

# 다중 바이오 인증에서 특징 융합과 결정 융합의 결합\*

이 경 희†  
수원대학교 전기공학과

## Combining Feature Fusion and Decision Fusion in Multimodal Biometric Authentication\*

Kyunghee Lee†

Department of Electrical Engineering, The University of Suwon

### 요 약

본 논문은 얼굴과 음성 정보를 사용한 다중 바이오 인증에서, 특징 단계의 융합과 결정 단계의 융합을 동시에 수행하는 다단계 융합 방법을 제안한다. 얼굴과 음성 특징을 1차 융합한 얼굴·음성 융합특징에 대해 Support Vector Machines(SVM)을 생성한 후, 이 융합특징 SVM 인증기의 결정과 얼굴 SVM 인증기의 결정, 음성 SVM 인증기의 결정들을 다시 2차 융합하여 최종 인증 여부를 결정한다. XM2VTS 멀티모달 데이터베이스를 사용하여 특징 단계 융합, 결정 단계 융합, 다단계 융합 인증을 비교 실험한 결과, 제안한 다단계 융합에 의한 인증이 가장 우수한 성능을 보였다.

### ABSTRACT

We present a new multimodal biometric authentication method, which performs both feature-level fusion and decision-level fusion. After generating support vector machines for new features made by integrating face and voice features, the final decision for authentication is made by integrating decisions of face SVM classifier, voice SVM classifier and integrated features SVM classifier. We justify our proposal by comparing our method with traditional one by experiments with XM2VTS multimodal database. The experiments show that our multilevel fusion algorithm gives higher recognition rate than the existing schemes.

**Keywords:** multimodal biometrics, feature fusion, decision fusion

## 1. 서 론

최근 사람의 지문, 얼굴, 음성, 홍채, 정맥 등과 같은 고유의 신체적 특성과 서명, 걸음걸이 등의 행동적 특성을 신원 인증에 이용하는 바이오 인식(biometrics) 기술 분야가 우수한 보안성을 이유로 새로

운 신원 인증 시스템으로서 크게 주목받고 있다. 바이오 인식 시스템은 사용자 자신이 자신임을 확인 받는 인증(authentication, 1:1)과 데이터베이스에서 사용자가 누구인지를 찾아내는 인식(recognition, 1:N)으로 나누어진다. 단일 바이오 정보를 이용한 인식 기술은 이상적인 환경에서는 신뢰할 만하지만, 실제 상황에서는 환경에 매우 민감하다는 단점이 있다. 예로 음성 인식은 소음 환경에서는 급격하게 인식률이 떨어지며, 얼굴 인식 또한 조명의 세기 및 방향, 사용자의 포즈에 큰 영향을 받는다. 이러한 단일 바이오 인식의 한계를 극복하고자 여러 종류의 바이오 인식 기술을 통합 사용하여 상호보완성을 제공하는 다중

접수일(2010년 7월 30일), 수정일(1차: 2010년 10월 1일), 게재확정일(2010년 10월 7일)

\* 본 연구는 경기도의 경기도지역협력연구센터(GRRC) 사업의 일환으로 수행하였습니다. [GRRC수원2009-A1, 지능형 감시를 위한 다중생체인식 기술 연구]

† 주저자, khlee@suwon.ac.kr

바이오 인식에 관한 연구가 활발히 진행되고 있다.

두 개 이상의 단일 바이오 인식 시스템의 융합은 여러 단계에서 일어날 수 있는데, 세 가지 융합 단계로 나누어 볼 수 있다[1,2]. 각 센서로부터 얻어진 데이터들을 하나의 특징벡터로 융합하는 특징추출 단계(feature-level)에서의 융합과, 각각의 단일 바이오 인식 매칭 모듈로부터 출력된 매칭점수들을 융합하는 매칭점수 단계(matching score-level)에서의 융합과, 각각의 단일 바이오 인식 결정 모듈로부터 출력된 수락 또는 기각 결정들을 융합하는 결정 단계(decision-level)에서의 융합이 있다. 기존의 다중 바이오 인식 연구들은 세 가지 단계 각각에서의 융합을 주로 수행하고 있다. 본 논문에서는 얼굴과 음성 정보를 이용한 다중 바이오 인증 시스템에서, 특징추출 단계에서의 융합과 결정 단계에서의 융합을 동시에 사용하는 다단계 융합 방법을 제안한다. XM2VTS 멀티모달 데이터베이스[3]를 사용하여 얼굴과 음성 특징을 추출하여 실험한 결과, 제안한 다단계 융합 방법이 가장 우수한 성능을 보였다. 본 논문의 2장에서는 Support Vector Machines(SVM)을 이용한 얼굴 및 음성 인증 방법에 대하여 기술하고, 3장에서는 다단계 융합을 이용한 다중 바이오 인증 시스템에 대해 살펴보고, 4장에서는 성능 실험 및 결과 분석을 설명하고, 마지막으로 5장에서 본 연구의 결론에 대하여 서술한다.

## II. SVM(Support Vector Machines)을 이용한 얼굴 및 음성 인증

### 2.1 SVM을 이용한 얼굴 인증

얼굴 인식 분야에서 널리 사용되고 있는 주요 소분석(Principal Component Analysis : PCA)[4,5]을 통한 특징을 추출하여 SVM 분류기로 얼굴 인증을 수행한다. SVM[6]은 V. Vapnik 등에 의해 제안된 방법으로 구조적 위험 최소화 개념에 기반한 최적의 선형 결정 평면(optimal linear decision surface)을 찾는 이진 분류 방법이다. 결정 평면은 학습집합의 원소들의 가중화된 조합으로 나타내며, 이러한 학습집합의 원소들을 *support vectors* 라고 부르고, 이것은 두 클래스 간의 경계면을 나타낸다.  $N$ 개의 학습데이터들  $(\mathbf{x}_i, y_i)$ 가 있다고 할 때, 여기서  $\mathbf{x}_i \in R^d$  이고  $y_i$ 는 분류될 레이블을 나타낸다( $y_i \in \{-1, 1\}$ ). 선형분

리 데이터를 가정할 경우, 최대 여백 분류의 목표는 support vectors의 거리가 최대화되는 초평면(hyperplane)에 의해 두 개의 클래스를 분리하는 것이다. 이러한 초평면은 최적 경계 초평면(optimal separating hyperplane)이라고 불리며, 식(1)과 같이 나타낼 수 있다. 여기서 계수들  $\alpha$ 와  $b$ 는 2차 프로그래밍 문제의 해이며, support vectors에 대한  $\alpha$ 는 0이 아닌 값을 갖는다.

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i (\mathbf{x}_i^T \mathbf{x}) + b \quad (1)$$

선형분리가 불가능한 데이터인 경우에는, 입력벡터를 선형 초평면이 발견되는 고차원 특징 공간으로 비선형 매핑을 수행해주는 커널(kernels) 함수  $K(\mathbf{x}, \mathbf{x}_i)$ 를 사용하고, 결정 평면은 다음 식(2)와 같이 나타낼 수 있다.

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (2)$$

다음의 커널 함수들이 SVM에 종종 사용된다. 다항식(polynomial) 커널:

$$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^T \mathbf{x}_i + 1)^p \quad (3)$$

RBF(Radial Basis Function) 커널:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) \quad (4)$$

tangent hyperbolic 커널:

$$K(\mathbf{x}, \mathbf{x}_i) = \tanh(\alpha \cdot \mathbf{x}^T \mathbf{x}_i + \beta) \quad (5)$$

이진 분류기인 SVM을 얼굴 인증에 사용하기 위하여, 사용자 본인(genuine)의 영상들과 사칭자들(imposters)의 영상들 각각에서 추출한 상위 50개의 PCA 계수를 특징 벡터로 사용하여 학습한다. 학습을 통하여 생성된 SVM 분류기는 사용자 본인 여부를 판단할 수 있는 얼굴 인증기가 된다. 얼굴 SVM 인증기는 식(4)의 RBF커널을 사용하여 생성한다.

### 2.2 SVM을 이용한 음성 인증

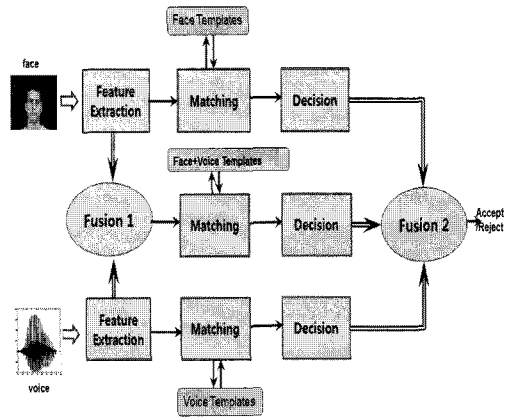
사람의 음성을 이용하는 화자 인식은 비교적 값이

싸고 손쉽게 구할 수 있는 마이크를 사용하며 사용자의 거부감이 적지만, 주위 소음에 민감하다는 단점이 있다. 본 논문에서 음성 정보를 이용한 인증에서는 사용자 본인의 음성 파일들과 사칭자들의 음성 파일들 각각에서 24개의 특징을 추출하여, 이를 학습시켜 음성 SVM 인증기를 생성한다. 음성 특징으로는 12차 MFCC(Mel-Frequency Cepstral Coefficient) [7]와 delta MFCC로 구성된 24차원 벡터를 사용하고, 20msec 윈도우를 10msec씩 이동하면서 추출한다. 모든 음성 데이터는 16kHz로 샘플링되고 16bit로 양자화 한다. SVM을 위한 특징 벡터로는 모든 프레임의 시간평균을 사용한다. 음성 SVM 인증기 또한 식(4)의 RBF커널을 사용하여 생성한다.

### III. 다단계 융합(Multilevel Fusion)을 이용한 다중 바이오 인증

다중 바이오 인식에서의 두 개 이상의 단일 바이오 인식 시스템의 융합에는, 각 센서로부터 얻어진 데이터들을 하나의 특징벡터로 융합하는 특징추출 단계에서의 융합과, 각각의 단일 바이오 매칭 모듈로부터 출력된 매칭점수들을 융합하는 매칭점수 단계에서의 융합과, 각각의 단일 바이오 결정 모듈로부터 출력된 수락 또는 기각 결정들을 융합하는 결정 단계에서의 융합이 있다. 특징추출 단계에서의 융합은 다중 센서 입력으로부터 추출된 바이오 특징벡터를 연결함으로써 새로운 특징벡터를 구성할 수 있다. 매칭점수 단계에서의 융합은 가중화된 합 융합, 가중화된 곱 융합과 매칭점수를 다시 특징으로 하여 의견벡터를 생성하여 또 하나의 후분류기(post-classifier)를 사용하여 최종 결정을 하는 방법 등이 있다. 결정 단계에서의 융합은 단일 바이오 인식기의 결정들을 다수결(majority voting), 순위결합(ranked list combination), AND, OR 논리 연산 등을 사용하여 통합한다. AND 융합은 낮은 FAR(False Acceptance Rate)이 요구되는 환경에 유용하며, OR 융합은 낮은 FRR(False Rejection Rate)이 요구되는 환경에 유용하다.

본 논문에서는 얼굴 인증과 음성 인증을 융합한 다중 바이오 인증을 이루기 위하여, [그림 1]과 같이 특징추출 단계에서의 1차 융합과 함께 결정 단계에서의 2차 융합을 동시에 수행하는 다단계 융합 방법을 제안한다. 1차 융합은 얼굴 인증과정에서 추출된 특징 50개와 음성 인증과정에서 추출된 특징 24개의 값들을 각각 -1.0에서 1.0사이의 값으로 정규화한 후 74개의



(그림 1) 제안하는 다단계 융합 다중 바이오 인증 시스템

특징값을 갖는 하나의 벡터로 연결하는 특징 융합으로, 이 얼굴·음성 융합특징에 대한 SVM 인증기를 추가로 생성한다. 이 때 융합특징 SVM 인증기도 RBF 커널을 사용하여 생성한다. 특징 집합은 원시 얼굴, 음성 데이터에 관하여 매칭점수나 최종결정보다 더 많은 정보를 가지고 있기 때문에, 특징 단계에서 이중의 얼굴, 음성 정보의 융합은 각 사용자에게 대해 더욱 풍부한 정보를 갖게 된다.

이제, 사용자 인증에서는 얼굴과 음성 정보가 입력되면 얼굴 인증과정, 음성 인증과정과 더불어 1차 융합으로 생성된 얼굴·음성 융합특징 인증과정도 수행하게 된다. 세 가지 인증과정의 결정들을 다시 2차 융합하여 최종 사용자 인증 여부를 결정한다. 2차 융합에서는 얼굴 SVM 인증기의 결정, 음성 SVM 인증기의 결정, 융합특징 SVM 인증기의 결정들을 다수결 방법 또는 AND 논리연산을 사용하여 통합한다.

### IV. 실험 결과 및 분석

#### 4.1 실험 방법

제안한 방법의 성능을 분석하기 위하여 XM2VTS 멀티모달(multimodal)데이터베이스[3]를 사용하여 실험하였다. XM2VTS 데이터베이스는, 얼굴영상은 2장씩 한 달 간격으로 4번 촬영한 295명에 대해 정면 영상 8장씩으로 구성되어 있고, 음성파일은 295명에 대하여 세 가지 문장을 2번씩 한 달 간격으로 4번 녹음하여, 사람별로 세 가지 문장 각각에 대해 8개의 파일로 구성되어 있다. 본 논문에서는 원본 데이터 파일에 오류가 있어 실험이 불가능한 8명을 제외하고 287명에 대하여 실험하였다. 음성데이터는 세 가지 문장

(표 1) 얼굴+음성을 이용한 다중 바이오 인증 실험 결과 (음성Set A)

융합방법		얼굴	음성	얼굴+음성 특징융합	얼굴+음성 결정융합 (OR)	얼굴+음성 결정융합 (AND)	얼굴+음성 다단계융합 (AND)	얼굴+음성 다단계융합 (MAJ)
에러율(%)								
SVM (RBF) $\sigma=0.2$	FAR	0.91	6.74	1.72	7.57	0.09	0.07	0.85
	FRR	11.15	19.08	5.92	1.57	28.66	29.79	5.40
	TER	1.85	7.86	2.10	7.02	2.68	2.77	1.27
SVM (RBF) $\sigma=0.3$	FAR	0.75	4.90	1.59	5.60	0.04	0.03	0.73
	FRR	11.85	21.78	6.27	2.09	31.53	32.75	6.10
	TER	1.76	6.43	2.01	5.28	2.91	3.01	1.22

에 따라 세 개의 데이터셋 A, B, C 로 나누어 실험에 사용하였다.

실험은 287명 각자에 대하여 학습데이터와 테스트 데이터를 구성하여 수행하였다. 개인별 1:1 인증기 생성용 학습데이터로는 사용자 본인(Genuine)의 4 개의 영상/음성 파일과, 40명 사칭자들(imposters)의 영상/음성 파일 1개씩을 사용하였다. 개인별 테스트데이터로는 학습에 사용되지 않은 4개의 사용자 영상/음성 파일과, 학습과정에 알려지지 않은 또 다른 사칭자들 40명에 대한 40개의 영상/음성 파일을 사용하여 성능 평가를 하였다.

얼굴 특징으로는 상위 50개 PCA계수로 이루어진 50차원 벡터를 사용하였고, 음성 특징으로는 12차 MFCC와 delta MFCC로 구성된 24차원 벡터를 사용하였다. 단일 바이오 인증과 비교하기 위하여, 단일 얼굴 인증 실험과 단일 음성 인증 실험을 수행하였다. 또 다중 바이오 인증의 단일 단계 융합과 비교하기 위하여, 특징 단계에서의 융합 실험과 결정 단계에서의 AND연산, OR연산을 이용한 융합 실험을 수행하였다. 본 논문에서 제안한 다단계 융합 실험에서는, 2차 융합과정에 다수결 융합과 AND연산을 이용한 융합으로 나누어 실험하였다. 얼굴 인증기, 음성 인증기, 얼굴+음성 특징융합 인증기 모두 RBF 커널을 이용한 SVM 분류기를 사용하였으며, 식(5)의  $\sigma$  값을 0.2와 0.3으로 하여 각각 실험하였다.

#### 4.2 결과 및 분석

비선형인 경우 표준 SVM 학습은 전체의 오분류율을 최소화하기 때문에, 바이오 인식 시스템의 성능 평가에 주로 사용되는 EER(Equal Error Rate) 측정에 적합하지 않다. 따라서 본 논문의 실험에서는 FAR, FRR과 함께, 잘못 인증된 데이터수를 총 테스트 데이터수로 나눈 TER(Total Error Rate)를 사용하여 실험결과들을 비교하였다. 음성 데이터셋 A, B, C 세 개로 나누어 실험한 결과는 [표 1, 2, 3]에 차례로 나타나 있다. 실험결과 음성 인증은 얼굴 인증에 비해 TER이 평균 6.0 높아 전체적으로 좋지 않은 성능을 보이고 있다. 얼굴과 음성을 특징 단계에서만 융합한 다중 인증 결과는, 단일 음성 인증보다는 평균적으로 FAR은 4.6, FRR은 17.4, TER은 5.7정도 모두 낮아졌으나, 단일 얼굴 인증보다는 FAR은 0.7 높아지고 FRR은 4.7 낮아져 TER은 0.2 높아졌다. 이는 음성 특징의 인식 성능이 매우 낮아서, 얼굴 특징에 융합한 경우 인식률의 개선이 일어나지 않고 오히려 다소 하락하게 되었다. 결정 단계에서의 융합 중 OR연산에 의한 융합은 OR연산자의 특성으로 FRR은 급격히 낮아진 반면에 FAR은 높아졌고, 결정 단계에서 AND연산에 의한 융합은 AND연산자의 특성으로 FAR은 급격히 낮아진 반면에 FRR은 높아지게 되었다.

(표 2) 얼굴+음성을 이용한 다중 바이오 인증 실험 결과 (음성Set B)

융합방법		얼굴	음성	얼굴+음성 특징융합	얼굴+음성 결정융합 (OR)	얼굴+음성 결정융합 (AND)	얼굴+음성 다단계융합 (AND)	얼굴+음성 다단계융합 (MAJ)
에러율(%)								
SVM (RBF) $\sigma=0.2$	FAR	0.91	7.00	1.65	7.81	0.11	0.10	0.78
	FRR	11.15	18.64	6.19	1.74	28.05	29.01	5.45
	TER	1.85	8.06	2.06	7.25	2.65	2.73	1.21
SVM (RBF) $\sigma=0.3$	FAR	0.75	5.34	1.49	6.01	0.08	0.07	0.69
	FRR	11.85	20.38	6.27	2.00	30.23	31.01	5.84
	TER	1.76	6.71	1.92	5.65	2.82	2.88	1.17

[표 3] 얼굴+음성을 이용한 다중 바이오 인증 실험 결과 (음성Set C)

융합방법 에러율(%)		얼굴	음성	얼굴+음성 특징융합	얼굴+음성 결정융합 (OR)	얼굴+음성 결정융합 (AND)	얼굴+음성 다단계융합 (AND)	얼굴+음성 다단계융합 (MAJ)
SVM (RBF) $\sigma=0.2$	FAR	0.91	7.26	1.54	8.08	0.09	0.08	0.85
	FRR	11.15	30.75	7.67	3.40	38.50	39.37	7.75
	TER	1.85	9.39	2.10	7.66	3.58	3.65	1.47
SVM (RBF) $\sigma=0.3$	FAR	0.75	5.50	1.34	6.20	0.04	0.04	0.73
	FRR	11.85	34.76	8.45	4.36	42.25	42.94	8.97
	TER	1.76	8.16	1.99	6.03	3.88	3.94	1.48

본 논문에서 제안한 특징 융합과 더불어 결정 융합을 함께 수행한 다단계 융합 실험에서는 결정 융합에 다수결을 사용한 결과가, 비교적 우수한 성능을 내고 있던 얼굴 인증의 결과보다 평균적으로 FAR은 0.1, FRR은 4.9, TER은 0.5 정도 모두 낮아져 더 우수한 결과를 내었다. 세 가지 문장의 음성 데이터셋 모두에 대해 제안한 방법이 우수한 성능을 보였다. 얼굴보다 성능이 좋지 않은 음성 데이터도 제안한 다단계 융합을 이용한 다중 바이오 인증으로 통합하면, 단일 얼굴 인증의 성능보다 더 우수해짐을 확인하였다.

V. 결 론

본 논문에서는 얼굴과 음성을 융합한 다중 바이오 인증 시스템에서, 특징추출 단계에서의 1차 융합과 함께 결정 단계에서의 2차 융합을 동시에 수행하는 다단계 융합 방법을 제안한다. XM2VTS 멀티모달 데이터베이스를 사용하여 실험한 결과, 제안한 다단계 융합 방법이 우수한 성능을 보였다. FAR의 감소는 물론 FRR도 동시에 감소된 것을 확인하였다. 얼굴 인증보다 성능이 낮은 음성 데이터이지만, 제안한 다단계 융합으로 얼굴 인증과 융합한 다중 바이오 인증을 하면 단일 얼굴 인증의 결과보다 더 우수해짐을 확인하였다. 따라서 성능이 좋지 않은 바이오 정보일지라도 제안한 방법으로 기존의 바이오 정보에 융합하여 다중 바이오 인증을 하게 되면, 기존의 단일 바이오 인증의 성능보다 더 우수한 결과를 낼 수 있음을 알았다. 또 이종의 바이오 정보를 이용한 다중 바이오 인증을 수행할 경우, 각각 상이한 정합 알고리즘들의 사용에 따라 메모리 사용 및 계산량이 크다는 문제점이 발생하는데, 본 논문에서는 얼굴, 음성, 얼굴·음성 융합특징에 대한 세 가지 인증 과정 모두, 동일한 분류 알고리즘인 SVM을 사용함으로써 자원이 제한된 스마트폰 등의 시스템에서도 다중 바이오 인증이 가능

한 이점이 있다.

참 고 문 헌

- [1] A. Ross and A. Jain, "Information Fusion in Biometrics," *Pattern Recognition Letters*, Vol. 24, No.13, pp.2115-2125, Sep. 2003.
- [2] A. Ross, K. Nandakumar, A. Jain, *Handbook of Multibiometrics*, Springer, May 2006.
- [3] XM2VTS Multimodal Database, <http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb>
- [4] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, Jan. 1991.
- [5] Harry Wechsler, *Reliable Face Recognition Methods-System Design, Implementation and Evaluation*, Springer, Nov. 2006.
- [6] K. Jonsson, J. Matas, J. Kittler, Y.P. Li, "Learning Support Vectors for Face Verification and Recognition," *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp.208-213, Mar. 2000.
- [7] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task," *10th International Conference on Speech and Computer*, Vol.1, pp.191 - 194, May 2005.

---

**〈著者紹介〉**

---



이 경 희 (Kyunghee Lee) 정회원

1993년 2월: 연세대학교 컴퓨터과학과 학사

1998년 8월: 연세대학교 컴퓨터과학과 석사

2004년 2월: 연세대학교 컴퓨터과학과 박사

1993년 1월~1996년 5월: LG소프트(주) 연구원

2000년 12월~2005년 2월: 한국전자통신연구원 선임연구원

2005년 3월~현재: 수원대학교 전기공학과 조교수

〈관심분야〉 바이오인식, 정보보호, 컴퓨터비전, 인공지능, 패턴인식