 극복 할 수 있기 때문이다. 따라서, 본 논문에서는 정지영상과 연속영상에서의 감정 추출의 장점을 살리기 위하여 두 가지 기법을 모두 사용하여 높은 인식률을 얻고자 한다. 영상의 동적 특성으로부터 감정을 인식하려는 기법으로는 광 플로우(optical flow), potential net 등의 기법이 있으나, 이들은 계산량이 많고 정지 화상에서 추출되는 특징값들을 이용할 수 없는 단점이 있다. 따라서, 본 논문에서는 HMM을 이용한 동적영상에서의 감정 추출과 정지영상에서의 감정추출의 융합으로 인해 향상된 인식률을 얻고자 한다.

우선 10명에 대해 얼굴의 시퀀스 영상을 얻는다. 이 시퀀스의 마지막 5번째 감정 영상으로 PCA를 이용하여 특징을 추출한다. 정지영상의 인식부에서는 검증영상과 학습영상의 유사도를 측정하기 위하여 유클리디안 거리법을 이용한다. 최종으로 정지영상에서의 인식은 학습영상과 검증영상간의 벡터 공간상의 거리가 가장 가까운 감정이 인식되는 것이다. 다음으로, 동영상에서의 감정인식의 인식부에서는 Forward 알고리즘을 이용하여 각각의 모델에 해당하는 관측벡터의 확률값을 구하게 된다. 동영상에서 최종 인식은 이 확률값이 가장 큰 감정이 인식된다. 정지영상 감정인식의 결과인 유클리디안 거리값과 동영상 감정인식의 결과인 최종 관측벡터의 확률값은 독립적인 둘 사이의 비교를 위해 아래의 식으로 정규화 과정을 거친다.

$$x' = \frac{x}{\max(X)}, \quad x \in X \quad (4)$$

식 (4)에 의해 정지영상과 동영상의 결과 값은 모두 0 과 1 사이로 정규화 된다. 최종 얼굴감정인식단계에서는 그림 10에서 보는 바와 같이 정지영상에 의한 매칭값과 동영상에 의한 매칭값을 합산하여 매칭도가 가장 우수한 감정을 인식하도록 설계하였다. 이렇게 함으로서 조명 등의 노이즈에 의해 정지영상의 매칭값이 어느 정도 떨어지더라도 동영상의 매칭값이 상대적으로 클 경우 정지영상의 매칭값을 어느 정도 보완해 줌으로서 향상된 성능을 보인다. 또한 이 융합방법은 동영상의 매칭값이 어느 정도 떨어지더라도 정지영상의 매칭값이 상대적으로 클 경우 동영상의 매칭값을 어느 정도 보완해 줌으로서 정지영상 또는 동영상만을 사용하는 감정인식에 비하여 향상된 결과를 나타낼 수 있다.

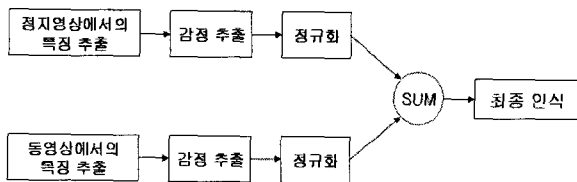


그림 10. 융합 과정  
Fig. 10. Fusion decision scheme

4. 실험 및 결과

본 논문에서는 제안한 알고리즘을 검증하기 위해 동영상 얼굴 감정 데이터베이스를 구축하여 실험하였다. 얼굴 감정

영상은 CANON 사의 power-shot A70모델의 디지털 카메라를 사용하였다. 데이터 베이스 습득방법은 초당 약 2.2매의 속도로 연사촬영을 하여 감정이 변해 가는 과정을 촬영하였다. 10명(남자5명, 여자5명)에 대해 6가지 감정(기쁨, 슬픔, 화남, 놀람, 공포, 혐오)의 변하는 과정을 취득하였으며 한 사람이 하나의 감정마다 6번씩 촬영하였다. 이때, 한번의 감정을 취득하는 데는 무표정에서 어떤 하나의 감정에 이르기까지의 과정을 시간에 따라 5번 연사 촬영하여 5개의 시퀀스로 나타내었다. 그림 11은 데이터 베이스를 위해 10명(남자5명, 여자 5명)에 대한 6가지 기본감정을 취득한 것 중 기쁨에 대한 영상의 일부를 나타냈다. 각 영상의 크기는 640×480으로 획득하였다.

본 데이터베이스는 전체 10명에 대해 6개의 감정을 5개의 시퀀스로 6번씩 촬영하였으므로 전체 1800장(10×6×5×6=1800)이 된다. 이 영상 중 절반(900장=180set)은 학습영상으로, 나머지 반(900장=180set)은 검증영상으로 구분하여 실험하였다. 정지영상은 10명(남성:5명, 여성:5명)으로 총360장이 된다. 이 영상 중 180장은 학습영상으로 180장은 검증영상으로 사용되었다.

정지영상에서의 얼굴 데이터베이스는 위에서 설명한 동영상 시퀀스의 마지막 감정 시퀀스를 이용하여 실험하였다. 동영상과 마찬가지로 10명에 대해서 6가지 감정(기쁨, 슬픔, 화남, 놀람, 공포, 혐오)을 사용하였으며 한 사람이 하나의 감정마다 6개의 영상을 취득하여 총 360장(10장×6개 감정×6장)의 영상을 사용하였다. 그 중 180장은 학습영상, 180장은 검증영상으로 사용하였다. 아래 영상은 6개의 기본감정영상이다. 그림 12에 정지영상에 대한 6개의 기본감정 영상의 일부를 나타냈다.

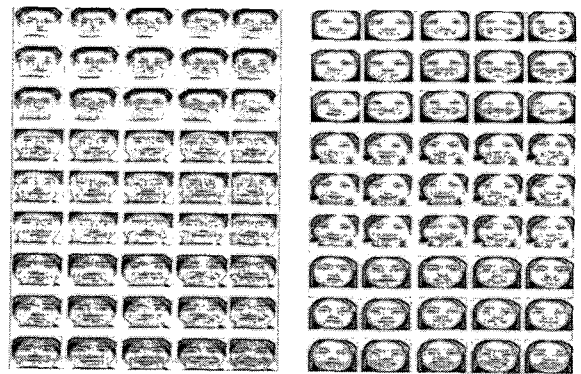


그림 11. 동영상에서의 얼굴감정 (기쁨)  
Fig. 11. Dynamic facial images (happy)



그림 12. 여섯 가지 감정(기쁨, 슬픔, 화남, 놀람, 공포, 혐오)  
Fig. 12. Six basic emotions (happy, sad, anger, surprise, fear and, dislike)

정지영상에서의 웨이블릿 대역별 인식결과를 표 1에 나타냈다. 표 1에서 보는 바와 같이 웨이블릿과 PCA를 사용한 경우 저대역에서 남성의 경우 88.89%의 인식률을 얻을 수 있었고, 여성의 경우 93.33%의 인식률을 얻을 수 있었다. 표 1에서 보는 바와 같이 웨이블릿 첫 번째 저대역은 이외의 대

역에 비해 월등히 좋은 인식률을 얻었으며, 이와 같은 결과로 정지영상에서 얼굴감정인식은 첫 번째 대역만을 이용하였다. 표 2에서는 얼굴표정을 이용한 감정별 인식결과를 나타냈다. 표 2에서 알 수 있듯이 “슬픔”에 대한 감정은 “화남” 감정으로 일부 오인식되어 다른 감정에 비해 좋지 않은 결과를 나타냈다.

동영상에서의 감정 인식결과를 알아보기 위한 전처리 과정으로 원영상(640×480)을 nearest neighbor interpolation 기법을 사용하여 160×120으로 재조정 하였다. 즉, 다해상도 분석을 위해서 학습 영상에 대해 이산 웨이블릿 변환을 3번 적용함으로써, 20×15의 4개의 해상도 영상을 얻을 수 있다. 여기에서 4개의 해상도 영역은 다해상도 분할 방법의 마지막 단계로 이전대역의 LL(저주파영역)영역의 정보만을 연속으로 분해하여 얻어진 영역들이다. 인식률의 정확성과 성별에 따라 감정의 표현이 다를 것을 고려하여 남, 녀 각각의 감정인식을 수행하였고 인식률을 구하였다.

얼굴 표정을 이용한 동영상에서의 감정별 인식률을 알아보기 위하여 실험한 결과를 표 3에 나타냈다. 표 3에서 보인 바와 같이 여성의 감정표현이 남성보다 2% 높게 나타남을 알 수 있다. 표 4에서는 동영상에서의 얼굴감정별 인식결과를 나타냈다. 표 4에서 알 수 있듯이 정지영상에서의 감정인식과 마찬가지로 “슬픔”에 대한 감정은 “화남” 감정으로 일부 오인식되어 다른 감정에 비해 좋지 않은 결과를 나타냈다.

표 1. 정지영상에서의 감정인식 결과

Table 1. Emotion recognition for static facial images [단위 : %]

구 분		LL	LH	HL	HH
인식률	남 성	88.89	48.88	60	47.77
	여 성	93.33	56.66	46.67	37.78

표 2. 정지영상에서의 얼굴 감정별 인식결과

Table 2. Recognition result per each emotion for static facial images [단위 : 개]

감정	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	27	1	1	1	0	0
슬픔	1	24	4	1	0	0
화남	0	1	29	0	0	0
놀람	0	1	0	27	1	1
공포	0	0	0	0	30	0
혐오	1	0	1	1	0	27

표 3. 동영상에서의 감정인식 결과

Table 3. Emotion recognition for dynamic facial images [단위 : %]

구 분		LL	LH	HL	HH	Total
인식률	남 성	85.56	76.67	77.78	70.00	90.00
	여 성	92.22	87.78	74.44	78.89	92.22

표 4. 동영상에서의 얼굴 감정별 인식결과

Table 4. Recognition result per each emotion for dynamic facial images [단위 : 개]

감정	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	28	0	0	1	1	0
슬픔	1	25	4	0	0	0
화남	0	1	28	1	0	0
놀람	0	2	0	27	1	1
공포	1	0	0	0	29	0
혐오	0	1	0	0	1	27

마지막으로 본 논문에서 제안한 정지영상 및 동영상에 이 용한 감정인식 결과를 살펴보면, 남성의 경우 표 5에서 나타 낸 것과 같이 94.44%의 인식률을 얻을 수 있었다. 표 6에서 는 여성의 경우의 얼굴 감정인식결과를 나타냈다. 여성의 경 우 97.78%로 남성 보다 얼굴감정 인식이 우수함을 알 수 있 다. 이와 같은 결과는 한국 남성의 경우 여성보다 감정표현 이 부족하기 때문이라 생각된다. 표 5와 표 6을 종합해 볼 때 표 7과 같이 전체인식률을 조사할 수 있었다. 표 7에서 보는 바와 같이 “슬픔”과 “놀람”에 대한 감정인식이 다른 감 정인식에 비해 인식률이 떨어짐을 알 수 있다. 그림 13에서 는 정지영상, 동영상, 정지영상과 동영상을 융합한 경우의 인 식결과를 비교하여 나타냈다. 그림 13에서 알 수 있는 바와 같이 본 논문에서 제안된 정지영상과 동영상을 융합하여 감정 인식한 경우 인식률이 96.11%로서 정지영상과 동영상을 각각 인식한 결과보다 나은 인식률을 보였다. 최종인식은 정지영상 과 동영상의 최종 결과를 정규화 과정을 통하여 서로 부족한 부분을 상호보완 함으로써 향상된 인식률을 얻을 수 있었다.

표 5. 남성에 대한 융합 후 인식 결과

Table 5. Final recognition result for man. [단위 : 개]

구 분	기쁨	슬픔	화 남	놀 람	공 포	혐 오
기쁨	15	0	0	0	0	0
슬픔	1	13	0	1	0	0
화 남	0	0	15	0	0	0
놀 람	0	1	0	13	1	0
공 포	1	0	0	0	14	0
혐 오	0	0	0	0	0	15

표 6. 여성에 대한 융합 후 인식 결과

Table 6. Final recognition result for woman. [단위 : 개]

구 분	기쁨	슬픔	화 남	놀 람	공 포	혐 오
기쁨	15	0	0	0	0	0
슬픔	0	14	1	0	0	0
화 남	0	0	15	0	0	0
놀 람	0	0	0	14	0	1
공 포	0	0	0	0	15	0
혐 오	0	0	0	0	0	15

표 7. 최종 인식 결과  
Table 7. Final recognition result

[단위 : 개]

구분	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	30	0	0	0	0	0
슬픔	1	27	1	1	0	0
화남	0	0	30	0	0	0
놀람	0	1	0	27	1	1
공포	1	0	0	0	29	0
혐오	0	0	0	0	0	30

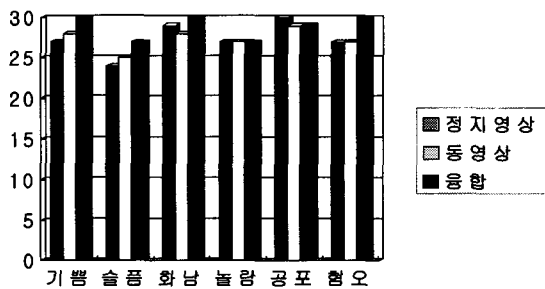


그림 13. 각각의 감정에 따른 인식률 비교  
Fig. 13. Emotion recognition for static, dynamic and fusion images

### 5. 결 론

본 논문에서는 얼굴을 휴먼 인터페이스로 구현하기 위해 HMM을 이용하여 얼굴의 감정인식에 대하여 연구하였다. 감정인식은 심리학자인 Ekman과 Friesen의 연구에 의해 문화에 영향을 받지 않고 공통으로 인식하는 6개의 기본 감정인 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 바탕으로 감정인식을 실험하였다. 사용된 데이터베이스는 10명(남자5명, 여자5명)에 대해 6가지 감정(기쁨, 슬픔, 화남, 놀람, 공포, 혐오)의 변하는 과정을 취득하였으며 한 사람이 하나의 감정마다 6번씩 촬영하였다. 이때 한번의 감정을 취득하는 데는 무표정에서 어떤 하나의 감정에 이르기까지의 과정을 시간에 따라 5번 연속 촬영하여 5개의 시퀀스로 나타내었다. 동영상에서는 이 데이터베이스의 시퀀스를 이용하여 실험하였고 정지영상의 감정인식에서는 감정이 완성된 마지막 시퀀스만을 이용하여 인식하였다.

실험결과, 정지영상의 감정인식에서는 웨이블릿과 PCA를 사용하여 인식하였다. 이때, 남성의 경우 88.89%의 인식률을 얻을 수 있었고 여성의 경우 93.33%의 인식률을 얻을 수 있었다. 동영상에서의 감정인식에서는 웨이블릿을 통하여 4개의 대역으로 분할된 영상을 HMM 과정을 거쳐 각각에 대한 인식률을 구하고 4개 대역의 확률값의 합에 의해 최종 확률값을 얻었다. 이때, 남성의 경우 90.00%, 여성의 경우 92.22%의 인식률을 얻을 수 있었다. 최종인식은 정지영상과 동영상의 최종 결과를 정규화 과정을 통하여 서로 부족한 부분을 상호보완 함으로써 인식률은 남성 94.44%, 여성 97.78%로서 정지영상과 동영상을 각각 인식한 결과보다 향상된 인식률을 얻을 수 있었다.

향후 과제로는 정지영상과 동영상을 이용한 감정인식을 융합하는 과정에 있어 정규화를 이용한 방법 이외에 좀 더

세밀한 과정을 연구하여야 할 것이며, 또한 다양한 인식방법과의 비교를 통하여 제안된 방법의 타당성을 검증할 필요가 있다.

### 참 고 문 헌

- [1] P.Ekman and W.V. Friesen. *Emotion in the human face System*. Cambridge University Press, San Francisco, CA, second edition, 1982.
- [2] David S. Bolme, *Elastic Bunch Graph Matching*, Thesis of the Degree of Master, Colorado University, Summer 2003.
- [3] M. Turk, A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.
- [4] P. Penev, J. Atick, "Local feature analysis: a general statistical theory for object representation", *Network : Computation in Neural Systems*, Vol. 7, pp. 477-500, 1996.
- [5] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, 1997.
- [6] C. Padgett, G. Cottrell, "Representing face images for emotion classification", *Advances in Neural Information Processing Systems*, Vol. 9, MIT Press, 1997.
- [7] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, "Comparison between geometry based and gabor-wavelets-based facial expression recognition using multi-layer perceptron", *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 454-459, 1998
- [8] Katsuhiko Matsuno and Saburo Tsuji. "Recognizing human facial expressions in a potential field. In Proc CVPR, pp 44-49, 1994.
- [9] H. Ushida, T. Takagi, and T. Yamaguchi. "Recognition of facial expressions using conceptual fuzzy set." In Proc. CVPR, pp594-499, 1993.
- [10] 전명근, 생체인식(Biometric) 총론, 한국정보통신교육원, 2002.
- [11] Rodrigo de Luis-García, Carlos Alberola-López, Otman Aghzout and Juan Ruiz-Alzola "Biometric identification systems" *Signal Processing*, Vol. 83, Issue 12. pp. 2539-2557, December 2003.
- [12] Lin Hong, Anil Jain, "Integrating Faces and Fingerprints for Personal Identification", *IEEE Trans. on Pattern analysis and machine intelligence*, Vol. 20, No. 12, 1998.
- [13] Stephane Mallat, *A wavelet tour of signal processing*, Academic press, 1999.
- [14] Rabiner, LR, "A tutorial on hidden markov models and selesced applications in speech recognition" *proceedings of the IEEE*, Vol. 77, Issue 2, pp. 257-286, 1989.

저 자 소 개



**이대종(Dae Jong Lee)**

1995년 : 충북대학교 전기공학과(학사)  
1997년 : 충북대학교 전기공학과(공학석사)  
2002년 : 충북대학교 전기공학과(공학박사)  
2003년~현재 : 충북대학교 컴퓨터정보통신연구소  
2004년~2005년 : University of Alberta, Post doc

관심분야 : 음성신호처리, 서명인식, 다중생체인식



**이경아(Kyung Ah Lee)**

2002년 : 충북대학교 전기전자공학부(학사)  
2004년 : 충북대학교 제어계측공학과(석사)  
2004년~현재 : 다산네트웍스

관심분야 : 감정인식, 얼굴인식



**고현주(Hyoun Joo Go)**

1999년 : 한밭대학교 제어계측공학과(학사)  
2002년 : 충북대학교 제어계측공학과(공학석사)  
2002년~현재 : 충북대학교 제어계측공학과 박사과정

관심분야 : Biometrics, Computer vision, 감정인식



**진명근(Myung Geun Chun)**

1987년 : 부산대학교 전자공학과(학사)  
1989년 : 한국과학기술원 전기 및 전자공학과(공학석사)  
1993년 : 한국과학기술원 전기 및 전자공학과(공학박사)  
1993년~1996년 : 삼성전자 자동화연구소 선임연구원

2000년~2001년 : University of Alberta 방문교수  
1996년~현재 : 충북대학교 전기전자 컴퓨터공학부 교수

관심분야 : Biometrics, 감정인식, 지능시스템