논문 2010-47SP-5-21

영역기반 주성분 분석 방법과 보조정보를 이용한 얼굴정보의 비트열 변환 방법

(A Study on A Biometric Bits Extraction Method Using Subpattern-based PCA and A Helper Data)

이 형 구*, 정 호 기*

(Hyunggu Lee and Ho Gi Jung)

요 약

생체인식은 개인의 유일하면서 변화하지 않는 생체의 특징을 이용하여 개인의 본인 여부를 판별하는 방법으로써 널리 사용되어 왔다. 생체정보의 고유 불변한 특징을 저장하는 것은 개인정보의 노출에 따른 보안상의 문제점을 갖고 있으며 이를 해결하기 위해 제안된 방법이 가변생체인식 (cancelable biometrics)이다. 가변생체인식은 생체정보의 도난이나 도용으로부터 강인하며 재생성 가능한 생체템플릿을 제공하는 생체 인식방법이다. 본 논문에서는 변환 생체인식의 한 가지 방법으로써 얼굴 생체정보의 새로운 이진화 방법을 제안한다. 얼굴 생체정보의 이진화를 위한 특징추출은 얼굴정보의 부분적 변화에 강인한 영역기반 주성분 분석 (Subpattern-based PCA) 을 이용하였으며 이로부터 얻어진 특징을 보조정보에 기반한 방법으로 이진화 하였다. 획득된 이진비트열은 영역기반 주성분 분석의 사용으로 여러 얼굴 영역의 고려와 함께, 선택된 주성분 개수만큼의 계수들에 대한 이진화 값들을 포함하고 있다. 이러한 서로 다른 얼굴영역의 여러 주성분들에서 추출된 이진비트열중 구분력이 좋은 비트 값들을 선택하였으며, 선택된 비트 값들은 이진화를 위한 보조 정보가 노출된 경우에서도 원 얼굴특징벡터보다 향상된 인식성능을 보여준다.

Abstract

Unique and invariant biometric characteristics have been used for secure user authentication. Storing original biometric data is not acceptable due to privacy and security concerns of biometric technology. In order to enhance the security of the biometric data, the cancelable biometrics was introduced. Using revocable and non-invertible transformation, the cancelable biometrics can provide a way of more secure biometric authentication. In this paper, we present a new cancelable bits extraction method for the facial data. For the feature extraction, the Subpattern-based Principle Component Analysis (PCA) is adopted. The Subpattern-based PCA divides a whole image into a set of partitioned subpatterns and extracts principle components from each subpattern area. The feature extracted by using Subpattern-based PCA is discretized with a helper data based method. The elements of the obtained bits are evaluated and ordered according to a measure based on the fisher criterion. Finally, the most discriminative bits are chosen as the biometric bits string and used for authentication of each identity. Even if the generated bits string is compromised, new bits string can be generated simply by changing the helper data. Because, the helper data utilizes partial information of the feature, the proposed method does not reveal privacy sensitive biometric information of the user. For a security evaluation of the proposed method, a scenario in which the helper is compromised by an adversary is also considered.

Keywords: Biometric bits string extraction, cancelable biometrics, Subpattern-based PCA

* 정회원, 연세대학교, 생체인식연구센터
(Yonsei University, Biometric Engineering Research Center)

※ 본 연구는 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지정 연세대학교 생체인식연구센터의 지원을 받아 이루어졌습니다(R112002105080020(2010)).
접수일자: 2010년2월17일, 수정완료일: 2010년5월24일

I. 서 론

생체인식은 개인마다 유일하며 고유한 생체정보를 이용하여 본인의 인증을 하는 방법이다. 원래의 생체정보 자체가 노출될 경우 이를 대신할 새로운 생체정보를 만들어 내는 것은 불가능 하며 따라서 생체인식에서 생체정보 자체에 대한 안전한 운용방법이 필요하다. 이러한 문제점에 대한 해결책으로 제시된 것이 가변생체인식 (Cancelable Biometrics)이며 이것은 원래의 생체정보를 비가역적으로 변환하여 이를 생체인식에 이용하는 방법이다[1]. 본 논문에서는 변환생체인식 중에서 생체정보를 이진화하는 방법에 그 초점을 맞추고 있으며 새로운 생체정보 이진화 방법을 제안한다. 제안하는 생체정보 이진화 방법은 영역기반 주성분 분석 (Subpattern-based PCA[2])으로부터 얻어진 여러 얼굴 영역에서 추출된 주성분 분석 (PCA)의 계수들을 보조정보에 기반한 방법[3]으로 이진화 한 후, 구분력이 좋은 특징 성분들을 선택하여 인식에 이용하는 방법이다. 영역기반 주성분 분석 방법의 사용으로 인해 얼굴 영역에서의 부분적 변화에 강인하며, 보조정보에 기반한 방법으로 보다 일관성 있는 이진화 값의 추출이 가능하다. 이진화 된 성분 중 구분력이 좋은 특징의 선별 방법은 이진화 된 계수가 위치한 얼굴 영역 및 주성분 분석 방법에서의 계수 성분에 따라 개별적으로 평가되고 선별되었다.

일반적으로 생체정보 이산화 방법은 개인정보가 도난당한 경우에서의 성능이 원래의 생체정보가 갖는 성능보다 떨어지는데, 이것은 생체정보 이산화 과정에서 원래의 생체정보가 가지는 구분력이 보존되지 않기 때문이다. 따라서 본 논문에서는 개인정보가 도난당한 경우를 고려하여, 이진화 된 성분을 선별 하였으며 선택된 이진 특징 값은 개인정보가 도난당한 경우에서도 원 생체정보 보다 향상된 인식성능을 보여준다. 그 밖에도 제안방법은 보조정보에 기반한 이진화 방법을 이용하여 보조정보로부터 원 생체정보로의 복원이 힘들며, 이진화 값이 '0'또는 '1'의 값을 갖도록 보조정보를 바꾸어 간편하게 새로운 이진화 템플릿을 생산 및 변경할 수 있다. 이러한 가변생체인식 방법으로써의 생체정보 이진화 방법이 가져야 할 특성은 다음과 같이 정리할 수 있다.

i) 변환된 생체정보는 원 생체정보 보다 향상된 인식성능을 가져야 하며, 보조정보가 도난 된 경우에서도 원래의 생체정보와 비교하여 크게 떨어지지 않는 인식성능을 가져야 한다. ii) 변환된 생체정보는 원 생체정보와 다르며 다수의 변환방법을 제공하여야 한다 (변환성/재생산성). iii) 보조정보와 변환된 생체정보로부터 원 생체정보의 복원이 쉽지 않아야 한다 (불가역성). 본 논문의 제안방법은 위 조건들을 모두 만족하며 이를 실험적으로 검증하였다.

기존 생체정보 이진화 방법

생체정보 이진화 방법은 크게 키 생성 (key generation)방법과 키 결합 (key binding)방법으로 분류하였다. 키 생성방법은 생체정보로부터 고유한 이진비트열을 얻는 과정이며 따라서 해당 생체에 대하여 새로운 키의 생성이나 변경이 불가능한 방법이다. 반면 키 결합 방법은 원래 생체정보에 어떠한 변환을 가하거나 보조정보를 생성하여 변경 가능한 이진비트열을 추출하는 방법을 말한다.

가. 키 생성 (key generation)방법

Vielhauer et al.[4]은 서명인식방법에서의 생체정보 이진화 방법을 제안하였다. 이 방법은 각 개인의 입력된 서명 특징 성분들에 대하여 각 특징의 값이 분포할 수 있는 범위에 대한 정보를 저장하고 이를 이용해 생체정보를 이진화 하는 방법이다. 저자들은 50개의 특징 성분들을 이용하였고 11명의 사용자들에 대해 인식률 기준으로 0%의 FAR과 7.05%의 FRR을 갖는 인식성능을 보여주었다. Qi Han et. al은 유사한 방법으로 각 개인마다 비균등한 이진화 영역을 정의하여 생체정보를 이진화하는 방법을 제안하였고 양자화 단위를 개인의 표준편차를 근거로 결정하였다[5]. 또한 Chen et. al은 가능도비 (likelihood ratio)를 이용하여 개인마다 생체정보 이진화를 위한 구간을 정의하였고 이에 따라 생체정보를 이진화하였다[6]. 이 방법을 확장시켜 저자들은 가능도비에 의한 확률 분포에 근거하여 본인에 대한 검출율을 높이기 위한 비트 할당방법을 제안하였다[7]. 이러한 키 생성 방법들은 개인마다 생체정보 이진화를 위한 전체 구간을 저장해야 하는 번거로움이 있으며 이러한 이진화 방법 자체가 생성되는 키의 변경 및 재생성을 지원하지 못하는 단점이 있다.

Tuyls et al.[8]은 각 개인의 생체특징들 중에서 신호대 잡음 비가 높은 특징들을 선택하여 이진화 하는 방법을 제안하였다. 이진화 방법으로는 특징 값의 평균

값을 기준으로 각각의 특징 값을 '0'또는 '1'의 값으로 이진화하였고 이진화 된 값은 오류정정부호를 통해 보정되었다. 그들의 방법은 지문 데이터베이스에 대해 실험되었고 40비트 특징에 대하여 인식률 오류 기준으로 4.2%의 오류율을 보였다. 그러나 이 방법은 생성된 이진비트열값의 변경 및 재생성을 고려하지 않았다.

나. 키 결합 (key binding)방법

Teoh et al.[9~10]은 BioHash로 명명된 변환생체인식 기술로써의 생체정보 이진화 방법을 제안하였다. 저자들은 생체정보의 이진화 이전에 재생성 가능한 생체정보의 변환을 위한 투영 행렬 (projection matrix)을 생성한다. 이 투영행렬은 난수발생기 (pseudo-random numbers generator)를 이용하여 각 개인마다 다르게 지정되었다. 입력 생체특징을 투영행렬을 통해 투영하여 변환시킨 다음 이 변환특징을 평균 값에 근거하여 이진화 하는 방법이며, 투영행렬을 변경시켜 새로운 이진화 값을 재생성 할 수 있다. 이 방법은 원래의 생체정보로부터 투영행렬을 변경함에 따라 여러 가지의 이진비트열을 생성할 수는 있으나 사용자가 지정한 이진비트열 값을 생성하도록 투영행렬을 정의하지는 못한다. 즉, 키 값을 지정할 수 없으며 또한 투영행렬 자체에 성능의 의존성이 큰 단점이 있다. 저자들은 이 방법을 발전시켜 선형 판별 분석 (Linear Discriminant Analysis (LDA))에 의해 추출된 특징을 이산화 하는 방법에 적용하였다[11]. 해당 논문에서 저자들은 이산화를 위한 특징 추출 방법으로 높은 인식성능을 갖는 선형 판별 분석방법을 이용해 투영행렬에 대한 의존성을 줄이고자 했으며 또한 여러 개의 투영행렬을 이용하여 성능을 향상 시켰다.

Linnartz 와 Tuyls는 보조정보 (helper data)를 이용한 생체정보의 이진화 방법을 제안하였는데 이 방법은 워터마킹 (watermarking) 등에서 이용되는 정보 은닉을 위한 Quantization Index Modulation (QIM)라는 방식에 기반한 것이다[3, 12]. 이 방법은 특징의 통계적인 분포정보로부터 양자화 단위 q 를 결정하고, 등록 시 비밀 비트정보인 S의 값이 '0' 또는 '1' 인지의 여부에 따라 보조정보 W 를 생성한다. 검증 시 보조정보 W 와 생체정보 X의 합으로부터 '0' 또는 '1'의 비트를 생성하게 되며 W의 값은 다음과 같이 정의된다:

$$W=\begin{cases}(2n+1/2)q-x & ,\text{if } S=1\\(2n-1/2)q-x & ,\text{if } S=0\end{cases} \quad (1)$$

여기서 n 은 $-q< W< q$ 을 만족 시키는 자연수이며 양자화 단위 q 는 각 개인의 생체정보 X 로부터 얻어진 표준편차를 이용해 결정된다. 이 방법은 보조정보 W 를 결정하기 위한 비밀 비트정보인 S의 값을 바꾸어 보조정보 W 를 다시 생성함으로써, 새로운 생체 템플릿을 제공한다. 그러나 저자들은 해당 논문에서 실제 생체정보를 이용한 실험 결과를 보이지 않았다.

기존의 생체정보 이진화 방법들은 연속적인 생체정보를 이진화 함으로써 정보의 손실이 발생하고 이로 인한 성능의 저하가 불가피하다. 또한 이진화를 위한 보조정보가 공격자에게 노출되어 원래의 생체정보에 대한 추측에 이용되는 등 보안상의 문제점을 갖고 있다. Teoh et. al.의 방법[10, 12]에서 보여지듯이 개인의 보조정보로써의 투영행렬이 도난 되었을 경우 인식 성능의 큰 저하와 함께 원래 생체정보에 대한 보안상의 문제도 발생한다[13].

본 논문에서는 영역기반의 주성분 분석 방법을 통해 부분적인 변화에 강인한 특징을 얻고 이를 이진화한 특징 중 구분력이 좋은 특징들을 선택하여 보조정보가 도난 되었을 경우에서도 향상된 성능을 갖는 생체정보 이진화 방법을 제안한다. Ⅱ 장에서는 제안방법을 설명하고, Ⅲ장에서 제안방법의 성능과 보안성에 대해 the PIE face database[14]를 통해 평가한다.

Ⅱ. 제안된 얼굴생체정보의 이진화 방법

이 장에서는 본 논문에서 제안된 얼굴생체정보의 이진화 방법에 대해 설명한다. 제안방법은 크게 다음과 같은 세 가지 과정으로 요약된다.

a. 입력된 얼굴영상에 대하여 영역기반 주성분 분석 방법을 이용한 특징 추출과정

b. 추출된 특징을 보조정보를 이용하여 이진비트열로 변환하는 과정

c. 획득된 이진비트열중 구분력이 좋은 성분을 선별하는 과정

첫 번째 과정은 이진화에 이용되는 특징을 영역기반 주성분 분석 방법을 통해 생성하는 과정이다[2]. 영역기반 주성분 분석 방법은 입력 얼굴영상을 일정한 크기의 영역으로 구분하고, 각 영역에 대하여 주성분 분석 방법을 적용하여 영역마다 특징을 추출하는 방법이다. 각 영역에서 생성된 특징 벡터들은 묶어져서 하나의 특징 벡터를 만들게 된다. 일반적인 외형기반 특징 추출 방

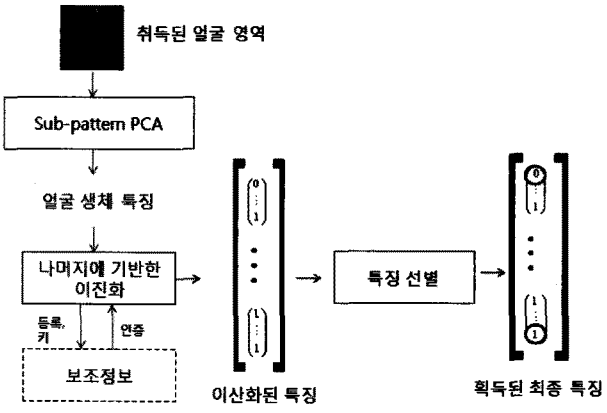

그림 1. 제안방법의 전체과정
Fig. 1. Overall block diagram of the proposed method.

법 (Appearance based feature extractor)은 얼굴 영역의 부분적인 변화가 획득되는 전체 특징 값들에 영향을 주는 경향이 있으나 영역 기반 주성분 분석 방법에서는 이러한 부분적 변화를 분리하여 처리함으로써 향상된 성능을 보여준다.

두 번째 과정에서는, 영역기반 주성분 분석 방법으로 생성된 특징을 보조정보에 기반한 이진화 방법을 통해 이진화 한다. 이 과정은 Linnartz와 Tuyls의 보조정보를 이용한 생체정보의 이진화 방법을 따른다[3]. 등록과정에서는 각각의 특징 값에 대하여 '0' 또는 '1'의 비트 단위 이진화를 위한 보조정보가 생성된다. 검증 시, 입력된 생체정보로부터 특징이 추출되면, 보조정보를 이용하여 각각의 특징마다 이진 비트 값을 얻고 전체 특징들로부터 이진 비트열을 생성한다.

마지막 단계에서는, 생성된 이진 비트열 중에서 구분력이 좋은 성분만을 선별하며 인식에 이용한다. 또한 이진비트열의 구분력에 따른 선별 과정에 있어 보조정보가 도난된 경우를 고려하여 개인의 보조정보가 노출된 경우에서도 이진화된 특징의 인식 성능이 원래의 생체정보가 가지는 인식 성능보다 향상 되었다. 생성된 얼굴생체템플릿의 도난이 발생할 경우 할당되는 비트의 값을 달리하여 보조정보를 변경함으로써 새로운 이진비트열을 생성할 수 있다. 그림 1은 제안하는 방법의 전체과정을 보여준다.

1. 영역기반 주성분 분석 방법 (Subpattern-based PCA)을 이용한 특징 추출

얼굴 영상의 영역을 중복되지 않는 동일한 크기의 영역으로 분리한 후 각각의 영역에 대하여 주성분 분석을 수행한다.

$$S_i\Phi_i = \Phi_i\Lambda_i,\ 1 \le i \le k$$

전체영역의 개수는 k 개 이며 각 영역마다 공분산행렬 S_i 가 정의되고 이에 따라 고유값 행렬 Λ_i 와 고유벡터 행렬 Φ_i 가 정의 된다.

입력 영상 x 를 동일한 k 개의 영역으로 구분한 뒤 각각의 영역 $x_i, 1 \le i \le k$에 대한 주성분 특징을 다음과 같이 계산할 수 있다.

$$p_i = \Phi_i^T x_i,\ 1 \le i \le k \tag{2}$$

최종적으로 입력 영상 x 에 대한 영역기반 주성분 분석 방법의 추출된 특징은 각 영역들에서 생성된 특징 벡터를 모아서 다음과 같은 하나의 특징 벡터로 만들어진다.

$$p = [p_1^T\ p_2^T \cdots p_k^T]^T \tag{3}$$

2. 보조정보에 기반한 이진 값 생성

영역기반 주성분 분석 방법으로 추출된 특징 벡터 p 의 각 요소 별로 보조정보를 생성하여 이진화에 이용한다. 특징 벡터를 이루는 요소 중 하나를 x 라 하고 수식 (1) 로부터 생성된 보조정보를 W 라 할 때 다음의 식에 따라 이진화 값을 생성한다.

$$bit(x, W) = \begin{cases} S=1, \text{if } 2nq \le x + W < (2n+1)q \\ S=0, \text{if } (2n-1)q \le x + W < 2nq \end{cases} \tag{4}$$

여기서 n 은 자연수이며 통계적 특성인 표준편차 값을 근거로 양자화 단위 q 가 결정된다. 식 (5)에서와 같이, 양자화 단위 q 는 각 사람의 표준편차 값을 $q_j, 1 \le j \le C$ 라 할 때 모든 사람들의 경우에 대해 평균을 취한 뒤, 이 값에 실험적으로 결정된 상수 Con 을 곱하여 얻어진다.

$$q = Con \times \frac{1}{C}\sum_{j=1}^{C} q_j,\ Con : constant \tag{5}$$

양자화 단위의 기준으로 각 사람으로부터 획득된 표준편차 값을 이용하지 않고 이것의 모든 사람에 대한 평균값을 사용한 이유는 각 사람마다 표준편차를 정확히 구하기 위한 충분한 수의 영상 취득이 어려우며 각 개인마다 발생할 수 있는 변이를 특정하기 어렵기 때문이다. 따라서 모든 사람에 대한 평균 값을 기준치로 이

용하였다.

3. 구분력이 좋은 이진 성분을 선별하는 방법

이번 절에서는 영역기반 주성분 분석에 의해 추출된 특징을 이진화한 값 중에서 구분력이 좋은 성분들을 선별하여 이용함으로써 성능을 높이는 방법에 대해 설명한다. 이진화 된 각각의 성분들이 가지는 구분력에 대한 기준은 fisher criterion의 개념에 기반한다. 즉, 동일인에 대한 이진화 값의 변화가 적으면서 타인간의 이진화 값의 차이가 큰 성분을 선택하는 방법이다. 그러나 본 논문에서 이용한 이진화 방법 자체가 각 개인으로부터 정의된 보조정보를 이용하므로 구분력에 대한 평가 기준은 개인의 보조정보에 대한 고려를 포함하여야 한다. 보조정보를 이용한 생체정보 이진화에서는, 개인마다 다르게 정의된 보조정보를 이용할 경우 매우 좋은 구분력을 갖게 된다. 그러나 이진화를 위한 보조정보가 노출된 경우 성능의 저하가 발생한다. 따라서 보조정보를 이용한 생체정보 이진화를 평가할 때 보조정보가 도난 된 경우에서의 구분력을 기준으로 평가하여야 한다. 이에 대한 방법을 본 논문에선 수식 (6) 과 같이 정의한다. 수식 (6)에서 분자성분은 서로 다른 타인에게서 추출된 이진화 값들의 차이를 평가하며, 분모성분은 동일인의 서로 다른 입력들로부터 추출된 이진화 값들의 일관성에 대하여 평가한다. 함수 $bit(x, W)$ 는 어떤 사람의 생체특징 x 와 그 사람의 생체특징으로부터 만들어진 보조정보 W 를 이용하여 비트 값을 만드는 함수이며 수식 (4) 에 정의되었다. 수식 (6)의 분모에서는 동일인의 서로 다른 입력들에 대하여 그 사람의 보조정보를 이용하여 얻어진 비트 값들의 차이를 해밍거리(Hamming distance)공간에서 측정하였다. 식 (6)에서의 함수 $distHD(,)$ 는 거리 측정의 단위로 비트 값들 간의 해밍거리를 이용한다는 것을 의미한다. 식 (6)의 분자에서는 서로 다른 두 명의 생체정보들에서 비트 값들을 추출할 때 두 명중 한 명의 보조정보가 다른 사람에게 노출 되었다고 가정하고 비트 값들을 얻은 뒤 그 값들 사이의 거리를 측정 하였다. 이와 같이, 본 논문에서는 개인의 보조정보가 노출된 경우에서 좋은 구분력을 가지는 특징을 식 (6)에 근거하여 선별하였다.

* 제안방법의 비가역성 과 재생산성

(Non-invertibility and revocability of the proposed method)

변환생체인식 방법은 변환된 생체정보나 변환방법을 알더라도 원래의 생체정보를 복원하기 쉽지 않아야 한다. 제안방법에서는 생성된 이진비트열과 보조정보를 알더라도 원 생체정보의 복원이 힘든데, 이것은 식 (1)에서 양자화 단위 q의 계수에 해당되는 정수 부분 n 은 저장하지 않고 나머지에 해당되는 성분을 이용해 보조정보를 생성하였기 때문이다. 식 (4)에서도 확인할 수 있듯이 보조정보와 해당 비트 값에 대하여 이를 만족하는 생체정보 x의 범위는 무수히 많은 자연수 n 에 대하여 가능하다. 따라서 보조정보를 이용한 생체정보 이진화 방법은, 보조정보가 노출되었을 경우뿐만 아니라 가변템플릿으로써의 이진화된 비트열이 도난 당한 경우에서도 원래 생체정보의 추출이 어렵다.

변환생체인식 방법에서는 저장된 변환생체템플릿이 노출되더라도 새로운 변환생체템플릿을 다시 생성할 수 있어야 하며, 제안방법의 재생산성은 보조정보를 이용한 이산화 방법에 의해 제공된다. 식 (1)의 보조정보 생성 시 할당되는 비트 값은 임의로 선택 가능하며 이것의 가짓수는 이진비트열이 N 차원 일 경우 2^N 가지의 조합이 가능하다. 예를 들면, 선택된 이진비트열의 차원이 80차원인 경우 $2^N = 2^{80} \cong 10^{24}$ 가지의 변환생체템플릿을 재생산할 수 있다.

III. 실 험

1. 실험 조건 및 평가 방법

제안된 얼굴생체정보의 이진화 방법을 검증하기 위해서 the PIE face database가 이용되었다[14]. 실험에 이용된 것은 정면 포즈에 대하여 조명변화와 일부 표정변화를 포함한 영상들이며, 등록시, 30명에 대하여 각 사람당 5장의 영상을 무작위로 선택하여 이용하였다.

또한 각 사람의 영상들로부터 특징벡터계수들의 표준편차를 구하고, 이것을 모든 사람에 대해 평균하여 이진화 과정의 양자화 단위로 이용하였다. 각 사람의 5

$$Distinctiveness = \frac{BetweenScatter}{WithinScatter} = \frac{\sum_{j \neq i} \sum_{m \in C_j} \sum_{n \in C_i} distHD(bit(x_{m,j},\ W_i), bit(x_{n,i},\ W_i))}{\sum_{j} \sum_{m \neq n,\ m,n \in C_j} distHD(bit(x_{m,j},\ W_j), bit(x_{n,j},\ W_j))} \quad (6)$$

그림 2. 사용되어진 the PIE face database 영상의 예시
Fig. 2. Examples of the illumination variation in images obtained from the PIE face database.

장의 영상으로부터 얻은 특징계수들을 평균하여 이진화를 위한 보조정보 생성에 이용하였으며, 이진화를 위한 비트 할당은 '0' 과 '1' 값들 중 무작위로 선택하였다. 실제 구현 시, 비트 할당은 각 개인의 ID를 근거로 할당하는 방법도 가능하다. 평가 시 각각의 사람으로부터 무작위로 15장의 영상을 선택하고 30명으로부터 전체 15x30=450장의 영상을 제안방법의 검증에 이용하였다. 획득된 영상들은 64x64크기로 입력되었으며, 영역기반의 주성분 분석 방법으로 특징을 추출하기 위해 16x16 크기의 총 16개의 영역으로 구분하고 각 단위영역(16x16)에 대하여 주성분 분석 방법을 적용하였다. 각 단위 영역마다 주성분 분석을 통하여 10개의 특징을 추출하였고 16개의 영역에서 전체 160개의 특징계수를 영역기반 주성분 분석 방법에 의해 추출하였다. 영역기반 주성분 분석 방법에서 얻어진 160개의 특징계수는 보조정보에 기반한 이진화 방법에 의하여 160차원의 이진비트열로 변환되었으며 이중에서 구분력 기준으로 80개를 선별하여 최종적인 이진비트열로 선택하였다. 생성된 이진비트열의 성능을 평가하기 위해 검증 (verification)에서의 Equal Error Rate (EER), 개인정보가 노출되었을 경우에서의 EER, 그리고 가변생체템플릿으로써의 재생산성 (revocability)을 이용하였다.

2. 인식성능 평가 (performance evaluation)

제안된 방법으로 생성된 이진비트열에 대한 인식성능을 일반적인 주성분 분석 방법과 영역기반 주성분 분석 방법에 대해서 비교하였다. 등록 시 각 사람당 하나의 보조정보를 생성하여 이진 비트열 생성에 이용하였다. 개인의 보조정보 생성 시 임의의 비트 값 을 할당하였기 때문에, 이러한 보조정보를 이용해 생성된 각 개인의 이진비트열은 매우 좋은 구분력을 가진다. 표 1.에서 보여지듯이 생성된 이진비트열은 원 생체특징의 성능과 비교하여 매우 뛰어난 성능을 보여준다.

실제로 개인마다 서로 다른 비트할당을 하여 보조정보를 생성하고 이로부터 생성되는 서로 다른 개인들의

표 1. 제안된 이진화 방법의 인식 성능
Table 1. Performance of the proposed method for the PIE face database.

	입력특징벡터		제안 이진화 방법	
	PCA, 160차원	영역 기반 PCA, 160 차원	영역기반 PCA의 이진화, 160차원	영역기반 PCA의 이진화 및 특징 선택, 80차원
EER (%)	48	41.6	0.7	2.1

비트 값들은 서로 독립적이다. 이 때 서로 다른 개인의 이진비트열을 비교하여 만들어지는 분포 (imposter distribuuuuuuu는 이항 분포의 형태를 갖게 되고, 이항 분포에서 독립 수행의 횟수에 해당하는 것이 비트열의 크기가 된다. 독립적인 N 비트들로 만들어지는 이항 분포는 다음의 식을 따른다[11, 15].

$$f(x) = \frac{N!}{\lambda!(N-\lambda)!} 0.5^{\lambda}(1-0.5)^{(N-\lambda)}$$

N 은 비트 수 이며, $x = \lambda/N, 0 \le x \le 1$ 은 총 수행 횟수 중 0.5의 확률로 일어난 사건 횟수의 비율이며, 본 논문에서는 서로 다른 두 사람으로부터 얻어진 이진비트열들 사이의 해밍거리를 말한다. 또한 이 경우, 서로 다른 두 사람간의 이진비트열을 비교하여 만들어지는 분포의 평균은 0.5 이며 표준편차 값은 $\sqrt{0.5(1-0.5)/N}$ 과 같다. 따라서 비트열의 크기가 커질 경우 타인간의 이진비트열 비교로 만들어진 분포의 평균은 해밍거리상에서 0.5로 고정되고 표준 편차 값은 작아진다. 개인마다 서로 독립적인 특징이 추출되었을 경우, 해당 특징의 성능은 타인간의 거리 비교로 획득된 분포에 크게 영향을 받는다. 이러한 개념은 Teoh et. al.[11]이 지적했으며, 결과적으로 독립적인 N 비트 특징의 인식 성능은 N 비트의 크기가 커짐에 따라 향상된다. 이것은 표 1.의 제안방법에서 이진비트열의 차원을 줄이기 전과 줄인 후의 성능이 차이가 나는 것을 설명해 준다.

표 1.에서 보여지는 결과는 각 개인 특징의 이진화에 이용되는 보조정보가 개인마다 독립적으로 지정되었기 때문에 가능하다. 따라서 이러한 보조정보가 노출되었을 경우에 대한 고려 역시 필요하며 다음의 단락에서 상세히 다루었다.

3. 개인정보가 도난 되었을 경우의 인식 성능 (performance under the stolen-token scenario)

변환생체인식에서 일반적으로 개인정보가 노출되었을 경우 보안상의 문제점이 발생하는데 이것은 개인마다 다르게 선택된 개인정보가 공격자에게 노출되어 각 개인의 변환생체템플릿에 대한 공격에 이용되기 때문이다. 본 논문에서는, 각 개인의 보조정보가 공격자에 해당되는 다른 모든 사람들에게 노출되었다고 가정하였다. 한 사람의 이진비트열을 추측하는데 그 사람의 노출된 보조정보와 다른 모든 사람의 생체정보가 이용되었다.

표 2.의 결과에서 보여지듯이 개인의 보조정보가 노출된 상태에서의 이진비트열이 가지는 성능은 보조정보가 노출되지 않은 경우보다 좋지 못하나, 원 생체정보의 성능과 비교하였을 경우 보다 향상된 인식 성능을 보인다. 제안된 이진화 방법은 개인정보가 노출된 경우에서도 22.9%와 19.9%의 EER을 각각 영역기반 주성분 분석의 특징 과 영역기반 주성분 분석의 선택된 특징에서 보여준다. 위 결과에서 이진비트열 성분 중 구분력이 좋은 성분들을 선택하여 보다 향상된 인식성능을 얻음을 확인 할 수 있다.

표 2. 개인정보가 노출된 상황에서의 제안된 이진화 방법의 인식 성능

Table 2. Performance of proposed method for the PIE database under stolen-token scenario.

	입력특징벡터		제안 이진화 방법	
	PCA, 160차원	영역 기반 PCA, 160 차원	영역기반 PCA의 이진화, 160차원	영역기반 PCA의 이진화 및 특징 선택, 80차원
EER (%)	48	41.6	22.9	19.9

4. 변환성 평가 (Revocability evaluation)

제안된 얼굴생체정보의 이진화 방법은 재생산 가능한 이진비트열을 생성한다. 재생산된 이진비트열들은 서로 정합되어선 안 되며 이를 평가하기 위한 방법이 변환성 평가이다. 변환성의 평가를 위해 동일인에 대하여 여러 개의 보조정보를 무작위로 할당하고 평가입력들로부터 서로 다르게 변환된 이진비트열이 생성되는지 확인하였다. 본 논문에서는 개인당 10개의 보조정보를 독립적으로 할당하였다. 이 경우 동일인에게서 서로 다른 보조정보를 이용해 생성된 비트열들로부터 만들어진 분포를 pseudo genuine distribution이라 한다. 일반적으로 생체인식에서 이용되는 분포들로는 genuine distribution과 imposter distribution이 있다.

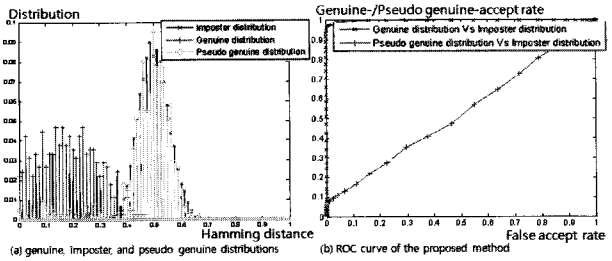

그림 3. 제안된 방법의 Genuine, imposter distribution과 ROC curve를 통한 변환성 측정결과

Fig. 3. Genuine, imposter, pseudo genuine distributions and ROC curve between pseudo genuine Vs imposter distributions for the proposed method.

본 논문에서의 genuine distribution은 동일인으로부터 동일한 보조정보를 이용하여 얻어진 분포이며 imposter distribution은 서로 다른 보조정보로부터 획득된 타인간의 이진비트열을 비교하여 만들어진 분포이다. 높은 변환성을 갖기 위해선 pseudo genuine distribution이 imposter distribution과 비슷하며 genuine distribution과는 차이를 보여야 한다. 제안방법의 이진비트열은 이러한 조건을 만족하고 이것은 그림 3.의 (a)에서 확인할 수 있다. 이상적인 변환성을 갖는 경우는 pseudo genuine distribution과 imposter distribution가 비슷하므로 이것들의 Receiver Operating Characteristic (ROC) curve는 그림 3.의 (b)와 같이 서로 비례하며 변화하게 된다.

Ⅳ. 결 론

본 논문은 변환생체인식의 한 방법으로써 얼굴 생체정보에서 이진비트열을 생성하는 방법을 제안했다. 제안 방법은 높은 보안성을 가지며 원래의 생체정보 보다 향상된 인식 성능을 갖는 새로운 방법이다. 생체정보 이진화를 위해 이용된 특징 추출방법은 조명변화나 표정 변화와 같은 부분적 변화에 강인한 영역 기반 주성분 분석 방법을 사용하였다. 추출된 특징을 보조정보를 이용하여 이진화하고 최종적으로 구분력을 기준으로 특징을 선별해서 향상된 성능을 가지는 이진비트열을 생성하였다. 개인의 보조정보를 이용한 생체정보 이진화 방법은 동일인에 대해 발생할 수 있는 변화를 상쇄하며 동시에 타인 사이의 구분력을 잃지 않도록 이용되었다. 변환 생체템플릿의 변경 및 재생산은 보조정보에 기반한 이진화 방법에서 지원하며 이것은 할당되는 비트 값을 변화하여 생체 템플릿을 변경 및 재생성 함으로써

가능하다. 개인마다 서로 다른 개인정보를 지정하고 이를 이용했을 경우 일반적으로 뛰어난 인식 성능을 보이며 본 논문의 경우 역시 좋은 성능을 보인다. 그러나 생체정보 변환 방법에서는 개인정보가 도난당한 경우 성능의 저하가 발생하며, 본 논문에서는 이를 고려하여 구분력이 좋은 특징을 선별하였다. 제안방법을 평가하기 위해 개인정보가 도난당하였을 경우를 가정하였고, 이진화 된 생체특징들 중에서 사람간의 구분력이 높은 이진 특징만을 선별하여 보다 강인한 이진비트열을 생성 할 수 있음을 실험적으로 보였다. 제안방법은 개인정보가 노출된 경우에서도 원래의 얼굴생체정보를 복원하기 쉽지 않으며, 원 얼굴생체로부터 획득된 특징정보와 비교하여 향상된 인식성능을 보여준다. 제안방법은 개인정보의 노출이 적으며 구현방법과 생성된 템플릿의 변경이 간단하여 실제 생체보안 시스템으로의 적용이 용이하다.

참 고 문 헌

[1] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, pp. 614-634, 2001.

[2] S. Chen and Y. Zhu, "Subpattern-based Principle component analysis," *Pattern Recognition*, vol 37, no 5, pp 1081-1083, 2004

[3] Jean-paul Linnartz and Pim Tuyls, "New Shielding Functions to Enhance Privacy and Prevent Misuse of Biometric Templates," *AVBPA*, pp. 393-402, 2003.

[4] C. Vielhauer and R. Steinmetz, "Handwriting: feature correlation analysis for biometric hashes," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 542 - 558, special issue on Biometric Signal Processing, 2004.

[5] Qi Han, Zhifang Wang, and Xiamu Niu, "A Non-uniform Quantizing Approach to Protect Biometric Templates," *International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP* '06., pp. 693-698, 2006.

[6] Chen C., Veldhuis R.N.J., Kevenaar T.A.M., and Akkermans A.H.M., "Multi-Bits Biometric String Generation based on the Likelihood Ratio," *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2007. BTAS 2007., 2007.

[7] Chen C., Veldhuis R.N.J., Kevenaar T.A.M., and Akkermans A.H.M., "Biometric Quantization through Detection Rate Optimized Bit Allocation," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.

[8] Pim Tuyls, Anton H. M. Akkermans, Tom A. M. Kevenaar, Geert Jan Schrijen, Asker M. Bazen, and Raymond N. J. Veldhuis, "Practical Biometric Authentication with Template Protection," *International conference on: Audio- and Video-Based Biometric Person Authentication. AVBPA*, Vol. 3546, pp. 436-446, 2005.

[9] Andrew Teoh, David Ngo and Alwyn Goh, "Biohashing: Two Factor Authentication Featuring Fingerprint Data And Tokenised Random Number," *Pattern Recognition*, Vol. 37, Issue 11, pp. 2245-2255, 2004.

[10] Andrew B. J. Teoh, Yip Wai Kuan, and Sangyoun Lee, "Cancellable biometrics and annotations on BioHash," *Pattern Recognition*, Vol. 41 , Issue 6, pp. 2034-2044, 2008.

[11] Andrew B.J. Teoh, Alwyn Goh, and David C.L. Ngo, "Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs," *IEEE transactions on a Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, 2006.

[12] Brian Chen and Gregory W. Wornell, " Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423-1443, 2001.

[13] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 113, 2008.

[14] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, 2003.

[15] J. Daugman, "The Importance of Being Random: Statistical Principles of Iris Recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279-291, 2003.

저 자 소 개

이 형 구(정회원)-교신저자
2003년 연세대학교 전기전자 공학과 학사 졸업.
2005년 연세대학교 전기전자 공학과 석사 졸업.
2010년 연세대학교 전기전자 공학과 박사 과정.
<주관심분야 : 생체인식, 컴퓨터 비젼, 패턴인식>

정 호 기(정회원)
1995년 2월 연세대학교 전자공학과 학사
1997년 2월 연세대학교 전자공학과 석사
2008년 8월 연세대학교 전기전자 공학부 박사
1997년 2월~2009년 4월 (주)만도 중앙연구소 책임연구원
2009년 5월~2010년 2월 연세대학교 생체인식 센터 전임연구원
2010년 3월~현재 연세대학교 전기전자공학부 연구교수
<주관심분야 : 컴퓨터비전 응용, 지능형 자동차, 지능형 감시시스템>