다중 생체 특징 기반 퍼지볼트

손호식*, 노용만*
*한국정보통신대학교
e-mail : yro@icu.ac.kr

Multi-biometric feature based fuzzy vault

Ho sik, Sohn*, Yong man, Ro*
*Information and Communications University

요 약

생체 암호 시스템에서 키로 사용하는 인간의 생체 특징은 외부 환경이나 인간적 요소를 포함하고 있기 때문에 항상 같은 개수, 같은 값의 데이터를 얻을 수 없는 불확실성을 가지고 있다. 퍼지볼트 체계 (Fuzzy vault scheme) [1]는 이러한 불확실성을 가지고 있는 데이터의 특성을 효과적으로 반영할 뿐만 아니라, 등록된 생체 데이터의 보안을 보장해 주는 알고리즘으로서 얼굴, 지문이나 홍채와 같은 단일 생체 특징으로의 적용 방법이 소개되어 왔다 [2,4,5]. 본 논문에서는 퍼지볼트 시스템의 인식 성능을 높이기 위해 이러한 단일 생체 데이터의 불확실성을 보완할 수 있는 다중 생체 특징 (얼굴과 지문) 데이터를 퍼지볼트 체계에 적용하는 방법을 제안하고 실효성을 검증한다.

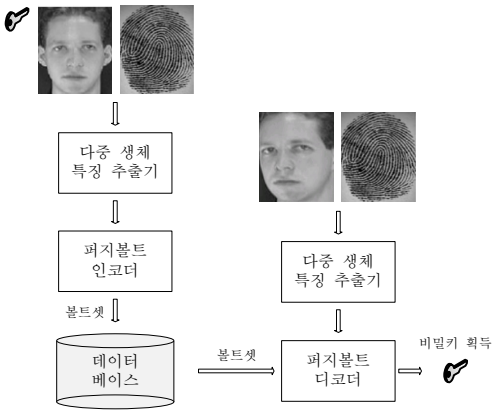
1. 서론

현재 가장 보편적으로 사용되고 있는 패스워드 기반 인증 (Password-based authentication) 시스템은 키 관리와 보관의 어려움, 분실의 위험과 같은 취약점을 가지고 있다 [2]. 생체 암호 시스템은 각 개인의 고유한 생체 특징을 이용하여 이러한 키 기반 암호 시스템의 단점을 보완하고 대체할 수 있는 암호시스템이다. 퍼지 볼트 체계 (Fuzzy vault scheme)는 생체 특징 값과 키를 결합하여, 생체 데이터와 키의 안전성을 보장해주는 생체 암호화 알고리듬으로서, 얼굴, 지문과 홍채 같은 단일 생체 데이터로의 적용 방법이 소개되어 왔다 [2,4,5].

생체 데이터는 외부환경과 인간적 요소에 매우 큰 영향을 받으므로, 항상 같은 개수와 같은 값으로 특징 값을 추출할 수 없다는 단점을 가지고 있다. 예를 들면 지문의 경우 올바른 지문 특징 점 (Minutiae)을 획득하기 위해선 정확한 얼라인먼트 (alignment)가 필요하다 [3]. 또한 얼굴의 경우도 조명이나 자세 등 여러 외부환경에 따라 특징이 달라지므로 매번 정확한 특징 값을 추출한다는 것은 거의 불가능하다. 퍼지볼트 체계가 생체 데이터의 특성을 반영하고 있지만, 이와 같이 불완전성을 가지고 있는 생체 데이터는 퍼지볼트 생체 암호의 정확성에 큰 영향을 끼친다.

본 논문에서는 단일 생체 데이터를 사용하는 기존의 퍼지볼트 체계 응용 방법과 달리 두 생체 데이터를 동시에 사용하여 불완전성을 가지고 있는 생체 데이터를 보완하고 정확성을 높이는 다중 생체 특징 기반 퍼지볼트 시스템을 제안한다. 논문의 구조는 다음과 같다. 2 장에서는 퍼지볼트 시스템의 등록, 인증 절차에 대해 설명하고, 3 장에서는 시스템의 입력이 되는 두 생체 데이터인 얼굴과 지문 특징 점을 추출하는 과정, 그리고 4, 5 장에서는 퍼지볼트 인코딩, 디코딩 과정을 설명하며, 마지막으로 6 장에서는 제안하는 방법의 우수성을 비교실험을 통하여 검증하였다.

2. 퍼지볼트 암호 시스템



(그림 1) 다중 생체 특징 기반 퍼지볼트 시스템.

(그림 1)은 제안하는 퍼지볼트 시스템의 등록, 인증 절차를 나타내는 그림이다. 이 시스템은 등록자의 다중 생체 데이터 (얼굴, 지문)와 비밀 키를 퍼지볼트 인코더의 입력으로 하여, 볼트셋 (Vault Set)을 생성하고 데이터베이스에 저장한다. 퍼지볼트의 장점은 생체 데이터가 등록과 동시에 다른 도메인의 데이터로 변환되므로, 등록 과정 중에 어떠한 생체 정보의 누출도 없다는 것이다 [3].

이 시스템에서 피인증자는 인증을 받기 위하여 자신의 다중 생체 데이터를 입력한다. 이 입력이 시스템에 등록되어 있는 생체 데이터와 일치하는 사람의 것이라면, 등록 시 사용하였던 비밀 키와 동일한 키를 출력한다. 시스템은 이 비밀 키를 비교함으로써 피인증자의 인증 여부를 결정한다. 다음 장에서는 이 생체 데이터를 추출하는 과정과 퍼지볼트 인코딩, 디코딩 과정에 대해 차례로 설명한다.

3. 다중 생체 데이터 입력

퍼지볼트 체계의 입력으로 사용되는 데이터는 16 비트 길이의 생체 특징 데이터 집합이다. 본 논문에서 사용한 생체 특징 값은 얼굴과 지문에서 추출한 16 비트의 데이터 집합이다. 얼굴 특징 값 추출의 경우 [4]의 방법을 사용하였다. [4]의 방법은 얼굴 영상을 Principal Component Analysis (PCA)를 사용하여 생성한 $H\times1$ 얼굴 특징 벡터와 개체 마다 가지고 있는 자신의 고유한 두 개의 $H\times I$ 난수 벡터 (R_1, R_2)의 각 열에 대해 유클리디언 (Euclidean) 거리를 구한다. 여기서 H 는 얼굴 공간에서 가장 큰 고유치를 가지고 있는 고유 벡터의 개수를 의미하고, I 는 추출하려는 얼굴 특징의 개수를 의미한다. 위에서 구한 거리 (b_{1i}, b_{2i}, $i=1\ldots\ldots I$)는 8 비트로 양자화 (b'_{1i}, b'_{2i}, $i=1\ldots\ldots I$)한 후, 두 값을 연접하여 ($b'_{1i}|b'_{2i}$), I 개의 16 비트 얼굴 특징 값으로 사용한다.

지문의 경우 추출된 지문 특징 점을 8 비트 2 차원 좌표로 표현하였다. 추출된 지문 특징 점의 좌표 (x, y)는 연접하여 16 비트의 지문 특징 값으로 사용한다.

위와 같은 방법으로 추출된 16 비트의 얼굴과 지문 특징 값들은 모두 퍼지볼트 시스템의 입력으로 사용된다. 다음 장에서는 이 다중 생체 데이터와 비밀키를 결합하여 볼트셋을 생성하는 인코딩 과정과 피인증자의 인증 여부를 결정하는 디코딩 과정을 설명한다.

4. 다중 생체 특징 기반 퍼지볼트 인코더

퍼지볼트 인코더는 비밀키 (K)와 다중 생체 특징 추출기를 통해 추출된 N 개의 16 비트 생체 특징 값을 입력으로 사용하여 볼트셋을 생성한다. 생성된 볼트셋은 동일한 사람의 생체 특징 값으로 디코딩 하였을 때에만 등록에 사용한 비밀 키와 동일한 키 (K')를 출력한다. 다음은 인코딩 과정을 나타낸다.

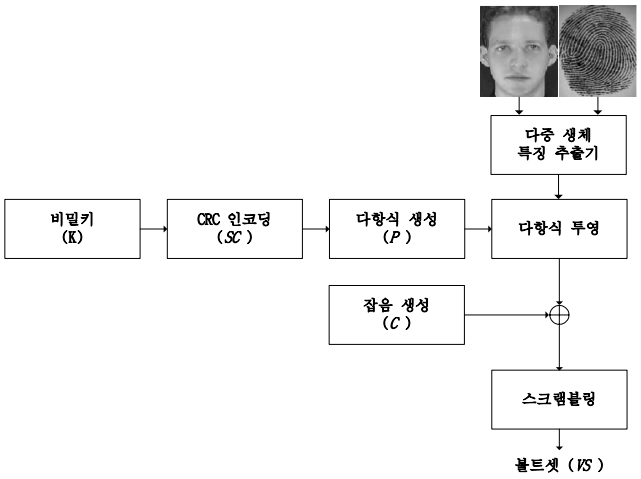
1. $D\cdot16$ 비트의 비밀키 (K)를 CRC (Cyclic Redundancy Check) 인코딩하여 SC 를 생성한다. 생성된 SC 는 $(D+1)\cdot16$ 비트이며, D 는 단계 3 의 다항식 $p(x)$의 차수이다.
   (CRC=16, $g(x)=x^{16}+x^{15}+x^2+1$)

2. SC 를 MSB 부터 순서대로 16 비트 단위로 나누어 $D+1$ 개의 계수 C_D~C_0 를 만들고 다음과 같이 다항식을 생성한다.
   $P(x)=c_Dx^D+c_{D-1}x^{D-1}+\ldots c_1x+c_0$ (1)
   *단계 2 부터의 모든 연산은 갈로아 필드 (Galois field), GF(2^{16}) 안에서 이루어 진다.

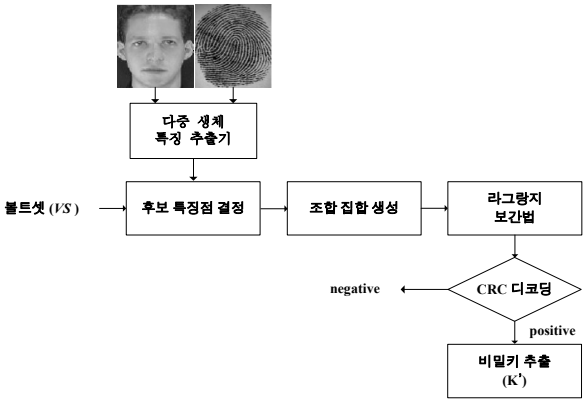
3. 추출된 N 개의 16 비트 생체 데이터 (U)를 다항식 $p(x)$에 투영시킨다. 생체 데이터와 다항식에 투영된 값으로 이루어진 N 쌍의 데이터 집합을 진집합 (Genuine set), G 는 다음과 같다.
   $G=\{(u_i, p(u_i)), i=1,\ldots,N\}$

4. 진집합 G 를 보호하기 위해 임의의 난수를 가짜 특징 값으로 삽입한다. 이때 가짜 특징 값으로 이루어진 잡음 집합 (Chaff set), Ch 는 다음과 같다. $Ch=\{(a_i, b_i), i=1,\ldots,M\}$
   생성된 잡음 집합의 원소는 $p(a_i)\neq b_i$ 의 조건을 만족해야 한다.

5. 생성된 집합 G 와 Ch 의 원소를 모두 섞어 볼트셋, $VS=\{(v_i, w_i), i=1,\ldots,N+M\}$를 생성한다.



(그림 2) 퍼지볼트 인코더



(그림 3) 퍼지볼트 디코더

5. 다중 생체 특징 기반 퍼지볼트 디코더

퍼지볼트 디코더는 볼트셋을 피인증자 (Query user)의 N 개 다중 생체 데이터로 디코딩 하였을 때 일치하는 특징 값이 $D+1$ 개 이상이면, 인코딩에 사용한 비밀 키와 동일한 키를 출력한다 (D: 다항식의 차수). 디코딩 과정은 (그림 3)과 같으며, 각 세부 단계는 다음과 같다.

1. 피인증자의 생체 특징 값과 볼트셋의 원소를 비교하여, 후보 특징 점을 결정한다. 특징 점 비교는 사용자의 각 특징 점을 포함하는 w 크

기의 윈도우 안에 속하는 볼트셋의 원소를 후보 점으로 선택한다.

2. 후보로 선택된 Q 개의 특징 점들에 대해 D+1 개를 선택하는 모든 조합 ($E={}_{Q}C_{D+1}$)의 집합 ($L_1 \sim L_E$)을 생성한다.
3. 각 집합에 대해 라그랑지 보간법 (Lagrange interpolation)을 이용해 다항식을 생성한다.
   $P'(x)=c'_D x^D+c'_{D-1}x^{D-1}+\ldots c'_1 x+c'_0$
4. 생성된 다항식의 모든 계수 ($C'_D \sim C'_0$)를 연접하여 $(D+1)\cdot 16$ 비트의 SC*를 생성한다.
5. 생성된 SC*를 CRC 디코딩하여 나온 16 비트를 확인한다. 이 값이 0 인 경우에 대해서만 SC*의 LSB 16 비트를 제외한 $D\cdot 16$ 비트 (K')를 등록에 사용한 비밀키 (K)와 비교한다. 이때 두 개의 키가 일치하는 경우 본인으로 인증한다.

6. 실험 결과

본 실험에서는 얼굴 또는 지문 한 종류의 생체 특징을 사용하였을 경우의 퍼지볼트 체계 성능과 얼굴과 지문 두 생체 특징을 사용하였을 때의 결과와 비교하여, 제안한 방법의 우수성을 검증한다.

얼굴 특징 점을 추출하기 위해 ORL 데이터 베이스 [6]의 영상을 사용하였다. ORL 데이터 베이스는 40 개 개체에 대해 각각 10 장의 영상으로 이루어져 있으며 본 실험을 위해 10 개의 객체를 선택하였다. 각 객체에 대해 10 장의 영상 중 5 장은 트레이닝에 사용하고, 나머지 5 장은 테스트 집합으로 사용하였다. 아래의 <표 1>은 식 (1)의 차수 (D) 변화에 따른 인증률 (GAR : Genuine Accept Rate)과 오인증률 (FAR : False Accept Rate)을 나타낸다. 실험에서 추출한 특징 점의 개수는 10 개이며, 잡음 집합의 원소 개수는 100 개를 사용하였다. 또한 후보 특징 점을 결정 할 때 사용한 윈도우의 크기는 3 이다.

<표 1> 얼굴 특징 점을 사용한 퍼지볼트 성능

D	2	3	4	5
GAR(%)	84.0	48.0	36.0	24.0
FAR(%)	0.22	0.0	0.0	0.0

지문의 경우 10 개 객체에 대해 각각 8 장의 모의 지문 데이터를 생성하였다. 8 개의 지문 데이터 중 1 개를 선택하여, 인코딩에 사용하였으며, 나머지 7 개의 데이터는 테스트에 사용하였다. 각 지문의 특징 점 개수는 19, 잡음 집합의 원소는 190 개를 사용하였다.

<표 2> 지문 특징 점을 사용한 퍼지볼트 성능

D	7	8	9	10
GAR(%)	95.71	82.85	77.14	70.0
FAR(%)	0.95	0.63	0.63	0.0

다음은 얼굴, 지문 특징 점 모두를 사용하였을 때의 인식률과 오인식률을 나타낸다. 10 개의 개체에 대하여 각각 얼굴 영상 5 장, 지문 5 장을 임의로 조합하여, 100 회의 실험을 하였다. 퍼지볼트 시스템의 등록과 인증을 위해 29 개 (지문 특징 점 19 개, 얼굴 특징 점 10 개)의 특징 점과 290 개의 잡음 원소를 사용하였으며, 후보 특징 점 결정을 위한 윈도우 크기는 3 이다.

<표 3> 얼굴-지문 특징 점을 사용한 퍼지볼트 성능

D	7	8	9	10
GAR(%)	99.28	97.88	93.36	85.74
FAR(%)	1.3	0.81	0.82	0.12

<표 3>의 얼굴, 지문 두 개의 특징 점을 사용한 퍼지볼트의 인증률과 지문 특징 점만을 사용한 경우의 인증률 <표 2>를 비교해 보면, 제안한 방법에서 전체적으로 오인증률이 조금 증가하지만, 그에 비해 인증률은 매우 크게 증가한 것을 확인할 수 있다. 이는 한 종류의 생체 특징을 사용한 퍼지볼트 체계에 비해 두 종류의 생체 특징을 사용하는 제안한 방법의 인식 성능이 더 우수하다는 것을 뒷받침해줄 수 있다.

마지막으로 암호 시스템을 구성함에 있어 가장 중요한 고려 사항인 안전성을 각 경우에 대해 고려해 보면 다음과 같다.

<표 4> 무차별 공격에 필요한 평균 연산량

	얼굴	지문	얼굴+지문
평균 연산양	$4.85\text{x}10^6$	$1.91\text{x}10^{10}$	$8.40\text{x}10^9$

<표 4>는 단일 생체 데이터와 제안한 방법에 대한 안전성을 비교 결과를 나타낸다. 비교 환경은 실험과 같으며, 다항식의 차수는 얼굴의 경우 4 차, 지문은 8 차, 얼굴과 지문 모두를 사용한 경우 8 차를 사용하였다. 안전성은 공격자가 볼트셋을 공격하여 비밀 키를 얻기 위해 걸리는 평균 연산 양을 통해 비교하였다.

세 가지 경우의 결과를 보면, 제안한 방법의 경우 공격자에게 필요한 평균 연산 양이 지문 특징 값만 사용하였을 때의 경우 보다 감소한 것을 알 수 있다. 하지만 ${}_{319}C_9/{}_{29}C_9 \approx 8.40\text{x}10^9$ 의 평균 연산 양은 3.4 Hz 프로세스 컴퓨터에서 약 70 년이나 걸리는 시간 양이다. 따라서 제안하는 방법은 지문 생체 데이터만을 사용하였을 때보다는 공격에 필요한 평균 연산 양이 조금 감소하지만, 이 시간 양 역시 공격자가 볼트셋 공격에 성공하기에는 불가능한 양이라고 할 수 있다.

7. 결론

생체 특징 값은 외부 환경이나 인간적 요소에 의해 매우 큰 영향을 받는 데이터이다. 따라서 인간의 생체 특징을 항상 같은 개수, 같은 값으로 추출하는 것은 거의 불가능한 일이며, 생체 암호 시스템의 인증률을 결정하는 중요한 요인 중 하나이다. 본 논문에서는 이러한 생체 데이터의 취약성을 보완하기 위해 단일 생체 특징이 아닌, 얼굴, 지문 두 종류의 생체 특징을 사용하였다. 얼굴, 지문 두 종류의 생체 특징을 모두 사용하는 방법은 불확실성을 가지고 있는 두 종류의 데이터를 서로 보완하는 역할을 하게 되어, 전체 시스템의 인증률을 높이는 결과를 가져왔다. 실험결과에서 알 수 있듯이, 제안한 방법은 피인증자의

생체 특징 점이 퍼지볼트 시스템에 등록되어있는 등록자의 템플릿과 일치할 확률을 증가시켜, 전체적으로 더 높은 인증률을 나타내었다.

참고문헌

[1] A. Juels, M. Sudan. A fuzzy vault scheme. Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on 2002 Page(s):408

[2] K. Nandakumar, A.K. Jain, and S. Pankanti. Fingerprint-Based Fuzzy Vault : Implementation and Performance. Information Forensics and Security, IEEE Transactions on Volume 2, Issue 4, Dec. 2007 Page(s):744-757

[3] U. Uludag and A. K. Jain, “Securing fingerprint template: Fuzzy vault with helper data,” in Proc. CVPR Workshop Privacy Research Vision, New York, Jun. 2006, p. 163.

[4] Y. Wang and K.N. Plataniotis. Fuzzy Vault for Face Based Cryptographic key Generation. Biometrics Symposium, 200711-13 Sept. 2007 Page(s):1-6

[5] Y.J Lee, K. Bae, S.J. Lee, K.R. Park and J. Kim. Biometric key Binding: Fuzzy Vault Based on Iris Images. Lecture Notes in Computer Science, Volume 4642/2007, Page(s):800-808

[6] ATT Laboratories Cambridge, ORE face databse, www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html