Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.6, Dec. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.6.1427

생체 정보와 다중 분류 모델을 이용한 암호학적 키 생성 방법*

이 현 석,[1†] 김 혜 진,[1] 양 대 헌,[1] 이 경 희[2‡]
[1]인하대학교, [2]수원대학교

Cryptographic Key Generation Method Using Biometrics and Multiple Classification Model*

Hyeonseok Lee,[1†] Hyejin Kim,[1] DaeHun Nyang,[1] KyungHee Lee[2‡]
[1]Inha University, [2]The University of Suwon

요 약

최근 생체 인증 시스템이 확대됨에 따라, 생체 정보를 이용하여 공개키 기반구조(Bio-PKI)에 적용하는 연구들이 진행 중이다. Bio-PKI 시스템에서는 공개키를 생성하기 위해 생체 정보로부터 암호학적 키를 생성하는 과정이 필요하다. 암호학적 키 생성 방법 중 특성 정보를 숫자로 정량화하는 기법은 데이터 손실을 유발하고 이로 인해 키 추출 성능이 저하된다. 이 논문에서는 다중 분류 모델을 이용하여 생체 정보를 분류한 결과를 이용하여 키를 생성하는 방법을 제안한다. 제안하는 기법은 특성 정보의 손실이 없어 높은 키 추출 성능을 보였고, 여러 개의 분류 모델을 이용하기 때문에 충분한 길이의 키를 생성한다.

ABSTRACT

While biometric authentication system has been in general use, research is ongoing to apply biometric data to public key infrastructure. It is a significant task to generate a cryptographic key from biometrics in setting up a public key of Bio-PKI. Methods for generating the key by quantization of feature vector can cause data loss and degrade the performance of key extraction. In this paper, we suggest a new method for generating a cryptographic key from classification results of biometric data using multiple classifying models. Our proposal does not cause data loss of feature vector so it showed better performance in key extraction. Also, it uses the multiple models to generate key blocks which produce sufficient length of the key.

Keywords: Biometrics, Cryptosystem, Key-binding scheme, Multiple classifiers

I. 서 론

생체 정보는 도용이 어렵고 분실의 위험이 없기 때문에 이와 관련된 연구가 활발히 진행되어왔다. 특히 생체 정보들 중, 경제적인 측면과 편리성 때문에 얼굴 이미지를 이용한 사용자 인증 방법이 최근 주목 받고 있다[1]. 이에 따라, 실생활에서도 얼굴 이미지를 이용한 인증 시스템이 확대되고 있다. 애플은 최근 얼굴 이미지를 이용한 사용자 인증을 도입하였고, 알리바바는 얼굴 이미지와 추가 입력 정보를 결

Received(07. 13. 2018), Modified(1st: 09. 14. 2018, 2nd: 10. 15. 2018), Accepted(10. 16. 2018)
* 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.B0717-16-0114, 비대면 본인확인을 위한 바이오 공개키 기반구조 기술 개발).
† 주저자, lhs@isrl.kr
‡ 교신저자, khlee@suwon.ac.kr(Corresponding author)

제 수단으로 사용하는 시스템을 출시하였다.

이처럼 생체 인증 시스템이 확대됨에 따라, 생체 정보를 공개키 기반 구조에도 적용하기 위한 연구들 또한 활발히 진행되고 있다[2]. 대표적으로 FIDO 연합(Fast IDentity Online Alliance)은 사용자 인증 프레임워크인 FIDO 프로토콜의 표준안을 제안하였다. FIDO 프로토콜은 비밀번호 분실, 해킹에 대한 위험성을 줄이고, 부인 방지 특성을 통해 온라인 환경에서 은행거래, 신용결제 등에 사용될 수 있다는 장점을 갖고 있다. FIDO 프로토콜은 공개키 기반구조로, 생체 정보로부터 암호학적 키를 생성하는 과정을 포함한다.

생체 정보로부터 추출한 특성 정보를 이용하여 암호학적 키를 생성하는 기법들은 대부분 특성 정보를 숫자로 정량화(quantization)하는 작업을 기반으로 한다. 하지만 이러한 방법은 최적화된 정량화 방법이 알려지지 않아 임의의 임계값을 기준으로 0 또는 1로 변환하는 경우가 많다. 이 과정 중에서 데이터가 과도하게 변환되기 때문에 좋은 성능을 기대하기 어렵다.

이 논문에서는 다중 분류 모델을 사용하여 각 분류기로부터 도출한 인덱스(index)를 기반으로 암호학적 키를 생성하는 방법을 제안한다. 이 기법은 위에서 언급한 정량화 과정을 거치지 않고 분류기 고유의 클래스 판정 방법을 이용함으로써 분류기의 성능을 최대로 이용할 수 있다.

논문의 구성은 다음과 같다. 2장에서는 관련 연구들을 정리하고 3장에서는 다중 분류기를 이용한 암호학적 키 생성 기법을 소개한다. 4장에서는 제안한 기법의 실험 결과를 통해 기존의 다른 키 생성 방법과의 성능을 비교하고, 마지막으로 5장에서는 결론을 내린다.

II. 기존 연구

2.1 생체 정보를 이용한 암호학적 키 생성 기법

생체 정보로부터 키를 생성하는 방법은 크게 키 바인딩(key-binding)과 키 생성(key-generation) 기법으로 나눌 수 있다[3].

키 바인딩 기법은 생체 정보와 임의로 생성한 비밀 키를 맵핑하는 특수한 병합 과정을 필수로 한다. 비밀 키와 생체 정보를 병합한 정보를 담고 있는 보조 데이터(helper data)를 생성한 후 인증 시 이 보조 데이터로부터 키를 생성한다. 따라서 비밀 키와 생체 정보 사이에 연관성은 존재 하지 않으며, 새로운 보조 데이터를 생성함으로써 키를 변경할 수 있다.

키 생성 기법은 보조 데이터를 생체 정보로부터 유도하여 생성하고, 이 보조 데이터와 생체 템플릿 정보로부터 비밀 키를 생성한다. 보조 데이터가 생체 정보로부터 생성되기 때문에 비밀 키는 생체 데이터와 밀접한 연관이 있다. 비밀 키를 변경하기 위해서는 생체 정보를 위한 취소 가능 기법들을 사용해야 한다.

키 바인딩의 대표적인 기법에는 fuzzy commitment[4]와 fuzzy vault[5]가 있고, 키 생성 기법으로는 fuzzy extractor[6]가 대표적이다. Y.Wang 등[7]은 얼굴의 특성 벡터와 랜덤한 벡터간의 거리 벡터를 이진 벡터로 표현하고, fuzzy vault 방법을 적용하여 키를 만들어내는 방법을 제안하였다. H.Lu 등[8]은 임의로 생성한 암호학적 키와 이진화 시킨 얼굴의 특성 벡터를 이용하여 보조 데이터를 만드는 fuzzy commitment 기반의 키 복원 방법을 제안하였다. 김혜진 등[9]은 얼굴 특성 정보와 PIN(Personal Identification Number)과 같은 사용자 입력정보로부터 취소 가능한 키를 생성하는 방법을 제안하였다. Z. Jin 등[10]은 fuzzy commitment를 대신하여 GHE(Graph-based Hamming Embedding) 지문 템플릿을 이용한 새로운 ECC-free 키 바인딩 기법을 제안하였다.

2.2 분류기

전통적인 얼굴 인식에 사용되었던 분류기는 PCA(Principal Component Analysis)이다. 이 방법은 데이터들 간의 분산 차이를 최대화하는 축을 찾아서 데이터들의 특성 벡터를 생성하고, 이를 비교하여 클래스를 구분한다. PCA를 기반으로 한 2-Dimensional PCA[11]는 벡터화 과정 없이 2차원 이미지를 직접 연산에 이용하여 적은 연산량으로 빠르게 특성 벡터들을 생성한다.

Local Feature Descriptor 방식은 이미지를 일정 크기의 구역으로 나누고 각 구역의 특성 정보를 추출하여 병합하는 방식으로 특성 정보를 추출한다. LBP(Local Binary Pattern)[12]은 각 픽셀들과 그 주변의 이웃 픽셀들의 값의 차이에 따라 특성 벡터를 생성한다. LBP는 얼굴 인식 분야를 포함한 이미지 인식 분야에서 좋은 성능을 보였고, 이를 기반

으로 한 다양한 응용 기법들이 제시되었다. 이 중, LFI(Local Edge/Corner Feature Integration)[13] 기법은 이미지에 광원의 영향을 최소화해주는 Frei-Chen 필터를 적용해 Edge 맵과 Corner 맵을 생성한다. 생성된 맵으로부터 LBP 방식의 특성 벡터를 만들어 병합하는 방법으로, 다양한 광원에 노출된 환경에서도 높은 인식률을 보여주었다.

III. 제안하는 기법

이 논문에서는 ITU-T X.1088[14] 표준을 참고하여 생체 정보와 비밀 데이터를 이용해 암호학적 키를 생성하는 프레임워크를 제안한다. [14]에서 비밀 데이터에 해당하는 요소로는 사용자의 PIN을 이용하였고 키 생성 요소 중 하나로 사용된다. 제안하는 프레임워크는 대칭 키를 생성하는 단계까지의 과정을 포함하고 있으며, 필요에 따라 [14]와 같이 PKI(Public Key Infrastructure) 과정을 추가하여 사용할 수 있다.

이 논문에서는 사용자를 등록하는 과정에서 여러 개의 모델을 생성하는 방법을 제안한다. 기존 연구와 달리 다중 모델을 사용하는 이유는 충분한 길이의 키를 생성하고, 보다 안정적으로 키를 생성하기 위함이다. 하나의 모델을 사용할 경우 제한된 길이의 키를 생성할 수밖에 없지만, 다중 모델을 사용함으로써 각각의 모델로부터 추출한 문자열을 연결하여 충분한 길이의 키를 생성할 수 있다. 또한, 여러 개의 모델 중 일부 모델에서 오분류 하더라도 정확히 분류한 다수의 모델 분류 결과로부터 본래의 키를 복구할 수 있다.

Fig. 1.은 생체 정보를 이용하여 사용자를 등록하기 위해 키를 생성하는 과정과 사용자를 인증하기 위해 키를 복구하는 과정을 도식화한 그림이다. 먼저 등록 과정에서 입력한 사용자의 생체 정보를 전처리 과정을 통해 가공한 후, 모델을 만들기 전에 각 모델로부터 분류될 결과인 인덱스를 생성한다. 생성한 인덱스를 이용하여 다중 모델을 만들고, 인덱스를 연결하여 인덱스 문자열을 생성한다. 인덱스 문자열을 ECC(Error Correcting Code) 인코딩(encoding)하여 ECC를 생성하고 모델과 함께 저장한다. 마지막으로 인덱스 문자열과 입력한 PIN을 해시하여 키를 생성한다.

인증 과정은 등록 과정과 유사한 과정을 거친다. 인증하고자 하는 사용자의 생체 정보(Biometrics')를 전처리 후, 모델에 입력하여 분류 결과를 추출한다. 추출한 인덱스 문자열을 저장한 ECC를 이용하여 디코딩(decoding)한 후 인증과정에서 입력한 PIN(PIN')과 해시하여 키를 복원한다.

3.1 전처리

생체 정보는 본질적으로 노이즈가 있기 때문에 수집할 때마다 동일한 데이터를 얻기 힘들다[3]. 수집한 데이터의 노이즈가 키 생성 과정에서 미치는 영향을 줄이고 분류기의 성능을 높이기 위해 등록자의 생체 정보를 가공한다. 생체 정보의 종류와 분류기의 특성에 따라 이미지 전처리 과정에 다양한 기법이 사용될 수 있다. 이 논문에서는 다양한 생체 정보 중 얼굴 이미지를 이용하여 실험하였고, 전처리 기법으로 이목구비 중심의 이미지 재단(cropping) 및 히스토그램 균일화(histogram equalization), 이미지 감마 보정(Gamma correction), 가우시안 필터(Gaussian filter) 등 다양한 필터를 적용하였다.

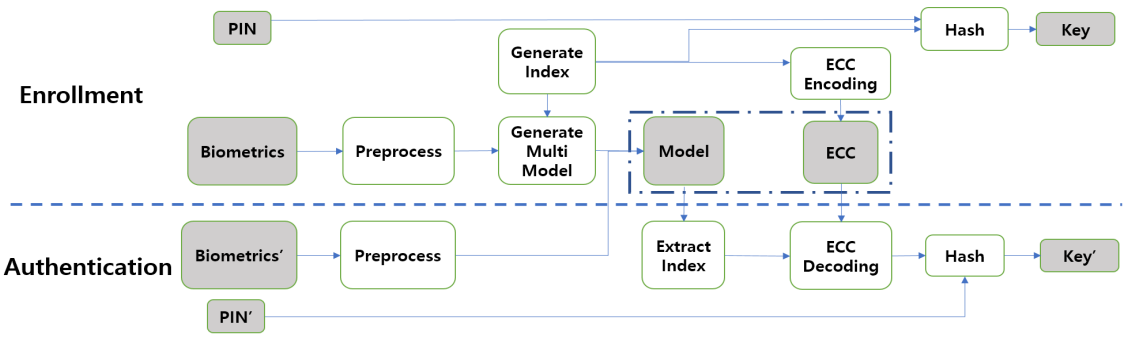

Fig. 1. Flow of proposed method

3.2 인덱스 생성

2.1에서 소개한 키 바인딩 기법을 이용하기 위해 모델 생성 전 인덱스를 생성한다. 인덱스란 모델에 입력한 클래스들을 분류한 결과로, 각 클래스를 분류한 결과를 해당 클래스의 인덱스라 한다. 생성하고자 하는 모델의 수를 m, 하나의 모델에서 분류 가능한 클래스 수를 c라 했을 때, 각 모델에 대하여 등록하고자 하는 사용자(genuine user)의 인덱스를 0에서 c-1의 범위에서 임의로 선택하여 총 m개의 인덱스를 생성한다. 임의로 생성한 인덱스와 사용자의 생체 정보 사이엔 연관성이 존재하지 않으므로, 키가 노출된 경우에도 인덱스를 재생성함으로써 새로운 키를 만들 수 있다.

m개의 인덱스를 이용하여 모델을 만든 후, ECC와 키를 생성하기 위해 인덱스를 이용하여 인덱스 문자열을 생성한다. 인덱스 문자열이란 m개의 인덱스를 비트 형태로 바꾸어 하나로 연결한 문자열이다. 인덱스 하나의 길이가 $\log_2 c$이므로 m개를 연결한 인덱스 문자열의 길이는 $m \times \log_2 c$이다. 모델 수 m이 3, 클래스 수 c가 8인 경우를 예를 들면, 임의로 선택한 3개의 인덱스를 각각 7, 1, 0이라 가정했을 때 이 3개의 인덱스를 이진수 111_2, 001_2, 000_2로 변환하고, 이진수들을 연결하여 인덱스 문자열 111001000_2을 생성한다.

3.3 다중 모델 생성

Fig. 2.는 사용자 등록 과정 중 모델 생성 과정에 관한 그림이다. 전체 사용자 수를 n이라 가정했을 때 모델 생성 과정은 다음과 같다. 하나의 모델을 구성하기 위해, 등록하고자 하는 사용자의 생체 정보와 전체 사용자 n명 중 임의로 선택한 c-1명의 생체 정보를 선택한다. 선택한 총 c명의 생체 정보를 이용하여 모델에서 사용하는 분류 기법에 맞는 특성 정보를 생성한다. i번째 모델 M_i 생성 시 등록하고자 하는 사용자의 생체 정보를 입력했을 때, 3.2에서 설명한 바에 따라 미리 생성한 m개의 인덱스 중 i번째 인덱스 idx_1^i가 분류 결과로 도출되도록 모델을 구성한다. 다른 사용자의 생체 정보를 입력 했을 때엔 idx_1^i를 제외한 0에서 c-1 범위의 인덱스가 분류 결과가 되도록 모델을 구성한다. 예를 들어 Fig. 2.에서 m=3, c=8이고 이전 단계에서 생성된 인덱스가 순차적으로 7, 1, 0이라 하면, M_2에서 등록하고자 하는 사용자의 idx_1^2는 1이다. 따라서 등록하고자 하는 사용자의 특성 정보는 모델의 1번째 클래스에 위치하고 다른 사용자들의 특성 정보는 1을 제외한 나머지 클래스에 위치한다. 모델 생성 시 사용되는 다른 사용자들과 인덱스를 임의로 선택하므로, 다른 사용자들의 조합과 인덱스가 동일한 모델은 생성되지 않는다.

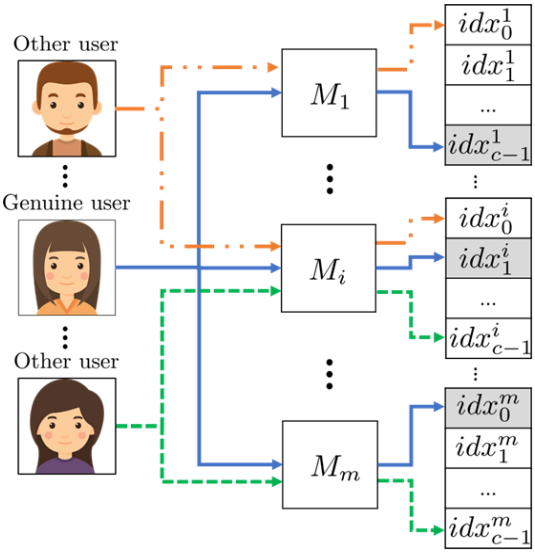

Fig. 2. Process of generating multiple classifier model

하지만 등록하고자 하는 사용자의 생체 정보가 모든 모델에 공통적으로 존재하여 모델이 노출되었을 시 중복된 생체정보로부터 인덱스를 유추할 수 있다. 이를 방지하기 위해 강전일 등[15]이 제시한 GPT 기법을 이용하여 모델을 보호하였다. GPT 기법은 특성 벡터를 추출하기 위한 투영 행렬(projection matrix)을 보호하기 위한 기법이다. 이 기법은 투영 행렬의 행들을 임의로 치환하여 본래의 행렬을 쉽게 복원하지 못하도록 함으로써 기존 행렬을 보호한다. 이 논문에서는 DiaPCA 분류기와 LFI 분류기를 사용하여 실험하였는데 DiaPCA의 경우 투영 행렬에 GPT를 적용하였고, LFI의 경우 기법의 특성상 투영 행렬이 존재하지 않아 생체 정보가 특성 벡터에 반영이 되기 때문에 특성 벡터를 행렬 형태로 만들어서 GPT를 적용하였다.

모델 생성 시 실험에 사용한 2가지 분류기 이외에

기존의 전통적인 기계학습 방식의 분류기와 최근 높은 성능으로 주목 받고 있는 인공 신경망 기반의 딥러닝(deep learning) 등 다양한 분류기를 사용할 수 있다.

3.4 인덱스 추출

인덱스 추출은 사용자 인증 과정에서 인증하고자 하는 사용자의 생체 정보를 전처리 후 모델에 입력하여 각 모델로부터 인덱스를 추출하는 과정이다. 모델 생성 과정과 마찬가지로 인증하고자하는 사용자의 생체 정보를 이용하여 모델에서 사용하는 분류 기법에 맞는 특성 정보를 생성한다. 이렇게 생성한 특성 정보와 모델에 포함된 c명의 특성 정보들 간의 거리를 비교하여 가장 유사한 클래스의 인덱스를 추출한다.

특성 정보간의 거리 비교 방법은 분류 기법에 적합한 비교 방법을 사용할 수 있으며, 이 논문의 실험에서는 LFI 특성 정보의 경우 카이-제곱 거리(Chi-square distance) 비교, DiaPCA 특성 정보의 경우 맨해튼 거리(Manhattan distance) 비교 방법을 적용하였다.

3.5 해시

인덱스 문자열을 입력한 PIN과 함께 해시하여 키를 생성한다. 다중 모델을 사용함으로써 키의 엔트로피가 증가하지만, 모델의 개수가 늘어남에 따라 사용자를 등록하는 과정에서 소요되는 시간 역시 증가하므로 사용자 편의성이 저하된다. 따라서 충분한 안정성을 확보할 때까지 모델 수를 늘리기엔 한계가 있다. 이를 보완하기 위해 생체 정보와 PIN을 이용하는 2요소(two-factor) 방식을 이용하였다. 이 논문에서는 SHA-256 해시 함수를 사용하였고, 해시 결과로 생성된 키 길이는 256비트이다.

3.6 ECC 인코딩 & 디코딩

앞서 서술한 바와 같이 생체 정보는 일정하지 않은 데이터이므로 인증 시 생체 정보 오분류로 인해 등록했던 인덱스 문자열과는 다른 인덱스 문자열이 생성될 가능성이 존재한다. 이러한 경우를 보완하기 위해 ECC를 사용하여 동일한 인덱스 문자열이 만들어지도록 보정한다. 등록 시 ECC 인코딩 과정을 통해 인덱스 문자열에 대한 ECC를 생성하고 인증 시 ECC를 이용하여 디코딩 후 인덱스 문자열을 복구한다. 이 논문에서는 ECC 중 하나인 BCH 코드를 사용하였다.

Table 1. BCH code parameters

Index string bit	Message bit(k)	Recoverable Message bit(t)
16	16	3
24	24	7
32	36	5
40	45	3
48	50	13
64	64	10
80	85	6
96	99	23
128	131	18
160	163	12
192	193	43
256	259	30
320	322	22
384	393	79

BCH 코드에선 사용가능한 파라미터가 고정되어 있기 때문에 정해진 길이의 메시지(k)만 인코딩 할 수 있다. 메시지의 길이와 인덱스 문자열의 길이가 일치하지 않는 경우 인덱스 문자열에 0을 추가로 패딩(padding)하여 인코딩 할 수 있는 메시지 비트로 만들었다. 복구 가능한 메시지 비트(t)는 메시지의 전체 비트 중 복원할 수 있는 비트의 크기를 나타낸다. 실험에 사용한 BCH 코드의 파라미터는 Table 1에서 확인할 수 있다.

IV. 실험 결과

실험은 다음과 같은 환경에서 수행하였다. 운영체제는 Ubuntu 16.04, CPU는 i5-4570 중 2개의 코어만을 사용하였고, 메모리는 6GB, 프로그래밍 언어는 C++11을 사용하였다.

기존에 연구되었던 기법[9]는 광원과 포즈의 변화가 적은 Essex Faces94 데이터베이스에서 좋은 성능을 보였지만, 다양한 포즈나 광원이 반영된 데이터베이스에 해당 기법을 적용한 실험 내용이 존재하지 않는다. 이에 Extended Yale B[16][17] 데이터베이스를 사용하여 다양한 광원에 노출된 환경에서

기존의 기법과 제시한 기법의 성능을 비교하였다. 또한, 다양한 인종 및 다수의 클래스로 구성된 데이터베이스에서의 성능을 실험해보기 위해 MUCT[18] 데이터베이스를 이용하여 실험을 진행하였다. 각 실험은 무작위성의 영향을 줄이기 위해 동일한 모델 수와 클래스 수에서 10번 반복 수행하여 평균값을 기재하였다.

[9]논문에서 분석한 공격 타입은 총 3가지로 PIN과 얼굴 이미지를 임의로 입력하는 공격, 노출된 PIN과 임의의 얼굴 이미지를 입력하는 공격, 노출된 얼굴 이미지와 임의의 PIN을 입력하는 공격이다. 이 기법은 PIN이 틀렸을 경우 사실상 FAR(False Acceptance Rate)이 0에 수렴하기 때문에 PIN이 노출된 상태에서 얼굴 위조자 공격에 대한 실험만을 수행하였다.

4.1 분류기에 따른 성능 분석

Table 2와 Table 3은 Extended Yale B 데이터베이스에서 LFI 분류기와 DiaPCA 분류기를 사용했을 때의 FRR(False Rejection Rate), FAR, 등록 시간 및 인증 시간을 나타낸다. 등록 및 인증 시간은 전처리 과정을 제외한 시간만을 기록하였다.

Table 2. Performance of the proposed scheme using LFI classifiers with Extended Yale B database

Model number	Class number	FRR (%)	FAR (%)	Enrollment time (sec)	Authentication time (sec)
1	4	0.26	75.3	0.028	0.002
	8	0.00	47.8	0.039	0.004
	16	0.00	30.5	0.055	0.007
	32	2.89	48.4	0.111	0.015
8	4	0.00	6.49	0.064	0.018
	8	0.00	5.92	0.107	0.033
	16	0.00	0.13	0.289	0.059
	32	3.16	0.00	0.713	0.113
16	4	0.00	3.23	0.102	0.035
	8	0.00	3.11	0.216	0.074
	16	0.00	0.04	0.546	0.119
	32	6.58	0.00	1.322	0.222
32	4	0.00	2.34	0.166	0.071
	8	0.00	1.79	0.302	0.124
	16	0.00	0.00	0.987	0.233
	32	0.79	0.00	2.619	0.445
64	4	0.00	1.28	0.307	0.137
	8	0.00	0.94	0.647	0.261
	16	0.00	0.00	2.044	0.484
	32	2.37	0.00	5.426	0.908

Table 3. Performance of the proposed scheme using DiaPCA classifiers with Extended Yale B database

Model number	Class number	FRR (%)	FAR (%)	Enrollment time (sec)	Authentication time (sec)
1	4	1.05	73.94	0.047	0.000
	8	2.63	47.43	0.057	0.000
	16	3.42	29.43	0.075	0.000
	32	3.16	49.52	0.122	0.000
8	4	0.26	11.42	0.132	0.002
	8	1.84	8.69	0.197	0.002
	16	4.74	0.83	0.349	0.003
	32	10.00	0.14	0.643	0.003
16	4	1.32	6.51	0.200	0.004
	8	1.58	6.75	0.354	0.005
	16	6.05	0.48	0.651	0.005
	32	9.21	0.12	1.244	0.007
32	4	0.79	5.28	0.382	0.008
	8	2.89	4.25	0.695	0.012
	16	6.58	0.23	1.309	0.012
	32	8.95	0.07	2.481	0.014
64	4	0.79	4.06	0.779	0.017
	8	1.05	3.83	1.444	0.034
	16	8.95	0.14	2.683	0.029
	32	7.89	0.06	5.147	0.031

데이터베이스 중에서 과도한 광원에 노출된 이미지는 제외하였고 각 사용자 당 27장의 이미지를 선별하여 사용하였다. 본 실험에 사용한 이미지의 크기는 128 × 128 픽셀(pixel)이고 훈련용 이미지는 각 사용자의 이미지 27장 중 임의로 10장을 선택하였다. 테스트용 이미지는 훈련용 이미지로 사용하지 않은 이미지 중 임의로 1장을 선택하여 수행하였다. LFI의 파라미터로 각 셀의 크기는 12 × 12 픽셀, 패딩의 크기는 상하좌우 2 픽셀로 설정하였다. DiaPCA의 파라미터로는 특성 정보 생성 시 선택하는 고유벡터(eigenvector)의 수를 8로 설정하였다. 모델 수 m이 8, 각 모델 당 클래스 수가 4일 때 LFI 분류기와 DiaPCA 분류기를 사용했을 때의 FAR은 각각 6.49%, 11.42%로 높은 수준이지만 클래스 수가 증가할수록 FAR이 낮아지는 것을 확인할 수 있다. 이는 모델의 클래스 수가 많아질수록 무작위하게 한 클래스를 선택했을 때 등록한 사용자일 확률이 줄어들기 때문이다. 클래스 수가 늘어날수록

FAR은 감소하는 반면, 일정 시점 이후로 FRR이 증가한다. 클래스 수가 많아질수록 다른 사용자의 생체 정보 중 등록하고자 하는 사용자의 생체 정보와 비슷한 극단 값(outlier)이 존재할 확률이 많아지기 때문이다. 이 실험에서는 클래스 수가 16일 때 가장 좋은 성능을 보였다.

Table 2와 Table 3을 비교하면 상대적으로 LFI 분류기를 사용했을 때의 성능이 DiaPCA 분류기를 사용했을 때 보다 성능이 좋다는 것을 알 수 있다. 이는 [13]에 의하면 LFI 분류기가 다양한 광원에 노출된 환경에서 좋은 성능을 보이기 때문이다. 반면 DiaPCA 분류기를 사용했을 때 등록 및 인증 과정에서 소요되는 시간이 더 적다는 것을 알 수 있다. 이는 LFI 분류기의 경우, 얼굴 전체 이미지를 일정 크기의 여러 이미지로 나누어 각각으로부터 특성 정보를 생성하고 이를 병합하여 최종 특성 정보를 만들기 때문에 연산량이 많은 반면, DiaPCA 분류기는 이미지를 나누지 않고 얼굴 이미지로부터 직접 특성 정보를 생성하기 때문에 상대적으로 연산량이 적기 때문이다.

4.2 기존 논문과의 성능 비교

김혜진 등[9]은 단일 DiaPCA 분류기를 이용하여 특성 벡터를 생성하고 이를 정량화 과정을 통해 키로 사용하는 방법을 제시하였다. 이 방법은 특성 벡터의 값들을 0과 1로 정량화한 뒤, BCH code를 이용해 오류를 보정하는 방법으로, 2.1에서 설명한 키 생성 방법에 해당한다. Fig. 3. (a)는 DiaPCA 분류기를 [9]논문에서 제시한 기법에 적용했을 때의 성능과, DiaPCA 및 LFI 분류기를 이 논문에서 제시한 기법에 적용했을 때의 성능을 ROC(Receiver Operating Characteristic) 곡선을 이용해 비교한 그림이다. (b)는 (a)의 데이터를 가시적으로 확인하기 위해 로그 스케일(scale)로 그린 그래프이다. 제시한 기법을 사용한 실험은 Table 2, Table 3의 모델 수 32의 자료를 기반으로 그래프로 나타내었고, 각 자료에서의 클래스 수는 라벨에 표시하였다.

[9]논문에서 제시한 기법은 논문에 따르면 FACE94 데이터베이스에서 높은 성능을 보였다. 하지만 동일 기법을 Extended Yale B 데이터베이스에 적용했을 때엔 Fig. 3. (a)에서 확인할 수 있듯이 낮은 성능을 보인다. 이는 [9]논문에서 사용한 DiaPCA 분류기가 다양한 광원에서 낮은 분류 성능을 보이기 때문이다. 반면 동일한 DiaPCA 분류기를 사용했음에도 제시한 기법으로 구현했을 경우 높은 성능을 보인다. 이는 앞서 기술했다시피 이 논문에서 제안하는 기법은 [9]논문에서 제안하는 기법과는 달리, 정량화 과정에서 발생할 수 있는 문제가 발생할 여지가 없고, 여러 모델을 사용함으로써 일부 모델에서 발생할 수 있는 오류를 복구할 수 있기 때문이다.

Fig. 3. (b) 그래프에 의하면 제시한 기법을 사용한 경우 LFI 분류기를 적용했을 때의 성능이 DiaPCA 분류기를 적용했을 때의 성능보다 우수하다. 이는 LFI 분류기가 다양한 광원에서 더 좋은 성능을 보이기 때문이다.

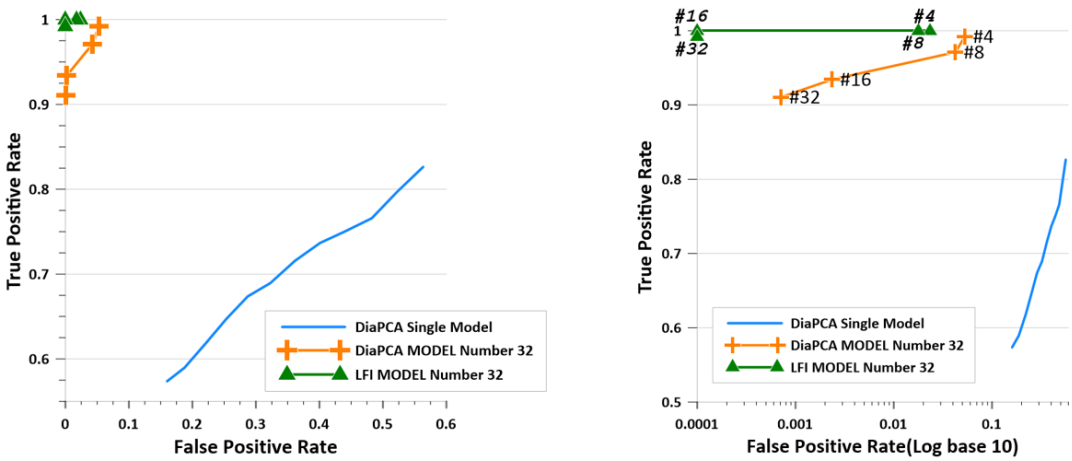

Fig. 3. ROC curve of 3 models (a) linear base and (b) log base 10

4.3 MUCT 데이터베이스 실험

제안한 기법을 다양한 인종 및 다수의 클래스로 구성된 데이터베이스에서의 성능을 실험해보기 위해 MUCT 데이터베이스를 이용하여 실험을 진행하였다. Fig. 6.를 통해 확인할 수 있듯이 Extended Yale B 데이터베이스와 비교하여 다양한 포즈로 구성되어 있다. 분류기는 LFI를 사용하였고, LFI의 파라미터로 각 셀의 크기는 6 × 6 픽셀, 패딩의 크기는 상하좌우 1 픽셀로 설정하였다.

각 사용자 당 이미지는 15장이고 이미지 크기는 64 × 64 픽셀이다. 훈련용 이미지는 각 사용자의 이미지 15장 중 임의로 10장을 선택하였다. 테스트용 이미지는 훈련용 이미지로 사용하지 않은 이미지 중 임의로 1장을 선택하여 수행하였다.

Table 4에서 보듯이 Extended Yale B 데이터베이스에 적용하였을 때와 비교하였을 때 전체적으로 FAR, FRR, 등록 및 인증 시간에서 비슷한 양상을 보였다.

Table 4. Performance of the proposed scheme using LFI classifiers with MUCT database

Model number	Class number	FRR (%)	FAR (%)	Enrollment time (sec)	Authentication time (sec)
1	4	0.10	74.4	0.079	0.002
	8	1.36	48.6	0.075	0.004
	16	1.06	30.3	0.100	0.007
	32	2.06	49.6	0.167	0.014
8	4	0.00	7.36	0.112	0.017
	8	0.00	7.62	0.150	0.030
	16	1.06	0.51	0.360	0.058
	32	6.73	0.02	0.836	0.111
16	4	0.00	3.25	0.146	0.033
	8	0.00	4.17	0.219	0.060
	16	0.30	0.21	0.613	0.114
	32	4.57	0.00	1.561	0.220
32	4	0.00	2.29	0.221	0.067
	8	0.00	2.19	0.372	0.123
	16	0.05	0.08	1.166	0.229
	32	2.01	0.00	3.030	0.437

4.4 저장 공간 및 사용성 분석

다중 모델을 사용함으로써 3장에서 언급한 이점을 얻을 수 있는 반면, 단일 모델을 사용하였을 때와 비교하여 저장 공간과 등록 및 인증에 소요되는 시간에서 오버헤드(overhead)가 발생한다.

본 실험에서 가장 좋은 성능을 보였던 모델 수 16, 클래스 수 16을 사용하였을 때 LFI 분류기의 경우 29MB, PCA 분류기의 경우 10MB의 저장 공간이 소요되었다. 가장 많은 저장 공간을 필요로 하였던 모델 수 64, 클래스 수 32을 사용하였을 때에도 LFI, PCA 분류기의 경우 각각 118MB, 83MB로 현재 GB단위의 저장 공간 크기를 감안하면 허용할 수 있는 크기이다.

Table 2, Table 3 및 Table 4에서 보듯이, 모델의 수가 늘어날수록 사용자 등록 및 인증에 소요되는 시간 또한 모델 수에 비례하여 증가하지만 인증 시간의 경우 연산량이 많은 LFI 분류기를 사용했을 경우에도 1초를 넘기지 않았다. 다만 등록 시간의 경우 Table 2에서 확인할 수 있듯이 모델 수 64, 클래스 수 32인 경우 5초를 초과하였다. 이처럼 너무 많은 모델을 사용할 경우 사용성이 저하될 수 있으므로 사용자 편의성을 고려한 모델 수 선택이 필요하다.

4.5 오분류 이미지 분석

Fig. 4.는 모델 분류 과정에서 오분류를 일으키는 이미지들의 예시이다. 오분류의 원인으로 각도, 빛, 표정 등이 존재하는데 LFI 분류 기법은 각도의 변화에 취약하다. 분류 기법의 특성상 Fig. 5.에서 보듯이 각도의 변화에 따라 얼굴 특징 점의 위치가 변화하게 되고, 이로 인해 이미지 간의 거리 비교 시 동일 인물의 이미지임에도 상당한 차이를 유발하기 때문이다. PCA 분류 기법의 경우에도 PCA 기법을 이용하여 생성되는 Eigenface 특징들이 다수의 이미지들을 통해 정면 얼굴 중심으로 특징이 생성되기 때문에 다양한 포즈에 취약하다. Fig. 6.은 실험에 사용된 특정 사용자의 이미지 중 오분류를 유발하는

Fig. 4. Examples of images that cause misclassification

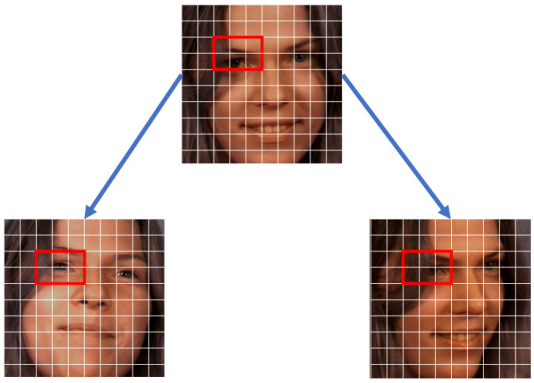

Fig. 5. Changes of facial features by poses

(a) (b)

Fig. 6. Images with (a) good performance and (b) relatively poor performance by poses

이미지와 그렇지 않은 이미지에 대한 예시이다.

4.6 보안성 분석

분류기를 1개만을 이용했을 경우 인덱스 문자열의 엔트로피는 $\log_2 c$ 로 매우 낮다. 반면 m개의 모델을 이용하여 엔트로피를 $m \times \log_2 c$ 으로 증가 시킬 수 있다. 인덱스 문자열을 해시하여 최종적으로 얻게 되는 키의 길이는 256비트지만, 실질적인 키의 엔트로피는 인덱스 엔트로피와 PIN의 자리 수 d에 의해 결정된다. PIN으로부터 얻을 수 있는 엔트로피는 $\log_2 10^d$ 이므로 최종 키의 엔트로피는 $m \times \log_2 c \times \log_2 10^d$ 이다.

공격자가 PIN을 획득해 무작위한 얼굴 이미지를 이용해 공격할 경우 PIN의 엔트로피를 제외한 인덱스 문자열의 엔트로피 $m \times \log_2 c$ 가 키의 엔트로피가 된다.

공격자가 얼굴 이미지를 획득하여 무작위한 PIN을 이용해 공격할 경우 올바른 PIN을 입력하면 정상적으로 복원된 모델로부터 인덱스 문자열을 생성할 수 있어 키를 추출할 수 있다. 따라서 키의 엔트로피는 PIN의 엔트로피와 같은 $\log_2 10^d$ 이다.

공격자가 ECC를 획득했을 경우에 대한 분석은 다음과 같다. ECC가 복구 가능한 비트(t)보다 2배 만큼의 정보를 내포하고 있어 2t만큼 키의 엔트로피가 감소한다. 예를 들어 인덱스 문자열의 길이가 256비트인 경우 Table 1에 따르면, BCH 코드의 복구 가능한 비트가 30비트이므로 키의 안정성은 256비트에서 60비트 감소한 196비트이다.

V. 결론 및 향후 연구

실험 결과 하나의 분류기만을 사용하여 키를 생성했을 때보다 여러 개의 분류기를 사용하였을 때 더욱 안정적인 키를 생성한다는 것을 확인할 수 있었다. [9]논문과 비교했을 때 동일한 분류기를 사용하였음에도 더 높은 성능을 보였고, 다양한 광원에 노출된 환경에서 좋은 분류 성능을 보이는 분류기를 사용함으로써 해당 환경에서 더욱 안정적인 키를 생성할 수 있었다.

실험에 사용한 데이터베이스가 다양한 광원 환경을 반영하고 있지만 포즈의 변화는 크지 않다. 향후 다양한 얼굴각도, 표정변화, 조도변화, 콘텍트렌즈 및 헤드벤드 착용 등이 반영된 KISA K-NBTC 데이터베이스에 대한 후속 연구를 진행할 예정이다.

References

[1] "Science, ICT Policy and Technology Trends," Korea Institute of S&T Evaluation and Planning, no. 84, Dec. 2016.

[2] Walter Scheirer, William Bishop, and Terrance Boult, "Beyond pki: The biocryptographic key infrastructure," IEEE International Workshop on Information Forensics and Security, Dec. 2010.

[3] Christian Rathgeb and Andreas Uhl, "A survey on biometric cryptosystems and cancelable biometrics," EURASIP

Journal on Information Security, vol. 5, no. 1, Dec. 2011.
〔4〕 Ari Juels and Martin Wattenberg, "A fuzzy commitment scheme," Proceedings of the 6th ACM conference on Computer and communications security, pp. 28-36, Nov. 1999.
〔5〕 Ari Juels and Madhu Sudan, "A fuzzy vault scheme," Designs, Codes and Cryptography, vol. 38, no. 2, pp. 237-257, Feb. 2006.
〔6〕 Yevgeniy Dodis, Leonid Reyzin and Adam Smith, "Fuzzy extractor: how to generate strong keys from biometrics and other noisy data," Advances in Cryptology - EUROCRYPT 2004, LNCS 3027, pp. 523-540, 2004.
〔7〕 Yongjin Wang and Konstantinos N. Plataniotis, "Fuzzy vault for face based cryptographic key generation," IEEE 2007 Biometrics Symposium, Sep. 2007.
〔8〕 Haiping Lu, Karl Martin, Francis Bui, Konstantinos N. Plataniotis and Dimitris Hatzinakos, "Face recognition with biometric encryption for privacy-enhancing self-exclusion," IEEE 2009 16th International Conference on Digital Signal Processing, July 2009.
〔9〕 Kim Hyejin, Choi Jinchun, Jung Changhun, Nyang Daehun and Lee KyungHee, "A method for generating robust key from face image and user intervention," Journal of the Korea Institute of Information Security and Cryptography, 27(5), pp. 1059-1068, Oct. 2017.
〔10〕 Zhe Jin, Andrew Beng Jin Teoh, Bok-Min Goi and Yong-Haur Tay "Biometric cryptosystems: A new biometric key binding and its implementation for fingerprint minutiae-based representation," Pattern Recognition, vol. 56, no. 8, pp. 50-62, 2016.
〔11〕 Daoqiang Zhang, Zhi-Hua Zhou and Songcan Chen, "Diagonal principal component analysis for face recognition," Pattern recognition, vol. 39, no. 1, pp. 140-142, Jan. 2006.
〔12〕 Timo Ahonen, Abdenour Hadid and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.
〔13〕 Almabrok Essa and Vijayan Asari "Local edge/corner feature integration for illumination invariant face recognition," VISUAL 2016 : The First International Conference on Applications and Systems of Visual Paradigms, pp. 13-18, Nov. 2016.
〔14〕 "Telebiometrics digital key framework (TDK) - A framework for biometric digital key generation and protection," ITU-T X.1088, May 2008.
〔15〕 Kang Jeonil, Nyang Daehun and Lee Kyunghee, "Two-factor face authentication using matrix permutation transformation and a user password," Information Sciences, 269(10), pp. 1-20, Jun. 2014.
〔16〕 Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643 - 660, Jun. 2001.
〔17〕 Kuang-Chih Lee, Jeffrey Ho, and David J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684 -

698, May. 2005.
[18] Stephen Milborrow, John Morkel, and Fred Nicolls, "The MUCT landmarked face database," Pattern Recognition Association of South Africa, pp. 32-34, Nov. 2010.

〈저자소개〉

이 현 석 (Hyeonseok Lee) 학생회원
2017년 2월: 인하대학교 컴퓨터정보공학과 졸업
2017년 3월~현재: 인하대학교 컴퓨터공학과 석사과정
〈관심분야〉 생체인증, 인공지능

김 혜 진 (Hyejin Kim) 학생회원
2016년 2월: 인하대학교 컴퓨터정보공학과 졸업
2018년 2월: 인하대학교 컴퓨터공학과 석사
2018년 3월~현재: 인하대학교 컴퓨터공학과 박사과정
〈관심분야〉 암호이론, 생체인증, 네트워크 보안

양 대 헌 (DaeHun Nyang) 종신회원
1994년 2월: 한국과학기술원 과학기술대학 전기 및 전자공학과 졸업
1996년 2월: 연세대학교 컴퓨터과학과 석사
2000년 8월: 연세대학교 컴퓨터과학과 박사
2000년 9월~2003년 2월: 한국전자통신연구원 정보보호연구본부 선임연구원
2003년 2월~현재: 인하대학교 컴퓨터공학과 교수
〈관심분야〉 암호이론, 암호 프로토콜, 인증 프로토콜, 무선 인터넷 보안

이 경 희 (KyungHee Lee) 정회원
1993년 2월: 연세대학교 컴퓨터과학과 졸업
1998년 8월: 연세대학교 컴퓨터과학과 석사
2004년 2월: 연세대학교 컴퓨터과학과 박사
1993년 3월~1996년 5월: LG소프트(주) 연구원
2000년 12월~2005년 2월: 한국전자통신연구원 선임연구원
2005년 3월~현재: 수원대학교 전기공학과 부교수
〈관심분야〉 바이오인식, 정보보호, 컴퓨터비전, 인공지능, 패턴인식