

지능형 로봇을 위한 감정 기반 휴먼 인터랙션 기법 개발

Development of Emotion-Based Human Interaction Method for Intelligent Robot

주영훈* · 소제윤*, 심귀보**, 송민국***, 박진배***

Young Hoon Joo, Jea Yun So, Kee Bo Sim, Min Kook Song, and Jin Bae Park

* 군산대학교 전자정보공학부

** 중앙대학교 전자전기공학부

*** 연세대학교 전기전자공학과

요 약

영상을 통한 감정 인식 기술은 사회의 여러 분야에서 필요성이 대두되고 있음에도 불구하고 인식 과정의 어려움으로 인해 풀리지 않는 문제로 남아 있다. 인간의 움직임에 이용한 감정 인식 기술은 많은 응용이 가능하기 때문에 개발의 필요성이 증대되고 있다. 영상을 통해 감정을 인식하는 시스템은 매우 다양한 기법들이 사용되는 복합적인 시스템이다. 따라서 이를 설계하기 위해서는 영상에서의 움직임 추출, 특징 벡터 추출 및 패턴 인식 등 다양한 기법의 연구가 필요하다. 본 논문에서는 이전에 연구된 움직임 추출 방법들을 바탕으로 한 새로운 감정 인식 시스템을 제안한다. 제안된 시스템은 은닉 마르코프 모델을 통해 동정된 분류기를 이용하여 감정을 인식한다. 제안된 시스템의 성능을 평가하기 위해 평가데이터 베이스가 구축되었으며, 이를 통해 제안된 감정 인식 시스템의 성능을 확인하였다.

키워드 : 감정 인식, 패턴 인식, 은닉 마르코프 모델 (HMM), 퍼지 분류기

Abstract

This paper is to present gesture analysis for human-robot interaction. Understanding human emotions through gesture is one of the necessary skills for the computers to interact intelligently with their human counterparts. Gesture analysis is consisted of several processes such as detecting of hand, extracting feature, and recognizing emotions. For efficient operation we used recognizing a gesture with HMM(Hidden Markov Model). We constructed a large gesture database, with which we verified our method. As a result, our method is successfully included and operated in a mobile system.

Key Words : Emotion recognition, Hidden Markov Model(HMM), Pattern recognition, Fuzzy classifier

1. 서 론

지능형 로봇의 휴먼 인터랙션을 위한 감정 인식은 음성 인식에 위한 방법과 얼굴 영상을 이용한 인식 방법이 많이 연구 되고 있다. 감정 인식 기술은 영상을 기반으로 하는 방법과 음성을 이용하는 방법 그리고 두 가지를 모두 이용하는 방법으로 연구가 진행 중이다. 그러나 음성 인식에 의한 방법은 로봇 내부에서 발생하는 소음, 로봇의 주행 및 동작에서 발생하는 소음 등으로 인해 실제 인간의 음성을 인식하고 해석하는데 어려운 문제가 있다.

영상을 이용한 감정 인식 방법은 일반적으로 감정이 가장 많이 표현되는 얼굴 영상을 이용하는 경우가 대부분이다 [3-6]. 얼굴 영상을 이용하는 방법은 얼굴에서 특징점을 추출 다음 추출된 특징을 바탕으로 감정을 분석하는 방법이 일반적이다. 얼굴을 이용한 감정 분석은 얼굴에 풍부한 감정

표현으로 인해 그 성능이 우수하고 빠르다는 장점을 가진다. 하지만 얼굴을 나타난 감정은 얼굴 미세한 변화로서 나타나기 때문에 이를 감지하기 위해서는 매우 정밀한 얼굴에 대한 분석이 이루어 져야 한다. 따라서 얼굴 영상의 크기가 미세한 감정 표현을 인식 할 수 있을 정도의 크기가 되어야만 감정 분석이 가능한 단점이 있다. 따라서, 얼굴 영상을 이용한 감정 인식 방법 하나 만 사용할 경우 지능형 로봇에 일정한 크기 이상의 얼굴 사진을 제공하기 위해 로봇에 가까이 가야 하는 불편한 행위를 취해야 한다. 따라서 얼굴 정보 이외에 전체 행동을 바탕으로 감정을 인식하는 연구가 필요하다. 이를 위해, 보다 일반적인 다양한 영상을 바탕으로 감정을 분석하는 시스템 개발, 즉 영상 정보에서 얼굴 뿐 만 아니라 인간의 동작을 분석하여 종합적으로 감정을 분석해 내는 시스템이 개발이 필요하다.

인간의 움직임을 이용한 감정 인식은 그 인식 기법에 따라 기하학적 인식 기법, 고유공간을 이용한 인식 기법, 기타 변환을 이용한 인식 기법으로 나누어진다 [2], [5], [6]. 이중 보다 일반적인 인간의 움직임에 대한 감정 인식을 수행하기 위해 인간의 움직임을 분석하여 인식을 수행하는 기법이 연구되고 있다. 하지만 인간의 움직임을 이용한 감정 인식은 이를 위해 다양한 선행 기술이 수정 및 개발되어야 하는 어

접수일자 : 2006년 9월 5일

완료일자 : 2006년 10월 10일

감사의 글: 본 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2005-042-D00268)

려움이 있다. 이를 해결하기 위해 인간의 움직임 영상 분석에 대한 선행 기술들이 연구되었다 [1], [2].

본 논문에서는 기 개발된 선행 기술을 바탕으로 감정 인식 시스템을 위한 인간의 움직임 추출 기법과 은닉 마르코프 모델(HMM)을 통해 동정된 분류기를 통한 감정 인식 기법으로 제안한다. 인간의 움직임 추출 기법은 선행 연구된 기법들만 얻어진 인간의 움직임 구성 요소에 정보를 바탕으로 감정 분류의 성능을 높이기 위한 손의 움직임을 파악하는 방법이다. 또한, 이렇게 생성된 손의 움직임 정보를 은닉 마르코프 모델을 통해 동정된 분류기 기반 감정 인식 기법이 개발되었다. 본 논문에서는 지능형 로봇으로부터 얻어진 동영상을 바탕으로 얼굴을 이용한 감정 인식과 더불어 특정 동작을 이용한 통합적인 감정 인식 시스템을 개발하는 것을 목표로 한다.

본 논문의 구성은 다음과 같다. 2장에서는 영상으로부터 얼굴을 이용한 감정 인식 기법들에 대해서 살펴본다. 3장에서는 인간의 움직임 벡터 추출 방법과 추출한 움직이을 이용한 감정 인식 시스템에 대해 설명하고, 4장에서는 통합적인 최종 개발된 감정 인식 시스템의 성능이 평가된다. 마지막으로, 5장에서 논문을 결론 맺는다.

2. 얼굴 영상 분석

2.1 가상 얼굴 모형

정면 얼굴 영상을 이용한 감정 인식을 위해 얼굴 영상에 대한 세밀한 분석이 필요하다. 사람의 피부색을 이용한 얼굴 영역 검출은 얼굴색의 다양함과 주변 환경의 조도 변화에 따른 피부색 변화로 인해 매우 힘든 일이다. 기존의 피부색을 이용한 얼굴 영역 추출 방법은 대부분 고정된 피부색 범위를 찾고 이를 바탕으로 피부색을 분리하는 방법을 사용하였다. 하지만 이러한 방법은 주변 환경 및 인물의 변화에 강인하게 대처하지 못하는 단점을 가진다. 이러한 문제를 해결하기 위해 퍼지 모델 기반 색상 필터가 제안되었다[2]. 퍼지 색상 필터는 다음과 같은 규칙으로 구성된다.

$$R_i: \text{IF } x_1 \text{ is } M_{i1} \text{ and } x_2 \text{ is } M_{i2} \text{ and } x_3 \text{ is } M_{i3} \quad (1)$$

$$\text{THEN } y_i(x) = d_i$$

여기서, $[x_1, x_2, x_3]$ 은 얼굴 색상 정보이며 각각 색상의 색상, 채도, 명도 값을 갖는다. M_{i1}, \dots, M_{i3} 는 조건부 퍼지 집합이며, $y_i(x)$ 는 i 번째 규칙의 결론부 출력이며, d_i 는 결론부 변수를 나타내며, $x = [x_1, x_2, x_3]^T$ 는 입력 벡터를 나타낸다. 소속 함수는 다음과 같은 가우시안 소속 함수를 사용한다.

$$\mu_{M_j} = e^{-\frac{(c_j - x_j)^2}{v_j^2}} \quad (2)$$

여기서, c_j^i 와 v_j^i 는 i 번째 규칙의 j 번째 입력의 중심 값과 분산을 나타낸다. 최종 퍼지 규칙의 최종 출력 $Y(x)$ 은 다음과 같다.

$$Y(x) = \frac{\sum_{i=1}^l h_i(x) d_i}{\sum_{i=1}^l h_i(x)} \quad (3)$$

$$h_i(x) = \prod_{j=1}^3 \mu_{M_j}(x_j) \quad (4)$$

여기서, l 은 퍼지 규칙의 수이다. 색상 필터링에 소모되는 시간을 단축시키기 위해 퍼지 규칙 수는 적을수록 좋다. 최종 퍼지 규칙의 최종 출력은 명확한 얼굴 피부색에 대한 소속 정도를 나타내지 못하기 때문에 다음과 같은 후 처리 과정을 사용하여 명확한 피부색에 대한 소속 정도를 얻는다.

$$\hat{Y}(x) = \begin{cases} 1, & |Y(x) - 1| \leq |Y(x)| \\ 0, & |Y(x) - 1| > |Y(x)| \end{cases} \quad (5)$$

여기서, $\hat{Y}(x)$ 는 입력 픽셀이 x 일 때 색상 필터가 적용된 최종 픽셀의 값이다. 본 논문에서는 퍼지 색상 필터가 피부색에서는 1의 값을 가지고 피부색이 아닌 곳에서는 0의 값을 가지도록 동정된다. 최종적으로 유전 알고리즘을 사용하여 조건부 소속 함수의 파라미터 c_j^i 와 v_j^i 그리고 결론부 파라미터 d_i 가 동정된다.

퍼지 색상 필터를 이용해 얼굴의 후보 영역을 찾은 후 정확한 얼굴 영역을 추출하기 위해서는 후처리 과정이 필요하다. 본 논문에서는 기존 연구에서 개발된 히스토그램 분석 기법을 사용하였다 [2].

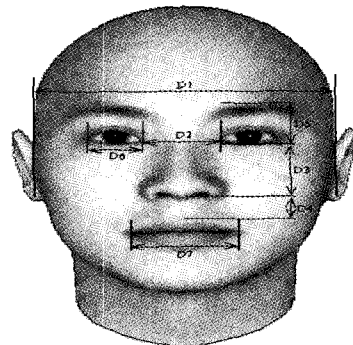


그림 1. 가상 얼굴 모형
Fig. 1 Model of virtual face

가상 얼굴 모형은 얼굴 영역 분석의 정확도를 높이고 수행 시간을 줄이기 위해 제안된 얼굴 구성 요소간의 정보를 기반으로 한 얼굴 모형이다. 그림 1은 가상 얼굴 모형이 가지고 있는 얼굴 구성 요소간의 정보 종류를 나타낸다. 얼굴 감정이 주로 눈썹, 눈 그리고 입의 모양 및 변화에 민감하기 때문에 세 구성 요소의 특징을 중점으로 가상 얼굴 모형은 구성된다. 가상 얼굴 모형은 그림 1에 표시된 거리 사이의 간격을 $D1$ 의 1로 축소한 비율을 가진다[2].

2.1.2 가상 얼굴 모형을 이용한 구성 요소 추출

가상 얼굴 모형을 이용해 얼굴 구성 요소를 추출하기 위해서 히스토그램 분석 방법을 사용한다. 히스토그램 분석을 이용한 얼굴 구성 요소 추출하기 위해서는 가상 얼굴 모형을 이용해 다음과 탐색 공간 설정이 필수적이다. 탐색 공간 R1은 코와 입의 위치를 찾기 위해 사용되는 탐색 공간이며, R2와 R3은 눈과 눈썹의 위치를 찾기 위해 탐색 공간이다. 탐색 공간의 폭과 길이는 추출된 얼굴 가로 길이와 가상 얼굴 모형의 정보를 바탕으로 결정된다. 자세한 탐색 공간에 내용은 기존 연구 논문에 자세히 언급되어 있다 [2].

2.2. 얼굴 영상을 통한 감성 인식 알고리즘

2.2.1. 특징 벡터 추출

얼굴 영상이 분석된 후 분석된 결과 중에서 감성 인식에 사용될 특징 벡터를 추출하는 과정이 필요하다. 감성 인식을 위한 특징 벡터는 최대한 감정이 많이 표현되며 감정별 편차가 큰 특징을 추출하는 것이 바람직하다. 이를 위해 본 논문에서는 감성 변화에 따라서 변화하는 얼굴 구성 요소를 눈 영역, 입술 영역 그리고 보조 영역으로 나누어서 특징 벡터를 생성하는 방법을 제안한다. 그리고 정확한 특징 벡터 추출을 위해 기하학적 정보 및 형태적 정보를 모두 사용하는 방법을 사용한다. 표 1-3은 각 영역별 특징 벡터 추출 내역을 나타낸다.

표 1. 눈 영역 특징 벡터
Table 1 feature vector in eye-region

특징 벡터	설명	크기
X_eyebrow	양 눈썹 사이 간격	1×1
X_eyeeeyebrow	눈썹과 눈 사이의 거리	1×1
X_eyenose_1, X_eyenose_2	코와 눈과의 거리	2×1
Shape_eye	눈 모양 템플릿과 에러	4×1

표 2. 입 영역 특징 벡터
Table 2 Feature vector in mouth-region

특징 벡터	설명	크기
X_mouth_width/X_mouth_height	입 크기 비율	1×1
X_nosemouth	코와 입술의 거리	1×1
Shape_mouth	입의 모양 템플릿과 에러	6×1

표 3. 보조 영역 특징 벡터
Table 3 feature vector in auxiliary region

특징 벡터	설명	크기
Shape_wrinkleseye	양미간 주름 모양	1×1
Shape_wrinklecheek1, Shape_wrinklecheek2	볼 주름 모양	2×1

추출된 특징 벡터는 그 영상의 크기와 개개인의 얼굴 크기가 다른 관계로 정규화가 필요하다. 본 연구에서는 추출된 얼굴 영역 중 가장 확실한 추출이 가능한 눈가로 길이 $D1$ 을 이용하여 모든 특징 벡터 값을 정규화 한다. 템플릿 변형을 통한 비교는 변형에 많은 시간을 필요로 하기 때문에 본 연구에서는 다음과 같은 방법을 통해 유사도를 산출한다.

$$similarity = |e_{width}| + |e_{height}| + |e_{ratio}| + |e_{pixel}|$$

여기서, $|e_{width}|$ 는 템플릿의 가로길이 오차, $|e_{height}|$ 는 세로 길이 오차, $|e_{ratio}|$ 는 가로 세로 비 오차 그리고 $|e_{pixel}|$ 는 내부 픽셀 수 오차는 나타낸다.

2.2.2. 퍼지 분류기 기반 감성 인식 시스템

인간의 감정은 매우 애매모호한 성격을 가지며 이를 수학

적인 방법으로 다루기가 매우 힘들다. 이러한 어려움을 해결하기 위해, 본 논문에서는 퍼지 이론을 이용한 감성 인식 시스템을 설계하기 위해 퍼지 분류기를 제안한다. 퍼지 분류기는 퍼지 모델 기반 분류기로 다음과 같은 규칙으로 구성된다.

$$R_i: \text{If } x_1 \text{ is } M_{i1} \text{ and } \dots \text{ and } x_m \text{ is } M_{im} \text{ Then } y_i(x) = a_{i1}x_1 + \dots + a_{im}x_m + b_i$$

여기서, $x = [x_1 \dots x_m]$ 는 입력 벡터이고 x_j 는 j 번째 특징 입력이다. M_{i1}, \dots, M_{im} 는 조건부 퍼지 집합이며, a_{ij} 와 b_i 는 결론부 파라미터를 나타낸다. 소속 함수는 식 (2)와 같은 가우시안 소속 함수를 사용한다. 퍼지 분류기의 최종 출력은 다음과 같이 계산된다.

$$Y(x) = \frac{\sum_{j=1}^l h_j(x) (\sum_{i=1}^m a_{ji}x_i + b_j)}{\sum_{j=1}^l h_j(x)} \quad (6)$$

데이터의 최종 클래스 값은 최종 출력과의 최소 에러를 갖는 클래스 값으로 선택된다. 최종 출력 (6)는 다음과 같은 행렬식으로 나타낼 수 있다.

$$Y(x) = H^T(Ax + B) \quad (7)$$

여기서,

$$H = \begin{bmatrix} h_1(x) \\ \vdots \\ h_i(x) \\ \vdots \\ h_l(x) \end{bmatrix}, A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{im} \\ \vdots & \ddots & \vdots \\ a_{l1} & \dots & a_{lm} \end{bmatrix}, B = \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_l \end{bmatrix}$$

제안된 감성 인식을 위한 퍼지 분류기는 입력이 19개이며 규칙수가 복수인 매우 복잡한 형태를 가진다. 기존에 많이 사용된 유전 알고리즘의 경우 변수의 자유도가 너무 많이 적절 한 해를 찾기 매우 힘들다. 따라서 본 논문에서는 선형 행렬 부등식을 이용한 최적화 방법을 사용하여 퍼지 분류기를 동정한다. 선형 행렬 부등식을 이용한 동정 방법은 조건부와 결론부를 모두 동정 할 수 있을 뿐만 아니라, 분류기의 입력의 수가 많아도 좋은 성능을 가지는 분류기를 동정 할 수 있다는 장점을 가졌다. 조건부 소속 함수 동정은 다음과 같은 식을 만족시키는 V_i 와 c_i 를 찾는 문제로 정리될 수 있다.

$$(x - c_i)^T V_i^T V_i (x - c_i) = 0, \forall x \quad (8)$$

식 (8)을 만족시키는 V_i 와 c_i 를 찾는 연구가 선형 연구에서 수행되었다 [2]. 본 논문에서는 [2] 에 제안된 정리 1을 이용하여 조건부의 소속함수를 동정하였다.

조건부의 소속함수가 결정되면 이를 이용하여 결론부의 파라미터를 동정 할 수 있다. 데이터 x 가 주어졌을 때, 퍼지 분류기의 목표값을 Y_d 라 하면 최적의 분류기는 다음 등식을 만족 하여야 한다.

$$Y_d = H^T(Ax + B) \quad (9)$$

식 (9)을 만족 시키는 A 와 B 를 찾는 연구 또한 선형 연구에서 수행되었다. 본 논문에서는 [1]의 정리 1을 이용하여

결론부 파라미터를 동정하였다. Y_d 는 감정 인덱스 값으로 각 감정을 나타내는 고유값이다. 일반적으로 정수 범위에서 적절한 값을 선택한다. 최종적으로 최종 출력 $Y(x)$ 과 각 감정의 인덱스 값 Y_d 의 차이가 가장 작은 감정이 인식된 감정으로 선택된다.

3. 인간의 움직임 추출 방법

3.1 혼합 실루엣 추출 방법

뼈대의 특징점을 뽑아내는 것은 인간의 실루엣을 바탕으로 한다. 실루엣을 만드는 세 가지 방법이 연구되어 왔다. 시간에 따른 차이 [3], 배경에서의 변화 [5], 그리고 최적의 흐름이다 [6].

그러나 이미지에서 움직임을 추출하는 최적의 접근법은 복합 실루엣 추출 기법이다 [8]. 복합 실루엣 추출 기법은 시간에 따른 차이에 기반을 둔 방법이다. 시간에 따른 차이 방법의 단점을 해소하기 위해서 다음과 같은 움직임 영역 모델과 누락되는 실루엣 정보 보상법이 제안되었다.

먼저 연속적인 이미지를 (10)과 같이 정의한다.

$$I(x,y;t) = [I_r(x,y;t), I_g(x,y;t), I_b(x,y;t)]^T \quad (10)$$

$$= [p(t)]^T.$$

시간의 변화도는 두 이미지 사이의 시간의 차이이다. 시간과 공간의 변화도는 모서리의 정보를 가지며 다음의 (11)와 같이 정의한다.

$$I_t = \frac{\partial I}{\partial t}$$

$$I_s = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \quad (11)$$

변화도 정보를 모서리의 핵심 값으로 구하고자 한다면, 변화도 정보는 자연수로 정의되어야 한다. 이러한 관점에서 다음의 (12)과 같이 간단히 변화도 정보를 정의한다.

$$\bar{I}_t = \| I_t \|$$

$$\bar{I}_s = \| I_s \| \quad (12)$$

모바일 시스템에서 시간의 변화도 \bar{I}_t 의 단점을 보완하기 위해 여러 방법이 도입 되었다. 카메라 자체의 문제로 인해서 우리는 시간의 변화도의 합 S_t 를 다음의 (13)와 같이 정의한다.

$$S_t = \int_{\forall p \in I} f(\bar{I}_t, \gamma_t) dp$$

$$f(\bar{I}_t, \gamma_t) = \begin{cases} 1, & \text{if } \bar{I}_t > \gamma_t \\ 0, & \text{else} \end{cases} \quad (13)$$

여기서 γ_t 는 색상 이미지를 이진 이미지로 옮기는 최소값이다. S_t 가 0이 될 때, 예전의 시간의 변화도는 현재의 시간의 변화도로 치환된다.

실루엣 정보의 부족함과 정보의 왜곡을 극복하기 위해서 움직임 영역 모델을 이용한다. 지적된 문제점을 해결하기 위해서 공간의 변화도를 이용하고 오래된 시간의 변화도 정보의 연속적인 합을 이용한다. 결국 정확한 실루엣 정보를 얻기 위해서는 공간의 변화도와 시간의 변화도의 적절한 합이

필요하고 이는 다음 (14)와 같이 정의한다.

$$\eta \bar{I}_s(x,y) + (1-\eta) \bar{I}_t(x,y) \quad (14)$$

여기서 η 는 불록합 합인 파라미터이다. 정확한 실루엣 정보를 얻기 위해서 η 값을 잘 정의하여야 하고, 이를 위해 본 논문에서는 움직임 영역 모델을 사용한다.

움직임 영역 모델은 초기 모델은 0에서 출발한다. 그리고 다음과 같은 규칙으로 갱신해 나간다.

$$R(x,y;t+1) = \begin{cases} R(x,y;t) + \gamma_i, & \bar{I}_t(x,y;t+1) > \gamma_t \\ R(x,y;t) + \gamma_d, & \bar{I}_t(x,y;t+1) < \gamma_t \end{cases}$$

$$\gamma_i = \begin{cases} \gamma_i, & \bar{I}_s(x,y;t+1) > \gamma_s \\ \gamma_i, & \bar{I}_s(x,y;t+1) < \gamma_s \end{cases}$$

$$\gamma_d = \begin{cases} \gamma_d, & \bar{I}_s(x,y;t+1) > \gamma_s \\ \gamma_d, & \bar{I}_s(x,y;t+1) < \gamma_s \end{cases}$$

γ_s 는 공간의 이미지의 이진화 파라미터이다. 따라서 복합 실루엣은 다음의 (15)와 같이 계산된다.

$$\bar{I}_b(x,y) = \eta \bar{I}_s(x,y) + (1-\eta) \bar{I}_t(x,y) \quad (15)$$

$$\eta = \begin{cases} \eta_R, & R(x,y) > \gamma_R \\ \eta_B, & R(x,y) < \gamma_R \end{cases}$$

여기서 η_R 는 움직임 영역의 불록합 파라미터이고, η_B 는 배경부분의 불록합 파라미터이다.

3.2 적응 뼈대 모델(Adaptive skeleton model)

움직임을 표현 하기위해 흔히 쓰이는 모델중의 하나가 뼈대 모델이다. 뼈대 모델은 가장 작은 정보만을 가지고 인간의 움직임을 표현한다. 본 논문에서 사용할 뼈대모델은 15개의 특징점을 가진다. 점 F_1 에서부터 F_{15} 사이의 각각의 거리를 S_1, \dots, S_{13} 으로 표현한다.

뼈대모델은 머리의 위치와 크기를 정하는 것에서부터 출발한다. 많은 연구자들에 의해서 머리의 위치를 찾는 다양한 방법들이 연구되어왔다 [9].

얼굴의 크기를 이용하여 특징점 간의 거리를 정규화 시키면 우리는 인간의 몸통 부분에 해당하는 특징점 (F_1, F_2, F_3, F_4, F_9 , and F_{10})의 대략적인 위치를 추정할 수 있다. 이러한 특징점에서 스네이크 알고리즘과 유사하게 외부와 내부에 각각 해당하는 에너지 함수를 정의할 수 있다. 그리고 에너지 함수의 값을 최소화하게 나머지 특징점들의 위치를 구할 수 있다. 내부 에너지 함수는 특징점들간의 기하학적인 관계와 거리로부터 계산되고, 이때 거리는 다음의 (16)과 같이 정의한다.

$$E_i^{int} = |S_i - \|f_j - f_k\|| \quad (16)$$

여기서 f_j 와 f_k 는 S_i 와 연결된 추정된 특징점들이다. 또한 몇 개의 특징점들 사이에서는 기하학적인 관계가 정의 되어야 한다. 이러한 기하학적인 관계는 다음과 같이 표현한다.

$$g(1, f_i, f_j) = \begin{cases} 1, & f_{ix} > f_{jx} \\ 0, & f_{ix} > f_{jx} \end{cases}$$

$$g(0, f_i, f_j) = \begin{cases} 1, & f_{iy} > f_{jy} \\ 0, & f_{iy} > f_{jy} \end{cases}$$

외부 에너지 함수는 실루엣 이미지에서의 픽셀값들을 이

용하여 정의한다. 예를 들어 머리 부분의 특징점에 대한 외부에너지 함수는 다음의 (17)과 같다.

$$E_i^{ext} = - \sum_{x=f_{ix}-k}^{f_{ix}-k+m} \sum_{y=f_{iy}-k}^{f_{iy}-k+m} \bar{l}_b(x,y) \quad (17)$$

여기서 k 는 특징점과 실루엣 사이의 거리이고, $2m$ 은 탐색 영역의 넓이다.

결국 에너지 함수 (18)는 볼록함 파라미터 k 를 이용하여 내부에너지 함수와 외부에너지 함수의 합으로 구성된다.

$$E_i = k\bar{E}_i^{ext} + (1-k)\bar{E}_i^{int} \quad (18)$$

여기서 볼록함 파라미터 k 는 인위적으로 정할 수 있다. k 의 값이 클수록 추정된 특징점들은 실루엣으로 이동하게 되지만 특징점의 모양이 어긋나게 된다.

이러한 에너지함수의 값을 작게 하기 위해서 다음과 같은 순서로 진행된다. 먼저 특징점 근처의 모든 픽셀에서의 에너지함수 값을 계산한다. 그리고 가장 작은 에너지함수 값을 가지는 픽셀을 새로운 특징점으로 정의한다. 만약 모든 픽셀에서 같은 에너지함수를 가지게 되면, 몸의 중심에서 가장 멀리 떨어진 픽셀을 새로운 특징점으로 정의한다.

3.3 색상 정보를 바탕으로 한 손의 움직임 추출 방법

실루엣 정보만으로 손의 특징점을 찾는 것은 충분하지 않다. 손과 팔꿈치 부분의 경우에는 움직임 동작에서 많은 변화가 나타남으로 이들의 정확한 위치를 찾는 것은 어려운 문제이다. 손의 위치를 찾기 위해서 추가적으로 색상 정보를 이용한다. 손의 위치를 찾은 이후에는 기하학적인 관계를 이용하여 팔꿈치의 위치를 찾는다. 피어(Peer)의 색상 모델은 동영상에서 색상을 찾는데 사용된다. 피부 색상을 가지는 이미지를 I_c 라고 정의하자. 그러면 복합 피부 영역은 다음의 (19)와 같이 계산된다.

$$\bar{l}_c = \bar{l}_c I_c \quad (19)$$

손의 위치를 찾기 위해서는 얼굴 근처에 나타나는 피부색상 지역을 제거한다. 얼굴 영역은 적응 뼈대 모델로부터 구해진다. 얼굴 부근의 피부색상 지역을 제거하고 나서 손의 위치는 평균값 이동 탐색 알고리즘에 의해 구한다. 평균값 이동 탐색 알고리즘은 최고값을 가지는 히스토그램 값을 최소화 시키는 것이다.

색상 기반 손 탐색 알고리즘은 색상과 움직임 정보만을 사용하기 때문에 근처에 다른 사람의 손이 오는 경우 이를 구분하지 못하는 단점을 가진다. 따라서 본 논문에서는 이를 극복하기 위해 어깨선 추출을 이용한 보다 정확한 손 위치 파악 알고리즘을 개발하였다. 어깨선 추출 알고리즘은 목 바로 아래 어깨 점을 시작으로 원형의 탐색 공간을 각 원형으로 돌아서면서 탐색을 하는 방법으로 어깨선을 추출한다. 이때 탐색 영역안의 각 각들은 R1 과 R2 영역을 가지는데 이 영역에서 다음과 같은 에너지 함수 (20)에 의해서 에너지 값을 계산하게 된다. 최종적으로 에너지 값이 가장 큰 각으로 탐색을 해 나가는 방법을 취한다.

$$S = \left(\sum_{\forall x \in R_2} f(x) - \sum_{\forall x \in R_1} f(x) \right) / r + \sum_{\forall x \in L} f(x) \quad (20)$$

4. 감성 인식 기법 및 모의 실험

본 논문에서 제안하는 감성 기반 휴먼 인터랙션 기법은 크게 영상에 나타난 사용자의 상태에 따라 두 가지 방법으로 사용자의 감정을 인식한다. 첫 번째, 로봇으로부터 얻어진 영상에 해석 가능한 크기와 해상도를 가진 얼굴 영상이 포함될 경우 얼굴 분석을 통한 감정 인식 기법을 사용한다. 두 번째, 그 이외의 얼굴 영역에 대한 분석이 불가능해 질 때, 로봇과 사람간의 거리가 멀거나 다른 요소로 인해서 영상으로부터 얼굴의 이미지 추출이 용이하지 않을 때는 전체 사람의 움직임 동작을 분석함으로써 상태를 예측한다.

실험에 사용된 영상은 CCD 카메라로부터 얻은 320×240 크기의 24bit의 컬러 영상이 사용되었다. 가상 얼굴 모형으로 얻어진 구성 요소를 바탕으로 제안된 특징 벡터 추출 기법을 사용하여 특징 벡터를 추출하게 된다. 변환된 총 150 개의 얼굴 영상에 대한 특징 벡터를 이용하여 퍼지 분류기가 동정되었다. 최종 동정된 퍼지 분류기는 5개의 규칙을 가지며 19개의 입력을 받아 1개의 인식된 감정을 출력한다. 동정된 퍼지 분류기를 바탕으로 최종적으로 감정 인식 프로그램이 개발되었다. 개발된 감정 인식 프로그램은 그림 2와 같다. 개발된 프로그램은 얼굴 영상이 주어졌을 때, 제안한 얼굴 영상 분석 기법들을 걸쳐 얼굴 구성 요소를 분석하여, 이를 이용해 특징 벡터를 생성한다. 그리고 최종적으로 퍼지 분류기를 통해 최종 감정을 판별한다.

모의실험은 크게 제안된 인간의 움직임 추출 알고리즘 평가와 제안된 은닉 마르코프 모델로 동정된 감정 인식 분류기의 성능 평가로 나누어 수행되었다. 움직임 추출을 위해 3장에서 소개된 적응 뼈대 모델과 색상 기반 손의 움직임 추출 알고리즘을 이용한다. 개발된 감정 인식 프로그램은 그림 3과 같다. 인간의 손 움직임 추출과 추출된 움직임을 바탕으로 은닉 마르코프 모델을 이용해 동정된 분류기를 설계하고 성능을 평가 한다.

표 4와 표 5는 각각의 실험 결과를 나타낸다. 결과를 살펴 보면 얼굴의 경우 특징 벡터 추출율이 78.67% 인 반면, 추출된 특징 벡터를 이용한 감정 인식률은 98.06%로 매우 높은 것을 확인 할 수 있다. 움직임을 통한 감정 인식의 결과를 살펴 보면 전체 인식률이 64%로 얼굴을 사용하였을 때 보다 낮음을 확인할 수가 있다. 우리는 얼굴을 통한 감정의 인식이 가능한 경우는 얼굴을 통한 분석을 우선시해야한다. 얼굴을 통한 감정인식이 불가능한 경우에만 움직임을 바탕으로 감정을 인식한다.

전체적인 혼합 시스템에서는 각각의 전체 감정 인식률인 75%와 65%를 바탕으로 가중치를 각각 0.53과 0.47로 계산 할 수 있다.

$$\text{전체감정 인식률} = 0.53 * \text{얼굴을 통한 감정의 인식률} + 0.47 * \text{움직임을 통한 감정의 인식률}$$

최종적으로 얼굴 영상을 이용한 감정인식과 더불어 동작 인식을 통한 감정 인식 기법을 통합 시스템을 구축하였다. 인터랙션 통합 시스템은 영상 분석을 통해 얼굴 영상이 일정 크기 이상으로 해석이 가능하면 얼굴 영상을 통한 감정 인식을 수행한다. 이와 같이 개발된 감정인식 알고리즘이 향후 지능 로봇에 탑재된다.

5. 결 론

감정 인식은 많은 분야에서 각광받고 있는 기술 중에 하나이다. 지능형 시스템의 개발이 가속화됨에 따라 감정 인식 기술 개발의 필요성이 커지고 있다. 이에 본 논문에서는 인간의 얼굴과 움직임 분석을 통한 감정 인식 기법을 제안 하였다. 제안된 알고리즘은 영상에서 얼굴과 움직임 추출 기법과 감정 분류 기법으로 구성 된다. 추출된 인간의 얼굴 및 움직임 특징 벡터에 분류기를 이용하여 감정을 인식하였다. 최종적으로 모의실험을 통해 전체 알고리즘 수행과정을 살펴 보았으며, 알고리즘이 잘 동작함을 확인하였다.

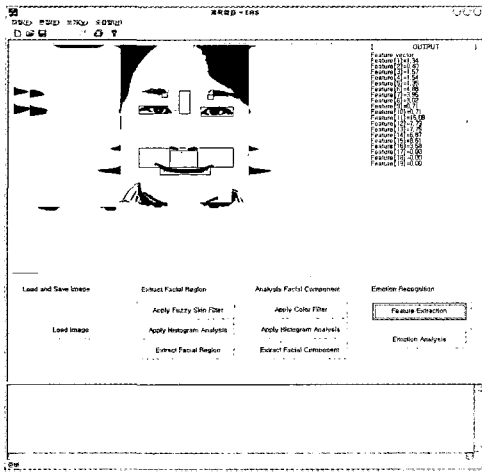


그림 2. 얼굴 영상 분석
Fig.2 Facial image analysis

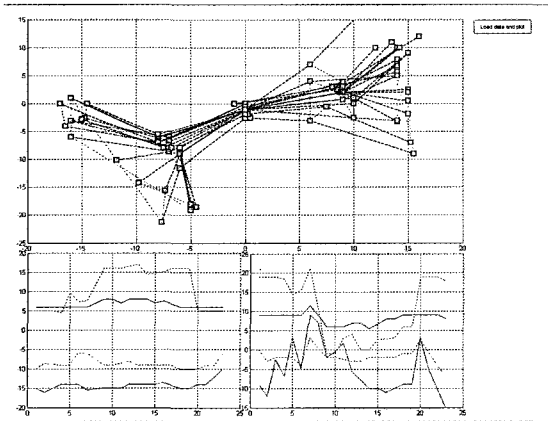


그림 3. 움직임 영상 분석
Fig 3. Human hand image analysis

표 4 얼굴인식에 의한 실험 결과
Table 4 Experiment results for facial recognition

특징 벡터 추출율	78.67%
특징 벡터를 이용한 감정 인식율	98.06%
전체 감정 인식 율	77.33%
각 감정 별 감정 인식율	
슬픔	75.06%
화남	78.67%
놀람	75.54%
역겨움	78.67%

표 5 제스처 인식에 의한 실험 결과
Table 5 Experiment results for gesture recognition

인간의 움직임 추출율	98.06%
손 영역 추출율	73.84%
전체 감정 인식 율	64%
각 감정 별 감정 인식율	
슬픔	62%
화남	67%
놀람	61%
역겨움	66%

참 고 문 헌

- [1] J. K. Aggarwal and Q. Cai., "Human Motion Analysis: A Review," Computer Vision and Image Understanding, pp. 428-440, 1999
- [2] B. Fan, Z.-F. Wang, "Pose estimation of human body based on silhouette images," International Conference on Information Acquisition Proceedings., pp. 296-300, June, 2004.
- [3] Haritaoglu, R. Cutler, D. Harwood and L. Davis, "Backpack: Detection of people carrying objects using silhouettes," Computer Vision and Image Understanding, pp. 385-397, No. 3, 2001
- [4] Haritaoglu, D. Harwood, and L. Davis, "A real time system for detection and tracking people" Journal of Image and Vision Computing, 1999.
- [5] Blake, M. Isard, and D. Reynard, "Learning to track curves in motion of contours," Artificial Intelligence, pp.101-133, 1995
- [6] Anderson, P. Burt, and G. van der Wal., "Change detection and tracking using pyramid transformation techniques," In Proceedings of SPIE - Intelligent Robots and Computer Vision, Vol 579, pp. 72-78, 1985
- [7] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," International Journal of Computer Vision, pp.42-77, 1994.
- [8] 김문환, 박진배, 주영훈, 조영조, 지수영, 김혜진, "지능형 로봇시스템에서 하이브리드 실루엣 추출 방법을 이용한 인간의 몸추출" 한국 퍼지 및 지능 시스템학회 추계학술대회, pp. 257-260 2005
- [9] 김문환, 박진배, 주영훈, 조영조, 지수영, 김혜진, "적응 뼈대 모델을 이용한 인간의 특징점 추출" 한국 퍼지 및 지능시스템학회 추계학술대회, pp. 257-260 2005

저 자 소 개

주영훈(Young hoon Joo)

2006년 제 16권 제 3호 참조

소제윤(Jae Yun So)

2006년 제 16권 제 3호 참조

심귀보(Kwee-Bo Sim)

2006년 제 16권 제 3호 참조

송민국(Min Kook Song)

2006년 제 16권 제 3호 참조

박진배(Jin Bae Park)

2006년 제 16권 제 3호 참조