

스마트 기기에서의 사용자 인증을 위한 얼굴 및 제스처 정보를 활용한 사용자 인증

최현석*, 손명규**, 박혜영***
*경북대학교 전자전기컴퓨터학부
**대구경북과학기술원
***경북대학교 IT대학 컴퓨터학부
e-mail:choihs@ee.knu.ac.kr

User authentication using face and gesture information for various smart devices

Hyunsoek Choi*, Myoung-Kyu Sohn**, Hyeyoung Park***
*School of Electrical Engineering and Computer Science,
Kyungpook National University
**Daegu Gyeongbuk Institute of Science and Technology
***School of Computer Science and Engineering,
Kyungpook National University

요 약

다양한 스마트 기기의 출현과 함께 정보보호를 위한 보다 강건한 사용자 인증 시스템에 대한 요구도 증대되고 있다. 하지만 스마트 기기별로 호환성을 유지하면서도 사용자 인증을 수행하기 위해서는 각 기기에서 공통적으로 제공되는 센서의 활용이 필요하며, 영상 기반 다중 생체 인식에 기반을 둔 사용자 인증 시스템은 이에 대한 대안이 될 수 있다. 본 논문에서는 전통적으로 사용자 인증에 사용되고 있는 얼굴 인식과 더불어 영상 기반의 제스처 인식을 함께 사용함으로써 호환성을 유지하면서도 강건한 사용자 인증 시스템을 제안하였다. 그리고 제스처 인식 데이터베이스의 하나인 ChaLearn 데이터에 적용하여 인식 성능을 평가하였다. 그 결과 기존의 스마트 기기에서 가속도계, 자이로스코프 또는 터치 패널에 의한 제스처 인식이 아니라 영상 기반의 제스처 인식을 사용하여 호환성의 확보뿐만 아니라 사용자 인증 성능 또한 개선할 수 있음을 확인하였다.

1. 서론

최근 금융기관이나 포털사이트 및 온라인 서비스 업체에 대한 악의적인 해킹을 통해 개인정보가 유출되는 사고가 빈번히 발생하고 있다. 이에 따라 사회 전반에 걸쳐 정보보호에 대한 관심이 고조되고 있다. 더불어 최근에는 온라인을 통해 정보에 접근할 수 있는 다양한 스마트 기기의 출현으로 인해 정보보호를 위한 사용자 인증 시스템의 강화 요구가 증대되고 있다. 특히 정보보호를 위해 사용자 인증인 비밀번호, 키, 보안카드와 같은 인증방법은 노출이나 도난, 분실 등의 문제로 한계를 지니고 있다. 이에 반해 신체적 특성 또는 행동학적 특성 기반의 생체인식에 기반을 둔 인증방법은 기존 방법을 대체할 수단으로 기대를 모으고 있다[1].

얼굴 인식은 신체적 특성을 이용한 사용자 인증 방법으로 지문과 더불어 전통적으로 많이 사용되는 기술이다. 특

히 얼굴은 지문이나 홍채와 같이 전용의 장비 또는 인식 장비와의 접촉을 필요로 하지 않아 생체 정보 획득의 거부감이 없고, 웹캠이나 스마트 기기에 부착된 카메라를 활용할 수 있다는 장점이 있다. 하지만 인식의 정확도가 조명, 포즈, 표정, 폐색에 많은 영향을 받을 수 있으며, 허가된 사용자의 얼굴 사진을 이용하여 인증 과정을 우회할 수 있다는 단점이 있다[2].

제스처 인식은 행동학적 특징에 기반을 사용자 인증 방법으로 초기 걸음걸이 분석에서 발전하여 인간-컴퓨터 상호작용 수단으로 로봇이나 시스템의 제어를 위한 손 제스처, 수화 인식까지 확대되고 있는 기술이다. 최근에는 스마트 기기에 부착된 가속도계(accelerometer)와 자이로스코프(gyroscope) 또는 터치 패널을 이용한 사용자 인증 시스템이 개발되었다[3,4]. 이들 방법들은 각 기기별로 가속도계와 자이로스코프 또는 다중 접촉 가능한 터치 패널을 장착하고 있어야 하기 때문에 스마트 기기에 따라 사용자 인증에 대한 호환성 문제가 존재한다[1].

다양한 스마트 기기별로 호환성을 유지하면서도 사용자 인증을 수행하기 위해서는 각 기기에서 공통적으로 획득할 수 있는 정보를 필요로 한다. 최근 출시된 스마트폰, 태블릿PC, 스마트TV 등의 사례를 통해 대부분의 스마트

*** 교신 저자

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학IT연구센터육성 지원사업(NIPA-2012-H0301-12-2004) 과 교육과학기술부에서 지원하는 대구경북과학기술원 기관고유사업(12-IT-03)에 의해 수행되었음

기기에서 영상 센서를 기본으로 장착하고 있음을 확인할 수 있었다. 본 논문에서는 스마트 기기에 장착된 영상 센서를 통해 획득된 영상 정보를 분석하여 개인의 신체적 특징인 얼굴 정보와 행동학적 특징인 제스처 정보를 추출함으로써 스마트 기기에 따른 사용자 인증 시스템의 호환성을 확보함과 더불어 다중 생체 인식 방법을 구성함으로써 단일 생체 인식에 비해 보다 강건한 사용자 인증 시스템을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서 얼굴 정보와 제스처 정보를 함께 활용한 사용자 인증 시스템을 제안하고, 3장에서는 실험을 통해 제안된 사용자 인증 시스템의 성능을 분석하였다. 마지막 4장에서는 실험 결과를 토대로 본 연구의 의의와 향후 연구 과제에 대하여 논의하였다.

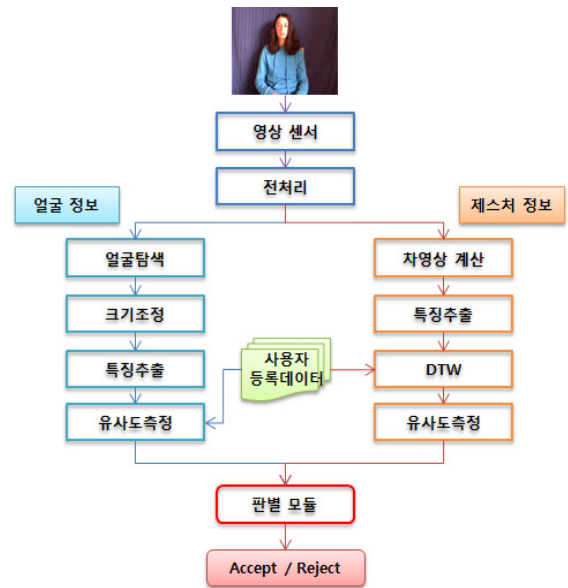
2. 제안한 사용자 인증 시스템

본 논문에서는 비디오 영상으로부터 앞서 언급된 방법으로 획득된 얼굴 정보와 제스처 정보를 함께 활용한 사용자 인증 방법을 제안한다. 그림 1에 제안한 사용자 인증 시스템의 구성을 나타내었다. 먼저 전처리 단계에서는 입력된 컬러 영상을 그레이 스케일 영상으로 변환한 후 얼굴 정보와 제스처 정보를 별도의 모듈로 나누어 처리하게 된다.

얼굴 정보 처리 모듈에서는 우선 비올라-존스 얼굴 검출기[5]를 사용하여 영상에서 얼굴을 탐색하고, 이로부터 특징 추출을 수행하기 위해 얼굴 영상의 크기를 32×32 로 조정하였다. 이후 얼굴 인식을 위해서는 먼저 조명 변화에 강건한 히스토그램 기반의 지역적 특징 기술자[6]를 추출하고, 강건한 얼굴 인식을 수행하기 위해 특징 기술자로 표현된 얼굴 데이터에 대해 통계적 학습을 수행하는 알고리즘을 적용한다[7]. 마지막으로 등록된 사용자 얼굴 데이터와의 유사도를 측정하여 최종 판별 모듈의 입력을 사용한다.

제스처 정보 처리 모듈에서는 입력되는 비디오 영상을 프레임 단위로 나누어 순차적으로 차영상을 계산하고, 이에 대해 통계적 특징 추출을 수행하는 알고리즘[8]을 적용한다. 이는 영상 획득 장치가 고정되어 있을 경우, 차영상에서는 움직임이 있는 부분에서의 변화가 그렇지 않은 부분에 비해 작게 나타나게 되며, 더불어 인식 과정에서의 연산 복잡도와 저장 공간의 효율을 위해 특징 추출이 필수적이기 때문이다. 본 연구에서는 각 프레임 단위의 영상에 대해 차영상을 계산한 후 주성분분석(principle component analysis)을 수행하여 특징 추출을 수행하였다. 마지막으로 길이가 다른 두 제스처의 유사도를 측정하기 위해서 동적 시간 정합(dynamic time warping)을 수행하여 제스처 데이터의 길이를 동일하게 조정된 후 상호 정보량(mutual information)에 기반을 둔 거리 척도를 통해 두 제스처 간의 유사도를 계산하여 최종 판별 모듈의 입력으로 사용한다.

판별 모듈에서는 얼굴 및 제스처 정보로부터 획득된 유



(그림 1) 제안한 사용자 인증 시스템

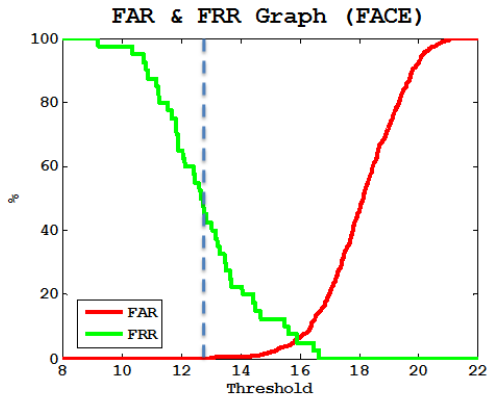
사도를 바탕으로 현재 입력된 영상의 사람이 등록된 사용자(genuine)인지 사칭자(impostor)인지를 판단한다. 입력된 얼굴 및 제스처 정보가 모두 각각의 문턱치(threshold)보다 낮을 경우 '수락(accept)'하여 사용자로 인증하게 되며, 두 개의 정보 중 어느 하나라도 문턱치보다 높은 경우 '거부(reject)'하게 된다.

3. 실험 결과

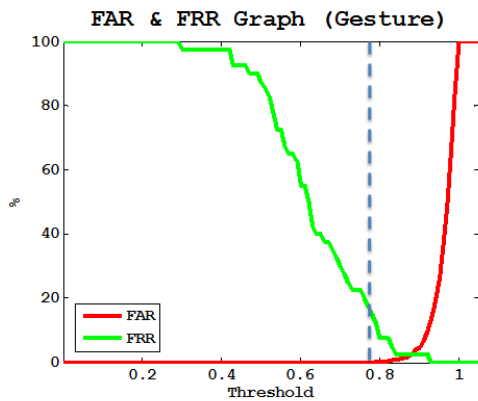
제안한 사용자 인증 시스템의 적절성을 확인하기 위해 제스처 인식을 위한 벤치마크 데이터인 ChaLearn 데이터를 사용하였다[9]. 이 데이터는 다수의 사용자로부터 획득된 컬러 및 깊이 영상으로 구성되어 있으며, 각 사용자마다 9개에서 13개의 제스처 정보를 가지고 있다. 실험에서는 이 중 얼굴이 다른 20명의 사용자에 대해 학습 및 테스트 데이터로 각 2개씩, 총 4개의 컬러 영상을 사용하였다.

이어서 테스트 단계에서는 20명 중 한명을 사용자라고 하고, 나머지 19명에 대해서는 사칭자로 간주하여 전체 20명에 대해 각 2개의 영상을 테스트 데이터로 사용하여 등록된 사용자 데이터와의 유사도를 측정하고 이의 분포를 관찰하였다. 먼저 학습 단계에서는 20명에 대한 40개 데이터에 대해 통계적인 특징 추출 과정을 수행하여 특징을 추출하기 위한 변환 행렬을 얻는 학습을 수행하였다.

그림 2는 얼굴과 제스처에 대해 개별적으로 판별 기준을 적용한 경우 FAR(False Accept Rate) 및 FRR(False Reject Rate) 그래프를 나타낸 것으로, 그림에서 수직선은 FAR이 0%가 되는 문턱치를 의미한다. 그림 2.(a)에서 얼굴에 대한 판정 기준이 12.28인 경우 FAR은 0%, FRR은 42.5%이며, 그림 2.(b)에서 제스처에 대한 판정기준이 0.78인 경우 FAR은 0%, FRR은 15%이다. FAR과 FRR은 일반적으로 판정 기준에 따라 변동되는 것으로 응용에 따라 조정이 가능한데, 실험에서는 보안이 엄격한 경우를 가정



(a) 얼굴에 대한 FAR/FRR 그래프



(b) 제스처에 대한 FAR/FRR 그래프

(그림 2) 얼굴 및 제스처에 대한 FAR/FRR 그래프

하고 FAR이 0%가 되도록 시스템을 설계하였다.

한편 얼굴 정보에 대해 FAR을 0%로 한 경우 FRR이 매우 높게 나타나는 문제가 발생하였다. 이는 얼굴 정보를 사용한 경우 분류 문제에는 적절하나, 인증 문제에서는 각 사람이 가지는 다양한 변형에 민감하기 때문에 발생하는 문제라 판단된다. 제스처 정보의 경우 FAR이 0%인 경우 15%의 결과를 보여주어 얼굴에 비해 인증 문제에 더 적합한 모습을 보여 주었다. 이는 각 사용자 별로 상이한 제스처를 사용함으로써 인해 클래스 간 분산이 클래스 내 분산에 비해 크게 발생하기 때문이라 판단된다.

표 1은 인증에 사용한 데이터에 따른 FAR/FRR 결과를 정리한 것으로 FAR이 0%가 되도록 문턱치를 설정한 경우의 FRR을 계산한 값으로, 얼굴과 제스처 정보로 개별적으로 사용한 경우와 두 정보를 함께 사용한 경우의 FRR을 나타내었다. 두 정보를 함께 사용한 경우 두 개의 정보 모두 '수락'인 경우에만 수락이며, 어느 하나의 정보라도 '거부'인 경우에는 최종 판별을 '거부'로 하여 FRR을 계산한 것으로 하나의 정보만 사용한 경우보다 FRR이 개선됨을 확인할 수 있었다.

4. 결론

본 논문에서는 영상 기반의 얼굴 및 제스처 정보를 활용하여 사용자 인증을 위한 다중 생체 인식 시스템을 제

<표 1> 인증에 사용한 데이터에 따른 FAR/FRR 비교

	얼굴	제스처	얼굴+제스처
FAR	0	0	0
FRR	42.5	15.0	10.0

안하였다. 스마트 기기에 공통적으로 장착된 영상 센서를 사용함으로써 다양한 스마트 기기에서 사용자 인증을 위한 호환성과 함께 다중 생체 정보를 활용함으로써 강건한 사용자 인증 시스템을 구성할 수 있음을 실험을 통해 확인하였다. 특히 기존에는 사용되지 않았던 영상 기반의 제스처 정보를 활용하여 사용자 인증을 수행할 수 있음을 보여주었으며, 이는 추후 보다 정교한 사용자 인증을 위해 제스처 정보가 충분히 활용될 수 있음을 보인 사례라 할 수 있겠다. 실제 응용을 위해서는 얼굴 정보와 제스처 정보가 가진 장점들을 결합하여 통합할 수 있는 방법과 인증의 정확성을 높이기 위한 정교한 판정 모듈에 대한 추가적인 연구가 필요하다고 생각된다.

참고문헌

[1] 이상민, “최근 인증기술 관련 현황”, 지급결제와 정보기술 제47호(2012년 1월), 금융결제연구소, 2012

[2] R. Gross, S. Baker, I. Matthews and T. Kanade, “Face Recognition Across Pose and Illumination”, Handbook of Face Recognition, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, 2004

[3] Dennis Guse, “Gesture-based User Authentication for Mobile Devices using Accelerometer and Gyroscope”, Master Thesis, Berlin Institute of Technology, 2011

[4] N. Sae-Bae, K. Ahmed, K. Isbister and N. Memon, “Biometric-Rich Gestures: A Novel Approach to Authentication on Multi-touch Devices”, Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, pp. 977-986

[5] P. Viola, M.J. Jones, “Robust Real-Time Face Detection”, Int. Journal of Computer Vision, Vol .57, No. 2, pp. 137-154, 2004

[6] N. Dala, B. Triggs, “Histograms of Oriented Gradients for Human Detection”, Int. Conf. on Computer Vision and Pattern Recognition, 2005

[7] J. Seo, H. Park, “A Robust Face Recognition through Statistical Learning of Local Features”, Int. Conf. on Neural Information Processing, 2011

[8] H. Choi, H. Park, “An Efficient Framework for Vision-Based Gesture Recognition”, Int. Conf. on Computer and Application, 2012

[9] ChaLearn Gesture Dataset (CGD 2011), ChaLearn, California, 2011