한국지능시스템학회 논문지 2008, Vol. 18, No. 1, pp. 20-26

다중 센서 융합 알고리즘을 이용한 사용자의 감정 인식 및 표현 시스템

Emotion Recognition and Expression System of User using Multi-Modal Sensor Fusion Algorithm

염홍기 · 주종태 · 심귀보*

Hong-Gi Yeom, Jong-Tae Joo, and Kwee-Bo Sim

중앙대학교 전자전기공학부

요 약

지능형 로봇이나 컴퓨터가 일상생활 속에서 차지하는 비중이 점점 높아짐에 따라 인간과의 상호교류도 점점 중요시되고 있다. 이렇게 지능형 로봇(컴퓨터) - 인간의 상호 교류하는데 있어서 감정 인식 및 표현은 필수라 할 수 있겠다. 본 논문에서는 음성 신호와 얼굴 영상에서 감정적인 특징들을 추출한 후 이것을 Bayesian Learning과 Principal Component Analysis에 적용하여 5가지 감정(평활, 기쁨, 슬픔, 화남, 놀람)으로 패턴을 분류하였다. 그리고 각각 매개체의 단점을 보완하고 인식률을 높이기 위해서 결정 융합 방법과 특징 융합 방법을 적용하여 감정 인식 실험을 하였다. 결정 융합 방법은 각각 인식 시스템을 통해 얻어진 인식 결과 값을 퍼지 소속 함수에 적용하여 감정 인식 실험을 하였으며, 특징 융합 방법은 SFS(Sequential Forward Selection) 특징 선택 방법을 통해 우수한 특징들을 선택한 후 MLP(Multi Layer Perceptron) 기반 신경망(Neural Networks)에 적용하여 감정 인식 실험을 실행하였다. 그리고 인식된 결과 값을 2D 얼굴 형태에 적용하여 감정을 표현하였다.

Abstract

As they have more and more intelligence robots or computers these days. so the interaction between intelligence robot(computer) - human is getting more and more important also the emotion recognition and expression are indispensable for interaction between intelligence robot(computer) - human. In this paper, firstly we extract emotional features at speech signal and facial image. Secondly we apply both BL(Bayesian Learning) and PCA(Principal Component Analysis). lastly we classify five emotions patterns(normal, happy, anger, surprise and sad) also, we experiment with decision fusion and feature fusion to enhance emotion recognition rate. The decision fusion method experiment on emotion recognition that result values of each recognition system apply Fuzzy membership function and the feature fusion method selects superior features through SFS(Sequential Forward Selection) method and superior features are applied to Neural Networks based on MLP(Multi Layer Perceptron) for classifying five emotions patterns. and recognized result apply to 2D facial shape for express emotion.

Key Words : Emotion Recognition and Expression, Speech Signal. Facial Image. Feature Fusion, Decision Fusion

1. 서 론

컴퓨터나 지능형 로봇 기술들이 점점 발전함에 따라 인간과의 상호교류에 대한 연구도 활발히 진행되고 있다. 이러한 연구는 인간에게 보다 편리하고 정확한 서비스를 제공하기 위해서 이루어지고 있으며 그 중에 인간의 감정을 인식하고 표현해주는 기능들은 필수라 하겠다. 그리고 이 기능들을 통해 인간-컴퓨터(지능형 로봇) 사이의 감정적인 교류가 가능해 질 것이라 생각된다. 그리고 이러한 감정 정보를 연구하는 분야는 크게 인식 부분과 표현 부분으로 나누어 질 수 있다. 그래서 본 논문에서도 감정 인식 및 표현 부분에 대한 연구가 이루어 졌으며 인간으로부터 감정을 인식할 수 있는 매개체로는 음성 신호, 얼굴영상, 제스처, 생체신호, 피부온도 등이 존재한다. 그 중에 음성신호와 얼굴영상을 이용한 연구가 가장 활발히 이루어지고 있다. 하지만 대부분의 기존 연구가 특정 한가지의 매개체만을 이용하였고 이것은 각각 매개체의 단점을 보완할 수 없다는 단점을 가지고 있고 실제로 인간들이 감정을 인식할 때는 여러 가지 매개체를 고려하여 감정이 인식된다. 그러므로 컴퓨터나 로봇에서도 여러 매개체들을 고려하여 감정을 인식하는 연구가 이루어져야 할 것이다.

감정 인식에 관한 기존의 연구들로는 다음과 같은 것들이 있다. 먼저 음성 신호를 이용한 감정 인식의 연구로 Lee C.M. et al와 New T.L. et al은 음성 신호로부터 특징을 추

접수일자 : 2008년 1월 15일
완료일자 : 2008년 2월 10일
* 교신 저자
감사의 글 : 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-313-D00493). 연구비지원에 감사드립니다.

출하는 방법으로 13차와 12차 MFCCs(Mel Frequency Cepstral Coefficients)를 사용하였으며 감정별 패턴 분류는 HMM(Hidden Markov Model)을 이용하였고[1][2] J. Nicholson은 8개의 감정(기쁨, 슬픔, 놀람, 화남, 혐오, 분노, 증오, 평상)들의 특징들을 추출했고 그 특징들은 운율적인 특징과 음성적인 특징들로 분류하였다[3]. 다음으로 Mase et al은 얼굴 영상에 지역별로 11개의 windows를 형성한 후 이 windows별로 근육의 움직임 정도를 파악하여 특징을 추출하였다. 그리고 K-nearest neighbor 규칙을 이용하여 감정별 패턴을 분류하였다[4].

이밖에 제스처 및 피부 온도를 이용한 연구사례로는 다음과 같은 연구가 존재한다[5][6]. 하지만 이 연구들은 각각 매개체들의 단점들을 보완할 수 없으므로 최근에는 감정 융합 방법을 이용하여 감정 인식 실험이 많이 이루어지고 있다. 감정 융합 방법으로는 크게 결정 융합 방법과 특징 융합 방법이 있다. 전자는 각각 인식 시스템을 통해 인식된 결과 값을 이용하는 방법이고, 후자는 각각의 매개체에서 특징들을 추출한 후 감정 융합이 이루어지는 방법이다. 현재 이와 관련 연구 사례로는 다음과 같은 것들이 있다. Mingli Song은 특징 융합 방법으로 Hidden Markove Model(HMM)을 이용하여 음성과 얼굴 영상에 대한 감정 인식 실험을 하였으며, De silva는 결정 융합 방법으로 퍼지 룰 베이스를 이용하여 음성과 얼굴영상에 대한 감정 인식 실험을 하였다[7][8]. 그리고 Busso는 두 가지 방법에 대해 실험하고 비교 설명하였다[9]. 그리고 그 결과 특정 한 가지 매개체를 이용하는 경우보다 다양한 매개체를 이용할 때가 감정 인식률이 높음을 알 수가 있었다.

한편 휴먼-로봇 인터페이스를 위해서는 인식된 감정을 표현하는 기능은 필수라 할 수 있겠으며 이와 관련 방법으로는 기구적인 얼굴 형태를 이용하는 방법과 2D 및 3D의 얼굴 형태로 표현하는 방법이 있다. 기구적인 얼굴 형태를 이용하는 방법은 다양한 표정을 구현하기 힘들고, uncanny valley 현상을 유발할 수도 있다. 그리고 2D 및 3D의 얼굴 형태를 이용하는 경우에는 다양한 감정을 쉽게 구현 할 수 있으며 제작비용도 저렴하다는 장점이 있다. 하지만 기구적인 얼굴 형태에 비해 생동감이 떨어지는 단점이 있다. 현재 이와 관련 연구로는 인간의 얼굴 근육의 움직임을 모방하여 26가지 근육의 움직임을 취할 수 있도록 기구적으로 제작하여 6가지 감정으로 표현할 수 있는 연구를 Fumio Hara가 하였으며[10], 3D 얼굴 형태를 이용한 연구로는 KAIST에서 개발한 아미(AMI) 로봇이 대표적인 예이다[11].

본 논문에서는 음성신호 및 얼굴영상에 대한 감정 인식 방법에 대해 2절에서 설명하였으며, 3절에서는 인식 결과 값과 특징들을 각각 결정 융합 방법과 특징 융합 방법에 적용하여 감정별 패턴을 분류하는 방법에 대해서 설명하였다. 그리고 4절에서는 동적 감정 공간에서 2D 얼굴 형태를 이용하여 감정을 표현하는 방법에서 설명하였다. 그리고 5절에서는 실험 결과들을 비교하였으며, 마지막으로 6절에서 결론을 언급한다.

2. 음성 신호 및 얼굴 영상을 이용한 감정인식

2.1 음성신호를 이용한 감정인식

본 논문에서는 마이크를 통해 입력된 음성신호에 대해서 6가지의 특징들을 (피치의 최대치 및 평균치, 소리의 크기, 섹션개수, Increasing Rate(IR), Crossing Rate(CR)) 추출한 후 이것들을 Bayesian Learning(BL)에 적용하여 감정별 패턴을 분류하였다. Bayesian Learning은 사전확률을 이용하여 어떤 가설의 확률을 계산하는 방법이다. 그래서 본 논문에서는 400개의 음성 샘플들을 이용하여 각 감정과 특징들 간의 확률분포를 조사하여 사전확률을 계산하였다. 그리고 사용자의 확률 분포와의 유사정도를 파악하여 5가지 감정(평활, 기쁨, 슬픔, 놀람, 화남)으로 패턴을 분류하였다[12].

2.2 얼굴영상을 이용한 감정인식

본 논문에서는 얼굴 영상을 이용하여 감정을 인식하기 위해서 피부톤 축적 알고리즘과 GRAY 형태 변환 방법을 이용하여 입, 눈과 눈썹들의 특징들을 추출하였다. 그리고 추출된 특징들은 다차원 특징 벡터로 구성되어져 있어서 패턴을 분류하기에 용이하지 않다. 그래서 정보를 유지하면서 저차원으로 특징 벡터를 축소시키는 방법이 필요한데 본 논문에서는 이 방법으로 Principal Component Analysis (PCA)을 사용하였다.

PCA 알고리즘을 통해 고유 데이터 벡터를 구한 후 유클리안 거리를 통해 학습 데이터와 입력 데이터간의 거리를 비교하여 그 거리가 최소가 되는 표정이 입력과 가장 유사한 표정이므로 그 학습데이터의 감정을 결과로 결정하게 된다[13].

3. 센서 융합 방법을 이용한 감정 인식

3.1 결정 융합 방법을 이용한 감정인식

결정 융합 방법은 각각의 감정 인식 시스템을 통해 얻어진 결과들을 이용하는 방법으로써 구현이 쉽다는 장점이 있으나 각각의 매개체의 단점을 보완하기에는 부족한 면이 있다. 다음 그림 1은 이러한 결정 융합 방법의 일반적인 실행 절차를 보여주고 있다.

본 논문에서는 결정 융합 방법을 하기 위해서 S-모양의 퍼지 소속 함수를 이용하였다. 다양한 소속 함수 중 S-모양의 소속 함수를 사용하는 이유는 가중치를 통해 인식률을 조절할 수 있기 때문이다. S-모양의 퍼지 소속 함수식은 식(1)과 (2)와 같으며 음성 신호와 얼굴 표정에 대한 5가지 감정별로 가중치를 구하게 된다.

$$w_s = \frac{1}{1+\exp[-a(x_s - c_s)]} \tag{1}$$

$$w_i = \frac{1}{1+\exp[-a(x_i - c_i)]} \tag{2}$$

여기서 w_s, w_i는 각각 음성 신호와 얼굴 영상에 대한 가중치이며, x_s, x_i는 각각 매개체의 입력 데이터들을 통해 감정을 인식한 결과 값이다. 그리고 c_i, c_s는 각각 매개체의 학습 데이터들을 감정을 인식한 결과 값들을 감정별로 Code book으로 형성한 후 평균을 구한 결과이며 이 값은 실험을 반복함에 따라 입력데이터가 학습 데이터로 등록됨으로써 변하게 된다. 마지막으로 a는 소속 함수의 기울기 정도를 나타내는데 이 값은 0.01~0.1사이로 값을 변화시켜 가중치의 결과 값이 가장 좋은 것을 선택하였는데, 실험 결과 0.05가 가장 우수한 결과를 보였다.

이와 같은 방법으로 가중치를 구한 후 식 (3)과 같이 각각의 매개체를 통해 얻어진 결과 값에 곱을 취하여 각각의 감정 상태에 대한 출력이 나오게 된다.

$$
\begin{aligned}
O_{normal} &= w_{i(normal)} I_{normal} + w_{s(normal)} S_{normal} \\
O_{happy} &= w_{i(happy)} I_{happy} + w_{s(happy)} S_{happy} \\
O_{surprise} &= w_{i(surprise)} I_{surprise} + w_{s(surprise)} S_{surprise} \\
O_{sad} &= w_{i(sad)} I_{sad} + w_{s(sad)} S_{sad} \\
O_{anger} &= w_{i(anger)} I_{anger} + w_{s(anger)} S_{anger}
\end{aligned} \tag{3}
$$

식 (3)에서 I는 얼굴 영상에서의 감정 출력이고, S는 음성 신호에서의 감정 출력이다. 그리고 식 (4)와 같이 인식된 감정들 중 최대값을 선택하여 감정 인식 결과로 나타낸다.

$$
\begin{aligned}
&System\ Output = \\
&\quad Max\{O_{normal}, O_{happy}, O_{surprise}, O_{sad}, O_{anger}\}
\end{aligned} \tag{4}
$$

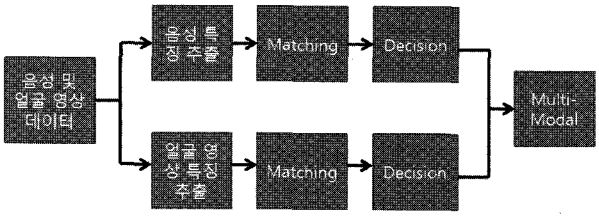


그림 1. 결정 융합 방법
Fig. 1. The Decision Fusion Method

3.2 특징 융합 방법을 이용한 감정인식

특징 융합 방법은 결정 융합 방법보다는 구현이 어렵지만 각각 매개체의 단점을 보완할 수 있다는 장점을 가지고 있다. 다음 그림 2는 이러한 특징 융합 방법의 일반적인 실행 절차를 보여주고 있다.

본 논문에서는 이러한 특징 융합 방법을 실험하기 위해서 음성 신호와 얼굴 영상에서 특징들을 추출하였는데 그 결과 각각 6가지와 5가지의 특징 벡터를 추출할 수 있었다. 하지만 11가지의 특징 벡터를 모두 고려하면 차원의 저주에 빠질 위험성이 크고 인식 속도가 느려지는 단점이 생길 수 있으므로 특징 선택 방법을 통해 우수한 특징들을 선택하게 된다. 이와 같은 특징 선택 방법으로는 여러 가지가 존재하지만, 본 논문에서는 Sequential Forward Selection(SFS) 방법을 이용하였다. SFS는 비어있는 집합에 순차적으로 특징들을 추가한 후 목적함수에 대입하여 그 결과가 가장 우수한 것들을 특징들로 선택하는 방법이다. 본 논문에서 사용된 목적함수는 식 (5)와 같다.

$$
\begin{aligned}
y \;=\; & 3x_0 + x_1 + 4x_2 + 10x_3 - 5x_4 + 8x_5 + 7x_6 \\
& + 8x_7 - 10x_8 + 6.8x_9 + 7.3x_{10} - 5.2x_{11}
\end{aligned} \tag{5}
$$

이 식에서 y는 목적함수 결과 값이고 x_n은 특징들의 종류를 나타낸다. 그리고 목적 함수 파라미터들을 다음과 같이 표현한 이유는 학습 데이터로부터 각각의 특징들을 추출하여 그 크기가 5번째 안에 있는 것들은 reward(+0.1)을 주고 그 이후에 있는 것들은 penalty(-0.1)를 주었다. 이와 같은 실험을 100번 반복한 결과 값이다. 이와 같이 특징들이 결정되면 이 값들을 인공 신경망 중 Back-Propagation(BP)로 학습하는 Multi Layer Perceptron(MLP)에 입력으로 설정하여 감정별 패턴을 분류하였다.

본 논문에서 사용한 BP 알고리즘은 출력층의 오차 신호를 역전파하여 은닉층과 출력층간의 연결 강도와 입력층과 은닉층간으로 연결 강도를 변경하는 학습 방법으로 다양한 분야에 그 응용 범위가 넓다.

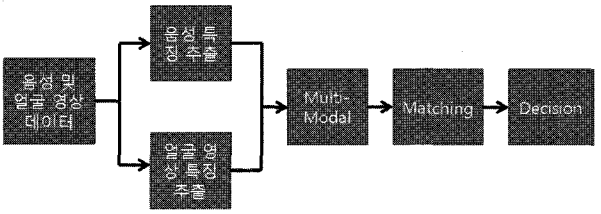


그림 2. 특징 융합 방법
Fig. 2. The Feature Fusion Method

다음 표1은 본 논문에서 사용된 초기 파라미터 설정 값들을 나타내고 있다. 표 1과 같은 초기 파라미터와 감정별 목표치 001~101, 초기 가중치는 -0.03 ~ 0.03로 랜덤하게 설정한 후 학습데이터들을 이용해 오차 범위보다 작아질 때까지 학습을 시킨다. 그리고 입력 데이터를 입력한 후 5가지 감정으로 패턴을 분류한다.

표 1. 신경망의 초기 파라미터 설정.
Table 1. Parameter setting of neural network

Parameter	Value
Hidden Units	13
Output Units	3
Learning Rate	0.005
Tolerance	0.1
Sigmoid Function	$1/1+e^{-3x}$

4. 감정 표현 알고리즘

본 논문에서 저자들이 개발한 감정 표현 알고리즘을 사용하였는데, 이 알고리즘은 먼저 감정 가중치를 입력 받게 되고 이것을 이용하여 감정별 공간을 구성한 후 영역을 계산하였으며, 이 영역의 크기에 따라 감정 표현 파라미터를 조절함으로써 감정을 표현할 수 있는 알고리즘이다. 이와 관련된 자세한 내용은 참고문헌 [14]에 자세히 설명되어 있다.

본 논문에서는 입력되는 가중치를 앞선 감정 인식기에서 획득된 감정 인식 결과 값(0~100)을 사용하였으며, 실제 컴퓨터 프로그래밍 과정에서는 감정 표현 알고리즘을 Avatar라는 클래스로 형성한 후 다음 표 2와 같이 InputArea 함수에 감정별로 인식된 결과 값을 입력하면 된다.

표 2. 감정 표현을 위한 컴퓨터 프로그래밍
Table 2. The computer programming for emotion expression

```
- CAvatar  avt;
- avt. InputArea(Happy, Anger, Sad, Surprise);
- avt. Display(pDC);
```

다음 그림 3은 음성 신호와 얼굴 영상을 이용한 감정 인식 및 감정 융합 실험을 모두 할 수 있고, 2D 기반 동적 감정 공간에 감정을 표현할 수 있는 기능을 가진 시스템을 나타낸다. 그리고 본 논문의 모든 실험은 이 시스템을 기반으로 이

루어졌다.

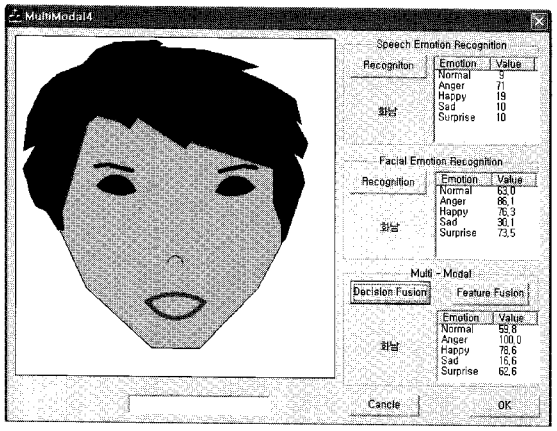


그림 3. 감정 인식 및 표현 시스템

Fig. 3. The Emotion Recognition and Expression System

4. 실험 결과

4.1 Database 구축

본 논문에서는 음성 신호와 얼굴 영상을 이용하여 감정 인식을 하기 위해 동일한 환경에서 감정별 Database를 구축한 후 실험을 하였다. 음성 신호의 경우 10명의 남성 대학원생들(나이:25~31)에게 5가지 감정별로 40개의 음성 샘플을 얻었으며, 얼굴 영상의 경우도 10명의 남정 대학원생들(나이:25~31)에게 5가지 감정별로 6개의 얼굴 영상을 촬영하여 샘플을 얻었다.

소리 및 영상의 크기가 피험자와 마이크 혹은 카메라의 거리에 따라 달라질 수 있으므로 그 거리를 30㎝로 고정하였으며 녹음된 형태는 11㎑, 16bit, mono이고 촬영된 형태는 320×240 pixels, 8bit, bmp이였다.

녹음된 문장들은 30개의 일상적이고 단순한 것들이었고 문장의 길이는 6~10음절로 제한하였으며 촬영될 공간도 배경이 전혀 없는 곳에서 안경을 착용한 사람은 벗고 실험이 이루어졌다.

4.2 음성신호와 얼굴영상을 이용한 감정 인식 실험

본 논문에서는 데이터베이스 구축된 대상자 외 10명의 피험자를 통해 동일한 환경에서 음성 신호와 얼굴 영상을 획득한다.

음성 신호의 경우 단어의 의미상에 감정적인 특징을 잘 나타내는 15가지 단어를 정하여 실험이 이루어졌으며 그 단어들은 표 3과 같다.

표 3. 실험에서 사용된 감정별 단어들

Table 3. The emotional words is used in an experiment

Words	Emotion
그래, 알았어, 안녕	평활
하하하, 아싸, 안녕	기쁨
건디기 힘들어, 죽겠어, 안영	슬픔
그러지 말랬지, 너 뭐야, 죽을래	화남
비명소리 (악~, 아~ 등)	놀람

마이크를 통해 입력된 음성 신호에 대해서 앞서 2절에서 설명한 방법에 의해 의미적인 특징을 배재한 운율적인 특징만을 고려한 특징(피치의 최대치 및 평균치, 소리의 크기, 섹션 개수, IR, CR)들을 추출한다.

앞서 설명한 봐와 같이 본 논문에서는 음성 신호에서 감정별 패턴을 분류하는 알고리즘으로 Bayesian Learning을 이용하였다. Bayesian Learning은 확률을 기반으로 하는 알고리즘으로써 먼저 구축된 Database를 통해 감정별 사전확률을 구하게 된다. 그리고 추출된 특징들을 입력 값으로 설정하고 알고리즘을 통해 사후확률을 구한 후 두 확률의 유사정도를 통해 5가지 감정별(평활, 기쁨, 슬픔, 놀람, 화남) 패턴을 분류하였다.

표 4는 피험자마다 50번의 실험을 하여 감정별 평균을 구한 결과이며 피험자에 따라 감정별 인식률의 결과가 약간씩 차이가 있는데 그 이유는 사람마다 감정별 표현 방식이 차이가 나기 때문이다.

표 4. 음성 신호의 감정 인식률 (%)

Table 4. The emotion recognition rate of speech signal

	평활	화남	슬픔	기쁨	놀람	평균
S1	65	75	63	48	42	59
S2	78	70	73	62	58	68.2
S3	80	63	62	68	68	68.2
S4	78	74	72	70	60	70.8
S5	85	90	47	68	54	68.8
S6	74	80	57	62	47	64
S7	55	92	63	59	38	61.4
S8	78	84	71	55	58	69.2
S9	85	84	73	65	65	74.4
S10	70	69	58	88	49	66.8
평균	74.8	78.1	63.9	64.5	53.9	67.08

얼굴영상의 경우에는 10명의 피험자로부터 카메라를 통해 획득하며 실험 조건은 Database 구축 때와 동일하게 하였다. 다음 그림 4는 본 논문에서 사용된 대표적인 얼굴 영상을 나타내고 있으며 피험자마다 동일한 감정에 대해 10번의 촬영을 하였으며 이를 바탕으로 50번의 감정 인식 실험이 이루어졌다. 이렇게 촬영된 얼굴영상에 대하여 피부톤 축적 알고리즘을 통해 눈, 눈썹, 입에 대한 특징들을 추출하게 되는데, 피부톤 축적 알고리즘은 피부색 영역만을 검출 한 후 2차원 평면상에 피부색 픽셀만을 축적하는 방법을 말한다. 이렇게 추출된 특징들을 영상처리에 용이한 GRAY 형태로 변환한 후 기본적인 히스토그램 평활화, Sobel 연산자 등을 통해 영상처리를 하게 된다.

이렇게 추출된 특징들은 다차원 특징 벡터로 구성되어 있는데 이것을 고유의 정보를 유지하면서 저차원 벡터로 축소시키기 위해서 본 논문에서는 PCA 알고리즘을 사용하였다. 그리고 유클리안 거리를 통해 5가지 감정별 패턴을 분류하였으며 그 결과는 표 5와 같다.

실험 결과를 보면 피험자마다 감정별 인식률의 차이가 별로 나지 않음을 알 수 있는데 이는 음성 신호 비해 사람마다 표현하는 방식의 차이가 크지 않기 때문이고 감정별 특징들을 비교적 정확하게 구분되어 있기 때문이다. 그리고 슬픔과 평활의 감정 인식률이 낮은 이유는 두 가지의 감정적인 표현의 차이점을 찾기 힘들기 때문이다.

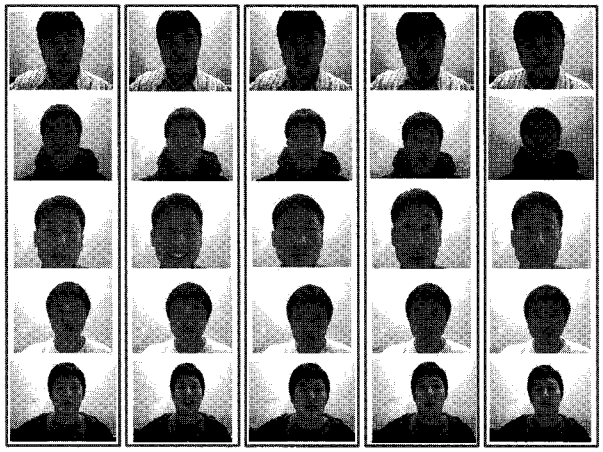


그림 4. 실험에 사용된 감정별 대표 얼굴 사진
Fig. 4. The representative facial image of various emotion used for experiments

표 5. 얼굴 영상의 감정 인식률 (%)
Table 5. The emotion recognition rate of facial image

	평활	화남	슬픔	기쁨	놀람	평균
S1	57	72	43	65	56	58.6
S2	53	68	46	70	60	59.4
S3	55	70	48	66	62	60.2
S4	50	66	50	70	64	60.0
S5	52	74	42	75	65	61.6
S6	48	68	43	65	54	55.6
S7	46	66	42	65	56	55
S8	48	68	44	63	60	56.6
S9	52	69	44	66	56	57.4
S10	50	68	48	68	58	58.4
평균	51.1	68.9	45.0	67.3	59.1	58.28

4.3 센서 융합 방법을 이용한 감정 인식 실험

앞선 3.1절에서 설명한 퍼지 소속 함수를 이용한 결정 융합 방법으로 감정 융합 실험이 이루어졌다. 그리고 그 실험 결과는 표 6와 같으며 실험결과 음성 신호와 얼굴 영상에서 인식률이 우수한 감정들(평활, 화남)은 인식률이 더 높게 나타났으며, 그렇지 못한 감정들(놀람, 슬픔)은 인식률이 여전히 낮게 나타났다. 이것은 결국 각각 매개체의 단점을 보완하지 못한다는 결론을 내릴 수 있다. 그리고 평균 인식률을 통해 음성신호와 얼굴 영상을 이용하는 경우보다 우수함을 알 수 있었다.

특징 융합 방법도 앞선 3.2절에서 설명한 SFS 특징 선택 방법과 Neural-Networks를 이용하여 실험을 하였으며 실험 절차는 다음과 같이 하였다. 먼저 음성 신호와 얼굴 영상의 감정 인식기에서 특징(11가지)들을 추출한 후 이를 SFS 특징 선택 방법에 적용하여 우수한 특징들을 선별한 후 Neural-Network의 입력 값으로 설정한 후 5가지 감정으로 패턴을 분류하였다.

본 논문에서는 앞선 실험과 동일한 환경에서 실험을 하였으며 실험 결과는 다음 표 7과 같다.

표 6. 결정 융합 방법의 감정 인식률 (%)
Table 6. The emotion recognition rate of decision fusion method

	평활	화남	슬픔	기쁨	놀람	평균
S1	72	86	58	64	56	67.2
S2	75	82	59	62	62	68
S3	80	80	62	66	65	70.6
S4	80	90	70	74	68	76.4
S5	76	85	64	65	60	70
S6	78	84	66	64	56	69.6
S7	73	92	60	59	56	68
S8	78	84	62	72	66	72.4
S9	74	84	63	70	64	71
S10	70	86	58	68	58	68
평균	75.6	85.3	62.2	66.4	61.1	70.12

실험 결과인 표 7과 표 4, 5를 비교해 보면 낮은 인식률을 보이던 감정들의(놀람, 슬픔) 인식률을 높였지만 전체적인 인식률 면에서는 앞선 결정 융합과 큰 차이가 없음을 알 수가 있다. 그리고 학습시 걸리는 시간이 오래 걸린다는 단점이 있으며 본 논문에서는 60만 번의 반복을 통해 오차율 허용범위에 도달할 수 있었다.

표 7. 특징 융합 방법의 감정 인식률 (%)
Table 7. The emotion recognition rate of feature fusion method.

	평활	화남	슬픔	기쁨	놀람	평균
S1	63	65	72	68	74	68.4
S2	75	72	68	72	68	71
S3	74	72	68	76	70	72
S4	72	65	72	68	70	69.4
S5	68	75	66	69	67	69
S6	78	74	70	64	72	71.6
S7	68	72	70	69	66	69
S8	72	70	65	72	66	69
S9	74	74	65	70	64	69.4
S10	73	85	58	65	78	71.8
평균	71.7	72.4	67.4	69.3	69.5	70.06

이와 같이 실험한 결과 감정인식 방법에 따른 감정별 평균 인식률은 그림 5와 같다. 그림 5에서 음성 신호의 평균 인식률은 약 66%이였으며, 얼굴 영상의 평균 인식률은 약 58%이였다. 그리고 결정 융합 방법과 특징 융합 방법의 평균 인식률은 약 70%이였다.

4.4 감정 표현 실험

평상시 표정에서 4가지 감정(기쁨, 화남, 슬픔, 놀람)으로 표현하기 위해서 다음 표 8과 같이 가중치를 조절하여 감정별로 매우, 보통, 조금으로 분리하여 표현하였으며 그 실험 결과는 그림 6과 같다.

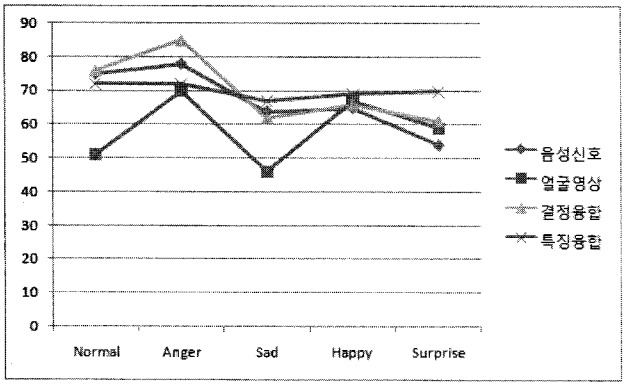


그림 5. 감정 인식 결과.
Fig. 5. The emotion recognition rate.

표 8. 감정표현에 사용된 가중치
Table 8. The weight which be used for emotion expression

	기쁨	화남	슬픔	놀람
매우	(0,100, 0, 0)	(100, 0, 0, 0)	(0, 0,100, 0)	(0, 0, 0,100)
보통	(10,100,10,10)	(100,10,10,10)	(10,10,100,10)	(10,10,10,100)
조금	(20,100,20,20)	(100,20,20,20)	(20,20,100,20)	(20,20,20,100)

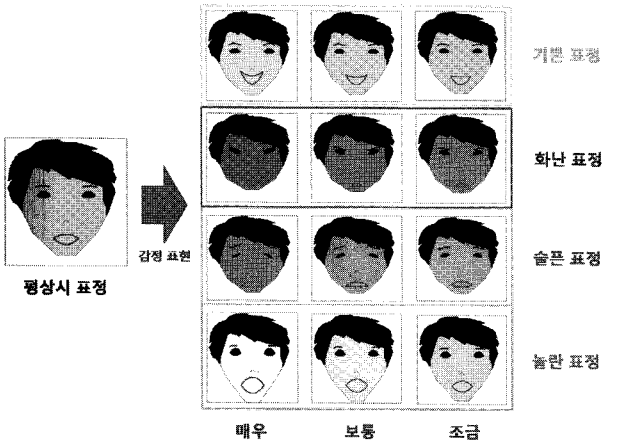


그림 6. 감정 표현에 대한 실험의 결과
Fig. 6. The result of experiment about emotion expression

5. 결 론

본 논문에서는 5가지 감정(평활, 기쁨, 화남, 슬픔, 놀람)에 대해 음성신호와 얼굴영상을 통해 감정 인식 실험을 수행 하였으며 그 인식 결과 음성신호를 이용하여 감정 인식을 한 경우 평균 인식률은 약 67%였으며, 얼굴 영상을 이용하여 감정 인식을 한 경우 평균 인식률은 약 58%였다. 즉 음성 신호를 이용하여 감정 인식을 한 경우가 얼굴 영상을 이용하여 감정인식을 한 경우보다 인식률이 좋음을 알 수 있었다. 그리고 S-모양의 퍼지 소속 함수 및 Neural-Networks를 이용한 결정 융합 및 특징 융합 방법 기반 감정 인식 실험을 하였으며 그 실험 결과 평균 인식률은 약 70%였다.

이 실험 결과 인간과 인간 사이에서의 감정 인식처럼 인간과 로봇간의 감정 인식에서도 다양한 매개체를 고려하는 것이 성능적으로 우수함을 알 수 있었다.

이렇게 인식된 결과 값을 동적 감정 공간에서 2D 얼굴 형태로 감정을 표현하는 시스템에 적용하여 사용자의 감정을 조절해 줄 수 있는 방법을 제안했다. 이와 같은 시스템을 구축함으로써 로봇과 인간과의 감정적인 인터페이스가 가능하게 되며 감정적인 제어가 필요한 상황에서의 사용자의 감정을 인식하여 조절해 줄 수 있는 기능을 할 수 있다.

차후 연구로는 획일화된 환경이 아니라 다양한 환경에서 실시간으로 감정을 인식하는 실험을 할 것이며 좀 더 다양한 결정융합 방법과 특징융합 방법을 실험하고 분석하여 우수한 감정 융합 방법을 제안할 것이다.

참 고 문 헌

[1] Lee C.M., Narayanan S.S. and Pieraccini. R., "Classifying emotions in human - machine spoken dialogs", *ICME'02*, vol. 1, pp. 737-740, 2002.

[2] New T.L., Wei F.S. and De Silva L.C., "Speech based emotion classification", *TENCON 2001*, vol. 1, pp. 297-301, 2001.

[3] J. Nicholson and K. Takahashi, R. Nakatsu, "Emotion recognition in speech using neural networks" Proc. of ICONIP, Vol. 2, 1996.

[4] Mase K., "Recognition of facial expression from optical flow", *IEICE Trans.*, vol. 74, no. 10, pp 3474-3483, 1991.

[5] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gesture", *Journal of Network and Computer Application*, pp. 1-12, 2006.

[6] Yoshitomi Y, S. I. Kim, Kawano T and Kilazoe T, "Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face", *Robot and Human Interactive Communication 2000*, pp. 178-184, 2000.

[7] Mingli Song, Jiajun Bu, Chun Chen and Nan Li, "Audio-visual based emotion recognition", *CVPR'04*, vol. 2, pp. 1020-1025, 2004.

[8] D. Silval and P. C. Nag, "Bimodal emotion recognition", *Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, pp. 332-335, 2000.

[9] Carlos Busso and Zhigang Deng et al, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information", *ICMI 2004*, pp. 205-211, 2004.

[10] Fumio Hara, "Artificial Emotion of Face Robot through Learning in Communicative Interactions with Human", Proceeding of the 2004 IEEE International Workshop on Robot and Human Interactive Communication, pp. 7~15, 2004.

[11] 양현승, 서용호, 정일웅, 한태우, 노동현, "서비스 로봇을 위한 감성 인터페이스 기술," 로봇공학회 논문지, 제1권, 제1호, pp. 58~65, 2006.

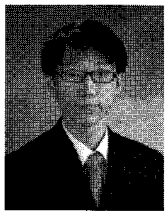
[12] C. H. Park and K. B. Sim, "Pattern Recognition

Methods for Emotion Recognition with speech signal", *International Journal of Fuzzy Logic and Intelligent Systems*", vol. 6, no. 2, pp. 150-154, 2006.

[13] Ho-Duck Kim, Hyun-Chang Yang, Chang-Hyun Park, and Kwee-Bo Sim, "Emotion Recognition Method of Facial Image using PCA ", *Journal of Korea Fuzzy Logic and Intelligent Systems Society(KFIS)*, vol.16, no.6, pp. 772-776, Dec. 2006.

[14] 심귀보, 변광섭, 박창현, "동적 감성 공간에 기반한 감정 표현 시스템," *한국퍼지 및 지능시스템학회 논문지*, 제15권, 제1호, pp18-23, 2005.

저 자 소 개



염홍기(Hong-Gi Yeom)

2008년 : 중앙대학교 전자전기공학부 공학사

2008년~현재: 중앙대학교 대학원 전자전기공학부 석사과정

관심분야 : Wearable robot, Application of Bio-signal 등

주종태(Jong-Tae Joo)

2006년 : 순천대학교 전기제어공학과 공학사

2008년 : 중앙대학교 전자전기공학부 공학석사

관심분야 : Embedded OS, Robotics, Digital 영상 처리, Machine Vision.

심귀보(Kwee-Bo Sim)

1990년 : The University of Tokyo 전자공학과 공학박사

1991년 ~ 현재 : 중앙대학교 전자전기공학부 교수

[제17권 7호 (2007년 12월호) 참조]

E-mail : kbsim@cau.ac.kr

Homepage URL : http://alife.cau.ac.kr