

신체 길이 및 자세를 이용한 TV 시청자 인식

온경운[○], 곽동현, 박태서, 황보선, 장병탁
서울대학교 컴퓨터공학부

{kwon, dhkwak, tspark, bshwang, btzhang}@bi.snu.ac.kr

TV Viewers Recognition using Personal Body Length and Posture

Kyoung-Woon On[○], Dong-Hyun Kwak, Tae-Suh Park, Bo-Sun Hwang, Byoung-Tak Zhang

¹School of Computer Sci. and Eng., Seoul National University

요 약

본 논문에서는 TV 시청자의 주 시청 자세 및 신체 길이 정보를 이용하여 실시간으로 TV시청자를 인식하는 방법을 제안한다. 이 방법은 기존의 TV 사용자 식별 및 인증 방식이 가지고 있는 여러 제약을 극복 가능하고, 실시간으로 수행 가능함에 따라 다양한 TV 시청자 맞춤형 서비스를 제공하는데 기반이 될 수 있다. TV 시청자의 주 시청 자세 및 신체 길이의 특징점들은 마이크로소프트 키넥트 기기를 이용하여 추출하고 이를 여러 가지 기계학습 모델을 이용해 학습하여 그 성능을 평가한다. 실험 결과에서는 두 가지 특징을 사용함으로써 높은 정확도로 시청자 인식이 가능함을 보인다. 또한 interactive TV의 다양한 서비스를 위해 본 연구진이 제안한 시청자 인식 모델의 사용 가능성에 대해 논의한다.

1. 서 론

TV 시청은 오늘날 많은 사람의 일상적인 활동이 되었다. 최근 수 년 동안 DTT (Digital Terrestrial Television), IPTV (Internet Protocol Television) 등 새로운 TV 방송 시스템이 발전하면서 TV 시청 활동에 많은 변화를 가져왔고, 이 시스템들은 시청자 개개인에 맞는 채널을 제공하는 새로운 개념을 소개하였다[1]. 이러한 개념은 시청자와 TV가 서로 상호작용하는 interactive TV(iTV)의 다양성을 증가시켰다.

하지만, 이러한 발전에도 불구하고, 가족 구성원 등 여러 사람이 TV를 시청하는 일상적인 환경에서는 어느 사용자가 시청하고 있는지 인식을 하지 못하기 때문에 시청자 개개인의 맞춤형 서비스를 제공하기 어렵다. 이러한 제약사항은 TV 서비스 제공자에게, 현재 누가 TV를 시청하고 있는지를 알려줌으로써 해결될 수 있으며, 그에 따라 개인 맞춤형 광고, 시청자별 선호 채널 제공, 맞춤형 콘텐츠 추천 등의 시청자 맞춤형 서비스를 제공할 수 있을 것이다.

이러한 시청자 인식을 위한 선행 연구로 마이크로소프트 키넥트를 이용하여 사용자가 허공에 손을 이용한 서명을 등록하여 식별하는 방식, 미리 동작을 정의하여 식별하는 방식 등이 연구되었다[2-3]. 하지만 이 두 방식은 시스템이 사용자의 미리 정해진 동작을 저장해야하는 단점이 있다.

사용자의 생체 측정 정보를 이용한 인증 및 식별도 다양하게 연구 되고 있다. Heusch와 연구진들은 환경 및 빛의 조건을 제어한 상태에서 EER (Equal Error Rate)이

2.4%가 되는 얼굴 인식을 통한 인증 방식을 개발하였다. 하지만 조건이 제어되지 않을 경우 EER은 13.49%로 그 성능이 저하되는 결과를 나타내었다[4].

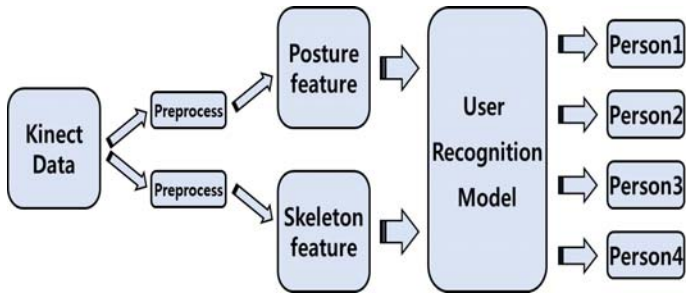
이 외에도 사람의 신체 크기도 생리학적 정보가 될 수 있는데, Araujo와 연구진들은 키넥트 센서를 이용하여 사람의 골격 정보를 통해 식별하는 방식을 개발하였다[5]. 또한 이 두 가지 방식을 조합하여 사람의 손을 흔드는 동작과 골격정보를 이용하여 사용자를 식별하는 방식도 연구되었다[6].

하지만 TV 시청 환경은 실시간으로 사용자가 변하고, 다수의 시청자가 동시에 시청하고 있는 경우가 많기 때문에 식별을 통해서 시청자 맞춤형 서비스를 제공하기가 어렵다. 또한 사용자는 손을 흔드는 등의 별도의 식별 과정 없이 서비스가 제공되길 원한다. 따라서, 본 연구는 TV 시청 환경에서 실시간으로 측정 가능한 시청자의 신체 길이와 현재 시청 자세를 이용하여 실시간으로 TV 시청자를 인식하는 방식을 제안하고자 한다.

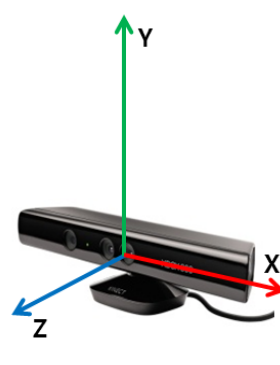
2. 연구 내용 및 방법

제안된 TV 시청자 인식 모델은 사람의 신체 길이와 주 시청 자세를 이용한다. 전체 구조는 [그림 1]에 도식화 되어있다. 먼저, TV위에 설치된 키넥트 기기를 이용하여 시청자의 신체 데이터를 생성한다. 그 후 해당 데이터를 가공하여 사람의 신체 길이와 주 시청 자세 데이터를 추출하고 이를 시청자 인식 모델을 이용하여 특정 사용자가 시청하고 있을 때 실시간으로 그 사용자가 누구인지를 인식한다.

2.1 데이터 전처리 및 특징점 추출



[그림 1] TV 시청자 인식 모델 구조



[그림 2] 키넥트의 좌표



[그림 3] 시청자 골격 정보

데이터 수집을 위해 마이크로소프트 키넥트 기기를 이용하여 기기 앞 사람의 관절 위치를 추정, 전신 골격 정보를 추출한다. 해당 정보는 [그림 2]에서 보는바와 같이 키넥트 기기를 중심으로 한 x, y, z좌표로 나타난다. 그러나, 사람의 TV 시청은 주로 앉은 자세로 행하여지므로 사람의 허리 아래 부분의 대한 좌표는 움직임의 차이가 많이 나지 않아 특정 시청자가 갖는 특징 즉, 시청자 인식에 주 특징점이 되지 못한다. 따라서 [그림 3]에서 보는 바와 같이 사람의 상체정보인 머리, 어깨 중심, 그리고 각 좌 우측 어깨, 팔꿈치, 손목 지점 좌표만 사용하도록 한다.

여기에서 해당 지점 좌표를 그대로 사용할 경우 그 값은 시청자가 TV와 떨어진 거리 및 앉은 위치에 따라 그 값을 달리하게 된다. 이러한 차이 또한 특정 시청자가 갖는 특징이 될 수도 있지만, 이는 상황 및 조건에 따라 매우 가변적이 될 수 있으므로 배제하여야 한다. 이에 따라 TV와 떨어진 거리 및 좌, 우로 치우친 정도의 영향을 제거하기 위해 어깨 중심 지점 좌표를 기준으로 하여 각 지점의 좌표를 변환하였다.

우리는 신체 길이의 정보를 추출하기 위해 각 지점사이의 거리를 계산하였다. 키넥트에서 초당 8프레임으로 얻어지는 데이터를 이용하여 40프레임인 5초 단위로 움직이는 이동 윈도우를 만들어 평균값을 추출하였다. 이러한 신호 처리를 통하여, 시청 중 자세의 변화나 이동으로 생기는 키넥트 기기 센서의 노이즈를 완화하였다.

또한 시청자의 시청 자세 정보를 추출하기 위해서 축이 되는 지점으로부터 연결된 다른 지점의 자세 각도를 단위벡터로 변환하였다. 예를 들어, 머리 지점의 자세 각도는 어깨 중심 지점에서 머리 지점으로 향하는 벡터의 단위벡터이고, 왼쪽 손목 지점의 자세 각도는 왼쪽 팔꿈치 지점으로부터 왼쪽 손목 지점으로 향하는 벡터의 단위벡터이다. 이러한 방식으로 위에서 제시한 모든 지점의 자세 각도가 결정되면, 상반신에 의해 특징지어지는 모든 자세를 추출 가능하다. 이 역시 40프레임인 5초 단위로 이동 윈도우를 만들어 평균값을 추출함으로써 센서

의 노이즈를 완화하고자 하였다. [표 1]은 특징점으로 추출된 각 지점의 거리 및 자세 각도를 보여준다.

2.2 학습 모델

주어진 데이터를 이용하여 시청자를 인식하기 위해 기계학습 모델을 이용하였다. Naive Bayes, SVM, Decision Tree(J48), Random Forest 4개의 기계학습 모델을 이용하여 데이터를 학습하고, 그 결과를 통해 전체적인 성능평가와 어떤 모델이 좋은 결과를 가져오는지 비교하였다.

성능평가는 10-fold cross validation과 66% percentage split 두 가지 방식의 정확도를 비교하였다. 10-fold cross validation의 경우 각 집단을 임의적으로 추출하기 때문에 훈련 데이터와 실험 데이터가 가깝게 위치할 가능성이 있다. 주 시청 자세 데이터의 경우 가까운 시간의 데이터는 비슷할 확률이 매우 높는데, 이는 실제보다 좋은 결과를 나타낼 가능성이 높음을 의미하므로 우리는 두 가지 성능평가 방법을 동시에 수행하여 검증을 수행하였다. Weka, 버전3.7.8이 주요 알고리즘 실험 도구로서 사용되었다.[7]

[표 1] 신체 길이 및 시청 자세 특징점

신체길이 (거리)	시청자세 (방향벡터)
어깨중심-머리	어깨중심→머리
어깨중심-왼쪽어깨	어깨중심→왼쪽어깨
어깨중심-오른쪽어깨	어깨중심→오른쪽어깨
왼쪽어깨-왼쪽팔꿈치	왼쪽어깨→왼쪽팔꿈치
왼쪽팔꿈치-왼쪽손목	왼쪽팔꿈치→왼쪽손목
오른쪽어깨-오른쪽팔꿈치	오른쪽어깨→오른쪽팔꿈치
오른쪽팔꿈치-오른쪽손목	오른쪽팔꿈치→오른쪽손목

3. 실험 결과 및 논의

실험은 대부분 가족 단위로 TV를 시청하는 환경임을



[그림 4] 가정의 거실을 모사한 시청자 인식 실험 환경

고려하여 4명의 피험자를 상대로 TV 예능 프로그램(무한도전 110, 111화) 두 편을 시청하였다. [그림 4]와 같이 실험실 환경을 일반 가정의 거실과 같이 구축하고 실험을 진행하였으며, 시스템 환경은 마이크로소프트 키넥트 SDK 1.8이 제공되는 API를 사용한 키넥트 기기를 사용하

[표 2] 신체 길이 데이터 실험 결과

	Naive bayes	SMO	J48	Random Forest
10-fold cross validation	64.1%	69.4%	96.4%	98.4%
66% percent split	64.1%	69.1%	95.8%	98.1%

[표 3] 시청 자세 데이터 실험 결과

	Naive bayes	SMO	J48	Random Forest
10-fold cross validation	75.4%	86.9%	98.1%	99.4%
66% percent split	75.3%	86.6%	97.8%	99.2%

[표 4] 신체 길이와 시청 자세 데이터 실험 결과

	Naive bayes	SMO	J48	Random Forest
10-fold cross validation	79.6%	94.3%	98.6%	99.7%
66% percent split	80.3%	94.2%	98.4%	99.5%

였다.

실험은 신체 길이 정보, 주 시청 자세 정보 각각을 이용한 방식과 위의 두 정보를 같이 사용한 방식, 총 3가지 방식으로 진행되었다. 실험 결과는 [표 2,3,4]와 같다.

실험 결과를 살펴보면, 골격 데이터보다 주 시청자세 데이터가 더 좋은 성능을 보이고 있음을 알 수 있었으며, 또한 이 두 데이터를 같이 사용하는 방법이 더 좋은 결과(어느 정도의 수치를 적어주기 바람)를 나타냄을 알 수 있다. 알고리즘간의 성능 비교 결과는 RandomForest, J48, SVM, Naivebayes 순으로 좋은 성능을 보이고 있음을 확인할 수 있었다.

4. 결론 및 향후 연구

본 논문에서는 TV 시청자의 주 시청 자세와 신체 길이 정보를 이용하여 실시간으로 TV 시청자를 인식하는 방법을 제안하였다. 실험 결과는 제안된 골격 데이터와 시청자세 데이터 정보를 모두 이용하여 Random forest 기계 학습 모델을 사용하였을 때, 결과적으로 높은 인식률을 나타내고 있음을 확인할 수 있었다. 제안된 인식 방법론은, 향후 interactive TV의 다양한 서비스에 활용될 수 있을 것으로 기대된다.

감사의 글

이 논문은 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2010-0017734-Videome.), 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원 지원(KEIT-10035348-mLife, KEIT-10044009)을 일부 받았음.

참 고 문 헌

- [1] Ardissono, Liliana, Alfred Kobsa, and Mark T. Maybury, eds. Personalized digital television, Kluwer Academic, 2004.
- [2] Tian, Jing, et al. "KinWrite: Handwriting-Based Authentication Using Kinect." Proc. of NDSS. Vol. 13. 2013.
- [3] Lai, Kam, Janusz Konrad, and Prakash Ishwar. "Towards gesture-based user authentication." Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. IEEE, pp. 282-287, 2012.
- [4] Heusch, Guillaume, and Sébastien Marcel. "A novel statistical generative model dedicated to face recognition." Image and Vision Computing, 28(1):101-110, 2010.
- [5] Araujo, Ricardo M., Gustavo Graña, and Virginia Andersson. "Towards skeleton biometric identification using the microsoft kinect sensor." Proceedings of the 28th Annual ACM Symposium on Applied Computing. ACM, pp. 21-26, 2013.
- [6] Hayashi, Eiji, Manuel Maas, and Jason I. Hong. "Wave to me: user identification using body lengths and natural gestures." Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp. 3453-3462, 2014.
- [7] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 10-18, 2009.