Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.29, NO.3, Jun. 2019 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2019.29.3.565

Index-of-Max 해싱을 이용한 폐기가능한 홍채 템플릿

김 진 아,[1†] 정 재 열,[1] 김 기 성,[2] 정 익 래[1‡]
[1]고려대학교 정보보호대학원, [2]대구가톨릭대학교

Cancelable Iris Templates Using Index-of-Max Hashing

Jina Kim,[1†] Jae Yeol Jeong,[1] Kee Sung Kim,[2] Ik Rae Jeong[1‡]
[1]Graduate School of Information Security, Korea University,
[2]Daegu Catholic University

요 약

최근에 생체인증은 다양한 분야에 사용되고 있다. 생체정보는 변경이 불가능하고 다른 개인정보와 달리 폐기할 수 없기 때문에 생체정보 유출에 대한 우려가 커지고 있다. 최근 Jin et al.은 지문 템플릿을 보호하기 위해 IoM(Index-of-Max) 해싱이라는 폐기가능한 생체인증 방법을 제안했다. Jin et al.은 Gaussian random projection 기반과 Uniformly random permutation 기반의 두 가지 방법을 구현하였다. 제안된 방법은 높은 매칭 정확도를 제공하고 프라이버시 공격에 강력함을 보여주며 폐기가능한 생체인증의 요건을 만족함을 보여주었다. 그러나 Jin et al.은 다른 생체정보에 대한 인증(예: 정맥, 홍채 등)에 대한 실험 결과를 제공하지는 않았다. 본 논문에서는 Jin et al.의 방법을 적용하여 홍채 템플릿을 보호하는 방법을 제안한다. 실험 결과는 이전의 폐기가능한 홍채인증 방법과 비교했을 때 더 높은 정확도를 보여주며 보안 및 프라이버시 공격에 강력함을 보여준다.

ABSTRACT

In recent years, biometric authentication has been used for various applications. Since biometric features are unchangeable and cannot be revoked unlike other personal information, there is increasing concern about leakage of biometric information. Recently, Jin et al. proposed a new cancelable biometric scheme, called “Index-of-Max” (IoM) to protect fingerprint template. The authors presented two realizations, namely, Gaussian random projection-based and uniformly random permutation-based hashing schemes. They also showed that their schemes can provide high accuracy, guarantee the security against recently presented privacy attacks, and satisfy some criteria of cancelable biometrics. However, the authors did not provide experimental results for other biometric features (e.g. finger-vein, iris). In this paper, we present the results of applying Jin et al.’s scheme to iris data. To do this, we propose a new method for processing iris data into a suitable form applicable to the Jin et al.’s scheme. Our experimental results show that it can guarantee favorable accuracy performance compared to the previous schemes. We also show that our scheme satisfies cancelable biometrics criteria and robustness to security and privacy attacks demonstrated in the Jin et al.’s work.

Keywords: Cancelable biometrics, biometric protection, cancelable iris templates, biometric recognition

I. 서 론

사용자인증이 중요한 문제로 대두되면서, 기존의 패스워드나 PIN을 이용한 방법의 문제점을 개선하기 위해 개인의 고유한 생체정보를 이용한 사용자인증이 다양한 응용 프로그램에 사용되고 있다. 생체인

Received(05. 03. 2019), Modified(06. 03. 2019), Accepted(06. 05. 2019)

† 주저자, cuteshoo@korea.ac.kr
‡ 교신저자, irjeong@korea.ac.kr(Corresponding author)

증은 목소리, 서명, 걸음걸이와 같이 무형의 특징을 기반으로 하는 방식과 지문, 얼굴, 홍채와 같이 유형의 특징을 기반으로 하는 방식으로 구분된다. 그 중 홍채는 생후 18개월 이후 완전하게 패턴이 완성되어 평생 변하지 않는 것으로 알려져 있으며 비접촉 방식이기 때문에 거부감이 없고 다른 생체정보에 비해 매칭 정확도가 뛰어나다. 홍채를 비롯한 생체인증이 사용되는 분야가 증가됨에 따라 생체 템플릿이 도난 당하거나 손상되는 경우에 보안 및 프라이버시에 치명적인 결과를 초래할 수 있기 때문에 생체 템플릿을 보호하기 위한 많은 방법들이 제안되었다[1, 2]. 이러한 방법은 생체 템플릿 암호화와 특징벡터 변환의 두 가지 유형으로 분류할 수 있다. 생체 템플릿 암호화는 서버가 암호화된 생체 템플릿을 복호화하여 매칭을 하기 때문에 서버가 생체정보를 알고 있어야 한다. 특징벡터 변환 방법은 폐기가능한 생체인식방법으로 템플릿의 보안과 프라이버시를 보장하며, 폐기가능한 템플릿이 손상되거나 도난당했을 때 같은 생체정보로부터 새로운 템플릿을 생성할 수 있다. 또한 원래의 생체정보를 복원하지 않고 보호된 템플릿을 이용하여 매칭을 하기 때문에 원래의 생체정보를 서버에 숨길 수 있다.

폐기가능한 생체 템플릿을 보호하는 방법은 다음과 같은 몇 가지 조건을 만족한다[3].

1) 비가역성(non-invertible) : 보호된 템플릿에서 원본 생체 템플릿을 만드는 것은 계산상 불가능해야 한다.
2) 폐기가능성(revocability) : 보호된 템플릿이 도난 또는 손상되었을 때 폐기가능 하거나 재생성이 가능해야 한다.
3) 비연결성(non-linkability) : 동일한 생체 특성에서 파생된 두 개 이상의 보호된 템플릿의 구분이 계산상 불가능해야 한다.
4) 성능 보존(performance preservation) : 보호된 템플릿의 정확도 성능은 보호되기 전의 템플릿의 정확도 성능을 보존해야 한다.

II. 관련 연구

이 섹션에서는 홍채 템플릿 보호에 대한 이전 연구를 요약한다. 생체정보는 불변하기 때문에 템플릿이 도난당하거나 손상되면 치명적인 결과를 초래한다. 폐기가능한 생체인식은 비밀번호와 같이 폐기 가능하며 어플리케이션마다 고유하게 함으로써 높은 보안 및 프라이버시 수준을 만족한다. 폐기가능한 생체인식의 개념은 Bolle et al.[4]에서 처음으로 소개되었으며, 그 이후로 많은 새로운 기술이 제안되었다. 폐기가능한 생체인증 방법은 일반적으로 1) Biometric Salting과 2) 비가역변환의 두 가지 범주로 나뉜다[5, 6].

2.1 Biometric Salting

Biometric Salting은 사용자 고유의 입력 요소(보조 데이터 등)를 생체정보와 혼합하여 원본 생체정보의 왜곡된 생체템플릿을 생성하는 방법이다.

Zuo et al.[5]는 GRAY-SALT 및 BIN-SALT 기법을 제안하였다. GRAY-SALT는 그레이 스케일 이미지를 픽셀 단위로 덧셈 또는 곱셈으로 결합하는 방법이다. BIN-SALT는 이진화된 홍채 코드에 XOR을 사용하여 키와 결합하는 방법이다. 두 방법은 홍채정보를 보조 데이터와 결합하여 폐기가능한 홍채 템플릿을 생성한다. 그러나 입력되는 이미지에 대한 회전을 고려하지 않았기 때문에 사전 정렬 과정이 없어 정확도가 상당히 저하되었다.

Chin et al.[7]과 Pillai et al.[8]은 홍채 이미지 전체를 사용하지 않고 분할된 랜덤 투영(random projection)을 사용하여 폐기가능한 홍채 템플릿을 생성하였다. 홍채 이미지 전체를 사용하면, 홍채 외의 영역에서 속눈썹이나 빛 반사 등과 같은 noise 등에 의해 정확도가 떨어지는 문제점이 있다. 따라서 홍채를 여러 개의 섹터(sector)로 분할하여 각 섹터의 Gabor 특징이 사용자의 고유한 Gaussian 랜덤 행렬(Gaussian random matrix)을 통해 낮은 차원으로 투영되고, 섹터 별로 투영된 결과 값을 연결함으로써 홍채 템플릿을 생성한다.

그러나, Kong et al.[9]은 동일한 랜덤 행렬을 다른 사용자들에게도 사용할 경우 정확도 성능이 크게 저하된다고 하였다. 또한 공격자에게 사용자 고유의 랜덤 행렬이 노출되는 경우(stolen-token 시나리오[10]) 폐기가능한 템플릿을 복원할 수 있다고 하였다. 따라서 Biometric Salting 방법에서 사용자 고유의 랜덤 행렬 등의 보조 데이터는 비밀로 유지되어야 한다.

2.2 비가역변환

비가역변환은 역변환이 불가능한 함수 또는 알고리즘을 이용해서 생체 템플릿을 변경하여 생체정보를 보호하는 개념이다. 변경된 템플릿으로 원래의 홍채 템플릿을 복원할 수 없기 때문에 안전하다.

Zuo et al.[5]은 GRAY-COMBO 및 BIN-COMBO라 하는 실수 값 또는 홍채 패턴에 적용할 수 있는 비가역변환 방법을 제안하였다. GRAY-COMBO는 랜덤 키를 통해 홍채 이미지를 행 단위로 시프트(Shift) 시킨 후 무작위로 선택된 두 행에서 덧셈 또는 곱셈 연산을 수행한다. 이 방법은 저화질의 홍채 이미지를 사용하는 경우 매칭 정확도가 떨어진다. BIN-COMBO에서는 이진화된 홍채 코드에 대해 XOR 또는 XNOR을 사용하여 수행한다. 두 가지 방법은 2개의 무작위로 선택된 행 간의 연산으로 변환되었기 때문에 비가역성을 만족한다. 두 방법 모두 홍채 이미지의 회전에 상관없이 행의 변화가 동일할 것이기 때문에 템플릿의 시프트된 행은 매칭할 때 정렬이 필요 없다. 이것을 'alignment free' 또는 'registration free'라고 한다. 그러나 두 가지 방법 모두 사용자 고유의 키를 사용하기 때문에 공격자에게 사용자 고유의 키가 노출될 경우 폐기가능한 템플릿을 복구할 위험이 있다.

Hämmerle-Uhl et al.[11]은 블록 재매핑(block remapping)을 사용한 방법을 제안하였다. 정규화된 홍채 이미지는 여러 개의 블록으로 분할되고 키를 사용하여 무작위로 치환된 후 새로운 이미지에 블록이 재매핑 된다. 이 방법은 원본 홍채 이미지의 재구성을 방지하고 비가역성을 만족한다고 하였으나, Jenish et al.[12]에 의해 도난 당한 템플릿으로부터 원래 홍채 이미지의 60%가 복원 가능 하다고 증명되었다.

Ouda et al.[13, 14]은 토큰을 사용하지 않는 Bio-encoding 스킴을 제안하였다. 여러 개의 샘플에서 플립(flip)될 확률이 낮은 비트를 consistence 비트로 추출한 뒤에 이를 이용하여 Biocode를 생성한 후 매칭하는 방법이다. 그러나 Lacharme[15]은 Biocode가 비가역성을 만족하지 않는다고 지적하였다.

Rathegeb et al.[16, 17, 18]은 블룸 필터(Bloom filter)를 사용한 방법을 제안하였다. 블룸 필터는 입력 값에 여러 개의 해시함수를 사용하여 만들어지는 비트 배열이며, 여기서는 해시함수를 사용하는 대신 이진코드를 십진수로 매핑하는 방법을 사용하였다. 이 방법은 정확도 성능이 원래의 성능과 비슷하다고 하였으나, Hermas et al.[19]은 낮은 복잡도로 인해 템플릿이 복원될 수 있음을 증명하였다. Bringer et al.[20] 또한 정확도 성능을 유지하기 위한 작은 키 스페이스로 인해 비연결성 공격이 가능하다고 지적했다. Gomez-Barrero et al.[21]은 최근 연구에서 블룸 필터 기반의 방법에서 cross-matching 공격을 막는 방법을 제시했다.

Dwivedi et al.[22]은 look-up 테이블과 매핑한 결과를 템플릿으로 생성하는 방법을 제안하였다. look-up 테이블은 폐기가능한 템플릿과 함께 보관되기 때문에 look-up 테이블과 매개변수가 노출되는 경우 홍채코드를 복구할 수 있다.

Lai et al.[23]은 Min-Hashing에 영감을 받아 실수 값의 홍채 특징을 Hadamard product와 모듈로 임계값 함수를 이용하여 이산 색인 해시코드로 매핑하는 IFO(Indexing-First-One) 해싱을 제안했다. Jin et al.[24]은 Gaussian 랜덤 행렬과 균일 랜덤 치환(uniform random permutation)을 사용하여 실수 값의 지문 특징을 최대 값의 이산색인 해시코드로 매핑하는 IoM(Index-of-Max) 해싱을 제안하였다. 두 방법 모두 주요 보안 공격에 강력함을 보여주었고 우수한 정확도 성능과 비가역성을 보여주었다. 본 논문에서는 Jin et al.이 제안한 방법을 홍채에 적용하여 높은 매칭 정확도의 성능으로 폐기가능한 홍채 템플릿을 생성한다.

III. 배경지식

이 섹션에서는 IoM 해싱의 기반이 되는 LSH(Locality Sensitive Hashing)에 대해 소개한다[25]. 또한 GRP(Gaussian Random Projection)-based IoM 해싱에서 참고한 RMF(Random Maxout Features)[26]와 URP(Uniform Random Permutation)-based IoM 해싱에서 참고한 WTA(Winner Takes All[27] 해싱에 대해 간략하게 소개하고자 한다. 논문에서 사용되는 표기법은 Table 1에 따른다.

3.1 LSH(Locality Sensitive Hashing)

LSH는 고차원에서의 데이터 유사성 검색을 최적화하기 위한 방법으로, 일반적인 해싱은 데이터를 고

Table 1. Notation

Notation	Description
H	LSH(Locality-Sensitive Hashing) family
h	LSH function $h \mathit{IN}\ H$
$S(X,Y)$	Similarity function defined on X and Y
$\mathbf{X}$	Iris vector $\mathbf{X} \in \mathbb{R}^d$
m	Number of Gaussian random matrices(GRP-based IoM)/ Number of Uniformly random permutation(URP-based IoM)
q	Number of Gaussian random projection vector
p	Order of Hadamard product
k	Window Size(URP-based IoM hashing)
$\boldsymbol{t}$	$\boldsymbol{t}_{GRP}$: GRP-based hashed code/ $\boldsymbol{t}_{URP}$: URP-based hashed code

르게 분포시키는 반면, LSH는 유사도가 높은 데이터는 같은 버킷에 담고 멀리 있는 데이터는 다른 버킷에 담는다.

정의 1[24] : LSH는 해시함수 h의 패밀리 H에 대한 확률 분포 P이며, $\mathrm{P}_{h\in H}[h(X)=h(Y)]=S(X,Y)$ 이다. 이 때 S는 X와 Y의 유사도 함수로 정의되고 U는 해시 공간이다.

if $S(X,Y)<R_1$ then $\mathbb{P}_{h\in H}(h_i(X)=h_i(Y)) \le P_1$

if $S(X,Y)>R_2$ then $\mathbb{P}_{h\in H}(h_i(X)=h_i(Y)) \ge P_2$

여기서 $\mathrm{P}_{h\in H}$는 확률을 나타내며 R_1과 R_2 $(R_1<R_2)$는 유사도 상수이고 P_1와 $P_2(P_1>P_2)$는 확률 상수이다.

3.2 RMF(Random Maxout Features)

Mroueh et al.[26]이 제안한 RMF는 Gaussian 랜덤 벡터를 투영한 후 각 set 마다 가장 큰 값을 기록하는 방법으로 논문에서는 GRP-based IoM 해싱에 사용된다. RMF는 최대 값을 기록하는 것과 달리, 본 논문에서는 최대 값의 인덱스를 기록한다.

정의 2[26] : w_j^l, $l=1...m$, $j=1...q$를 독립적인 랜덤 Gaussian 벡터라 하자. $x\in\mathbb{R}^d$에 대해서 maxout random unit $h_{l(x)}$는 다음과 같다.

$$h_{l(x)}=\phi(x,W^l)=\max_{j=1...q}< w_j^l x >, \quad l=1...m.$$

여기서 $w_j^l \sim \mathcal{N}(0,\ \mathbf{I}_d)$이고 $W^l=(w^{l_1}...w_q^l)$이다. 따라서 maxout random feature map ϕ 는 아래와 같이 정의된다.

$$\phi\ (x)=\frac{1}{\sqrt{m}}(h_1(x),...,h_m(x)).$$

3.3 WTA(Winner-Takes-All) Hashing

WTA 해싱은 빠른 유사성 검색에 사용되는 방법으로, 입력 벡터에 랜덤 치환(random permutation)을 적용하여 최대 값의 색인을 기록한다. WTA 해싱은 다음의 5단계로 이루어진다. 논문에서는 URP-based IoM 해싱에 사용된다.

1) 입력 벡터 $\mathrm{X}\in\mathrm{R}^d$에 랜덤 순열 H 을 수행한다.
2) 치환된 X에서 처음 k개 요소를 선택한다.
3) k개 요소에서 가장 큰 요소를 선택한다.
4) 가장 큰 요소의 인덱스를 비트로 기록한다.
5) 1 - 4 단계를 m번 반복하여 m길이의 해시코드를 생성한다.

IV. 제안하는 홍채 템플릿 생성 방법

IoM 해싱은 홍채코드를 실수 값으로 변환하여 입력으로 사용한다. 먼저 입력 벡터에 대해 간단하게 설명하고, Jin et al.[24]이 제안한 GRP-based IoM 해싱과 URP-based IoM 해싱에 대해 설명한 다음 각각의 매칭에 대해 설명한다.

4.1 입력 벡터 생성

4.1.1 홍채 코드 생성

Jin et al.은 MCC(Minutiae Cylinder Code)[28, 29]를 이용해 지문의 minutia를 추출하여 커널 행렬 계산을 통해 1×299 크기의 지문 벡터를 IoM 해싱의 입력으로 사용하였다. 논문에서는 홍채 코드를 생성하기 위해 Daugman[30]의 알고리즘을 사용하였다. 홍채 이미지에서 홍채와 눈동자의 경계를 찾아 노이즈를 제거한다. 분할된 홍채 영역은 Rubber Sheet Model을 사용하여 정규화한다. 그런 다음 Gabor 필터에 의해 홍채 특징을

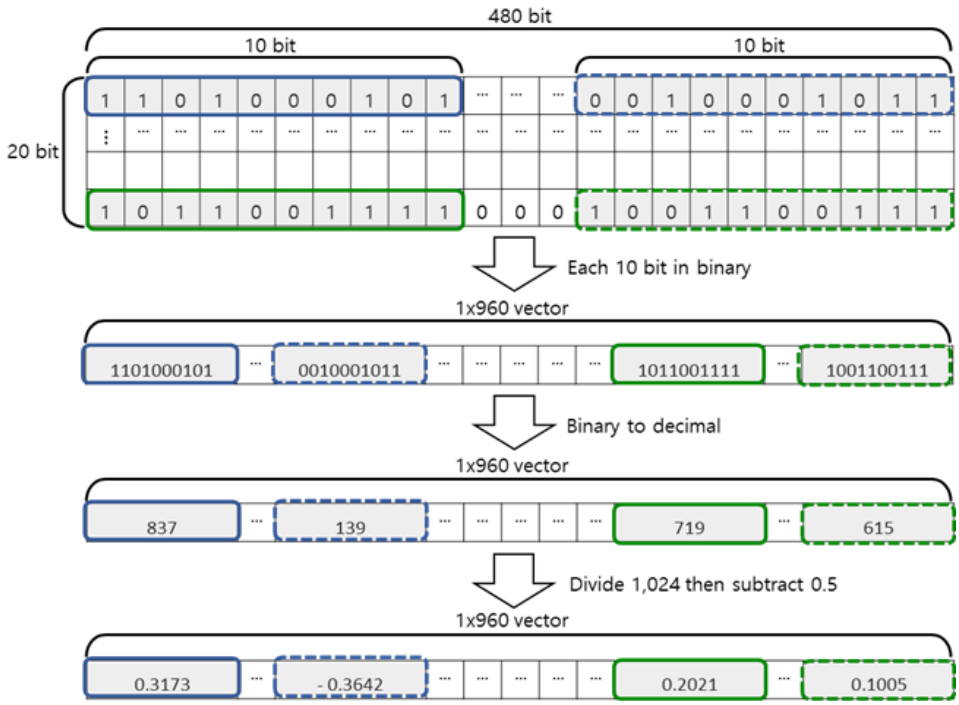

Fig. 1. Conversion of iris bits into real numbers

검출하여 홍채 코드를 생성하는데, 이 실험에서는 20×480(=9,600)bit로 생성하였다. 이렇게 생성된 코드는 0과 1을 가지는 이진 비트이며, 이를 실수값으로 변환하기 위해 다음과 같이 템플릿을 생성하였으며 Fig. 1에 나타냈다.

1) 20×480 크기의 홍채 코드 벡터를 10bit씩 끊어 열 자리의 이진수로 만든다.
2) 이진수를 0~1,023의 십진수로 변환한다.
3) 십진수를 1,024로 나누면 0~0.999 값이 되고 0.5를 뺄셈하면 -0.5000~0.4990 사이의 값을 가진다. 이렇게 최종 변환한 템플릿을 IoM 해싱의 입력 벡터로 사용하였다. 0.5를 뺄셈하여 음수와 양수가 혼재된 값을 가지게 하는 이유는 6.2절에서 자세히 설명하였다.

4.1.2 사전 정렬(pre-alignment)

쿼리로 사용되는 홍채 이미지는 항상 일정하게 정렬된 상태로 입력되지 않기 때문에 방향이 조금씩 달라질 수 있다. 따라서 시프트 연산 과정을 거쳐 템플릿을 생성하고 등록된 템플릿과 매칭을 수행한다. 시프트는 홍채코드를 실수화 하기 전에 20×480bit의 이진 코드 상태에서 진행된다. 이진 홍채코드에서 좌측으로 1~8bit, 우측으로 1~8bit 시프트하여 한 개의 쿼리 이미지당 -8에서 8만큼 총 17개의 템플릿을 생성한다(Shift 시키기 전 : 0 포함). 이렇게 생성된 17개의 템플릿은 실수로 변환된다.

4.2 IoM 해싱(Index-of-Max hashing)

IoM 해싱은 실수값의 특징 벡터를 입력으로 하여 최대 값을 갖는 인덱스를 가지는 순위 기반의 코드로 변환함으로써 다음과 같은 몇 가지 장점을 갖는다.

1) 해싱된 코드를 가지고 원본 특징벡터를 복원할 수 없기 때문에 비가역성을 만족한다.
2) 상대적인 순서에 의존적인 방법이기 때문에 특징벡터의 크기에 영향을 받지 않는다. 따라서 노이즈나 변형에 강력하다.
3) IoM 해싱의 크기 독립성으로 인해 해시 코드를 크기 불변(scale-invariant)하게 한다.

4.2.1 GRP-based IoM 해싱

GRP-based IoM 해싱의 알고리즘은 Jin et al. 의 알고리즘과 동일하며 Fig. 2에 나타냈다[24]. 먼저 실수화된 홍채 템플릿 벡터를 q-차원의 Gaussian 랜덤 행렬에 투영시켜 최대 값의 색인을 결과로 얻는다. 이 과정을 m번 반복하여 m개의 IoM 인덱스 집합을 생성한다.

Input Feature vector $\mathbf{X}$,
number of Gaussian random projection vector q, number of Gaussian random matrices m,

1. Create m Gaussian random matrices
   $W^i = (w_1^i, ..., w_q^i)\ \ i = 1, ..., m.$
2. Set i^{th} hashed code $\mathbf{t}_i = 0$
3. Execute random projection and store the maximum index in the projected feature vector.
   for $k = 1 : m$
   $\overline{\mathbf{X}}^k = \mathbf{W}^k\mathbf{X}$
   Find $\mathbf{X}_j^k = \max(\overline{\mathbf{X}}^k),\ \ j = 1, ..., q$
   Then $\mathbf{t}_i = j$ (j refers the index of $\overline{\mathbf{X}}^k$)
   End for

Output Hashed code
$\mathbf{t}_{GRP} = \{\mathbf{t}_i |\ \ i = 1, ...m RIGHT$
and $\mathbf{t}_{GRP} \in [1, q]$

Fig. 2. GRP-based IoM hashing

4.2.2 URP-based IoM 해싱

URP-based IoM 해싱의 알고리즘은 Jin et al.의 알고리즘과 동일하며 Fig 3에 나타냈다[24]. 먼저 실수화된 홍채 템플릿 벡터를 p개의 균일 랜덤 치환 seeds로 치환하여 벡터를 생성한다. p개의 치환된 벡터를 Hadamard product 하여 윈도우 사이즈 k내에서 가장 큰 값을 가지는 인덱스를 기록한다. 이 과정을 m번 반복하여 IoM 인덱스 집합을 생성한다.

4.3 매칭(matching)

IoM 해싱의 매칭은 유사도가 높은 $\boldsymbol{t}^e$와 $\boldsymbol{t}^q$간에 충돌 확률이 높은 순위 기반 LSH을 따른다. 두 개 홍채 벡터가 유사도가 높으면 해시코드의 충돌 확률이 높고, 유사도가 낮으면 해시코드의 충돌 확률이 낮다.

Input Feature vector $\mathbf{X}$, window size k,
p order of Hadamard product,
number of random permutation m

For each permutation set
$\theta_{(l,i)} | l = 1, ..., q,\ \ i = 1, ..., m.$
1. Perform permutation of elements in $\mathbf{X}$ based on $\theta_{(l,i)}$, $\hat{\mathbf{X}} = perm(\mathbf{X}), perm(.)$ is the random permutation function.
2. Set i^{th} hashed code $\mathbf{t}_i = 0$
3. Hadamard product vector generation and output hashed codes.
   for $j = 1 : k$
   Set $\overline{\mathbf{X}}(j) = \prod_{l=1}^{p} (\hat{\mathbf{X}}_l(j))$
   if $\mathbf{X}$
   Then $\overline{\mathbf{X}}(j) > \overline{\mathbf{X}}(t_i)$ then $t_i = j$
   End for

Output Hashed code
$\mathbf{t}_{URP} = \{t_i \in [1, k] | i = 1, ..., m\}$

Fig. 3. URP-based IoM hashing

등록된 해시코드 $\boldsymbol{t}^e = \{t_i^e | i = 1, ..., m\}$와 쿼리 해시코드 $\boldsymbol{t}^q = \{t_j^q | i = 1, ..., m\}$ 사이의 충돌 가능성은 유사도 $S(\boldsymbol{t}^e, \boldsymbol{t}^q)$로 표현된다. 즉, $\mathbb{P}[\boldsymbol{t}_i^e = \boldsymbol{t}_j^q] = S(\boldsymbol{t}^e, \boldsymbol{t}^q)$ for 1, ..., m이 된다.

4.3.1 GRP-based IoM 매칭

h를 $h : \mathbb{R}^d \rightarrow \{1, ..., q\}^m$ 로 정의되는 LSH 함수라 하자. $\mathbf{u}$, $\mathbf{v}$ IN $\mathbb{R}^d$ 이고
$h(\mathbf{u}) = \{\arg\max_{j=1...q} < \mathbf{w}^i, \mathbf{u} > |\ \ i = 1, ..., m\}$ 이다. 여기서 $\{\mathbf{w}_j \in \mathbb{R}^d | j = 1, ..., q\} \sim \mathcal{N}(0,\ \mathbf{I}_d))$이다.

GRP-based IoM 해싱에서의 유사도 $S_{GRP}(\boldsymbol{t}^e, \boldsymbol{t}^q) =$
$\mathbb{P}\left\{\arg\max_{j=1...q} < \mathbf{w}_j^i, \mathbf{u} > \ = \ \arg\max_{j=1...q} < \mathbf{w}_j^i, \mathbf{v} >\right\}$ 이다.

구현에서의 매칭 점수는 해시코드의 총 엔트리(m)에 대한 $\boldsymbol{t}^e$와 $\boldsymbol{t}^q$간의 요소별 뺄셈에 의해 '0'(충돌)이 되는 횟수이다.

4.3.2 URP-based IoM 매칭

URP-based IoM 해싱에서의 유사도는 일정한 범위 내에서의 상대적인 순서 측정이다.

$$S_{URP}(\boldsymbol{t}^e,\boldsymbol{t}^q)=\frac{\sum_{i=0}^{n-1}\binom{R_i(\boldsymbol{t}^e,\boldsymbol{t}^q)}{k-1}}{\binom{d}{k}}$$

구현에서의 매칭 점수는 해시코드의 총 엔트리(m)에 대한 $\boldsymbol{t}^e$와 $\boldsymbol{t}^q$간의 요소별 뺄셈에 의해 '0'(충돌)이 되는 횟수이다.

V. 실험 결과

논문에서는 실험을 위해 Windows 10, Intel CPU i7-4790(3.6GHz)와 8GB RAM 사양의 시스템에서 MATLAB Ver.R2017a 소프트웨어로 구현하였으며 CASIA v3 데이터셋을 사용하였다. CASIA v3는 320×280 해상도 이미지로, 249개의 subject로 이루어져 있으며 총 2,639개의 이미지가 있다. 249개의 subject는 각각 왼쪽과 오른쪽 눈으로 나누어져 있으며 좌우의 이미지 개수는 일정하지 않다. 왼쪽 눈과 오른쪽 눈은 홍채 패턴이 다르기 때문에 실험을 위해서 왼쪽 눈과 오른쪽 눈을 다른 subject로 구분하였으며, 7개 이상의 이미지를 갖는 subject만 사용하였다. 그래서 246 subject, 총 1,722(246×7)개의 이미지가 실험에 사용되었다. 각 subject의 7개의 이미지 중 처음 3개 이미지는 쿼리로 사용하고 나머지 4개의 이미지는 등록된 템플릿으로 사용하였다. 정확도는 EER(Equal Error Rate)로 측정하였다.

5.1 IoM 해싱의 파라미터

5.1.1 Gaussian 랜덤 매트릭스 개수 m, Gaussian 랜덤 투영 벡터 개수 q

GRP-based 매칭에서 Gaussian 랜덤 매트릭스 수 m과 Gaussian 랜덤 투영 벡터 수 q를 달리하였을 때 EER에 미치는 영향을 조사하였다. 실험에서 m과 q는 각각 2, 5, 10, 50, 100, 200, 300으로 변경해 가면서 EER를 측정하였고, 결과는 Fig. 4에 나타내었다.

1) m이 커질수록 EER이 큰 폭으로 감소되는 것을 확인하였으며, m=100 이상일 때 q에 상관없이 EER이 0.1 미만으로 나타났다. 따라서 GRP-based 매칭에서는 Gaussian 랜덤 매트릭스의 수 m이 정확도 성능을 결정하는 중요한 요소이며, 적절히 m을 선택하면 EER을 손상시키지 않고 계산비용 및 저장공간을 절약할 수 있다.
2) q는 5일 때 정확도가 가장 높게 나타났으며, m이 어느 수준(=100) 이상일 때 q가 정확도에 미치는 영향은 거의 없다고 판단된다.

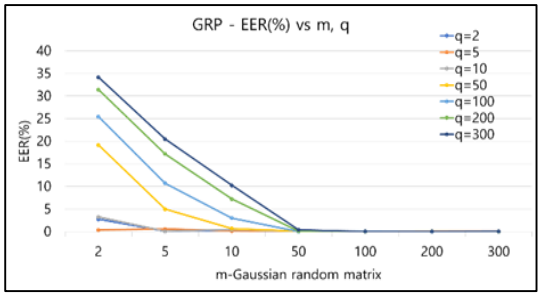

Fig. 4. EER for number of Gaussian random matrices m and number of Gaussian random projection vectors q

5.1.2 윈도우 사이즈 k, Hadamard Product Order p

URP-based 매칭에서 윈도우 사이즈 k와 Hadamard product order p를 달리하였을 때 EER에 미치는 영향을 조사하였다. 실험에서 k는 50, 100, 200, 300, 500, 700, 900으로 변경하였고 p는 2, 3, 4, 5로 변경해 가면서 EER을 측정하였고, 결과는 Fig. 5 에 나타내었다.

1) 윈도우 사이즈 k가 정확도에 미치는 영향은 거의 없었다. p=5일 때, k=50과 900에 대해 EER=0.34%, 0.39%로 각각 나타났으며

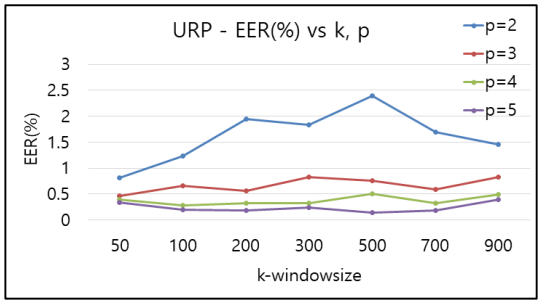

Fig. 5. EER for window size k and Hadamard product order p

큰 차이가 없었다.

2) Hadamard product order p에 대해서는 p가 커짐에 따라 정확도가 높아지는 것을 확인하였고 $p=5$일 때 정확도가 가장 높았다.

5.1.3 Uniform 랜덤 치환 개수 m

URP-based 매칭에서 uniform 랜덤 치환 수 m을 달리하였을 때 EER에 미치는 영향을 조사하였다. 실험에서 윈도우 사이즈 $k=100$, Hadamard product order $p=5$로 고정시키고 m은 2, 5, 10, 50, 100, 200, 300으로 변경해 가면서 EER을 측정하였고, 결과는 Fig. 6에 나타내었다. 정확도는 $m=50$일 때 EER=0.02로 가장 높았고 m이 50 이상에서는 정확도가 높은 수준으로 유지되었다. 따라서 URP-based 매칭에서는 uniform 랜덤 치환의 수인 m이 정확도 성능을 결정하는 중요한 요소이며 적절히 m을 선택하면 EER을 손상시키지 않고 계산비용 및 저장공간을 절약할 수 있다.

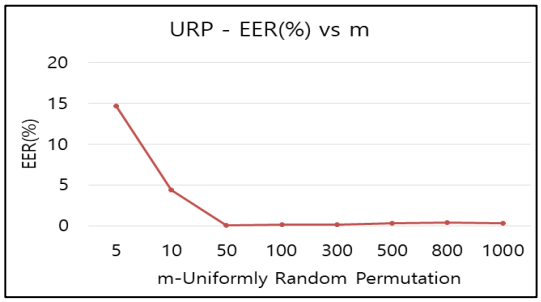

Fig. 6. EER for uniform random permutation m with $k=100$, $p=5$

5.2 정확도 평가

논문에서는 정확도 성능을 평가하기 위한 방법으로 EER을 사용하는데, 이는 FAR(False Acceptance Rate)과 FRR(False Rejection Rate)이 교차되는 지점으로 정의된다. FAR은 본인의 것이 아닌 생체정보를 본인의 것으로 잘못 판단할 확률이고 FRR은 본인의 생체정보를 본인이 아닌 것으로 잘못 판단할 확률이다. FAR과 FRR은 Fig. 7에서 보는 바와 같이 trade-off 관계에 있으며 보안 수준을 설정하는데 사용되는 임계 값에 따라 달라진다. Fig. 7는 GRP-based IoM 해싱에서 임계값에 따른 Error Rate를 그래프로 나타내었으며 URP-based IoM도 유사한 그래프를 가진다. 권한이 없는 사용자의 접근을 어렵게 만들기 위해 임계값을 늘리면 권한이 부여된 일부 사용자는 접근이 어려워지고 반대로 임계 값을 줄이면 권한이 없는 사용자가 접근할 수 있는 확률이 높아지기 때문에 적절하게 임계 값을 설정해야 한다.

이전 논문에서 제시된 폐기형 생체인증 방법과 논문에서 제안한 방법의 성능(EER)을 분석하여 Table 2에 비교하였다. 최근에 제안된 IFO hashing 보다 정확도 성능이 우수함을 볼 수 있다.

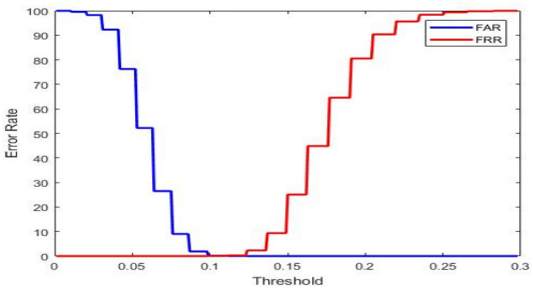

Fig. 7. FAR and FRR of IoM hashing

Table 2. Accuracy comparison between the cancelable iris authentication schemes (CASIA v3 database)

Methods	No. Iris images used	Lowest EER(%)
GRP-based IoM hashing(proposed)	1,722	0
URP-based IoM hashing(proposed)	1,722	0.14
IFO hashing[23]	868	0.54
Block Remapping[11]	2,653	1.30
Bio-encoding[13]	740	6.27
Adaptive bloom filter[16]	1,332	1.14
Bin-Combo[5]	1,332	4.41

5.3 효율성 평가

IoM 해시코드를 생성할 때와 매칭할 때의 효율성을 분석하기 위해 각각의 평균 소요 시간을 측정하였다. 해시코드를 생성할 때의 시간은 매개변수와 관련

이 있기 때문에 매개변수를 변경하여 실험을 진행하였다. GRP-based IoM 해싱에서 Gaussian 랜덤 투영 벡터 개수 q와 Gaussian 랜덤 매트릭스 개수 m을 달리하여 측정하였다. URP-based IoM 해싱에서 윈도우 사이즈 k는 매칭 정확도(EER)나 시간에 미치는 영향이 거의 없기 때문에 $k=100$으로 고정하고, Hadamard Product Order p와 Uniform 랜덤 치환 개수 m을 달리하여 측정하였으며 각각 Fig. 8과 Fig. 9에 나타냈다.

URP-based IoM 해싱보다 GRP-based IoM 해싱에서 해시코드를 생성하는 시간이 더 소요되었으며 각 해싱에서 매개변수 값을 증가할수록 소요되는 시간도 증가되었다. 5.1절에서 분석한 내용을 바탕으로 정확도를 참고하면 EER을 손상시키지 않고 효율성을 높일 수 있다.

매칭에 소요되는 시간은 1개의 쿼리로 입력된 해시코드에 대해 등록된 해시코드와 매칭하는데 걸린 시간을 측정하였다. 매개변수는 정확도가 높을 때의 값으로 고정시키고 10회에 걸쳐 평균 시간을 측정하였으며, Tabel 3에 나타냈다. GRP-based IoM과

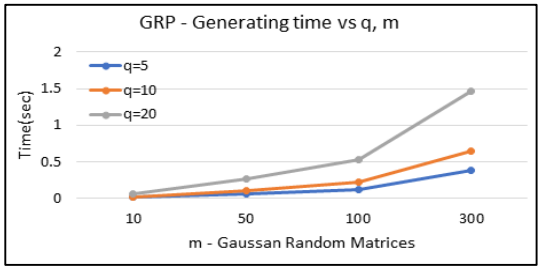

Fig. 8. Average Time Processed in Generating hashed code for number of Gaussian random matrices m and number of Gaussian random projection vectors q

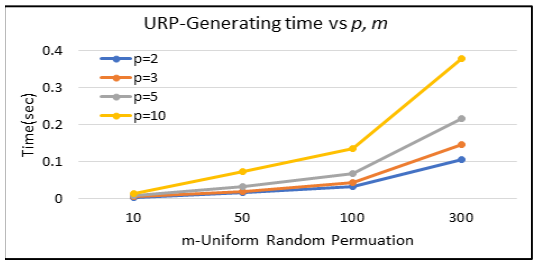

Fig. 9. Average Time Processed in Generating hashed code for Hadamard product order p and uniform random permutation m with $k=100$

Table 3. Average time processed in matching stages

GRP-based IoM	URP-based IoM
0.065 sec	0.058 sec

URP-based IoM의 매칭 시간은 큰 차이를 보이지 않았다.

VI. 프라이버시 및 보안 분석

이 섹션에서는 제안된 방법에 대한 비가역성, 비연결성, 폐기가능성 및 주요 보안 공격에 대한 가능성을 분석한다.

6.1 비가역성, 비연결성, 폐기가능성 분석

비가역성은 GRP에서의 랜덤 매트릭스와 URP에서의 치환 seeds의 유무에 상관없이 IoM 해시코드로부터 홍채코드를 복원할 때의 계산복잡도를 의미한다. 공격자가 해시코드와 토큰(예 : 랜덤 행렬, 치환 seeds)을 획득하였다고 가정하고, 해시 알고리즘과 매개변수(예 : m, k, p, q)를 알고 있다고 가정한다. 공격자가 이산 색인된 IoM 코드를 가지고 홍채코드를 알아내는 것은 불가능하다. 토큰 정보를 알고 있다 하더라도 토큰과 홍채코드 사이에 직접적인 연관이 없기 때문에 홍채코드를 복구하는데 아무런 도움이 되지 않는다. 최악의 경우 공격자가 입력벡터 X의 요소인 실수 값을 추측한다고 가정했을 때, 최소 값은 -0.5000이고, 최대 값은 0.4990이다. 입력 벡터 X의 각 요소는 홍채 코드의 열 개 비트를 붙여 한 개의 실수값으로 만들어졌기 때문에 공격자가 한 개 요소를 추측할 가능성은 9,991이 아니라 1,024($=2^{10}$)가 된다. 따라서 1×960의 한 개 입력벡터를 추측하려면 $2^{10\times960}=2^{9600}$ 번의 시도가 필요

Table 4. Complexity to invert single and entire feature component

Min value with four decimal precision	-0.5000
Max value with four decimal precision	0.4990
Possibilities for single feature component	$1,024=2^{10}$
Total possibilities for entire feature	$2^{10\times960}=2^{9600}$

하며 이는 계산 불가능하다.

6.2 ARM(Attacks via Record Multiplicity)

ARM은 공격자가 알고리즘과 관련된 매개변수를 알고 있는지 여부와 관계없이 손상된 여러 개의 템플릿을 이용하는 공격이다.

프라이버시에 대한 ARM은 여러 개의 손상된 템플릿에서 원래의 생체 정보를 복구할 확률이다. IoM 해싱의 경우 등록된 템플릿이 홍채의 feature 공간과 전혀 관계가 없는 순위 공간으로 변환되었기 때문에 원래의 생체정보를 추측하는 것은 불가능하다. 따라서 공격 복잡도는 Table 3의 비가역성 분석과 동일하다.

Security에 대한 ARM은 손상된 여러 개의 템플릿으로 pre-image 인스턴스를 생성하여 불법 접근하는 보안 공격 중 하나이다. 공격자가 feature 요소의 순위와 치환 seeds를 알고 있다면 가짜 feature 벡터를 생성할 수 있다. ARM 공격의 복잡도는 홍채 입력 벡터의 순위를 결정하는 복잡도로 간주할 수 있다. 예를 들어, 입력 벡터 $\mathbf{X}=\{x_a, x_b, x_c\}$에 대해, $p=2$이고 랜덤으로 치환된 feature 벡터 $\widehat{\mathbf{X}_1}=\{x_c,x_a,x_b\}$, $\widehat{\mathbf{X}_2}=\{x_b,x_c,x_a\}$가 있다고 가정하자. Hadamard product 연산에 의해 벡터 $\overline{\mathbf{X}}=\{x_bx_c, x_ax_c, x_ax_b\}$가 생성된다. x_ax_c가 가장 큰 값이라고 할 때, 두 가지 부등식 $x_ax_c > x_bx_c$ 와 $x_ax_c > x_ax_b$ 을 도출할 수 있다. 여기서 공격자는 $x_a > x_b$ 이고 $x_c > x_b$ 임을 알 수 있으며, 이 절차를 반복함으로써 전체 순위 정보를 알 수 있다. 실제 값으로 예를 들면, $\mathbf{X}=\{x_a, x_b, x_c\}=\{-0.3, 0.8,-0.5\}$라고 하자. $x_ax_c > x_bx_c$에 의해 $(-0.3)\times(-0.5) > (-0.3)\times(0.8)$ 이 된다. 그러나 $(-0.5) < (0.8)$이므로 $x_a < x_b$가 되며 앞서 언급한 $x_a > x_b$에 모순된다. 이러한 추측은 GRP-based 해싱에서도 동일하게 적용되며, IoM 해싱에서의 ARM 공격은 입력 벡터의 요소 값이 모두 음수 또는 양수인 경우에만 가능하다. 음수와 양수가 혼재된 값을 가지는 벡터는 크기에 따라 부등호를 추론하는 것이 어렵기 때문에 공격이 불가능하다.

6.3 무차별 대입 공격(Brute-Force Attack)과 거짓 수락 공격(False Accept Attack)

GRP-based 해싱에서, $m=100$, $q=5$일 때 가장 높은 정확도를 보였으며, 각 엔트리별 공격 복잡도는 $q=5 > 2^2$이 된다. $m=100$이므로 $2^{2\times100}=2^{200}$이상의 시도가 필요하다. URP-based 해싱에서, $m=50$, $k=100$일 때 가장 높은 정확도를 보였으며 IoM 해시코드는 1에서 100 사이의 값이 된다. 따라서 추측할 수 있는 복잡도는 각 요소별로 $k=100 > 2^6$ 이 되며 $m=50$이기 때문에 $2^{6\times50}=2^{300}$번 이상의 시도가 필요하다. 따라서 GRP-based 및 URP-based 해싱에서 무차별 대입 공격은 계산상으로 불가능하다.

무차별 대입 공격에서 맹목적으로 추측하는 것과는 달리 False Accept Attack은 무차별 대입 공격보다 적은 횟수의 시도로 공격이 가능하다. 이 공격은 임계값 기반의 인식 시스템에서 매칭 점수가 사전에 정의된 임계값보다 크면 접근권한이 주어지는 것으로 가능하다. 공격 복잡도는 Jin et al.에서 처럼 임계값과 매개변수(m, k, q)를 적절히 조절하여 공격 난이도를 높일 수 있다.

VII. 결 론

본 논문에서는 Jin et al.[24]이 제안한 IoM 해싱을 이용한 방법을 홍채에 적용하여 폐기가능한 홍채 템플릿 생성 방법을 제안하였다. GRP-based 및 URP-based IoM 해싱의 두 가지 구현을 통해, 이전의 논문에서 제안된 방법과 비교하여 높은 정확도를 보여주었다. IoM 해싱은 비연결성 및 폐기가능성을 만족하여 생체 템플릿을 보호하는 조건을 모두 충족시킨다. 또한 적절하게 매개변수를 조정하여 보안 및 프라이버시에 대한 주요 공격에 강력하게 대처할 수 있다.

References

[1] F. Hao, R. Anderson and J. Daugman, "Combining crypto with biometrics effectively," IEEE transactions on computers, vol. 55, no. 9, pp. 1081-1088, 2006.

[2] N. Ratha, J. Connell and R. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM systems Journal, vol. 40, no. 3, pp. 614-634, 2001.
[3] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," EURASIP Journal on Information Security, vol. 2011, no. 3, pp. 1-25, 2011.
[4] R.M. Bolle, J.H. Connell and N.K. Ratha, "Biometric perils and patches," Pattern Recognition, vol. 35, no. 12, pp. 2727-2738, 2002.
[5] J. Zuo, N.K. Ratha and J.H. Connell, "Cancelable iris biometric," in 19th International Conference on Pattern Recognition, pp. 1-4, 2008.
[6] V.M. Patel, N.K. Ratha and R. Chellappa, "Cancelable biometrics: A review," IEEE Signal Processing Magazine, vol. 32 no. 5, pp. 54-65, 2015.
[7] C.S. Chin, A.T.B. Jin and D.N.C. Ling, "High security iris verification system based on random secret integration," Computer Vision and Image Understanding, vol. 102, no. 2, pp. 169-177, 2006.
[8] J.K. Pillai, V.M. Patel and R. Chellappa, "Sectored random projections for cancelable iris biometrics," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1838-1841, 2010.
[9] A. Kong, K.H. Cheung, D. Zhang, M. Kamel and J. You, "An analysis of BioHashing and its variants," Pattern recognition, vol. 39, no. 7, pp. 1359-1368, 2006.
[10] A.K. Jain, K. Nandakumar and A. Nagar, "Biometric template security," EURASIP Journal on advances in signal processing, no. 113, 2008.
[11] J. Hämmerle-Uhl, E. Pschernig and A. Uhl, "Cancelable iris biometrics using block re-mapping and image warping," International Conference on Information Security, Springer, Berlin, Heidelberg, vol. 5735, pp. 135-142, 2009.
[12] S. Jenisch and A. Uhl, "Security analysis of a cancelable iris recognition system based on block remapping," in 18th IEEE International Conference on Image Processing, pp. 3213-3216, 2011.
[13] O. Ouda, N. Tsumura and T. Nakaguchi, "On the security of bioencoding based cancelable biometrics," IEICE TRANSACTIONS on Information and Systems vol. 94, no. 9, pp. 1768-1777, 2011.
[14] O. Ouda, N. Tsumura and T. Nakaguchi, "Tokenless cancelable biometrics scheme for protecting iris codes," in 20th International Conference on Pattern Recognition, IEEE, pp. 882-885, 2010.
[15] P. Lacharme, "Analysis of the iris codes bioencoding scheme," Int. J. Comput. Sci. Softw. Eng. (IJCSSE 2012), vol. 6, no. 5, pp. 315-321, 2012.
[16] C. Rathgeb, F. Breitinger and C. Busch, "Alignment-free cancelable iris biometric templates based on adaptive bloom filters," International Conference on biometrics (ICB), IEEE, pp. 1-8, 2013.
[17] C. Rathgeb, F. Breitinger, C. Busch and H. Baier, "On application of bloom filters to iris biometrics," IET Biometrics, vol. 3, no. 4, pp. 207-218, 2014.
[18] C. Rathgeb, C. Busch, "Cancelable multi-biometrics: Mixing iris-codes based on adaptive bloom filters,"

Computers & Security, vol. 42, pp. 1-12, 2014.
[19] J. Hermans, B. Mennink and R. Peeters, "When a bloom filter is a doom filter: security assessment of a novel iris biometric template protection system," International conference of the biometrics special interest group (BIOSIG), IEEE, pp. 1-6, 2014.
[20] J. Bringer, C. Morel and C. Rathgeb, "Security analysis of bloom filter-based iris biometric template protection," International conference on biometrics (ICB), IEEE, pp. 527-534, 2015.
[21] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch and J. Fierrez, "Unlinkable and irreversible biometric template protection based on bloom filters," Information Sciences, vol. 370, pp. 18-32, 2016.
[22] R. Dwivedi, S. Dey, R. Singh and A. Prasad, "A privacy-preserving cancelable iris template generation scheme using decimal encoding and look-up table mapping," Computers & Security, vol. 65, pp. 373-386, 2017.
[23] Y.L. Lai, Z. Jin, A.B.J. Teoh, B.M. Goi, W.S. Yap, T.Y. Chai and C. Rathgeb, "Cancellable iris template generation based on Indexing-First-One hashing," Pattern Recognition, vol. 64, pp. 105-117, 2017.
[24] Z. Jin, J.Y. Hwang, Y.L. Lai, S. Kim and A.B.J. Teoh, "Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing," IEEE Transactions on Information Forensics and Security, vol. 13, no. 2, pp. 393-407, 2018.
[25] M.S. Charikar, "Similarity estimation techniques from rounding algorithms," In Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, pp. 380-388, 2002.
[26] Y. Mroueh, S. Rennie and V. Goel, "Random maxout features," 2015, [Online]. Available: https://arxiv.org/abs/1506.03705.
[27] J. Yanik, D. Strelow, D.A. Ross and R.S. Lin, "The power of comparative reasoning," International Conference on Computer Vision, IEEE, pp. 2431-2438, 2011.
[28] R. Cappelli, M. Ferrara and D. Maltoni, "Minutia cylinder-code: A new representation and matching technique for fingerprint recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pp. 2128-2141, 2010.
[29] M. Ferrara, D. Maltoni and R. Cappelli, "Noninvertible minutia cylinder-code representation," IEEE Transactions on Information Forensics and Security, vol. 7, no. 6, pp. 1727-1737, 2012.
[30] J. Daugman, "How iris recognition works," in Proc. International Conference on Image Processing, IEEE, vol. 1, 2002.

〈저자소개〉

김 진 아 (Jina Kim) 학생회원
2007년 2월: 가천대학교(구.경원대학교) 전자거래학과 졸업
2016년 3월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 정보보호, 생체인증, 프라이버시 향상 기술

정 재 열 (Jae Yeol Jeong) 학생회원
2010년 2월: 고려대학교 수학과 졸업
2013년 8월: 고려대학교 정보보호대학원 정보보호학과 석사
2013년 9월~현재: 고려대학교 정보보호대학원 박사과정
〈관심분야〉 암호 프로토콜, 프라이버시 향상 기술, 생체인증

김 기 성 (Kee Sung Kim) 정회원
2009년 2월: 서울시립대학교 수학과 졸업
2011년 2월: 고려대학교 정보경영공학전문대학원 석사
2015년 8월: 고려대학교 정보보호대학원 박사
2018년 9월: 국가보안기술연구소 선임연구원
2018년 9월~현재: 대구가톨릭대학교 IT공학부 조교수
〈관심분야〉 암호 알고리즘, DB 암호화, 보안 프로토콜

정 익 래 (Ik Rae Jeong) 종신회원
1998년 2월: 고려대학교 전산학과 졸업
2000년 2월: 고려대학교 정보보호학과 석사
2004년 8월: 고려대학교 정보보호학과 박사
2008년 3월~현재: 고려대학교 정보보호대학원 조교수, 부교수, 교수
〈관심분야〉 프라이버시 향상 기술, 데이터베이스 보안, 생체인증