

# 구간 분할 및 HMM 기반 융합 모델에 의한 온라인 서명 검증

## On-line Signature Verification Using Fusion Of Segment Matching and HMM

양동화\*, 이대종\*<sup>\*\*\*</sup>, 전명근\*

\*충북대학교 전기전자 컴퓨터공학부

\*\*알바타대학교 전기 컴퓨터공학과

Dong-Hwa Yang\*, Dae-Jong Lee\*<sup>\*\*\*</sup>, Myung-Geun Chun\*

\*School of Electrical and Computer Engineering

Chungbuk National University

\*\*Dept. of Electrical and Computer Engineering, University of Alberta

E-mail : mgchun@chungbuk.ac.kr

### 요 약

기존의 참조서명과 입력서명을 비교하는 방법 중 분절 단위 비교 방법은 전역적 방법과 점 단위 방법에 비하여 우수한 장점을 가지고 있다. 그러나 분절 단위 비교 방법은 인식률과 직접적인 관계가 있는 분절의 불안정 문제점이 있다. 본 연구에서는 분절 단위 비교 방법을 이용한 서명검증의 신뢰도를 향상시키기 위해 두 가지 형태의 모델을 구축하였다. 우선 기존에 사용된 구간 분할 매칭 방법을 사용하여 매칭도를 산출하였다. 다음으로 서명의 분할된 영역을 주성분 분석 기법에 의해 특징 벡터를 산출한 후 HMM에 의해 서명 모델을 구축하였다. 산출된 두 특징을 융합하는 방법으로는 SVM 분류기를 사용하였다. 실험 결과 제안된 기법은 분절 단위 기반의 구간분할매칭 기법에 비해 우수한 성능을 나타냈다.

### 1. 서론

생체인식(Biometrics) 기술로 사용하고 있는 신체적 특징 중 서명은 오래 전부터 계약체결 등의 서류에 대한 증빙 목적으로 사용될 만큼 사용자에게 익숙한 인증수단 알려져 있다. 뿐만 아니라, 음성, 홍채, 얼굴 등의 다른 생체인식기법에 비해 소형 타블릿과 같은 가격 경쟁력을 지닌 입력 장비만을 필요로 함으로써 가격 면에서도 우위를 보이고 있다. 최근에는 타블릿 마저 필요치 않은 PDA의 등장과 활성화로 서명인식에 관한 연구가 주목 받고 있다[1][2]. 그러나, 서명인식의 경우 생체인식을 위한 다른 특징들과 달리 사칭자에 의해 전문 위조가 가능하고 동일인의 서명 조차도 외부 환경 또는 내적인 요인에 의해 서명의 변이 차이로 시스템의 성능이 저하되는 문제점을 지니고 있다[3].

현재까지의 온라인 서명검증방식은 참조서명(Reference Signature)과 입력서명(Input Signature)의 비교 방식에 따라 크게 진위적 특

징 비교, 점 단위 비교, 분절 단위 비교방법이 있다[4]. 이 중 분절 단위 비교방법은 참조서명과 입력서명을 각각 적절한 분절단위로 분할한 후, 분절 간의 최적의 대응관계를 찾는 방식으로, 정적정보 또는 동적정보를 이용하여 기준 분할이 이루어진다. 여기서, 정적정보는 굴곡점, 봉우리 등의 정적특징을 이용하여 분절단위로 기준 분할이 이루어지며[4][5], 동적정보는 최소 속력점, 압력 등의 동적특징을 이용하여 입력서명을 분절단위로 분할한다[6]. 이와 같은 분절단위 비교방법은 분절 간에 지역적 비교가 가능하며, 지역적으로 통계적 모델을 사용할 수 있어 분석이 용이한 장점을 지니고 있는 반면에 서명의 변이로 인해 모든 서명을 같은 분절단위로 분할하기 어려워 진서명간에도 분할점이 불안정할 수 있다는 문제점을 지니고 있다. 이러한 문제점은 진서명간의 비교하는 특징점 위치가 서로 상이하여 인식률과 직접적인 관계가 있다. 본 논문에서는 기존의 분절 단위 비교만을 이용한 방법에 대한 문

제점을 보완하기 위해 HMM 기법과 융합하여 서명인식의 신뢰도를 높이는 온라인 서명 검증 기법을 제안한다. 즉, 서명의 동적정보인 시간특성을 이용하여 분절단위 비교를 수행하였고, 정적정보인 서명의 X축 및 Y축 변이 특성을 확률 모델 기반인 HMM로 구축한 후 인식을 수행한다. 여기서, HMM의 모델형성을 위한 입력패턴으로는 얼굴인식시스템에서 우수한 특성을 보이고 있는 주성분 분석 기법(PCA:Principal Component Analysis)을 이용하여 정적서명의 특징을 추출하였다. 최종 융합단계에서는 구간분할에 의해 계산된 동적정보와 HMM에 의해 계산된 정적정보를 효과적으로 융합하기 위하여 서포트 벡터 머신 (SVM:Support Vector Machine) 분류기를 사용하였다.

본 논문의 구성은 다음과 같다. 2, 3장에서는 온라인 서명 검증 기법에 사용된 알고리즘을 설명한다. 4장에서는 실험 결과를 바탕으로 제안한 방법의 타당성을 검토한다. 마지막으로 5장에서 결론을 맺는다.

## 2. 구간분할 및 HMM 기반 매칭기법

분절단위 비교방식에 의하여 서명 데이터를 분할할 경우 참조서명과 입력서명 간에 비교되는 분할점, 즉 특징벡터를 안정적으로 추출하는 것은 검증성과 직결될 만큼 중요시 다루어지는 부분이다. 본 논문에서는 기존에 연구한 방법인 동일서명 간에 변이가 적은 1차원상의 Y축의 봉우리를 기준 분할점으로 선택한 후 분할점 간에 위치한 특징벡터인 봉우리·계곡점들을 비교함으로써 구간적으로 안정된 분할이 이루어지도록 제안된 방법을 사용하였다. 자세한 내용은 다음을 참조하길 바란다[6].

은닉 마코브 모델(Hidden Markov Model, HMM)은 신호에 포함된 특성들의 발생과정을 확률적으로 모델화한 것으로 적은 계산량으로 좋은 인식률을 얻을 수 있어 현재 음성인식이나 Gesture 시스템에 널리 사용되는 알고리즘이다.

HMM은 세가지 요소로 구성되어 있는데 상태(state)의 개수, 시간에 따른 상태의 변화를 결정하는 상태전이 확률분포, 그리고 각 상태에서의 출력 심볼의 확률분포이다. 각 상태들은 직접적으로 관찰 가능하지는 않으나 대신 각 상태들이 일정 확률을 가지고 만들어 내는 심볼을 보고 원래 상태를 추정하는 방법이다. 이러한 정의를 이용하여 HMM은  $\lambda = (A, B, \Pi)$ 로 나타낼 수 있다. 여기서,  $\Pi = \pi_{ij}$ : 초기 상태전이 확률분포,  $A = a_{ij}$ : 상태전이 확률분포,  $B = b_{ij}$ : 관측 심볼 확률분포 이다.

HMM을 이용하기 위해서는 다음 두 단계를 거치게 된다. 즉, 모델형성 과정과 형성된 모델을 이용하여 관측 심볼의 확률 값을 구하는 두 과정을 거치게 된다.

■ 단계 1 : 모델 형성과정 ( $P(O|\lambda)$ 를 최대화하는 모델 파라미터  $\lambda = (A, B, \Pi)$ 를 구하는 문제 )

■ 단계 2 : 모델 인식과정 ( 관찰된 심볼의 시퀀스  $O = O_1 O_2 \dots O_T$ 와 모델  $\lambda = (A, B, \Pi)$ 가 주어졌을 때 모델에 대한 심볼의 likelihood  $P(O|\lambda)$ 를 구하는 문제 )

본 논문에서 단계 1의 모델형성과정을 살펴보면 다음과 같다. 관측된 심볼의 시퀀스는 PCA를 이용하여 사용된 서명의 정적정보인 X축 Y축 특징점을 검출한 후 이를 이용하여 서명 특성을 대표하는 코드북을 형성한다. 각각의 서명별 시퀀스는 입력 서명과 대표 코드북과의 특징벡터를 비교함으로써 얻어진다. 이렇게 얻어진 관측 심볼의 시퀀스를 이용한 HMM 모델링은 Baum-Welch 알고리즘을 모델 파라미터  $\lambda = (A, B, \Pi)$ 를 설정하였다. 단계 2의 인식단계에서는 각각 설정된 서명별 HMM 모델과 관측된 심볼 시퀀스와의 부합확률은 forward 알고리즘을 이용하여 산출하였다.

입력 서명을 주성분분석기법을 이용하여 특징 추출하는 과정을 설명하면 다음과 같다. 타블렛으로부터 입력서명의 X축 및 Y축의 X, Y 좌표값에 대한 정보를 획득한 후 다음 식과 같은 서명의 입력값을 취득한다.

$$s_i = [x_1, y_1, x_2, y_2, \dots, x_l, y_l]^T \quad (1)$$

여기서,  $x_l, y_l$ 은 시간  $t_l$ 에서의 온라인 입력서명의 X축과 Y축의 좌표값을 각각 의미한다.

다음으로 서명의 구간별 확률모델 형성을 위해 입력서명을 분할한다. 즉, 입력서명  $s_i$ 는 다음과 같이 시간변이에 따라  $p$ 개의 구간으로 분할된다.

$$s_i = [s_i^1 : s_i^2 : \dots : s_i^p] \quad (2)$$

각 구간별로 PCA를 이용하여 특징벡터를 산출한다. 즉, 학습서명의 개수가  $n$ 개 일 경우 1차원 열벡터의 학습벡터를  $s_1, s_2, \dots, s_n$ 로 표시하면,  $k$ 번째 시간구간에 대한 PCA 특징벡터는 식(3)에 의해 얻어진다.

$$z_i^k = [v_i^k]^T (s_i^k - m_i^k) \quad (3)$$

여기서,  $m_i^k$ 는 사용된 학습서명에 대한  $k$ 번째 구간간의 평균벡터를 나타낸다. 위에서 설명한 바와 같이 각각 구간별 계산된 PCA 특징벡터의 천이과정을 HMM모델링 하였다.

## 3. 구간분할 및 HMM 기반 융합모델에 의한 온라인 서명 검증

본 논문에서는 기존의 구간 분할 매칭방법의 성능 향상을 위해 HMM과 융합한 새로운 온라인

인 서명 검증시스템을 제안하였다. 제안한 온라인 서명 검증 구성은 그림 1과 같이 데이터 획득, 전처리, 특징 추출 및 비교, 융합과정으로 이루어진다. 각각의 처리부에 대하여 알아보면 다음과 같다.

우선, 데이터 취득부에서는 와콤(WACOM)사의 인튜어스(intuos) 4×5사이즈의 타블렛을 이용하여 압력이 1이상인 서명의 X축, Y축 및 시간(T)에 대한 온라인 서명 데이터를 획득하였다. 여기서 타블렛이나 전자펜으로부터 획득한 데이터는 동일인의 서명간에도 서명 당시의 감정 또는 환경변화에 의해 서명의 미세한 변이가 발생하고 그로 인해 취득한 서명데이터의 사이즈가 동일하지 않다. 그러나 주성분분석기법에 의해 특징을 추출하기 위해서는 각 입력되는 데이터의 사이즈가 동일해야 하므로 본 논문은 리샘플링 과정을 거쳐 서명데이터의 사이즈를 동일화한다.

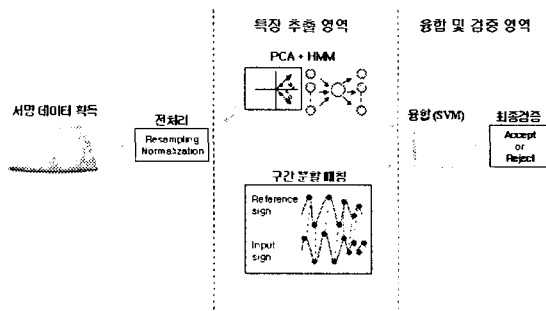


그림 1. 제안한 온라인 서명 검증시스템

본 논문에서는 안정된 리샘플링 과정을 통하여 서명마다 사용할 서명 데이터의 사이즈를 동일하게 하였다. 즉, 이러한 리샘플링 과정은 취득한 데이터의 순서에 따라 데이터 시퀀스 번호를 부여한 후 시퀀스 번호를 균등 분할하여 데이터를 재 취득하였다. 리샘플링 과정의 첫 번째로 각 서명데이터를 동일한 조건으로 균등 분할하기 위해 데이터 시퀀스 번호의 시작점을 0, 끝점을 1로 정규화하였다. 다음으로 샘플링 데이터 개수에 따라 정규화된 시퀀스 번호를 균등 분할 후 분할영역내의 정규화된 시퀀스 번호들 중 중앙에 위치한 데이터를 선택한다. 그림 2는 서명 데이터 개수가 각각 6개와 10개인 데이터를 2개로 리샘플링하는 과정을 나타내고 있다.

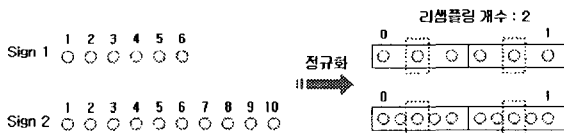
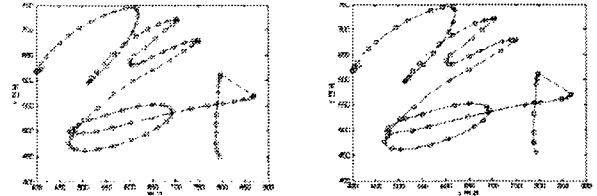


그림 2. 데이터의 시퀀스를 이용한 리샘플링과정

그림 3은 리샘플링 전·후의 서명 데이터 결과를 나타냈다. 그림 3에서 'o' 표시는 취득 데이터의 분포를 나타낸다. 리샘플링 전의 서명데이터를 나타낸 그림 3-(a)의 경우 취득 데이터의

수는 123개에서 리샘플링 과정을 통하여 취득 데이터의 수를 80개로 재취득 하였으며, 그 결과를 그림 3-(b)에 나타냈다. 그림에서와 같이 두 서명에 대하여 서로 유사한 위치를 갖는 데이터를 추출한 후에 각 서명간의 위치와 크기를 균일화 시키기 위해 데이터를 0과 1사이의 값으로 정규화 한다.



(a) 리샘플링 전 데이터 (b) 리샘플링 후 데이터

그림 3. 서명의 리샘플링 결과

서명의 특징추출은 2장에서 설명된 식 (2)에 의해 선형판별분석기법에 의해 특징벡터와 구간 분할 매칭에 의해 선택된 봉우리 및 계곡점에 대한 시간 T의 데이터를 얻는 과정이다. 각각의 알고리즘으로부터 구한 여러값은 현재 이진분류에 성능이 우수한 SVM 분류기를 적용하였다.

제안된 방법의 타당성을 검증하기 위하여 20명의 사람으로부터 각각 10개씩 서명 데이터를 획득하였고 충분한 학습과정을 통해 각 사람당 10개씩 위조서명을 취득하였다. 동일인의 서명에 대한 10개의 진서명 중 5개는 참조서명으로 사용하였고 나머지 5개의 진서명과 10개의 전문위조서명을 대상으로 제안된 알고리즘의 성능을 실험하였다. 그림 4에 사용된 본인서명과 위조서명의 일부분을 나타냈다. 서명의 형태만을 이용할 경우 육안으로도 진서명과 위조서명 간을 분별하는데 한계가 있음을 확인할 수 있다.

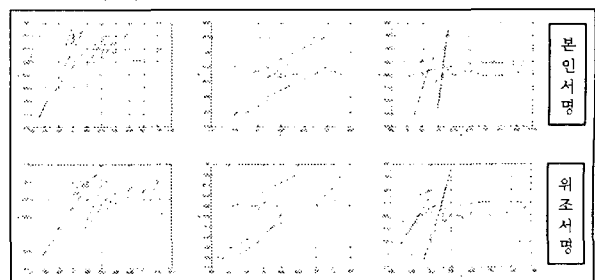


그림 4. 진서명과 위조서명의 일례

#### 4. 실험 결과

그림 5에 실험결과를 나타냈다. 그림 5-(a)는 구간분할매칭 방법에 의한 인식결과를 나타냈으며, 그림 5-(b)는 HMM 기반 모델에 의한 서명 인식결과를 나타냈다. 서명인식시스템의 성능지표로서 임계치의 변화에 대한 오거부율(False Reject Rate, FRR)과 오인식률(False Accept Rate, FAR)을 사용하였다. 여기서, 오거부율이란 본인의 서명이 거부된 오류율을 나타내고 오인식률은 타인의 서명이 본인으로 인식된 오류율을 각

각 의미한다. 표 1에서는 실험결과를 수치로 비교·분석하여 나타냈다.

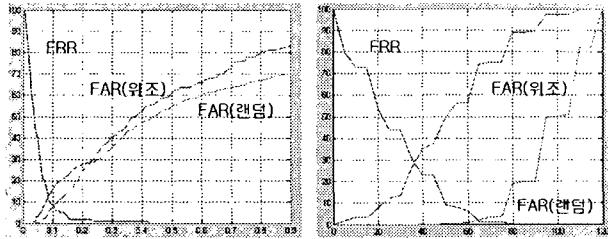


그림 5. (a)구간분할 매칭결과 (b)HMM 매칭결과

표 1. 적용 기법별 서명인식 결과

Method	Recog FRR	Recognition		
		FAR		FAR + FRR
		Random	Skill	
PPP	12%	12.5%	4.4%	28.9%
HMM	24%	33.5%	0%	57.5%
Fusion	7%	2.5%	0%	9.5%

### 5. 결론 및 향후과제

서명인식은 사용자에게 거부감이 없는 친숙한 생체인식 기술 분야 중에 하나이다. 그러나, 서명인식의 경우 생체인식을 위한 다른 특징들과 달리 사칭자에 의해 전문 위조가 가능하고 동일인의 서명조차도 외부 환경 또는 내적인 요인에 의해 서명의 변이 차이로 시스템의 성능이 저하되는 문제점을 지니고 있다.

따라서, 본 논문에서는 기존에 분절단위비교 방법의 분절추출에 대한 불안정 문제를 HMM기법과 융합함으로써 서명인식의 신뢰도를 향상시킬 수 있었다. 실험을 통해 제안된 방법의 타당성을 검증한 결과 HMM 기법에 의한 방법은 구간분할매칭방법보다 랜덤위조서명에 우수한 성능을 보인 반면에 구간분할매칭방법은 선형판별분석기법을 사용한 경우보다 전문위조서명에 우수한 성능을 보였다. 두 기법을 융합한 결정법칙을 적용한 결과 랜덤위조 및 전문위조에 대하여 향상된 인식결과를 나타냄을 확인할 수 있다.

향후 과제로서는 인식률을 높일 수 있는 방법과 온라인 서명을 실제로 PDA에 탑재하여 실험과 같은 인식률을 구해보는 것이다.

#### 감사의 글

본 연구는 한국과학재단 목적기초연구(R01-2002-000-00315-0) 지원으로 수행되었음.

### 6. 참고문헌

[1] 전명근, 생체인식(Biometric) 총론, 한국정보통신교육원, pp. 8-1~8-21, 2002.

[2] Chin-Chuan Han, "An on-line signature verification system using multi-template matching approaches", Security Technology, Proceedings. IEEE 33rd Annual International Carnahan Conference on , pp. 477-480, 1999.

[3] Ma Mingming, "Automatic on-line signature verification based on multiple models", Computational Intelligence for Financial Engineering, (CIFEr) Proceedings of the IEEE/IAFE/INFORMS Conference on , pp. 30 -33, 2000.

[4] Rhee T.H., "On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries", Proceedings. Sixth International Conference on , Document Analysis and Recognition, pp. 645 -649, 2001.

[5] Yue K.W. and Wijesoma W.S., "Improved segmentation and segment association for on-line signature verification," IEEE International Conference on, Systems, Man, and Cybernetics, Vol. 4, 2000.

[6] Sang-Yeun Ryu, Dae-Jong Lee, Myung-Geun Chun, "A Robust On-line Signature Verification System", International Journal of Fuzzy Logic and Intelligent Systems, vol. 3, no. 1, pp.27-31, 2003

[7] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegsmann, "Eigenfaces vs. Fisherfaces : recognition using class specific Linear Projection", IEEE Trans. on Pattern Analysis and Machine Intell., 19(7), pp. 711-720, 1997.

[8] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, JOHN WILEY&SONS, INC. Second Edition, 2002.

[9] 민준오, 류상연, 광근창, 전명근, "얼굴과 서명을 이용한 다중생체인식 시스템", BERC Biometrics Workshop, Vol.1, pp. 141~145, 2004.

[10] Viswanathan, R., Varshney, P.K., "Distributed detection with multiple sensors I. Fundamentals", Proceedings of the IEEE , Vol. 85, pp. 54-63, 1997.

[11] Krzysztofowicz, R., Long, D., "Fusion of detection probabilities and comparison of multisensor systems", IEEE Transactions on , Systems, Man and Cybernetics, Vol. 20, pp. 665-677, 1990.

[12] Ben-Yacoub, S., Abdeljaoued. Y., Mayoraz, E., "Fusion of face and speech data for person identity verification", IEEE Transactions on , Neural Networks, Vol. 10, pp. 1065 -1074, 1999.