

인간의 움직임 추출을 이용한 감정적인 행동 인식 시스템 개발

Emotional Human Body Recognition by Using Extraction of Human Body from Image

송민국* · 박진배* · 소제윤**, 주영훈**

Min Kook Song, Jin Bae Park, Je Yoon So and Young Hoon Joo

* 연세대학교 전기전자공학과

{s5ngm2n9k,jbpark}@control.yonsei.ac.kr

** 군산대학교 전자정보공학부

{ifuloveme, yhjoo}@kunsan.ac.kr

요 약

영상을 통한 감정 인식 기술은 사회의 여러 분야에서 필요성이 대두되고 있음에도 불구하고 인식 과정의 어려움으로 인해 풀리지 않는 문제로 남아 있다. 인간의 움직임을 이용한 감정 인식 기술은 많은 응용이 가능하기 때문에 개발의 필요성이 증대되고 있다. 영상을 통해 감정을 인식하는 시스템은 매우 다양한 기법들이 사용되는 복합적인 시스템이다. 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.

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.

영상을 이용한 감정 인식 방법은 일반적으로 감정이 가장 많이 표현되는 얼굴 영상을 이용하는 경우가 대부분이다 [3-7]. 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 나타낸 감정은 얼굴 미세한 변화로서 나타나기 때문에 이를 감지하기 위해서는 매우 정밀한 얼굴에 대한 분석이 이루어 져야 한다. 따라서 얼굴 영상을 이용한 감정 인식 방법 하나만 사용할 경우 지능형 로봇에 일정 크기 이상의 얼굴 사진을 제공하기 위해 로봇에 가까이 가야 하는 불편한 행위를 취해야 한다. 따

라서 얼굴 정보 이외에 전체 행동을 바탕으로 감정을 인식하는 연구가 필요하다.

인간의 움직임을 이용한 감정 인식은 그 인식 기법에 따라 기하학적 인식 기법, 고유공간을 이용한 인식 기법, 기타 변환을 이용한 인식 기법으로 나뉜다 [2], [5], [6]. 하지만 인간의 움직임을 이용한 감정 인식은 이를 위해 다양한 선행 기술이 수정 및 개발되어야 하는 어려움이 있다. 이를 해결하기 위해 인간의 움직임 영상 분석에 대한 선행 기술들이 연구되었다 [1], [2].

본 논문에서는 기 개발된 선행 기술을 바탕으로 감정 인식시스템을 위한 인간의 움직임 추출 기법과 은닉 마르코프 모델(HMM)을 통해 동정된 분류기를 통한 감정 인식 기법으로 제안한다. 인간의 움직임 추출 기법은 선행 연구된 기법들로 얻어진 인간의 움직임 구성 요소에 정보를 바탕으로 감정 분류의 성능을 높이기 위한 손의 움직임을 파악하는 방법이다. 본 논문에서는 지능형 로봇으로부터 얻어진 동영상 데이터를 바탕으로 얼굴을 이용한 감정 인식과 더불어 특정 동작을 이용한 통합적인 감정 인식 시스템을 개발하는 것을 목표로 한다.

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 얼굴 색의 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.

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

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.

$$\mu_{M_i} = e^{-\frac{(c_i - x)^2}{v_i}} \quad (2)$$

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

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

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

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.

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

여기서, $\Upsilon(x)$ 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_i^j 와 v_i^j 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].

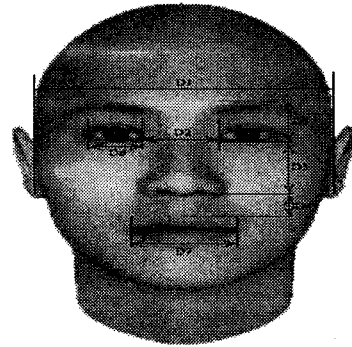


그림 1 가상 얼굴 모형
Fig. 1 Model of virtual face

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 D1의 1로 축소한 비율을 가진다 [1].

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. 자세한 탐색 공간에 내용은 기존 연구 논문에 자세하게 언급되어 있다 [2].

2.2. 얼굴 영상을 통한 감정 인식 알고리즘

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

추출된 특징 벡터는 그 영상의 크기와 개개의 얼굴 크기가 다른 관계로 정규화가 필요하다. 본 연구에서는 추출된 얼굴 영역 중 가장 확실한 추출이 가능한 눈가로 길이 D_1 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.

3. 색상 정보를 바탕으로 한 손의 움직임 추출 방법

실루엣 정보만으로 손의 특징점을 찾는 것은 충분하지가 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 (6)과 같이 계산된다.

$$\bar{I}_c = I_c \bar{I} \quad (6)$$

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 각들은이 영역에서 다음과 같은 에너지 함수 (7)에 의해서 에너지 값을 계산하게 된다. 최종적으로 에너지 값이 가장 큰 각으로 탐색을 해 나가는 방법을 취한다.

$$S = \left(\sum_{Vx \in R_2} f(x) - \sum_{Vx \in R_1} f(x) \right) / r + \sum_{Vx \in L} f(x) \quad (7)$$

4. 감정 인식 기법 및 모의 실험

실험에 사용된 영상은 CCD 카메라로부터 얻은 320×240 크기의 24bit의 컬러 영상이 사용되었다. 가상 얼굴 모형으로 얻어진 구성요소를 바탕으로 제안된 특징 벡터 추출 기법을 사용하여 특징 벡터를 추출하게 된다. 변환된 총 150 개의 얼굴 영상에 대한 특징 벡터를 이용하여 퍼지 분류기가 동정되었다. 개발된 감정 인식 프로그램은 그림 2와 같다. 제안한 얼굴 영상 분석 기법들을 걸쳐 얼굴 구성요소를 분석하여, 이를 이용해 특징 벡터를 생성한다. 그리고 최종적으로 퍼지 분류기를 통해 최종 감정을 판별한다.

인간의 움직임 추출 알고리즘 평가와 제안된 은닉 마르코프 모델로 동정된 감정 인식 분류기의 성능 평가로 나누어 수행되었다. 표 1과 2는 각각의 실험 결과를 나타낸다. 결과를 살펴보면 얼굴의 경우 특징 벡터 추출율이 78.67% 인 반면, 추출된 특징 벡터를 이용한 감정 인식률은 98.06%로 매우 높은 것을 확인할 수 있다. 움직임을 통한 감정 인식의 결과를 살펴보면 전체 인식률이 64%로 얼굴을 사용하였을 때 보다 낮음을 확인할 수가 있다.

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

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

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

5. 결론

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 살펴보았으며, 알고리즘이 잘 동작함을 확인하였다.

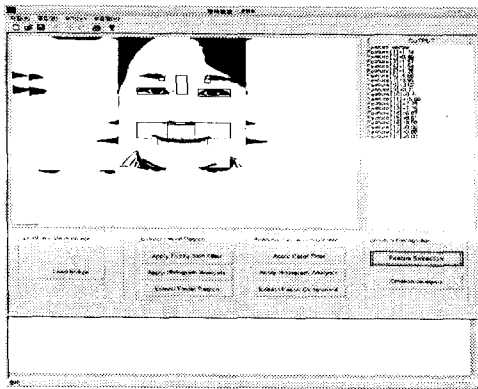


그림 2 얼굴 영상 분석
Fig.2 Facial image analysis

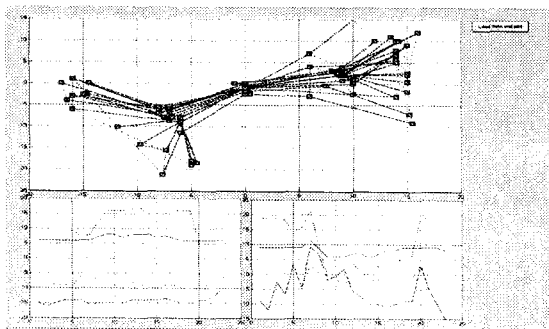


그림 3. 움직임 영상 분석
Fig 3. Human hand image analysis

표 1 감정 인식 실험 결과

Table 1 Experiment result on emotion recognition

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

표 2 감정 인식 실험 결과

Table 2 Experiment result on emotion xrecognition

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

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

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. Hawood 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.