

상반신 추적 기술 기반 감정 인식 시스템

*오지훈 **유선진 *이민규 ***임우택 ***안충현 *이상윤
 *연세대학교 전기전자공학과 **제주한라대학교 방송영상과 ***한국전자통신연구원
 syleee@yonsei.ac.kr

Emotion Recognition System based on Upper Body Tracking

*Jihun Oh **Sunjin Yu *Minkyu Lee ***Wootack Lim ***ChungHyun AHN *Sangyoung Lee

*Department of Electrical and Electronic Engineering, Yonsei University

**Department of Broadcasting & Film, Cheju Halla University

***Electronics and Telecommunications Research Institute

요약

Kinect를 통해 Color영상과 Depth영상을 받아온 후, 사람과 사람의 스켈레톤이 검출되도록 했다. 스켈레톤이 검출되면 머리 위치를 중심으로 얼굴 유효영역을 만들고, 효율적인 얼굴 검출로 사용자 인식이 가능하도록 했다. 스켈레톤 검출 및 추적을 통해, 4가지 감정에 대해 제스처를 정의했으며, 각 감정에 따른 제스처를 취했을 때 정의한 감정이 인식되는지 실험했다. 실험 결과, 제스처를 통한 감정 인식 성공률이 86 ~ 88% 나왔으며, 이 제스처 인식이 다른 감정인식 방법과 융합될 필요가 있다.

1. 서론

감정은 사람의 마음의 상태를 표출하는 가장 중요한 요소 중 하나이다. [1] 이 감정을 인식하는 기술이 가장 널리 활용될 수 있는 분야는 Human Computer Interaction(HCI) 시스템이다. 이러한 지능형 HCI 시스템에서는 얼굴표정, 시선, 제스처, 음성과 이들의 결합을 통하여 컴퓨터가 사용자의 감정을 인식하고 감정 상태에 적합한 작동 및 처리를 하기 때문에, 사용자의 만족을 극대화할 수 있다. 이와 같은 상황인지(context awareness)를 통하여 지금까지 기계 중심으로 설계되었던 인터페이스가 사용자 중심의 인터페이스로 전환되고 진정한 의미의 Ubiquitous computing environment를 구축할 수 있게 된다. 그렇기 때문에, 이 HCI 시스템에서 감정인식은 미래 기술에 많이 활용되고 있어, 많이 각광을 받고 있다.

감정인식 분야를 크게 인간의 얼굴표정이나 제스처가 담긴 영상 인식과 음성 인식을 통한 감정인식으로 나눌 수 있다. [2] 감정 인식 분야 중 표정 인식은 얼굴 인식에 비해 인식률이 낮으며, 다양한 표정을 인식하기가 쉽지 않다. 또 표정인식 만으로 감정인식을 하기에는 무리가 있으며, 복합적인 부분이 포함되어야 한다. 얼굴표정인식 시스템의 성능은 머리움직임, 몸짓, 포즈 등의 제스처 영상과 결합하여 더욱 개선될 수 있으며, 현재는 이 분야의 연구가 활발히 진행되고 있다.

본 논문에서는 상반신 부위들을 추적하여, 감정을 인식하는 방법을 제안하였다. 4가지 감정에 따른 제스처를 정의하여, 사용자가 그 정의된 제스처를 취하면 감정을 인식하는 시스템으로 구성하였다. Kinect 센서를 통해 실시간으로 Color영상과 Depth영상을 취득하여, 상반신 부위들의 추적을 통한 실시간 감정인식이 되도록 했다. 이 제스처를 통한 감정인식은 표정 인식만을 사용한 감정인식의 미흡한 점을

보완하여 더 성능이 좋은 감정인식이 가능하게 할 수 있다.

2. Background

Kinect는 Depth영상, Texture, 사용자 정보, 및 스켈레톤 정보 등을 만들어낼 수 있는 센서이다. [3] 이 Kinect의 기능들이 융합되면, Depth 영상에서 사람의 주요 부위들을 찾고, 그 부위들의 정보를 추출할 수 있다. 단, 사람은 Kinect 앞에서 전체적인 몸이 나온 상태로 몇 초동안 서 있어야 한다. Kinect를 통해 실시간으로 이 정보들을 얻을 수 있기 때문에, 실시간으로 사람의 주요 부위들을 추적하는 Full body tracking이 가능하다. OpenNI에 예제로 있는 Full body tracking에서 사람의 전체적인 몸이 들어오면 사람을 검출하고, 주요 몸 부위들이 스켈레톤으로 표시되며 추적된다. 만일 사람이 Scene 밖으로 나가고 빨리 들어오면, 그 해당 사람으로 인식하지만, 오래 머문 후 들어오면 새로운 사람으로 인식하게 된다. Kinect가 사람을 추적하는 뛰어난 센서라고 할지라도, Kinect는 Calibration을 하지 않고 사람의 스켈레톤을 검출할 수 없는 제약이 있다.

얼굴 검출은 영상에서 얼굴을 찾는 것으로, 영상을 얼굴과 얼굴이 아닌 클래스로 구분하는 이진 분류의 문제이다. [4] 얼굴은 개인별로 컬러, 크기, 모양이 다를 뿐 아니라 안경 착용 및 화장에 따라 매우 다양한 변화를 포함하는 클래스이다. 이러한 상황에 적합한 검출 알고리즘으로 AdaBoost를 들 수 있다. 이는 여러 약분류기(Weak classifier)들의 조합을 통해 성능이 우수한 강분류기(Strong classifier)를 생성해내는 방식이다. 이 검출된 얼굴 영역 영상들을 학습하여, 누구인지 인식하는 것이 얼굴 인식이다. 분류를 위한 대표적인 알고리즘은 LDA(Linear Discriminant Analysis)로서, S_B (Between-class

scatter)와 S_W (Within-class scatter) 사이의 관계를 이용해서 최적의 Projection을 찾는다. [5] 즉, 데이터들의 클래스들을 서로 잘 분류할 수 있도록, 각 클래스들의 중심 간의 거리를 최대화하는 Projection을 찾는 것이다. 이 Projection matrix를 등록 영상들에 곱해서 특징들을 구한다. 이 Projection matrix를 입력 영상에도 곱해서 특징들을 구한 후, 등록 영상들의 특징들과 비교를 시킨다. 입력 영상과 가장 가까운 거리를 갖는 등록 영상을 구해 인식을 확인한다.

3. Proposed Method

본 논문은 Kinect의 기능들을 융합하여 만들어진 Full body tracking 예제를 기반으로 하여, 사용자 인식과 제스처 인식을 통한 감정 인식을 설계하였다. 본 시스템은 아래와 같이 구성된다.

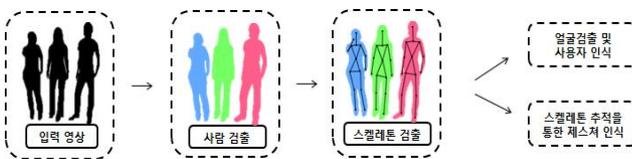


그림 1. 제안한 시스템 구성도

Full body tracking 예제에서 사람 각 부위의 스켈레톤을 추적할 때, 사람 부위 중 머리 부분 좌표를 활용하여 얼굴을 검출하도록 했다. 얼굴 검출 알고리즘으로 AdaBoost를 활용하였고, 사람 머리 위치를 중심으로 좌우 여백영역을 적절하게 고려하여, 얼굴 검출 유효영역을 만들도록 했다. 즉, 전체 Scene에서 얼굴을 검출하지 않고, 머리 부분 스켈레톤을 고려한 얼굴 유효영역에서만 얼굴을 검출하도록 했다. 이는 Color map에서만 얼굴 검출을 하지 않고, 스켈레톤 추적을 위한 Depth map도 활용하여 얼굴 오검출을 감소시키게 했다. 그 뿐만 아니라, 전체 Scene이 아닌 얼굴 유효영역만 찾기 때문에, Searching 영역을 감소시켜 속도를 개선시켰다. 이 효율적으로 검출된 얼굴 영상을 실시간으로 입력받아, SVM-DA(Support Vector Machine Discriminant Analysis) 알고리즘을 이용하여 얼굴을 인식시켰다. SVM-DA는 LDA의 약점을 보완한 것으로, 어떤 데이터 분포 형태에도 분류를 잘 할 수 있도록 각 클래스 간에 Margin을 최대화하는 Hyperplane을 찾는 것이다. [5] 이 얼굴인식 알고리즘을 활용하여, 실시간으로 매 프레임 얼굴 인식된 사용자를 저장했고, 50 프레임까지 저장된 사용자 중 제일 많이 나온 인식된 사용자를 최종 인식된 사용자로 설정했다. 즉, 감정 인식을 위해 유동적인 사용자 얼굴을 인식하는 것이 오차가 있기 때문에, 한 프레임의 인식이 아닌 여러 프레임의 인식을 종합하여 인식결과를 영상에 표시한 것이다. 사람의 스켈레톤이 검출되어 각 스켈레톤이 추적되면, 각 스켈레톤의 좌표를 알아내어 스켈레톤 간 위치관계를 파악할 수 있도록 했다. 이 스켈레톤의 위치관계를 통해, 각 감정에 따른 제스처를 정의했다. 감정은 총 4가지로, 감정들은 Glad(기쁨), Loneliness(외로움), Worried(걱정), Smell(불쾌감)이다. 먼저 양 손의 Y좌표가 머리의 Y좌표보다 높은 경우에 Glad를 정의했다. 다른 감정들도 다음과 같이 제스처를 정의했다. Loneliness는 양 손을 가슴에 모은 동작으로, 양 손이 어깨와 몸통 사이에 있는 경우이다. Worried는 오른 손의 위치가 어깨와 머리 사이에 있고, 오른 팔꿈치가 오른 손보다 앞에 있는 경우이다. Smell은 왼 손이 코를 만지는 동작으로, 왼 손이 코 근처에 가는 경우이다. 아래 그림

은 사용자 인식과 제스처 인식을 통한 감정인식의 결과를 보여준다.

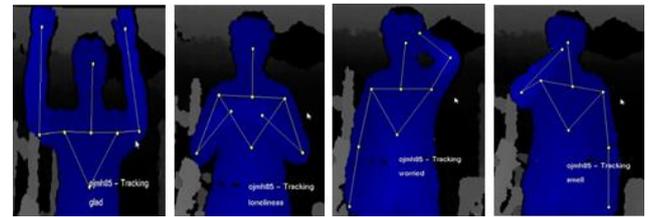


그림 2. 제스처 인식을 통한 4가지 감정인식 결과

4. Experimental Results

사람 5명에 대해서, 각 감정에 따른 정의된 제스처를 얼마나 잘 인식하는지 실험을 구성했다. 인식된 사용자가 4가지 감정에 맞는 제스처를 100번 취했을 때, 인식 성공률을 측정했다. 아래 표는 제스처를 통한 각 감정인식 성공률이다.

표 1. 제스처를 통한 감정 인식 성공률

사용자	사용자 인식	감정 인식 성공률 (100번 동작)				전체 감정 인식률
		Glad	Loneliness	Worried	Smell	
A	성공	96%	80%	88%	85%	87%
B	성공	95%	83%	85%	89%	88%
C	성공	97%	82%	90%	83%	88%
D	성공	96%	85%	89%	81%	88%
E	성공	97%	78%	88%	82%	86%

5. Conclusion

Kinect를 통해 Color영상과 Depth영상을 받아온 후, 먼저 사람의 스켈레톤을 검출시켰다. 그 스켈레톤을 추적시켜 각 감정에 따른 제스처 인식을 하도록 구현하였다. 실험 결과, 제스처를 통한 감정 인식 성공률이 86 ~ 88% 나왔다. 다른 감정인식 방법인 표정인식을 이 제스처 인식과 융합하면, 더욱더 효과적인 감정인식이 가능하다. 표정인식을 감정인식에 적용해보는 연구를 앞으로 진행하겠다.

6. Acknowledge

본 연구는 미래창조과학부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음

7. Reference

- [1] 김남수, "감정인식 기술의 현황과 전망," Telecommunications Review, 제19권5호, 2009년 10월
- [2] 신동일, "감정인식 기술 동향," 주간기술동향 통권, 1283호, 2007.2.14.
- [3] Chanjira Sinthanayothin, Nonlapas Wongwaen, Wisarut Bholsithi, Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene, Volume4, Number11, June 2012
- [4] 박지영, 이준호, "대용량 훈련 데이터의 점진적 학습에 기반한 얼굴 검출 방법," 정보과학회 논문지: 소프트웨어 및 응용, 31권, 7호, 949~958, 2004
- [5] Sang-Ki Kim, Youn Jung Park, Kar-Ann Toh, Sangyoun Lee, SVM-based feature extraction for face recognition, Pattern Recognition, Volume 43, Issue 8, August 2010, Pages 2871-2881.