

이동로봇에서의 2D얼굴 영상을 이용한 사용자의 감정인식

Emotion Recognition of User using 2D Face Image in the Mobile Robot

이동훈, 서상욱, 고광은, 심귀보

서울시 동작구 흑석동 221, 중앙대학교 전자전기공학부
E-mail: kbsim@cau.ac.kr

요 약

본 논문에서는 가정용 로봇 및 서비스 로봇과 같은 이동로봇에서 사용자의 감정을 인식하는 방법 중 한가지인 얼굴영상을 이용한 감정인식 방법을 제안한다. 얼굴영상인식을 위하여 얼굴의 여러 가지 특징(눈썹, 눈, 코, 입)의 움직임 및 위치를 이용하며, 이동로봇에서 움직이는 사용자를 인식하기 위한 움직임 추적 알고리즘을 구현하고, 획득된 사용자의 영상에서 얼굴영역 검출 알고리즘을 사용하여 얼굴 영역을 제외한 손과 배경 영상의 피부색은 제거한다. 검출된 얼굴영역의 거리에 따른 영상 확대 및 축소, 얼굴 각도에 따른 영상 회전변환 등의 정규화 작업을 거친 후 이동로봇에서는 항상 고정된 크기의 얼굴 영상을 획득 할 수 있도록 한다. 또한 기존의 특징점 추출이나 히스토그램을 이용한 감정인식 방법을 혼합하여 인간의 감성 인식 시스템을 모방한 로봇에서의 감정인식을 수행한다. 본 논문에서는 이러한 다중 특징점 추출 방식을 통하여 이동로봇에서의 얼굴 영상을 이용한 사용자의 감정인식 시스템을 제안한다.

Key Words : 특징점 추출, 얼굴 인식, 표정 인식, 감정 인식, 지능형 로봇

1. 서 론

인간의 감정을 인식하기 위한 연구는 음성, 영상, 생체신호 및 몸짓의 4가지 매체로 구분 지을 수 있으며 각각의 매체에 따라 독립적, 또는 혼합적으로 각 매체의 장점만을 살린 감정인식 연구가 진행 중이다. 1990년부터 2005년까지의 IEEE 논문으로 출간된 감정인식 연구를 통계해 보면 음성에 대한 연구가 가장 많고, 그 다음으로 영상 또는 음성과 영상을 조합한 연구, 그리고 생체 신호와 몸짓에 대한 연구가 소수 진행 중임을 볼 수 있다. 위와 같은 결과는 음성과 영상에 대한 신호의 추출이 생체신호나 몸짓보다 용이하고 신호의 분류 가능성도 더 좋기 때문으로 보여진다 [1]. 최근 급성장한 로봇 분야에서도 다른 생체인식 기술과 달리 감정인식 대상이 되는 인간에게 특별한 장치를 동원하지 않고, 어떤 행위나 동작을 강제적으로 요구하지 않으면서 자연스럽게 데이터를 획득할 수 있는 음성 및 영상을 이용하여 인간-로봇간의 인터페이스가 활발히 연구 중이며, 휴머노이드 로봇의 정점에서 부각되어 질 인간과 로봇 상호간의 감정의 교류에 대한 연구도 다방면에서 진행 중이다. 결국 로봇이 인간의 감정을 어떻게 인식 할 수 있는지가 가장 큰 문제인데, 로봇이 인간의 감정을 인식하는 것은 인간이 다른 인간의 감정을 인식 할 때 시각 및 청각에 의존하여 대다수의 정보를

획득한다는 사실에서 그 실마리를 찾을 수 있다. 특히 시각정보는 감정을 인식할 때 가장 큰 비중을 차지하고 있으며, 인간이 다른 인간의 감정을 인식하는데 얼굴 표정에 대한 시각 정보를 가장 많이 이용한다는 데에는 이견이 없을 것이다. 로봇에서도 이 얼굴 표정의 시각 정보, 다시 말해 카메라를 통해 획득되는 얼굴 영상정보를 이용하면 감정인식을 하는데 좋은 성능을 기대할 수 있다. 얼굴 표정은 다양한 감정이 표현되며 가장 민감하게 반응하는 부분이기도 하다. 이 표정은 얼굴의 여러 가지 특징(눈썹, 눈, 코, 입)의 움직임을 이용하여 감정을 표현하고 인식한다 [2]. 얼굴영상을 이용하여 감정을 인식하는 방법은 영상획득과 획득된 영상에서의 특징점 추출, 추출된 특징점을 통한 감정인식으로 구분 지을 수 있다. 영상 획득은 2차원 또는 3차원으로 획득되며 본 논문에서는 로봇에서 구현이 간편한 2차원 영상정보를 이용한다. 다음으로 획득된 영상에서의 특징점 추출 방법은 특징 기반 방법 중 얼굴의 색상정보, 즉 피부의 색상정보 및 모양, 히스토그램 등을 이용하는 방법 등 여러 특징점 알고리즘을 병렬적으로 추가하는 방식인 다중 특징점 추출 방식을 사용한다. 마지막으로 특징점 추출을 통하여 획득된 감정정보를 구별하여 인식하는 방법으로 인식알고리즘에서 가장 많이 사용되고 있는 다층 신경망 구조(Multi Layer Perceptron)를 이용하여 감정인식을 수행 한다.

2. 감정인식 시스템

2.1 감정인식 시스템

이동로봇에서 얼굴영상을 이용하여 감정을 인식하는 방법은 배경영상획득과 획득된 영상에서의 인간의 영역만을 검출하고, 검출된 영역에서 얼굴영역만을 찾은 후, 여러 알고리즘을 적용하여 특징점을 추출하고, 추출된 특징점을 통한 감정인식으로 구분 지을 수 있다. 영상 획득은 2차원 또는 3차원으로 획득되며 포즈 및 조명의 변화에 민감한 2차원의 문제를 개선하기 위하여 range image와 vertex point로 나뉘는 3차원 영상획득 방법도 있다. 하지만 현재까지 진행되어진 대다수의 연구는 손쉽게 영상을 획득할 수 있는 2차원 방법이 사용되어졌으며, 본 논문에서도 로봇에서 구현이 간편한 2차원 영상정보를 이용한다. 다음으로 획득된 영상에서의 특징점 추출 방법은 지식기반 방법, 특징기반 방법, 템플릿 매칭 방법 및 외형기반 방법 등이 알려져 있다. 본 논문에서는 특징기반 방법 중 얼굴의 색상정보, 즉 피부의 색상정보 및 모양, 히스토그램 등을 이용하는 방법 등 여러 특징점 알고리즘을 병렬적으로 추가하는 방식인 다중 특징점 추출 방식을 사용한다. 마지막으로 특징점 추출을 통하여 획득된 감정정보를 구별하여 인식하는 방법으로 인식알고리즘에서 가장 많이 사용되고 있는 다층 신경망 구조를 병렬적으로 이용하는 방식을 사용하여 감정을 인식한다.

2.2 움직임 추적알고리즘

본 논문에서는 USB CCD 카메라가 내장된 이동로봇을 사용하여 사용자의 감정인식을 수행한다. 그림 2는 본 논문에서 감정인식을 수행하게 될 연구용 로봇을 보여준다. 본 논문에서는 로봇에서 움직이는 인간의 얼굴 표정으로 부터 감정을 인식해야 하기 때문에 로봇이 수행해야 할 가장 첫 번째 임무는 움직이는 인간의 위치를 찾아 영상정보를 획득하는 것이다.

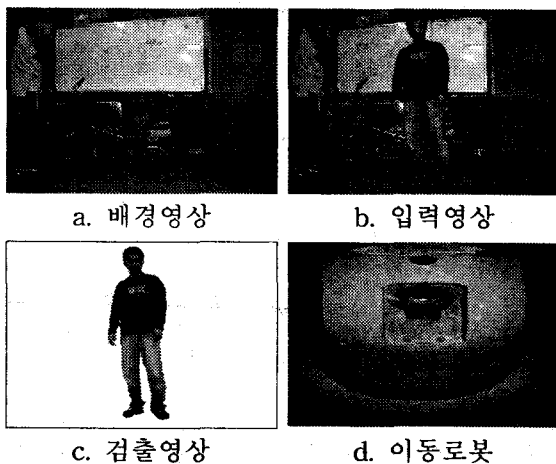


그림 2. 움직임 추적 및 이동로봇

이동로봇은 배경 영상으로부터 움직이는 인간의 영역을 검출해야 하는데 이 부분은 촬영

시점의 배경을 측정해 입력되는 영상과의 영상차이에 따라 검출된 배경 영상을 획득한 후 촬영시점의 배경영상에서 검출된 배경영상을 빼줌으로서 새롭게 입력된 인간의 영상만을 획득할 수 있다. 그림 2에 이동로봇과 이 로봇이 배경영상에서 움직이는 인간의 영상을 획득한 결과를 보여준다.

$$\text{Background Image} - \text{Input Image} = \text{Result Image} \quad (1)$$

2.3. 얼굴영역 검출

인식할 대상의 영상이 획득되면 획득한 영상에서 얼굴의 표정으로 감정인식을 수행해야 하기 때문에 배경이 복잡하더라도 얼굴영역을 정확히 찾아야만 한다. 이 문제를 해결하기 위해 여러 알고리즘이 연구되어 왔는데, 본 논문에서는 피부색 추출 모델[3]을 이용하여 얼굴영역을 검출한다. RGB 색상은 빛의 밝기나 강도에 민감하기 때문에, RGB 색상 모델을 그대로 이용하기에는 문제점이 따른다. 따라서 다음 식과 같은 강도 표준화를 통해 피부색의 편차를 감소시킬 수 있다.

$$s = \frac{R}{R+G+B} \quad (2)$$

$$t = \frac{G}{R+G+B} \quad (3)$$

피부색을 이용하여 얼굴영역을 검출할 때에는 손과 옷 등에 포함되어 있는 피부색이 문제가 된다. 이 부분은 수직·수평 축의 피부색 분포를 통하여 피부색 후보 영역 중 가장 큰 살색영역 즉 얼굴영역을 제외하고 나머지 부분은 제거한다. 그림 3은 수직·수평영역에서의 피부색영역을 검출한 후 이 중 가장 큰 면적을 가지는 얼굴영역만을 검출한 영상을 보여준다.

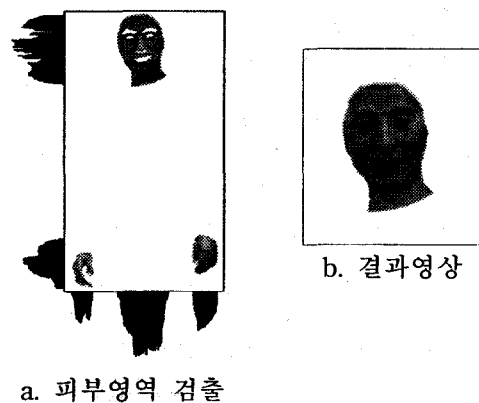


그림 3. 피부영역 검출 알고리즘

다음으로 여러 특징점 추출 방식을 적용하기 위해, 검출된 얼굴 영상만을 원래 색채로 변환하는 과정을 거치게 된다. 이 과정이 완료되면 얼굴영역에서 원하는 감정정보를 얻기 위해 특징점을 추출하게 되는데, 이 과정이 매우 중요하다. 얼굴의 여러 가지 특징점, 인간이 인식

하는 특징점 뿐만 아니라 인간이 인식할 수 없는 특징점까지도 감정인식을 위한 특징점 추출에 이용된다. 기존의 특징점 추출은 기하학적 방식이나 히스토그램 정보, 얼굴 구성요소의 크기 및 모양, 형판 매칭, 아이겐 벡터 등 다양한 방식이 연구되어 왔는데, 본 논문에서는 이러한 알고리즘을 병렬적으로 연결한 다중 특징점 추출 방식을 적용한다.

3. 다중 특징점 추출과 감정 분류

3.1. 얼굴영역의 정규화

2차원 영상에서 가장 문제가 되는 부분이 입력 영상의 밝기와 크기 및 회전에 대해 민감하다는 것이다. 본 논문에서는 이러한 문제점을 해결하기 위하여 RGB색채 모델을 HSI모델로 변환한다. RGB색채 모델을 HSI 색채 모델로 변환하는 식은 다음과 같다.

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B \geq G \end{cases}, \quad (4)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R,G,B)] \quad (5)$$

$$I = \frac{1}{3}(R+G+B) \quad (6)$$

여기서,

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{[(R-G)^2 + (R-B)(G-B)]^{1/2}}} \right\}$$

RGB 색채 모델로 변환 후에는 명도에 해당하는 I 부분을 사용하지 않음으로서 이차원 영상에서의 밝기로 인한 문제점을 해결할 수 있다. 다음으로 이동로봇에서 움직이는 인간의 영상을 획득할 때, 촬영시점에서의 로봇과 인간과의 거리에 따라 얼굴 영상의 크기가 변형되어 입력된 영상을 전처리 없이 사용하기가 힘들다. 이 부분은 얼굴영역이 검출된 후 입력 영상크기의 확대 및 축소 기법을 사용하여 해결할 수 있다. 영상의 확대를 위한 좌표 변환 식은 식 7과 같다. 여기서 x, y 는 원 영상의 좌표이고, x', y' 는 확대영상의 좌표, T_x, T_y 는 확대인수이다.

$$(x', y') = ((x \times T_x), (y \times T_y)) \quad (7)$$

영상의 확대로 인한 품질이 나빠지는 현상을 줄이기 위하여 주위 화소의 평균값을 이용하는 보간법 중 쌍 일차 보간법(Bilinear interpolation)을 사용하여 영상의 품질을 개선한다. 영상의 축소는 보간법을 사용할 필요가 없기 때문에 식 5의 영상축소를 위한 좌표 변환식으로 해결이 가능하다.

$$(x', y') = ((x \div T_x), (y \div T_y)) \quad (8)$$

영상의 회전변환은 Sine 함수와 Cosine 함수를 이용하여 좌표를 변환한다.

$$\begin{aligned} x_2 &= \cos(\theta) \times (x_1 - x_0) - \sin(\theta) \times (y_1 - y_0) + x_0 \\ y_2 &= \sin(\theta) \times (x_1 - x_0) - \cos(\theta) \times (y_1 - y_0) + y_0 \end{aligned} \quad (9)$$

여기서 x_0, y_0 는 회전중심점, x_1, y_1 은 원영상의 좌표, x_2, y_2 는 변환좌표이다. 영상의 회전변환시 영상의 확대와 마찬가지로 빈 홀이 생겨 영상의 품질이 나빠질 수 있다. 이 부분은 영상의 확대에서 사용했던 쌍 일차 보간법을 사용하여 품질을 개선할 수 있다. 이와 같은 방식으로 피부색을 이용하여 검출된 얼굴 영상을 크기 확대 및 축소, 회전을 거쳐 항상 고정된 크기인 256*256의 영상 이미지를 획득 한다.

3.2 다중 특징점 추출 알고리즘

다양한 특징점 추출 방식을 통합적으로 이용하기 위해, 본 논문에서는 다중 특징점 추출 방식을 제안한다. 그림 4는 다중 특징점 추출 방식의 개략도를 보여준다. 그림 4에서 알 수 있듯이 이동로봇에서 움직이는 인간의 영상을 배경이미지와의 차이를 이용하여 추적한 후 획득된 영상에서 얼굴 영상만을 검출한 후 검출된 얼굴 영상으로부터 다양한 특징점을 추출하여 감정 분류 시스템의 입력으로 사용하게 되며, 언급된 특징점 추출이외의 다른 알고리즘도 병렬적으로 추가될 수 있다.

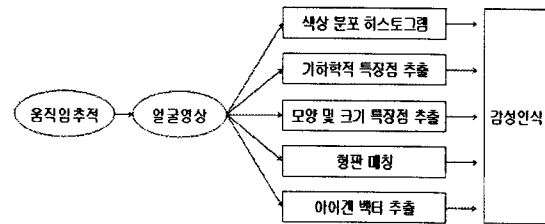


그림 4. 다중 특징점 추출 방식 개략도

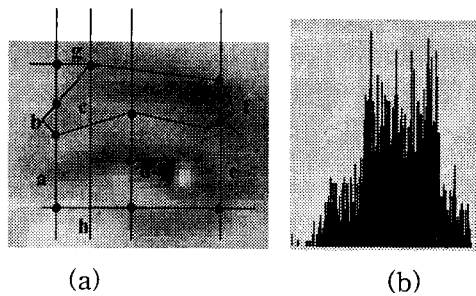


그림 5. (a) 눈썹의 특징점 추출에 사용된 파라미터 (b) 눈 및 눈썹의 색상 분포 히스토그램

본 연구에서는 기존에 가장 많이 사용되었던 기하학적 특징점 추출 방식과 감정 인식 분야에서는 사용되지 않았던 색상 분포 히스토그램 및의 두 가지에 대해 우선적으로 특징점을 추출한다. 기하학적 특징점 추출은 R. Brunelli와 T. Poggio가 제안한 방식을 사용한다 [5]. 이 알고리즘에서 사용된 특징점은 눈썹의 두께와 눈의 중심위치에서의 수직 위치, 왼쪽 눈썹의 표현

(11개), 코의 수직 위치 및 폭, 입의 수직위치, 폭, 입술의 높이, 턱의 모양을 나타내는 11개의 반경, 코 위치에서의 얼굴 폭, 코 끝과 눈 사이의 중심점에서의 얼굴 폭 등 총 35개의 기하학적 특징점을 추출한다[2]. 그림 5(a)는 이 중에서 눈썹 부분의 특징점 추출 모습을 보여준다. 다음으로 사용되는 특징점 추출 알고리즘으로 색상 분포 히스토그램을 이용하는데, 눈 및 눈썹, 코 끝 주변, 입 주변의 색상 분포 히스토그램의 모양을 추출한다. 그림 5(b)는 눈 및 눈썹 주변의 색상 분포 히스토그램 추출 모습을 보여준다.

3.3 병렬구조 다층 신경망을 이용한 감정 분류

본 연구에서는 다층 특징점 추출 방식을 사용하기 때문에 특징점을 추출하는 단계의 개수만큼 다층 신경망 구조가 필요하다. 또한 우선 4개의 감정; 기쁨, 슬픔, 놀람, 화남에 대하여 분류하기 때문에 신경망의 마지막 층에서는 4개의 노드를 갖게 된다. 여기에서 분류하는 감정 및 특징점 추가 여부는 유동적으로 변형이 가능하다. 본 논문에서는 히스토그램 정보와 기하학적 특징점 및 입과 눈썹, 눈의 모양과 크기의 특징점 등 세 가지를 사용하므로 그림 6과 같은 세 개의 다층 신경망이 필요하고 마지막 단계 4개의 노드와 가중치가 추가되어 최종적인 감정 분류를 수행하게 된다. 우선 60개의 얼굴 영상으로부터 교사 학습을 통해 신경망의 노드와 가중치가 업데이트 되고, 최종적으로는 새로운 입력 영상에 대해 감정을 분류할 수 있는 시스템을 형성한다. 그림 6의 블록 안의 구조는 일반적인 다층 신경망 구조를 나타내는데, 최종 단계 연결된 4개의 노드와 가중치를 업데이트하기 위해서는 세 단계의 학습이 필요하다. 먼저 히스토그램 정보로부터의 감정정보를 분류하기 위한 위쪽의 다층 신경망의 학습과 기하학적 특징점으로 부터의 감정정보 분류를 위한 학습, 모양 및 크기의 특징점으로부터 감정 정보 분류를 위한 가장 아래쪽의 신경망 학습이 이루어진다. 세단계의 학습이 완료된 다음 이 병렬 구조를 연결하는 최종의 노드와의 가중치를 위해 전체적인 학습이 다시한번 이루어져 4가지의 감정을 분류하는 전체 감정 분류 시스템을 구성한다.

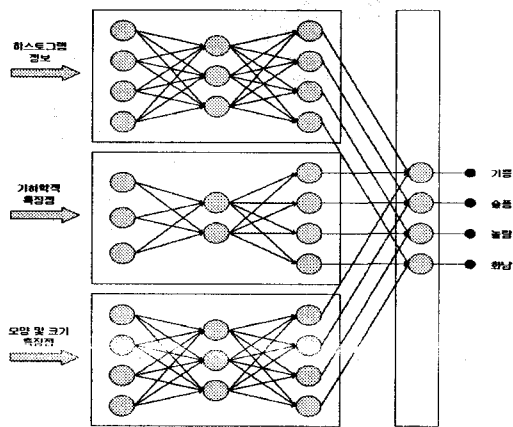


그림 6. 감정 분류를 위한 병렬구조 다층 신경망

4. 결론 및 향후 과제

본 논문에서는 이동로봇 이용하여 배경영상에서 움직임을 검출한 후 검출된 영역에서 피부색을 이용하여 얼굴 영역만을 검출하였다. 얼굴 영역의 감정 인식을 위해, 인간의 복잡한 감정 인식 시스템을 기반으로 한 다층 특징점 추출 방식이 본 연구에서 제안되었다. 이 방식은 여러 가지 다양한 특징점 추출을 수행하고 이를 병렬적으로 분류한 후에, 최종적으로 감정을 분류한다. 다층 특징점 추출 방식은 다양한 감정 정보를 입력으로 하고 통합적인 감정 분류를 수행하기 때문에 강인한 감정 분류를 수행할 수 있을 것으로 기대한다. 본 논문에서는 특징점 추출로서, 히스토그램 정보를 이용하였는데, 형판 매칭이나 아이겐벡터와 같은 다양한 특징점 추출 방식을 적용하는 것이 앞으로의 과제이다. 또한 우선적으로 4가지의 감정에 대해서만 감정 분류를 제안하였는데, 더 많은 감정을 분류하기 위해 신경망의 구조 변화와 실험이 필요하다.

감사의 글 : 이 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 수행된 연구입니다 (KRF-2005-042-D00268). 연구비 지원에 감사드립니다.

참고 문헌

- [1] C.-H. Park and K.-B. Sim, "The Novel Feature Selection Method Based on Emotion Recognition System," *Lecture Notes in Computer Science(LNCS) published in Springer Berlin/Heidelberg*, vol. 4115, pp. 731-740, Aug. 2006.
- [2] K.-S. Byun and K.-B. Sim, "Hybrid Feature Extraction for the Facial Emotion Recognition", *Proc. of ICCAS 2004(The International Conference on Control, Automation, and Systems)*, pp. 1281-1285, Aug. 2004.
- [3] R. S. Feris, T. E. Campos, R. M. Cesar Junior, "Detection and Tracking of Facial Features in Video Sequences", *Lecture Notes in Artificial Intelligence, Springer-Verlag press*, vol. 1793, pp. 197-206, Apr. 2000.
- [4] M. Pantic, L.J.M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, Dec. 2000.
- [5] R. Brunelli, T.Poggio, "Face recognition: features versus templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, Oct. 1993.
- [6] Eric Hamilton, JPEG File Interchange Format Version 1.02, Sep 1992.