

얼굴 표정 인식 기반 콘텐츠 선호도 추론 시스템

이연곤*, 조덕현[○], 장준익** 서일홍***

*한양대학교 대학원 지능형로봇학과

**한양대학교 대학원 전자컴퓨터통신공학과

***한양대학교 컴퓨터공학부

e-mail: yeon3710@yahoo.co.kr, {basecom, jijang}@incorl.hanyang.ac.kr, ihsuh@hanyang.ac.kr

Facial expression recognition-based contents preference inference system

Yeon-Gon Lee*, Durkhyun Cho[○], Jun Ik Jang**, Il Hong Suh***

*Department of Intelligent Robot Engineering, Hanyang University

**Department of Electronics and Computer Engineering, Hanyang University

***Department of Computer Science and Engineering, Hanyang University

● 요약 ●

디지털 콘텐츠의 종류와 양이 폭발적으로 증가하면서 콘텐츠 선호도 투표는 강한 파급력을 지니게 되었다. 하지만 콘텐츠 소비자가 직접 투표를 해야 하는 현재의 방법은 사람들의 투표 참여율이 저조하며, 조작 위험성이 높다는 문제점이 있다. 이에 본 논문에서는 콘텐츠 소비자의 얼굴 표정에 드러나는 감정을 인식함으로써 자동으로 콘텐츠 선호도를 추론하는 시스템을 제안한다. 본 논문에서 제안하는 시스템은 기존의 수동 콘텐츠 선호도 투표 시스템의 문제점인 콘텐츠 소비자의 부담감과 번거로움, 조작 위험성 등을 해소함으로써 보다 편리하고 효율적이며 신뢰도 높은 서비스를 제공하는 것을 목표로 한다. 따라서 본 논문에서는 콘텐츠 선호도 추론 시스템을 구축하기 위한 방법을 구체적으로 제안하고, 실험을 통하여 제안하는 시스템의 실용성과 효율성을 보인다.

키워드: 감정 인식(Emotion Recognition), 얼굴 표정 인식(Facial Expression Recognition), 인간-컴퓨터 상호작용(Human-Computer Interaction), 콘텐츠 선호도(Contents Preference)

1. 서론

최근 IT기술의 발달로 디지털 콘텐츠의 종류와 양이 폭발적으로 증가하고 있다. 무선통신기술과 스마트기기(스마트폰, 태블릿 PC 등)의 발달 및 확산은 누구나 쉽게 콘텐츠를 제작, 공유할 수 있는 환경을 만들었다. 여기에 유튜브, 페이스북, 싸이월드, 블로그 등 플랫폼 서비스의 등장은 디지털 콘텐츠의 증가와 확산을 촉진시키고 있다.

이러한 상황 속에서 페이스북의 성공요인 중 하나로 주목되고 있는 ‘좋아요 투표’는 간단하지만 강한 파급력을 지니게 되었다.[1] 이 투표를 통해 쌓인 선호도 통계는 콘텐츠 소비자와 생산자 모두에게 의미 있는 정보를 제공한다. 생산자에게는 콘텐츠에 대한 피드백을 제공하며, 소비자에게는 수많은 콘텐츠 중 어떤 콘텐츠를 소비할 것인지에 대한 정보를 제공하는 것이다. 위에서 언급한 여러 요인들에 의해 이 간단한 투표는 많은 서비스들에 적용되었으며, 온라인을 넘어 오프라인에 까지 영향을 끼치고 있다.

하지만 이 투표는 콘텐츠 소비자가 직접 투표 버튼을 눌러줘야 한다는 맹점이 있다. 이는 귀찮음을 유발하여 콘텐츠를 소비한 대부분의 사람들이 투표를 하지 않게 되는 결과를 낳는다. 또한 실제로 느끼는 감정과 상관없는 투표가 가능하여 조작이 쉽다는 문제점이 발생한다.

이러한 문제점들을 해결하기 위해 본 논문에서는 콘텐츠 소비자의 얼굴 표정을 인식하여 자동으로 선호도를 추론하는 시스템을 제안한다. 이와 관련된 연구로 TV 시청자의 생체신호를 TV 프로그램 추천 시스템에 활용한 연구가 있다.[2] 이 연구는 선호도 추론의 객관성은 확보할 수 있지만 별도의 장비를 장착해야하는 추가 비용과 번거로움이 더해지기 때문에 결코 좋은 해결책이라고 할 수 없다. 최근 대부분의 콘텐츠 소비자가 스마트기기를 통해 이루어지고 있으며, 스마트기기에는 대부분 전면 카메라가 달려있어서 콘텐츠 소비자의 얼굴 영상을 얻는 것이 용이하다. 또한 얼굴 표정에는 사람의 감정, 의도, 상태 등 상당히 많은 정보가 담겨 있으며, 대체로 문화권에 상관없이 보편적이다.[3] 따라서 본 논문에서 제

인하는 콘텐츠 선호도 자동 추론 시스템은 기존 수동 투표 시스템의 문제점을 해결하여 좀 더 효율적이고 신뢰도가 높으며, 보편적으로 적용이 가능한 서비스를 제공할 것으로 기대된다.

본 논문은 다음과 같이 구성되었다. 2장에서 제안하는 콘텐츠 선호도 추론 시스템을 설명하고, 3장에서 실험 결과를 보인 뒤, 4장에서 결론을 맺는다.

II. 콘텐츠 선호도 추론 시스템

본 논문에서 제안하는 얼굴 표정 인식을 이용하는 콘텐츠 선호도 자동 추론 시스템의 개요는 다음과 같다. 카메라로부터 얼굴 영상을 받아오면 1) 얼굴을 찾고, 2) 찾은 얼굴 내에서 주요 특징점들을 찾는다. 그리고 3) 얼굴 단위 움직임 인식 후, 4) 이를 바탕으로 감정을 인식한다. 5) 마지막으로 인식된 감정 정보를 기반으로 콘텐츠 선호도를 추론한다.

1. 얼굴 검출

얼굴 표정을 인식하기 위해선 먼저 입력 영상에서 얼굴을 찾아야 한다. 본 논문에서는 얼굴을 찾는 방법으로 Viola&Jones Detector[4]를 사용한다. 이는 Haar-like feature와 AdaBoost를 이용하는 방법으로 얼굴 찾기에 널리 쓰이는 방법이다.

2. 얼굴 특징점 추출

얼굴은 주로 눈, 눈썹, 코, 입을 통해서 감정을 표출한다. 즉, 감정 인식을 위해서는 이들의 위치와 모양을 알아야 한다. 얼굴 특징점 추출을 위해 본 논문에서는 AAM(Active Appearance Model)[5]을 사용한다. AAM은 얼굴이나 손 같이 변형 가능한 객체의 특징점을 추출하는 데 활용되는 통계적 모델로, 특징점 위치 정보를 모델링한 shape 모델과 픽셀 세기 정보를 모델링한 texture 모델로 구성된다. AAM을 이용하여 얼굴 특징점을 추출하는 과정을 <그림 1>에 나타냈다.



그림 1. AAM을 이용한 얼굴 특징점 추출 과정
Fig. 1. A process of facial feature extraction using AAM

3. 얼굴 단위 움직임 인식

얼굴에서 감정을 읽어내기 위해 선행되어야 할 것은 얼굴의 단위 움직임(Facial Action Unit; AU)을 인식하는 일이다. 얼굴 단위 움직임은 예를 들면 ‘눈썹이 올라갔다’, ‘입꼬리가 올라갔다’와

같은 것들이다.

의미 있는 얼굴 단위 움직임 체계는 심리학자들이 만들어놓았다.[6][7] 본 논문에서는 감정 및 선호도 파악에 매우 중요한 영향을 끼치는 얼굴 단위 움직임만을 선별하여 인식하고자 한다. <그림 2>와 <그림 3>에 인식할 얼굴 단위 움직임들이 명시되어 있다.

얼굴 단위 움직임 인식을 위한 방법으로는 다층 퍼셉트론 신경망[8]을 이용한다. 그 이유는 얼굴 단위 움직임은 한 순간에 한 가지만 발현되는 것이 아니라 여러 가지가 동시에 발현될 수 있는데, 이를 모두 인식하는 방법으로 신경망이 가장 적합하기 때문이다.

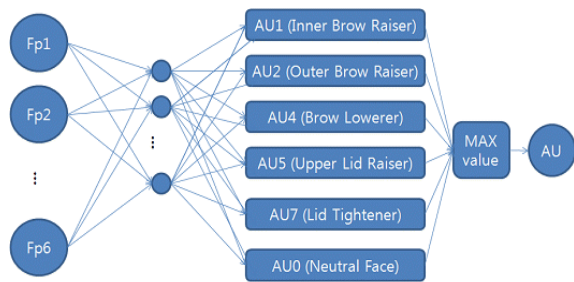


그림 2. 얼굴 위쪽을 위한 신경망
Fig. 2. Upper Face Neural Network

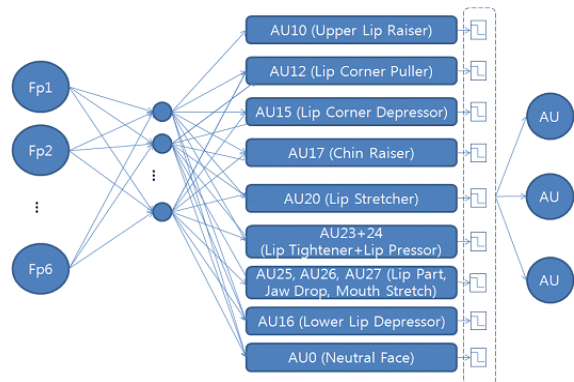


그림 3. 얼굴 아래쪽을 위한 신경망
Fig. 3. lower Face Neural Network

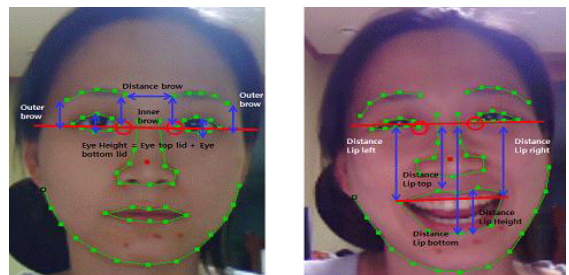


그림 4. 신경망 입력 파라미터. 왼쪽 사진은 얼굴 위쪽용 신경망의 입력 파라미터를 나타내고 있으며, 오른쪽 사진은 얼굴 아래쪽용 신경망의 입력 파라미터를 나타내고 있다.
Fig. 4. Input parameters of neural networks for upper face(left) and lower face(right)

보다 정확하고 효율적인 인식을 위해서 얼굴 위쪽 단위 움직임과 얼굴 아래쪽 단위 움직임을 나누어 인식한다. <그림 2>에 얼굴 위쪽용 신경망을, <그림 3>에 얼굴 아래쪽용 신경망을 나타냈다.

신경망의 입력으로는 전단계에서 추출한 얼굴 특징점 정보를 한번 더 가공한 파라미터가 들어간다. 이는 움직임에 초점을 맞추는 것으로, 기존 얼굴의 특징점과 현재 입력된 얼굴의 특징점을 비교하여 얻어진다. <그림 4>에 자세히 표현하였다.

4. 감정 인식

얼굴 영상에서 얼굴 단위 움직임들을 인식했다면, 이들의 조합으로 감정을 인식할 수 있다.

본 논문에서는 대표적 얼굴심리학자인 폴 에크먼이 개발한 FACS(Facial Action Coding System)[6]와 EMFACS(EMotional FACS)[7]를 근간으로 하여 <표 1>과 같이 6가지 기본 감정을 인식하는 규칙들을 만들었으며, 이들을 기반으로 감정을 인식한다.

표 1. 기본 감정 인식 규칙
Table 1. The rules of basic emotion recognition

감정	얼굴 단위 움직임
기쁨(Joy)	12
슬픔(Sadness)	1, 4 + 15
놀람(Surprise)	0, 1, 2, 5 + 25, 26, 27
공포(Fear)	2, 4, 5 + 20, 26
화남(Anger)	4, 7 + (23+24)
혐오(Disgust)	10, 16, 17, 20

5. 콘텐츠 선호도 추론

위의 단계들을 수행하게 되면, 콘텐츠를 소비하는 동안 매 순간 콘텐츠 소비자의 감정을 알 수 있다. 이제 이 감정 sequence를 활용하여, 콘텐츠 소비자의 해당 콘텐츠 선호도를 추론하면 된다.

어떤 콘텐츠의 평가에 모든 순간에 느낀 감정들이 동일하게 작용하지는 않는다. 따라서 본 논문에서는 중요한 순간의 감정에 가중치를 주었다. 가중 함수는 상황에 따라 두 가지 형태로 정의하여 사용한다.

첫 번째는 콘텐츠 소비자가 해당 콘텐츠를 좋아하는지 여부를 판단할 때의 가중 함수로, 단조 증가 선형 함수 형태를 띠도록 하였다. 콘텐츠 소비자의 만족도에 가장 큰 영향을 미치는 시간은 콘텐츠 후반부이기 때문이다. 두 번째는 콘텐츠 제작자의 의도에 적합한 지 여부를 판단할 때의 가중 함수로, 가우시안 함수 형태를 띠도록 하였다. 어떤 콘텐츠 제작자는 소비자가 그 콘텐츠를 좋아하는지 여부보다 자신의 목적에 부합했는지 여부를 더 알고 싶을 수 있다. 공포 동영상이 적절한 예가 될 것이다. 이때는 중요한 부분이 주로 콘텐츠 중간에 많이 나올 것이므로 이때의 감정을 중요하게 보는 것이 가장 적합할 것이다.

상황에 맞게 선택된 가중 함수와 매 프레임 별로 인식된 감정을 곱한 뒤, 각 감정들을 bin으로 하는 히스토그램을 쌓는다. 그리고 히스토그램에서 가장 높은 수치를 가지는 bin의 감정이 콘텐츠 소비자가 현재 콘텐츠를 소비하면서 얻은 주요 인상 혹은 감정이라고 판단한다.

고 판단한다.

콘텐츠 제작자의 의도에 맞는지 보기 위해서는 최종 결과 감정과 콘텐츠 제작자의 목적 감정을 비교해보면 되고, 콘텐츠 소비자가 해당 콘텐츠를 좋아하는지 여부를 판단하기 위해서는 기쁨 계열 감정이 최종 결과 감정으로 나왔는지 여부를 보면 된다.

본 논문에서 제안하는 시스템은 6가지 기본 감정에 추가로 무표정(neutral)도 인식이 가능하다. 따라서 무표정도 히스토그램의 하나의 bin으로 취급하였다. 사람들이 콘텐츠를 소비하면서 항상 웃거나 울거나 하는 표정을 짓지는 않기 때문이다. 다만 무표정에 대해서는 일정량의 패널티를 주었다. 이는 어떤 상황에선 싫다는 의미일 수 있지만, 어떤 상황에선 좋다는 의미가 될 수도 있는 중의적인 성격을 갖고 있기 때문이다.

III. 실험

본 논문에서 제안하는 시스템은 Windows 환경에서 C++와 OpenCV 2.0을 이용하여 구현되었으며, 실험은 랩톱PC에서 built-in camera를 활용하여 진행되었다.

실험에 사용된 AAM 모델은 동일 인물에 대해 여러 변화를 주어 직접 촬영한 50장의 영상에 각각 58개의 특징점을 찍어 학습시켰다. 또 다층 퍼셉트론 신경망을 학습시키기 위해서 직접 촬영한 얼굴 영상들과 JAFFE(Japanese Female Facial Expression)

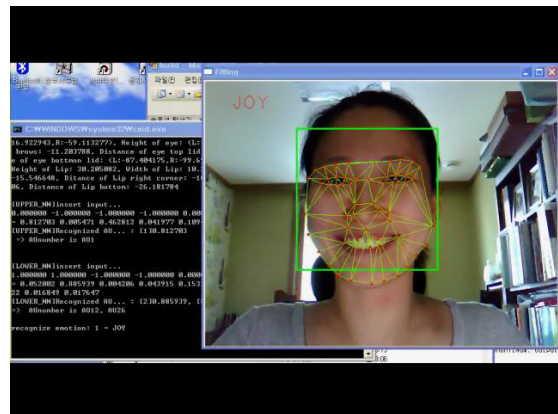


그림 5. 실험 결과
Fig.5. Experimental result

DB[9]에 직접 얼굴 단위 움직임을 라벨링한 데이터를 학습 데이터로 이용하였다.

실험은 실제 유튜브에서 동영상을 보는 영상을 촬영한 뒤에, 촬영한 영상을 구현한 시스템에 입력으로 넣어서 수행하였다. 3명의 피실험자가 3가지 동영상을 보면서 실험을 수행한 결과, 전반적으로 콘텐츠 소비자가 콘텐츠에 대해 직접 내린 평가와 유사한 결과가 나오는 것을 확인할 수 있었다. <그림 5>에 실험 결과 영상을 나타냈다.

IV. 결론 및 향후 계획

본 논문에서는 얼굴 표정 인식을 기반으로 하여 콘텐츠 선호도를 추론하는 시스템을 제안하였다. 제안하는 시스템은 기존의 수동 콘텐츠 선호도 투표 시스템의 문제점인 콘텐츠 소비자의 부담감과 번거로움, 조작 위험성 등을 해소함으로써 보다 편리하고 효율적이며 신뢰도 높은 서비스를 제공할 수 있도록 해준다. 이러한 점들과 콘텐츠 소비 중 스마트 기기를 통해 전면 얼굴 영상을 얻는 것이 용이하다는 점에서 본 논문에서 제안하는 시스템의 실용도는 높다고 할 수 있겠다.

향후에는 배경, 조명 등의 환경 변화가 있거나 표정 변화가 심해지면 얼굴 특징점 추출을 제대로 하지 못하는 문제점을 해결하여 좀 더 강인한 성능을 갖는 시스템을 개발할 계획이다. 또한 감정 인식하는 부분을 단순 규칙 기반 방법 대신 확률 기반 방법으로 변경하여 불확실한 상황과 표정에 적절히 대응하도록 할 계획이다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 로봇 특성화 대학원 사업단 지원사업의 연구결과로 수행되었음 (NIPA-2012-H1502-12-1002)

참고문헌

[1] Brian Carter, "The Like Economy: How Businesses Make Money With Facebook," Que Publishing, 2012.

[2] Seunghwan Lee, Jinhyuk Choi, Geehyuk Lee, Hankyu Lee, "AffecTV - TV Watcher's Preference Inference Based on Physiological Signal Analysis," HCI Korea, 2006.

[3] Paul Ekman, "Emotions Revealed," Henry Holt and Company, 2007.

[4] Paul Viola and Michael J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, pp. 137-154, 2004.

[5] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 23, pp. 681-685, 2001.

[6] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," Consulting Psychologists Press, 1978.

[7] P. Ekman and W. Friesen. "EMFACS-7: Emotional Facial Action Coding System." Unpublished manual, University of California, 1983.

[8] Y. Tian, T. Kanade and J. Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp.97-115, 2001.

[9] M.J. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.