얼굴표정과 음성을 이용한 감정인식
(An Emotion Recognition Method using Facial Expression and Speech Signal)

고 현 주 † 이 대 종 †† 전 명 근 †††
(Hyoun-Joo Go) (Dae-Jong Lee) (Myung-Geun Chun)

요 약 본 논문에서는 사람의 얼굴표정과 음성 속에 담긴 6개의 기본감정(기쁨, 슬픔, 화남, 놀람, 혐오, 공포)에 대한 특징을 추출하고 인식하고자 한다. 이를 위해 얼굴표정을 이용한 감정인식에서는 이산 웨이블렛 기반 다해상도 분석을 이용하여 선형판별분석기법으로 특징을 추출하고 최소 거리 분류 방법을 이용하여 감정을 인식한다. 음성에서의 감정인식은 웨이블렛 필터뱅크를 이용하여 독립적인 감정을 확인한 후 다중의사 결정 기법에 의해 감정인식을 한다. 최종적으로 얼굴 표정에서의 감정인식과 음성에서의 감정인식을 융합하는 단계로 퍼지 소속함수를 이용하며, 각 감정에 대하여 소속도로 표현된 매칭 값은 얼굴에서의 감정과 음성에서의 감정별로 더하고 그중 가장 큰 값을 인식 대상의 감정으로 선정한다.

키워드 : 감정인식, 음성인식, 얼굴인식, 웨이블렛 변환, 선형판별분석기법

Abstract In this paper, we deal with an emotion recognition method using facial images and speech signal. Six basic human emotions including happiness, sadness, anger, surprise, fear and dislike are investigated. Emotion recognition using the facial expression is performed by using a multi-resolution analysis based on the discrete wavelet transform. And then, the feature vectors are extracted from the linear discriminant analysis method. On the other hand, the emotion recognition from speech signal method has a structure of performing the recognition algorithm independently for each wavelet subband and then the final recognition is obtained from a multi-decision making scheme.

Key words : Emotion Recognition, Speech Recognition, Face Recognition, Wavelet Transform

1. 서 론

최근 정보화 사회의 발달로 컴퓨터가 급속도로 대중화되어 가고 있으며, 과거에 사용되는 곳이 한정되어 있었던 고기능의 개인용 컴퓨터들이 각 가정으로 확산되어 보급됨에 따라 인간과 컴퓨터의 상호작용은 능동적인 양방향성 인터페이스로 변화 되어가면서 좀더 자연스럽고 쉬운 형태로 발전하고 있다. 이러한 휴먼 인터페이스 기술은 사용자의 감정 상태를 추출, 인식하는 것을 목적으로 하고 있으며, 사용자의 감정상태에 대한 인식을 설계하기 위한 도구로 언어, 음성, 제스처, 시각, 청각 등을 이용하고 있다.

· 본 연구는 한국과학재단 한·일 국제 공동연구 지원으로 수행되었습니다.
† 학생회원 : 충북대학교 제어계측공학과
ghjswy@hanmail.net
†† 비 회 원 : 충북대학교 컴퓨터정보통신 연구소
leebigbell@hanmir.com
††† 비 회 원 : 충북대학교 전기전자컴퓨터공학부 교수
mgchun@chungbuk.ac.kr
논문접수 : 2003년 7월 21일
심사완료 : 2003년 3월 13일

이와 관련한 사람의 얼굴표정은 감정을 전달함에 있어 중요한 역할을 하는 생체인식 분야 중 하나로, 연구방법으로는 광학적 흐름 분석(optic flow analysis), 홀리스틱 분석(holistic analysis), 국부적인 표현(local representation) 등이 있다. 광학적 흐름 분석에는 Line이 얼굴 감정인식을 수행하기 위해 광학적 흐름 추정을 통한 얼굴의 모션 분석을 하였으며[1], 홀리스틱 분석 방법으로는 PCA(principal component analysis)[2], LFA(Local Feature Analysis)[3], LDA(Linear Discriminant Analysis)[4], ICA(Independent Component Analysis)등이 연구되어지고 있다. 국부적인 표현방법으로는 얼굴영상의 세부적인 영역을 다루는 Local PCA [5], Gabor 웨이블렛 표현방법[6] 등이 있다.

한편, 음성은 청각에 기반을 둔 효율적이고 자연스러운 방법으로 여기에 내포된 감정을 추출하려는 연구가 활발히 행해지고 있다. Fukuda는 음성신호의 템포와 에너지를 가지고 여섯 개의 기본감정에 대한 분류를 시도하였는데, 녹음실과 같은 외부 잡음이 전혀 없는 환경하에서 일본어와 이탈리아어에 대한 음성신호를 녹음한

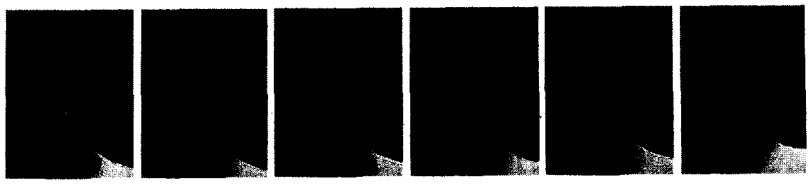

그림 1 여섯 가지 기본 감정(기쁨, 슬픔, 화남, 놀람, 혐오, 공포) 표현

후 감정 추출에 대한 연구를 하였다[7]. Moriyama는 음성신호의 피치(pitch)와 전력의 포락선 검출을 통하여 20개의 일본어 샘플에 대하여 실험하였고, 실험 결과 '화남' '슬픔' '놀람' 감정이 다른 감정보다 인식률이 비교적 높은 것으로 나타났다[8]. 또한, Silva는 음성신호의 피치와 HMM(Hidden Markov Model)을 이용하여 영어와 스페인어에 대하여 감정추출을 실험하였다[9]. 한편, 국내에서도 음성을 이용한 감정인식 연구가 활발히 진행되고 있는 요즈음, 우리나라 국악의 창에서 인간의 희로애락을 표현하는 음의 고저와 장단을 기본으로 하여 분석하는 연구가 행하여 졌으며[10], 화난 감정의 독특한 특성을 찾아내기 위해, 대화의 내용에 사용한 단어나, 톤(Tone), 음성신호의 피치(Pitch), 포만트 주파수(Formant Frequency), 말의 빠르기, 음질 등을 이용하는 연구도 진행 중에 있다[11].

그림 1은 심리학자인 Ekman과 Friesen의 연구에 의해 분류된 6개의 기본 감정인 기쁨, 슬픔, 화남, 놀람, 공포, 혐오를 얼굴 표정에 나타낸 것으로, 나라마다 개인마다 감정을 표현함에 있어 다소 차이는 있을 수 있으나, 가장 일반적인 감정상태를 나타내는 표정을 보인 것으로 본 논문의 연구를 위해 사용되어진 데이타베이스중 하나이다[12].

본 논문에서는 위 6개의 기본 감정을 바탕으로 한 얼굴표정과 목소리를 동시에 이용하여 사람의 감정을 인식하는 알고리듬을 제안하고자 한다. 얼굴표정을 이용한 감정인식에서는 이산 웨이블렛을 기반으로 하는 다 해상도 분석 기법을 사용하고, 각 해상도에서 얻어진 계수를 이용하여 LDA 기법으로 특징을 추출하고 최소 거리 분류 방법인 유클리디안 거리를 이용하여 감정을 인식한다. 음성에서의 감정인식은 웨이블렛 필터뱅크를 이용하여 독립적인 감정 확인을 한 후 다중의사 결정 기법에 의해 감정을 인식하는 구조로 이루어져 있다. 최종적으로 얼굴 표정에서의 감정인식과 음성에서의 감정 인식에 대한 매칭 결과 값을 퍼지 소속함수를 이용하여 변환하고 그에 의한 결과를 감정별로 더한 후 소속도가 가장 높은 감정을 대상 감정으로 인식하는 융합모델을 구성하였다. 논문의 구성은 2장에서 얼굴표정을 이용한 감정인식과, 3장에서 음성신호를 이용한 감정인식을 논하였다. 그리고 4장에서는 본 논문에서 제안한 얼굴표정과 음성을 동시에 이용한 감정인식으로 융합하는 과정을 설명하였으며, 제안한 알고리듬과 관련한 실험 및 고찰을 설명하였다. 마지막으로 5장에서 결론을 맺는다.

2. 얼굴표정을 이용한 감정인식

2.1 웨이블렛과 LDA를 이용한 얼굴영상의 분해

웨이블렛 변환(Wavelet Transform)은 비 주기적인 신호분리가 가능한 기저함수를 사용하여 신호를 해석하는 것으로 신호를 형성하고 있는 주파수가 다른 두 개의 사인함수와 하나의 델타함수를 "시간-스케일" 공간에 정확하게 분리해 낸다[13]. 2차원의 경우 웨이블렛은 아주 작은 비트율로 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, 이산 웨이블렛 변환을 영상신호에 적용하는 것은 영상을 공간상의 x축과 y축 방향으로 저대역 통과필터(LPF)와 고대역 통과 필터(HPF)를 사용하여 신호를 추출하는 것을 의미하는 것으로, 이산 웨이블렛 변환을 거친 신호는 총 네 개로 분리될 수 있다.

그림 2는 원 영상에 대해 한번의 웨이블렛 변환을 거친 후 LL1, LH1, HL1, HH1으로 분리된 영상을 보여준 것이며, LL1 영상에 대해 한번 더 웨이블렛 변환을 거친 LL2, LH2, HL2, HH2으로 분리된 영상을 보여주고 있다. 본 논문에서는 네 번의 웨이블렛 변환을 거친 LL4, LH4, HL4, HH4을 사용하였으며, 기저함수로는 Daubechies[13]를 사용하였다.

한편, 일반적으로 얼굴 영상은 매우 고차원의 데이타로 표현되기 때문에 특징 추출과 분류를 위해서는 저차원의 데이타로 표현되는 것이 요구된다. 얼굴 인식에서 주성분분석기법은 학습영상의 2차 통계적 특성을 이용하여 학습영상의 전체적인 특성을 표현하는 직교기저영상인 고유얼굴로 분해할 수 있으며, 이 고유얼굴의 선형조합으로 임의의 얼굴 영상을 표현하는 방법이며 입력

그림 2 웨이블렛 변환후 4개의 밴드로 분리된 영상

데이타를 저차원의 데이타로 표현하는 효과적인 방법이다. 그림 3은 PCA를 이용한 얼굴인식으로 특징벡터 $a_1, a_2 \ldots a_n$과 고유얼굴(Eigenfaces)의 선형적인 결합에 의해 얼굴영상들을 표현할 수 있다.

선형판별분석기법(LDA)은 클래스 내의 분산을 나타내는 행렬(Within-Scatter Matrix)과 클래스 간 분산을 나타내는 행렬(Between-Scatter Matrix)의 비율이 최대가 되도록 하는 선형 변환 방법으로, PCA 방법은 영상 공간에서 저차원의 특징 공간으로의 선형 사영을 기초로 하므로 전체 데이타 베이스의 모든 얼굴 영상을 최대화하는 사영 방향을 찾아낸다. 그러나 조명 조건과 얼굴 표정의 변화로 생기는 원하지 않는 변화도 포함되게 되므로 PCA 방법은 저차원의 기저벡터로부터 복원을 하는 관점에서는 최적의 방법이지만 조명이나 표정 변화가 있는 얼굴영상의 식별, 인식에서는 LDA가 우수한 인식성능을 나타내고 있다[14]. LDA 방법은 Fisherfaces를 기반으로 한 효율적인 인식방법으로, 얼굴인식에서 가장 많이 연구되어 지고 있으며, 얼굴 표정을 이용한 감정인식에도 이와 같은 방법이 사용되고 있다. 본 연구에서는 감정의 변화와 개성에 따라 얼굴 표정이 다양할 수 있으나 각 감정별로 얼굴영상에서 공통된 특징을 추출하고 이를 이용한 감정을 분류할 수 있다. 이를 위해 얼굴인식에 많이 사용되어 지고 있는 PCA, ICA를 이용한 감정인식[15] 뿐만아니라 LDA를 이용한 얼굴표정의 특징벡터로 감정인식에 적용해 보는 것을 제안하며, 네 번의 웨이블렛 변환을 거친 후의 영상인 LL4, LH4, HL4, HH4를 특징벡터를 얻기 위한 입력 영상으로 사용한다.

2.2 얼굴영상을 이용한 감정인식

본 논문에서 다 해상도 분석을 위해 학습이미지에 대해 이산 웨이블렛 변환을 네 번 적용하였으며, 원 이미지(640x480)에 대해 세 번째 대역으로부터 사이즈 40x30의 네 개의 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, 웨이블렛 함수에 의해 사영되어진, 수직, 수평, 대각선의 방향성을 가진 세 개의 영역으로 구성되어 있으며, 각각의 영역에 대한 정보들을 해상도별로 저장한다. 그리고 앞에서와 같은 웨이블렛 결과로 얻어진 LL4 영상(40x30)에 대해, 먼저 차원축소를 위해 PCA 방법을 적용한다. 이때, 고유벡터에 대해 남자 150개, 여자 170를 사용하였을 때, 인식 결과가 가장 우수하였다. PCA를 거쳐 차원 축소된 결과는 LDA 방법을 이용하여 fisherface를 얻을 수 있는데 이때, fisherface는 최대 5개(클래스-1)를 설정해 줄 수 있으며, 본 논문에서는 5개를 사용하였다. 그림 4는 LDA에 의해 얻어진 5개의 fisherface를 보여주고 있다.

위와 같은 환경 하에 얼굴영상만을 이용한 인식결과를 알아보기 위해 웨이블렛 변환을 세 번 거친 후의 영상을 이용하여 기존에 만들어 놓은 코드 북과 비교하여, 마지막 대역의 4개 영상에 대해 PCA를 이용한 인식률과 ICA를 이용한 인식률을 조사해 보았다. 표 1은 PCA와 ICA를 이용한 인식결과와 본 논문에서 채택한 LDA를 이용한 인식률을 보이고 있다. 표에서와 같이, PCA와 ICA를 이용한 방법보다 LDA를 사용하여 얼굴 감정인식을 실험해 보았을 때, 남자의 경우 90%, 여자

그림 4 LDA에 의해 얻어진 fisherface

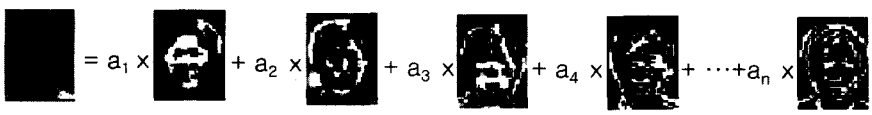

그림 3 PCA를 이용한 고유얼굴

의 경우 95%의 인식률로, 첫 번째 대역은 이외의 대역에 비하여 월등히 좋은 인식률을 획득 할 수 있었다. 이와 같은 결과는 앞에서 언급되었던 것과 같이 PCA, LDA를 이용하였을 때에는 저차원의 기저벡터로부터 복원을 하는 관점에서는 최적의 방법이 될 수 있으나, 조명이나 표정변화가 있는 얼굴영상의 식별, 인식에는 좋은 성능을 보이지 않았으며, LDA를 이용하는 방법은 얼굴표정의 변화를 클래스 내의 분산과 클래스간의 분산 비율이 최대가 되도록 선형 변환하는 방법으로 본 연구와 같은 감정인식을 위한 표정변화를 식별하고자 할 때 더욱 우수하다고 할 수 있다. 이와 같은 결과로 인해 LDA를 이용한 얼굴감정인식은 표 1에서와 같이 첫 번째 대역인 LL3을 사용하였다.

표 1 PCA, ICA, LDA를 이용한 얼굴 감정 인식 결과

(단위 : %)

	남 자	89.4	71.1	65	62.2
	여 자	88.8	73.3	67.2	60
	남 자	88.8	76.1	68.3	72.7
	여 자	92.2	78.3	67.7	67.2
	남 자		75	74.4	70
	여 자		70	64.4	75

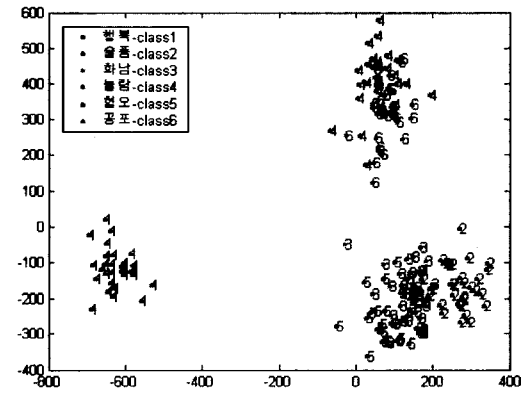

그림 5 감정별 분류된 특징벡터

그림 5는 LDA에 의해 얻어진 5차원의 특징벡터 중 두 개의 차원에 대해 감정별 분포를 보이고 있다. 그림에서 알 수 있듯이 감정별로 데이타들이 군집되어 있음을 알 수 있고, 이는 LDA를 거쳐 효과적으로 분류될 수 있다.

3. 웨이블렛 필터뱅크를 이용한 음성 감정인식

이산 웨이블렛 변환은 고역 통과 부분을 한 단계의 필터 뱅크로 구성하고, 저역 통과 부분을 계속적인 필터뱅크로 확장하는 옥타브 밴드(octave-band)구조와 고역 통과 부분도 필터뱅크로 확장하는 구조를 가지는 웨이블렛 패킷(wavelet packet)구조로 구현될 수 있다[16]. 그림 6과 그림 7은 옥타브 밴드 구조와 웨이블렛 패킷 구조를 보이고 있는데, 여기서 g[n]은 저역 통과 필터를 h[n]은 고역통과필터를 각각 나타내며, 마더 웨이블렛으로 부터 구성됨을 알 수 있으며, ↓2는 샘플의 개수를 1/2로 줄이는 데시메이션(decimation)을 나타낸다. 또한, 그림 8은 웨이블렛 패킷 구조로 필터뱅크를 통해 나오는 출력신호 ⓐ는 고주파의 고주파신호며, ⓑ는 고주파의 저주파신호, ⓒ는 저주파의 고주파신호, ⓓ는 저주파의 저주파 신호를 가지고 있으며, 각각의 신호에 대해 특징이 되는 신호를 선택해 사용할 수 있다.

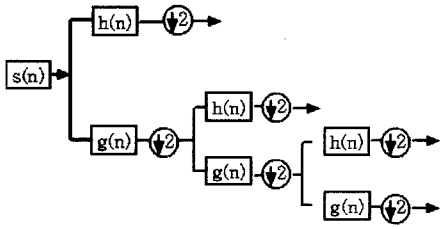

그림 6 웨이블렛 옥타브 밴드의 구조

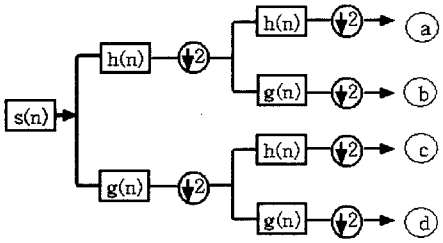

그림 7 웨이블렛 패킷의 구조

본 연구에서는 주파수 대역을 균등하게 분할하는 방식인 웨이블렛 패킷구조방식을 사용하였으며, 필터뱅크의 출력 수는 그림 7과 같은 4개의 필터뱅크로 구성되었고, 이 중 남자의 코드 북을 만드는 경우, 다양한 실험에서 매우 낮은 인식률을 보인 1개의 고주파대역을 즉, ⓐ 고주파의 고주파신호를 제외하고자 한다. 그리고 가장 널리 사용되고 있는 Daubechies 기저함수를 이용하여 신호를 해석하였다. 또한, 음성의 시작과 끝점을 찾아내는 방법으로 예측계수나 자기상관계수와 같은 음성특징 계수를 사용할 수 있으나, 본 논문에서는 현재 음성의 끝점을 위해 일반적으로 사용되어 지는 방법인 단구간 에너지와 단구간 영교차율(ZCR, zero crossing rate)을 이용하는 방법을 채택하였다.

이렇게 획득된 음성신호는 웨이블렛 필터링 기법을

이용하여 각각의 대역별로 출력 값을 획득할 수 있다. 이때, 4개의 웨이블렛 필터에서 출력되는 음성신호는 음성 분석부에서 특징벡터로 FFT기반 멜켑스트럼 계수를 구한 후 K-means 알고리듬을 이용하여 독립적인 코드북을 미리 만들어 놓는다. 이때, 감정인식률 향상을 위하여 남성화자와 여성화자용 코드북을 각각 만들었다.

그림 8은 본 논문에서 제안한 웨이블렛 필터뱅크를 이용한 감정인식기를 나타내었다. 그림에서와 같이 인식과정에서는 인식하고자 하는 음성신호가 입력되면 웨이블렛 변환하여 주파수별로 음성신호를 분할한다. 그리고 성별을 구분하여 미리 만들어 놓은 코드북과 비교하기 위해 저주파대역에서 피치를 이용한 성 식별을 분석한 후 음성 분석부에서 각각의 주파수 대역에 대한 특징벡터를 계산한다. 이와 같이 음성 분석부에서 계산된 특징벡터는 미리 뱅크별로 만들어 놓은 코드 북과의 거리를 계산한 후 독립적인 인식률을 산출한다. 그림 9는 다중 밴드 의사결정 방법에 대해 보이고 있다[17]. 각 대역별

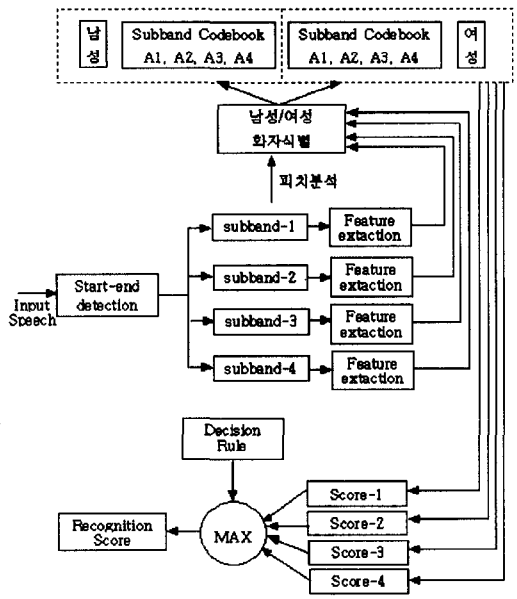

그림 8 웨이블렛 필터뱅크를 이용한 감정인식기

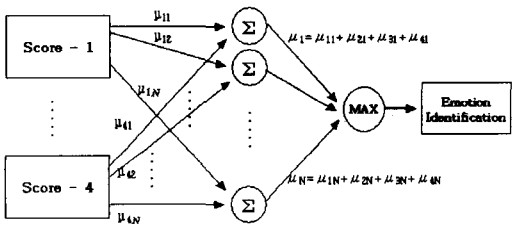

그림 9 다중 밴드 의사결정 방법

에서 산출된 인식률은 음성신호를 프레임으로 나누고 각각의 프레임에서 얻어진 특징벡터와 코드북과의 거리계산에 의하여 산출되기 때문에 어느 특정 감정에 대한 정보만을 가진 것이 아니라 인식하고자 하는 각각의 감정들에 대한 소속정도를 모두 가지고 있다.

이와 같은 시스템은 일반적으로 인식대상의 특징벡터의 대표 값들을 나타내는 코드북 및 특징벡터의 종류에 따라 인식률에 큰 차이를 보인다. 일반적으로 코드북 사이즈가 클수록 인식률이 향상되지만, 인식속도의 저하 및 메모리상의 문제점으로 인하여 사전에 코드북의 적정 사이즈를 결정할 필요가 있다. 또한 인식대상을 잘 표현해 줄 수 있는 특징벡터의 선정도 중요한데, 본 연구에서는 일반적으로 사용되는 멜켑스트럼 계수 13차를 이용하였으며, 코드북의 사이즈는 384×13로 4개의 대역으로 특징벡터의 종류별 인식률을 조사하였다.

표 2는 음성 신호만를 이용한 감정인식결과를 알아보기 위해 웨이블렛 변환을 두 번 거친 후 멜켑스트럼 계수를 구해 얻은 특징 값을 기존에 만들어 놓은 코드 북과 비교하여 네 개의 대역에 대해 인식률을 나타낸 것으로, 대역별 인식률이 모두 다르며, 남자의 경우 제일 낮은 값을 갖는 마지막 대역(A4)을 제외한 세 개의 대역을 이용한 경우 93.3%의 인식률을 얻을 수 있었으며, 여자의 경우 네 개의 대역을 모두 사용한 경우 가장 높은 93.3%의 인식률을 얻을 수 있었다. 이와 같은 결과는 A4대역은 가장 높은 고주파의 고주파 신호로 남자의 음성신호는 여자의 음성신호보다 낮은 주파수대역을 갖는 경우가 많으므로 높은 주파수 대역인 A4 대역을 제외하는 것이 좋은 성능을 보였다.

표 2 대역별 감정인식률 비교 (단위 : %)

구 분	Band				밴드3개	밴드4개
	A1	A2	A3	A4		
남 자	89	78	71	57	93.3	85
여 자	85	72	81	68	86	93.3

4. 얼굴표정과 음성을 이용한 감정인식

4.1 얼굴표정과 음성을 이용한 감정인식

본 논문에서는 이산 웨이블렛 변환과 LDA를 거쳐 얻은 감정별 얼굴영상에 대한 특징 값과 사람의 음성으로부터 얻어진 감정별 음성에 대해 웨이블렛 변환과 멜켑스트럽을 거쳐 얻어진 특징 값을 이용해 사람의 얼굴표정과 음성을 동시에 이용한 감정추출 알고리듬을 제안하고자 한다. 우선 대상자 20명에 대해 얼굴영상과 음성신호를 입력받아 음성신호의 피치를 검출하여 대상자를 여자와 남자로 나눌 수 있다. 그리고 나누어진 여자와

남자대상자별 음성신호와 얼굴영상에 대해 각각 코드북을 만든다. 그림 10은 20명의 화자에 대해 코드북을 만드는 과정을 보여주고 있다.

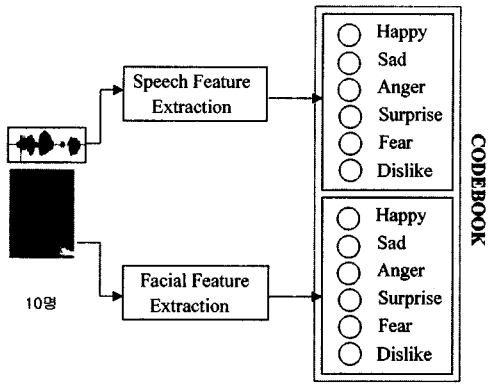

그림 10 화자 10명에 대한 코드북 생성 과정

위와 같이 코드북이 만들어졌을 때 인식하는 과정은 시스템의 입력부로 부터 대상 화자의 얼굴영상과 음성 입력이 들어오면, 음성신호의 피치를 조사하여 여자, 남자를 판명하고 판명된 결과로 음성신호와 얼굴영상을 분류하여 음성신호는 앞에서와 같은 방법으로 특징벡터를 구하고 기존의 코드북과 비교하여 감정별 소속 값 μ_{1i}(기쁨($i=1$), 슬픔($i=2$), 화남($i=3$), 놀람($i=4$), 혐오($i=5$), 공포($i=6$))로 표현할 수 있다. 또한, 얼굴영상에 대하여 앞에서와 같은 방법으로 특징벡터를 구하고 기존이 얼굴영상 코드북과 비교하여 감정별 소속 값 μ_{2i}(기쁨($i=1$), 슬픔($i=2$), 화남($i=3$), 놀람($i=4$), 혐오($i=5$), 공포($i=6$))로 표현할 수 있다. 이때, S-모양의 소속함수(S-Type Membership Function)를 이용하여 각 감정에 대해 0과 1사이의 소속도로 표현할 수 있게 되는데, 소속함수의 경계 값은 유사도의 분포를 확인하고 실험자의 경험에 의해 결정되어 질 수 있다[18]. 본 논문에서는 0.2, 0.8을 사용하였으며, 이렇게 얻어진 얼굴영상에서 대한 각 감정별 소속도와 음성 신호에서의 각 감정에 대한 소속도를 같은 감정별로 비교했을 때, 동일한 감정에 대해 두 개의 소속도 중 큰 값을 선택하는 방법과 각 감정에 대해 두 개의 소속도 값을 모두 더하는 방법이 있을 수 있다. 본 논문에서는 각 감정별 소속도를 모두 더했을 때(i의 감정 = $\mu_{1i}+\mu_{2i}$) 가장 큰 값을 인식 대상으로 선정하는 방법을 사용하였다. 이때, 최종 인식 값은 특정감정의 정보만 가지고 있는 것이 아니라 다른 감정에 대한 소속정보도 확인할 수 있다. 그림 11은 얼굴영상과 음성을 동시에 이용한 최종 인식 결정방법을 보이고 있다.

4.2 얼굴표정과 음성을 이용한 감정인식 실험 결과

제안한 방법의 유용성을 알아보기 위해 데이타베이스를 구축하고 실험장치를 준비하였다. 우선, 데이타베이스를 위해 연구실 학생 20명(남자 10명 여자 10명)에 대한 6가지 기본감정(행복, 슬픔, 화남, 놀람, 공포, 혐오)을 취득함에 있어, 얼굴영상은 기본감정이 깃 든 얼굴영상 640×480의 크기를 획득하였으며, 음성신호에 대하여는 "아! 그렇습니까"에 대해 각 감정별 음성신호를 획득하였다. 얼굴영상에 대한 데이타는 총 720개(20명×6개 감정×6장)의 영상으로, 여자에 대한 630개중 180장에 대하여 학습영상, 즉 코드북을 만드는데 사용하였으며, 나머지 180장에 대하여는 검증영상으로 구분하였으며, 장치로는 디지털카메라를 사용하였다. 또한, 음성신호에 대한 데이타로 한 사람이 하나의 감정마다 3개의 음성을 취득하여 총 360개(20명×6개 감정×3개)의 음성을 사용하였다. 이 중 여자에 대한 180개중 120개를 이용하여 코드북을 만들었으며, 나머지 60개를 검증음성으로 하고 음성 취득용 소프트웨어로 상용중인 Cool Edit 2000을 사용하였다. 또한, 음성은 샘플링 주파수를

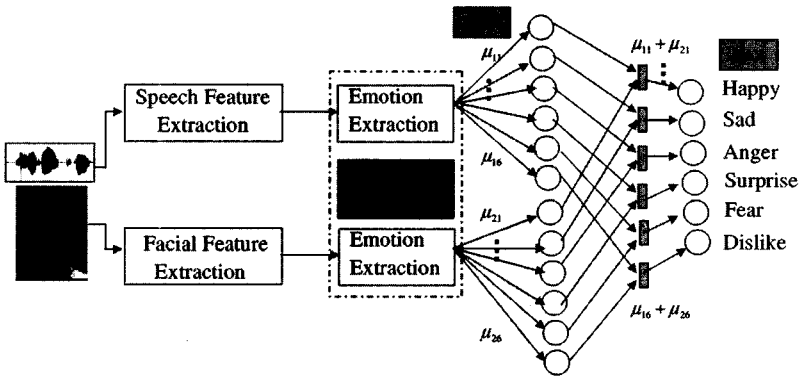

그림 11 얼굴영상과 음성을 이용한 최종 인식 결정방법

11.025kHz로 하였으며, 기준패턴인 코드북의 사이즈는 64로 하였다. 음성신호의 특징파라미터는 약 20ms 구간에서 음성신호가 정상이라는 가정아래 20ms의 프레임 단위로 구하게 된다. 그러나 본 논문에서는 10ms의 Hamming window를 사용하고, 프레밍 양 끝단의 신호 정보를 보상하기 위하여 5ms씩 중첩을 시켜서 윈도우를 이동시켰다. 이렇게 Hamming window를 사용하여 원 신호를 프레임 단위로 분할한 후 각각의 프레임에 포함된 데이타에서 13차의 멜켑스트럼 계수를 구하였다.

본 논문에서 제안한 얼굴표정과 음성을 동시에 이용한 감정인식 결과를 살펴보면, 남자의 경우 표 3에 나타낸 것과 같이 95%의 인식률을 얻을 수 있었다. 이와 같은 결과는 그림 11과 같이 최종 감정인식 결정 방법에 의한 것으로, 얼굴에서의 작은 매칭값의 차이로 오 인식된 결과를 음성에서 보안한 것이며, 음성에서의 작은 매칭값의 차이로 오 인식된 결과를 음성에서 보완한 결과라 할 수 있다.

표 3 남성화자에 대한 최종 인식 결과 (단위 : 명)

	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	10	0	0	0	0	0
슬픔	0	10	0	0	0	0
화남	0	0	10	0	0	0
놀람	0	0	1	9	0	0
공포	0	0	0	2	8	0
혐오	0	0	0	0	0	10

여자의 경우 표 4에 나타낸 것과 같이 98.3%의 높은 인식률을 얻을 수 있었으며, 이는 남자의 경우와 같이 얼굴을 이용한 감정인식과 음성을 이용한 감정인식에서 작은 매칭값의 차이로 오 인식된 경우, 최종 감정인식 결정 방법에 의해 보완함으로 얻어진 결과라 할 수 있다.

표 4 여성화자에 대한 최종 인식 결과 (단위 : 명)

	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	9	0	1	0	0	0
슬픔	0	10	0	0	0	0
화남	0	0	10	0	0	0
놀람	0	0	0	10	0	0
공포	0	0	0	0	10	0
혐오	0	0	0	0	0	10

표 3과 표 4의 결과를 종합해 볼 때 표 5와 같은 전체 인식률을 조사할 수 있다. 이는 기쁨감정에 대해 대상자 20명의 경우 1명의 대상자에 대해 화남으로 오 인식을 했으며, 놀람과 공포에 대하여도 각각 1명과 2명에 대하여 화남과 놀람으로 오 인식했음을 알 수 있다. 그러나 슬픔, 화남, 혐오와 같은 감정에 있어서 100%인식을 보였으며, 전체 96.65%로 비교적 우수한 성능을 보임을 확인할 수 있었다.

표 5 최종 인식 결과 종합 (단위 : 명)

	기쁨	슬픔	화남	놀람	공포	혐오
기쁨	19	0	1	0	0	0
슬픔	0	20	0	0	0	0
화남	0	0	20	0	0	0
놀람	0	0	1	19	0	0
공포	0	0	0	2	18	0
혐오	0	0	0	0	0	20

이와 같은 결과는 그림 12에서와 같이 얼굴을 이용한 인식률 남자의 경우 90%, 여자의 경우 95%와 음성을 이용한 인식률 남자의 경우 93.3%, 여자의 경우 93.3%의 인식률을 획득할 수 있었던 것에 대해 본 논문에서 제안한 얼굴영상과 음성을 동시에 이용한 감정인식은 최종인식결정방법에 의해 남자 95%, 여자 98.3%로 높은 인식률을 얻음으로 알고리듬의 유용성을 증명하였다.

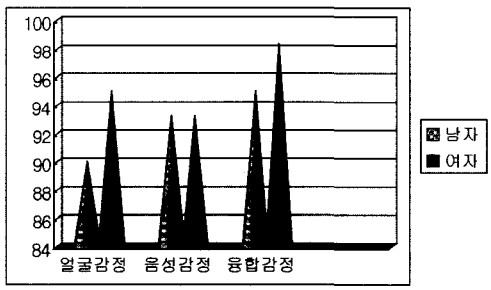

그림 12 최종인식결과 비교

5. 결 론

본 논문은 휴먼인터페이스 방법 중 하나로 사람의 얼굴표정과 음성을 동시에 이용한 감정인식을 연구한 것으로, 기본 감정으로 심리학자 Ekman과 Friesen가 제안한 기쁨, 슬픔, 화남, 놀람, 혐오, 공포를 사용하였다. 이와 같은 감정인식 연구로 Fumio Hara는 역전파 학습 알고리즘에 대한 feed forward 인공 신경망을 이용하여 여섯 가지 얼굴 표정을 인식할 수 있는 얼굴 로봇을 구현한 바 있으나, 특징점 사이의 거리를 이용한 것으로 특징점의 신뢰성 있는 추출문제와 표정 이외의 다른 다양한 얼굴 모양에 대해 신뢰성 있는 시스템을 구축하기에 어려운 단점을 가지고 있다[19]. 또한, Takagi와 Ushida는 개념적 퍼지 집합이라는 개념을 도입하여 거리를 기반으로 한 특징치를 표정인식에 사용하는 방법

으로 78.7%의 인식률을 얻었으며, Mastuno 등은 에지 영상에서의 potential field 개념과 Karluen-Loeve 변환을 사용하여 분노, 행복, 슬픔, 놀람 4가지 얼굴 표정에 대한 약 92%의 인식률을 획득하였다[20,21]. 이에, 본 연구에서는 얼굴영상을 이용한 감정인식을 하기 위해 이산 웨이블렛 변환을 이용하여 영상의 크기를 줄이고, 저주파의 대역을 사용함으로서 중요한 정보만을 이용하였으며, PCA를 이용, 차원을 축소하고 LDA과정을 거쳐 특징 점을 추출하여 코드북을 만들고, 인식부에서 기존에 만들어 놓은 코드 북과 입력으로 들어온 영상에 대해 유클리디언 거리를 이용하여 매칭도를 확인할 수 있었다. 또한, 음성을 이용한 감정인식을 하기 위해 웨이블렛 필터 뱅크를 사용, 저주파의 저주파에서부터 고주파의 고주파로 대역을 분리한 후 각 대역에서 얻어진 값을 멜켑스트럼을 이용 특징점을 추출하고, 앞의 얼굴에서와 같이 코드북을 생성, 인식부에서 기존에 만들어 놓은 코드북과 입력으로 들어온 음성에 대해 유사도를 측정하여 매칭도를 확인할 수 있었다.

향후 과제로는 음성을 이용한 감정에 있어서, 여러 가지 언어적 표현에 대한 다양한 실험을 통하여 알고리듬의 일반성을 더욱 높이는 연구가 필요할 것이라 생각된다.

참 고 문 헌

[1] J. Lien, T. Kanade, C. Li, "Detection, tracking, and classification of action units in facial expression," Journal of Robotics and Autonomous Systems, Vol. 31, No. 3, pp. 131-146, 2000.

[2] M. Turk, A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.

[3] 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.

[4] 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.

[5] C. Padgett, G. Cottrell, "Representing face images for emotion classification," Advances in Neural Information Processing Systems, Vol. 9, MIT Press. 1997.

[6] 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.

[7] V.Kostov and S.Fukuda, "Emotion in User Interface, Voice Interaction System," IEEE Intl Conf. on Systems, Man, Cybernetics Representation, no. 2, pp. 798-803, 2000.

[8] T. Moriyama and S. Oazwa, "Emotion Recognition and Synthisis System on Speech," IEEE Intl. Conference on Multimedea Computing and Systems, pages 840-844, 1999.

[9] L.C. Silva and P.C. Ng, Bimodal Emotion Recognition, Proceeding of the 4th International Conference on Automatic Face and Gesture Recognition, pp. 332-335, 2000.

[10] 김이곤, 배영철, "퍼지 로직을 이용한 감정인식 모델설계", 한국퍼지 및 지능시스템 춘계학술대회, 2000.

[11] 심귀보, 박창현, "음성으로부터 감성인식 요소 분석" 퍼지 및 지능시스템학회 논문지, 2001.

[12] P.Ekman and W.V. Friesen. "Emotion in the human face System," Cambridge University Press, San Francisco, CA, second edition, 1982.

[13] 강현배, 김대경, 서진근, "웨이블릿 이론과 응용", 대우학술총서, 2001.

[14] Hyung-Ji Lee, Wan-Su Lee, Jae-Ho Chung, "Face recognition using Fisherface algorithm and elastic graph matching," Image Processing, Volume: 1, 7-10 Oct. 2001.

[15] Fasel B, Luettin J, "Recognition of asymmetric facial action unit activities and intensities," Pattern Recognition, Proceedings. 15th International Conference on, Volume: 1, 3-7 Sept. 2000.

[16] Stephane Mallat, "A wavelet tour of signal processing," Academic press, 1999.

[17] 이대종, 곽근창, 유정웅, 전명근, "웨이블렛 필터뱅크에 기반을 둔 강인한 화자식별 기법", 정보처리학회논문지 C 제9-C권 제4호, 2002.

[18] Roger Jang, Chuen-Tsai Sun, "Neuro-fuzzy and Soft computing," Prentice-Hall International, 1997.

[19] H. Kobayashi and F. Hara. "Study on face robot for active human interface-mechanismsn of face robot an expression of 6 basic facial expression," In IEEE Int'1 Workshops on Robot and Human communication, pp.276-281, 1993.

[20] H. Ushida, T. Takagi, and T. Yamaguchi. "Recognition of facial expressions using conceptual fuzzy set," In Proc. CVPR, pp594-599, 1993.

[21] Katsuhiro Matsuno and saburo Tsuji, "Recognizing human facial expressions in a potential field," In Proc CVPR, pp 44-49, 1994.

고 현 주

1999년 한밭대학교 제어계측공학과(학사) 2002년 충북대학교 제어계측공학과(공학석사). 2002년~현재 충북대학교 제어계측공학과 박사과정. 관심분야는 Biometrics, Computer vision, 감정인식

이 대 종

1995년 충북대학교 전기공학과(학사). 1997년 충북대학교 전기공학과(공학석사) 2002년 충북대학교 전기공학과 (공학박사). 2003년~현재 충북대학교 컴퓨터정보통신연구소. 관심분야는 음성신호처리, 서명인식, 다중생체인식

전 명 근

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 전기전자컴퓨터공학부 교수. 관심분야는 Biometrics, 감정인식, 음성신호처리, 얼굴인식