
동적 서명의 특징 정보에 대한 통계적 분석에 관한 연구

김진환* · 조재현**

A Study on a Statistical Analysis of the Feature Information for the
Dynamic Signature Verification

Jin-Whan Kim* · Jae-Hyun Cho**

요 약

본 논문에서는 서명에서 생성되는 속도 성분(방향 정보, 거리 정보)에서 방향 정보만을 특징 정보로 서명DB에서 저장하고 거리(속력) 정보는 조정상수(w)를 이용함으로써 오류율에 영향을 주지 않으면서 처리속도를 개선하고, 특징 정보 크기도 줄일 수 있었다. 이를 위해 통계적 검정 T-test를 이용하여 확인하였다.

ABSTRACT

This paper is a research on the feature information using direction information and adjusting constant w for the dynamic signature verification. We could improved processing time and reduce signature database without the increase of error rate. We could confirmed these results by using statistical method T-test.

키워드

동적 서명인증, 특징 정보, 통계적 검정, T-test, 조정상수(w)

* 영산대학교 컴퓨터공학과
** 부산가톨릭대학교 컴퓨터공학과

접수일자 2009. 03. 11
심사완료일자 2009. 04. 24

I. 서 론

동적 서명인증시스템은 생체인증시스템 중에서 생리적(신체적) 특징을 가지는 보안시스템에 비해 여러 가지 다른 특성을 가지고 있다. 첫째는, 서명은 기준의 은행거래, 신용카드 등에서 이미 사용하고 있는 본인 확인 방법으로 친숙한 매체라는 것이고, 둘째는, 모바일 컴퓨팅, 유비쿼터스(ubiquitous) 사회와 더불어 널리 보급되고 있는 PDA(Personal Digital Assistants), 스마트폰(smart-phone)의 주요 입력 수단으로 펜을 사용하고 있다는 것이다. 셋째는, 비밀번호를 바꾸어 사용할 수 있듯이 서명도 필요할 때 언제든지 약간의 노력으로 자신의 서명을 바꾸어 사용할 수 있다는 것은 보안적인 측면에서는 매우 중요한 요소이다. 넷째는, 특정 펜 입력 장치에 상관없이 사용 가능하며, 마우스로도 간단한 도형을 이용하여 사용 가능하다. 다섯째는, 개인 서명의 숙련도에 따라 자신의 보안수준을 설정하여 사용할 수 있다. 여섯째는, 기후나 온도 등 외부의 환경변화에 따른 오작동이 거의 없다. 일곱째는, 개인 신체상의 변화에 따르는 영향을 적게 받는다. 여덟째는, 동적인 벡터 정보를 사용하므로 해킹에 의한 도용을 예방할 수 있다. 아홉째는, 여러 사용자가 동일한 서명으로 인증시스템을 공유하고자 할 경우, 매우 단순한 서명 패턴을 사용함으로써 그룹 사용자가 쉽게 활용할 수 있도록 하여 보안성, 편의성과 융통성 있는 시스템을 제공할 수 있다.

자필 서명은 서명자마다 그 특성이 다르고, 동일한 필기자의 경우에도 필기 자세, 감정 상태나 사용된 서명 입력 장치 등에 따라 조금씩 다르게 나타나며, 동일한 조건에서도 시간에 따라 다소 변하게 된다. 따라서 서명인증 시스템에서 주로 고려해야 할 요소는 많은 특징들 중에서 어떠한 특징을 실제 인증에 사용할 것인가 하는 것이고, 두 서명의 특징들을 어떻게 효과적으로 비교하여 정교한 상이도(dissimilarity)를 측정할 것인가 하는 문제이다.

두 서명의 특징을 비교하기 위해서는 RMS(Root Mean Square)방법[1,2]이나 DTW(Dynamic Time Warping)[3,4,5,6], Neural Network[7,8], HMM(Hidden Markov Model)[9,10,11,12] 등이 사용된다. RMS는 대용하는 두 함수를 위치에 따라 비교하여 그 차이를 누적시키는 선형 대용방법으로 약간의 차이나 왜곡에도 큰 상이도를 나타낸다. 이를 보완하기 위해서 상관관계(correlation)를 사용하기도 하나 이 방법도 미세한 왜곡

에 적절히 대응하지 못한다[6].

HMM은 은닉 관측열(hidden observation)을 갖는 Markov chain이다. 이는 관측 symbol을 구성하는 확률 함수 및 상태(state)와 상태 간의 전이 확률로 구성할 수 있다. 이 방법은 음성인식을 위한 패턴인식 기법으로써 가장 많이 사용한다. 그 이유로는 DTW나 신경망 등이 고립단어 수준의 인식기에서 좋은 성능을 나타내고 있으나 이를 대용량 어휘인식에 적용할 경우 해결하여야 할 많은 문제들을 지니고 있기 때문이다. HMM은 구성 모델 방식에 따라 소용량에서 대용량 어휘에 이르기까지 인식어휘 수에 따른 인식 단위 설정만으로 여러 종류의 인식기를 구현할 수 있다는 장점 때문에 많이 이용된다. HMM은 입력 파라미터의 형태에 따라 이산형 HMM과 연속형 HMM을 사용할 수 있다. 일반적으로 HMM의 선택은 이용 가능한 학습 자료와 모델의 정확성에 의해 결정된다. 연속형 HMM은 정확한 학자 모델화가 가능하지만, 학습 자료가 충분하지 않을 경우 학습과정이 수렴하지 않거나, 오히려 이산형 HMM에 비해 부정확한 모델을 만들게 되므로 적은 학습자료 환경에서는 이산형 HMM을 사용하는 것이 바람직하다. 그러나 비록 이산형 HMM이 연속형 HMM에 비해 적은 자료로도 학습이 가능하지만 시스템의 견고성(robustness)을 위해 여전히 많은 양의 학습 자료를 필요로 한다. 일반적인 이산형 HMM의 경우, 연관성 있는 벡터의 열을 양자화의 단위로 삼아 입력 패턴의 선형/비선형 종속성, 확률 밀도 함수의 형태 등에 포함된 중복성을 제거하여 특징 벡터의 수를 줄이는 벡터 양자화 (VQ, vector quantization)를 사용하여 각 특징 벡터를 하나의 코드워드로 변환하는 과정을 수행한 후, 코드 열을 HMM의 입력으로 사용하는 구조로 이루어져 있다.

Neural Network은 인간의 두뇌가 이루고 있는 뉴런(neuron)들의 결합인 신경망 조직을 모델링 하였다. 즉 인간 두뇌의 정보처리 방식을 모방하여 시스템이 어떤 업무를 수행할 수 있도록 훈련시키는 것이다. 인간의 두뇌가 대량의 복잡한 데이터를 효율적으로 병렬 처리할 수 있을 뿐만 아니라 학습 능력이 있다는 사실에 근거하여 신경회로망이 제안되어 여러 가지 방법에 응용되고 있다. 특히 신경망은 기본적으로 학습기능이 있어서 많은 가정이나 지식들이 병렬, 고속연산 등에 의해 처리되는 것을 필요로 하는 음성인식의 패턴인식의 처리에 있어서 우수한 성능을 나타낸다. 가변적인 가중치를 지닌 고리들로 연결

된 수많은 연산요소로 이루어진 방대한 병렬구조를 사용하기 때문에 많은 지식들을 탐색할 수 있는 것이다. 또한 주어진 출력을 내기 위해서 스스로 조정하고 맞추는 기능이 있다. 게다가 한번 학습되면 입력이 약간 바뀌더라도 출력은 영향을 받지 않는다. 이것은 잡음과 왜곡에 상당한 효과를 가지고 있다고 볼 수 있을 것이다. 학습기능과 병렬성을 보유하고 자기 조정(auto tuning)기능을 갖는 유연성이 뛰어난 계산 모델이지만 지식 표현이 구조적으로 분명하지 않고 전문가의 애매한 지식을 반영하기가 힘들며, 학습 시간이 걸리고 정확한 분류를 요구하는 문제에 대해서는 적용하기 힘든 단점도 있다.

DTW방법은 부분적인 왜곡에 대해 덜 민감하면서 정교하게 처리할 수 있고, 모든 가능한 정합 경로에 대해 가장 상이도(dissimilarity)가 낮은 경로를 효과적으로 찾아주는 방법으로 잘 알려져 있다. 일반적으로 두 서명의 길이는 선형적으로 변화하면서 전체 길이가 다를 수 있고, 비선형적으로 부분적인 길이가 다를 수 있으므로 서명인증을 위한 비교 알고리즘으로 우수한 성능을 낼 수 있다. 서명은 세월이 흐름에 따라 서명의 형태도 사람에 따라 약간씩 변하는 특징과 비밀번호를 바꾸어 사용하듯이 때로는 자신의 서명을 바꾸어서 사용해야 하는 경우가 있는데 이에 대해 효율적으로 대응하기 위해서는 DTW방식이 Neural Network이나 HMM 방식보다 우수함을 보이고 있다[3]. DTW는 적은 양의 데이터로도 학습이 가능하므로 학습과 데이터 수집이 간단하며, 특징 정보를 잘 선택하면 필요한 메모리 용량을 크게 줄일 수 있고, 연산량도 줄여서 처리속도를 향상시킬 수 있다.

기존의 동적 서명인증에 관한 연구에서는 다양하고 복잡한 특징 정보와 가중치를 사용하여 서명인증시스템에 적용하였으나, 특징 정보의 종류를 많이 사용한다고 해서 우수한 인증시스템의 성능을 보이는 것은 아니며, 특징 정보의 종류가 많을수록 적절한 가중치를 부여하는 것은 매우 어려운 일이다. 본 논문에서는, 서명의 여러 가지 중요한 특징(서명의 속도, 모양, 획 순서, 획 수, 서명 시간 등)을 잘 반영할 수 있는 특징 정보를 추출하였고, 실험을 통하여 적절한 가중치를 부여하였으며, 두 패턴을 비교하는 방법에서는 기존의 DTW 방법에서의 처리속도 문제와 정교성을 개선하여 제안된 DTW 방법[13]을 사용함으로써, 낮은 오류율(본인 거부율, 타인 수락률), 적은 양의 특징 정보, 빠른 처리 속도 등에서 성능을 개선할 수 있었다.

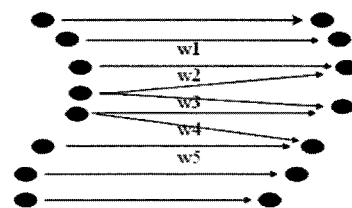
II. 서명 특징 정보

동적 서명인증시스템을 만들 때, 사용 가능한 특징 정보들을 잘 조합하여 사용하고 비교부에서 상이도를 계산하는 좋은 방법을 찾는 것이 무엇보다 중요하다고 볼 수 있다. 서명인증을 위하여 가장 좋은 특징 정보는 진서명의 변화 폭을 줄여 주고, 모조서명과의 변별력을 크게 해 주는 것이며, 여러 개의 특징 정보를 사용할 경우, 적절한 가중치를 부여하는 것 또한 중요하다고 볼 수 있다.

1. 다양한 특징 정보와 가중치

그림 1은 시간에 따른 서명의 동적인 특징 정보에 대한 가중치를 부여하는 것을 설명하고 있다. 어떤 특징 정보를 선택하여 대분류에 사용하고, 비교 알고리즘에 적용할 것인가 그리고 적절한 가중치를 부여할 것인가 하는 것은 매우 중요한 문제이다. 진 서명에 대해 모조가 쉬운 특징은 유효한 특징 정보에서 배제되어야 하고, 진서명 내에서 변화가 적은 특징이면서 다른 사람의 모조가 어려운 특징은 강조되어 큰 가중치를 부여할 필요가 있다.

서명 특징 정보	가중치
속도	w1
모양	w2
압력	w3
획(Stroke) 순서	w4
획 수	w5
전체 서명시간	w6
...	...



[서명A 특징정보][서명B특징정보]

그림 1. 특징 정보에 대한 가중치
Fig. 1 Weight of feature information

펜 입력 장치로부터 입력된 서명의 좌표점이 n 개의 점으로 구성되었다면, 그림 2의 서명정보는 아래의 식으로 표현될 수 있다.

$$\text{Sign}(i) = (x(i), y(i), t(i), p(i)), \quad i=1, 2, \dots, n \quad (2.1)$$

$x(i)$ 는 서명의 i 번째 x 좌표, $y(i)$ 는 서명의 i 번째 y 좌표, $t(i)$ 는 i 번째 좌표가 입력되었을 때의 시간, $p(i)$ 는 i 번째 좌표가 입력되었을 때의 압력 정보(펜을 누를 때의 세기 정보)를 나타낸다. 물론, 압력정보 $p(i)$ 는 압력을 감지할 수 있는 서명 입력 장치를 사용하였을 경우에 사용 가능한 정보이며, 이의 활용에 대해서는 적용 환경을 고려하여 판단해야 하며, 본 논문에서는 고려하지 않았다.

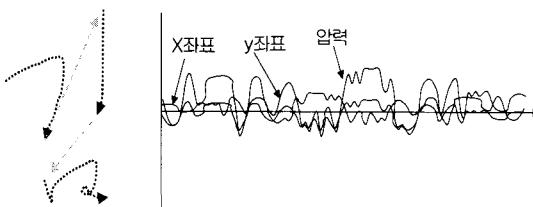


그림 2. 서명정보
Fig. 2 signature information

다양한 서명 특징들 중에서는 속도 정보가 가장 우수한 것으로 알려져 있으며, 본 연구에서도 실험 결과, 속도 정보가 가장 유용한 정보임을 알 수 있었다.

2. 제안된 특징 정보와 조정상수

본 연구에서 제안한 서명의 주요 특징 정보는, 서명의 Pen Down 시에 나타나는 속도정보, Pen Up 정보, 획 순서, 전체 획 수, 서명의 절대 크기 등이다. 기존의 연구에서도 잘 알 수 있듯이 서명의 속도 정보는 매우 중요한 정보이다. 즉, 서명의 대부분은 속도에 대하여 일관성을 유지하고 있음을 의미하는 것이다. 이 속도는 특징 정보로서 방향정보와 속력정보로 나눌 수 있는데, 서명의 방향에 대한 정보는 쉽게 누출될 가능성이 높은 특징을 가지고 있고, 속력에 대한 정보는 상대적으로 다른 사람에게 쉽게 누출이 되지 않아 보안성이 높은 특징을 지니고 있다. 본 논문에서는 서명에서 추출되는 두 점사이의 방향정보와 속력정보로 구분하였고, 속력정보를 이용하는 대신 조정상수(W)를 사용하는 방법을 제안하게 되었다. 이렇게 함으로써 전체 오류율에 영향을 주지 않으면

서 필요한 메모리 용량을 줄이고, 처리속도를 개선할 수 있었다. 이를 위해 통계적 방법 T-test를 이용하여 분석하였다.

III. 실험 및 결과 분석

300여명의 대기업 연구원, 100여명의 대학생, 일반인을 대상으로 수집된 진서명, 모조서명이 연구에 활용되었으며, 40가지의 서명 패턴에 대하여 1447개의 서명(진서명 750개, 모조서명 697개)을 8개의 표본서명으로 나누어 실험데이터로 사용하였고, PC (CPU: 2.1GHz, Main Memory: 256MB) Windows, iPaq PDA 환경에서 실험하였다.

현재 국내·외에서 연구, 개발된 다양한 동적 서명인증시스템이 소개되고 있지만, 이 기술을 객관적으로 평가하고 검증할 수 있는 기준이나 지침서가 없다. 그리고 사용 환경, 사용된 서명 데이터베이스(진서명, 모조서명), 사용된 특징 정보 및 가중치, 비교 알고리즘 등이 동일하지 않으면, 오류율(EER)의 절대 비교는 어려울 것이다. 특히, 진서명과 모조서명의 저장 형태에 따라서 실험결과의 오류율은 크게 차이가 날 수 있음은 기정사실이다. 객관적인 오류율을 평가하는 방법에는 현실적으로 다소 어려움이 있는 실정이며, 이를 위해서는 객관적으로 공인된 서명 데이터베이스(진서명, 모조서명)가 준비되어 있어야 한다. 본 논문에서는 자체적으로 구축한 서명 데이터베이스를 기준으로 성능을 분석하였다.

본 논문에서 제안하고 구현된 방법의 실험 및 통계적 분석 결과를 요약하면 표 1과 같다. 표 1은 제안된 DTW 방법과 특징 정보에 따른 오류율을 나타낸다.

동적 서명인증을 위한 좋은 특징 정보는 진서명의 변화 폭을 줄여 주고, 모조서명과의 변별력을 크게 해 주는 것이며, 여러 개의 특징 정보를 사용할 경우, 특징 정보들을 잘 조합하고 적절한 가중치를 부여하여 두 서명의 상이도를 계산하는 것이 무엇보다 중요하다. 그러나 실제로 여러 개의 서명 특징 정보에 대한 가중치를 효과적으로 부여하기란 매우 어려운 일이다. 표 1은 기존의 속도 정보(방향정보, 거리정보)와 제안된 특징 정보의 사용에 따른 오류율이다.

방법2는 제안된 DTW 방법에 방향정보와 거리정보를 이용하고, 적절한 조정상수(w)를 사용하였을 때의 오

류율이다. 제안된 방법(방법1)에 있는 w 는 해당 서명표본에서 가장 적절한 조정상수(w) 값이며, T 는 해당 EER에서의 판단 임계치(보안수준)를 나타낸다.

표 1. 제안된 DTW 방법과 특징 정보에 따른 오류율 (EER : %)

Table 1 Error rate by proposed DTW method

표본 방법	서명 표본1	서명 표본2	서명 표본3	서명 표본4	서명 표본5	서명 표본6	서명 표본7	서명 표본8
방법2	4.0	5.6	3.75	21.2	1.6	1.35	2.85	13.05
	w=50 T=30	w=40 T=30	w=60 T=30	w=60 T=30	w=60 T=36	w=70 T=37	w=60 T=37	w=60 T=45
	5.0	3.05	2.85	3.65	0.9	0	1.09	3.9
(제안된 방법)	w=40 T=20	w=30 T=15	w=40 T=20	w=70 T=35	w=60 T=30	w=60 T=30	w=60 T=30	w=60 T=30

방법 1: 제안된 DTW 방법 + 방향정보 + 조정상수(w)

방법 2: 제안된 DTW 방법 + 방향정보 + 거리정보 +

조정상수(w)

표 2는 아래의 귀무가설(H_0)과 대립가설(H_A)에 따른 방법1과 방법2의 통계적 가설검정을 위한 T-test 결과이다. 검정통계량 t 값에 대한 유의확률은 0.143으로 유의수준 0.05보다 크다. 따라서 유의수준 5% 하에서 ‘방법1과 방법2의 오류율의 차이가 존재하지 않는다.’라고 결론을 내릴 수 있다.

귀무가설 H_0 : 방법1 = 방법2

대립가설 H_A : 방법1 \neq 방법2

표 2. 방법1과 방법2의 오류율에 대한 T-test 결과

Table 2. T-test result by error rate

of method 1 and method 2

방법	표본 수	오류율 평균	오류율 표준편차	검정통계량 t	유의 확률
1	8	2.555	1.721		
2	8	6.675	6.941	1.630	0.143

결과에서처럼 본 논문에서 제안된 방법1과 방법2의 오류율을 비교해 보면 근소한 차이지만 방법1이 약간 우수함을 보이고 있으나, 위의 통계적 분석에 따르면, 서명에서 나타나는 두 점사이의 실제 거리 정보가 오류율에 큰 영향을 주지 않음을 알 수 있다.

IV. 결론

컴퓨터와 인터넷이 널리 보급됨에 따라서 업무의 효율과 생산성을 높이고, 멀리 떨어져 있는 사람들과 실시간으로 손쉽게 정보를 주고받는 등 사회에 주는 긍정적인 영향은 이루 말로 표현할 수 없을 것이다. 그러나 최근 들어 컴퓨터와 인터넷을 이용한 대형 사이버 범죄가 증가하고 있는 실정이다. 본 논문은 사이버 범죄, 신용카드 범죄 행위를 예방하고 방지하는 사용자 인증보안 기술로서, 자필 서명을 이용하여 본인여부를 확인하는 동적 서명인증 기술에 관한 것이다.

본 논문에서는 서명에서 생성되는 속도 성분(방향 정보, 거리 정보)에서 방향 정보만을 특정 정보로 서명 DB에서 저장하고 거리(속력) 정보는 조정상수(w)를 이용함으로써 오류율에 영향을 주지 않고, 처리속도를 개선하고, 특정 정보 크기도 크게 줄일 수 있었다. 이를 확인하기 위해 통계적 방법 T-test를 이용하여 분석하였다. 이러한 결과는 네트워크 전송속도 문제, 서명인증 처리속도에 민감한 환경이나 인증 서버의 메모리 용량에 민감한 즉, 방대한 양의 회원관리나 정보를 처리해야 하는 환경에서는 매우 중요하며 큰 의미를 가질 것이다.

참고문헌

- [1] G. Dimauro, S. Impedovo, G. Pirlo, "Component-oriented algorithms for signature verification", IJPRAI, Vol.8, No.3, pp. 771-794, 1994.
- [2] G.Dimauro, S.Impedovo, G.Pirlo, "A stroke-oriented approach to signature verification", in From Pixels to Features III - Frontiers in Handwriting Recognition, S. Impedovo and J.C.Simon eds., Elsevier Publ., pp.371-384, 1992.
- [3] Hansheng Lei, Srinivas Palla, Venu Govindaraju, "ER2: An Intuitive Similarity Measure for On-Line Signature Verification", Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04), pp.191-195, October 2004.

- [4] Sascha Schimke, Claus Vielhauer, Jana Dittmann, "Using Adapted Levenshtein Distance for On-Line Signature Authentication", Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2, pp.931-934, August 2004.
- [5] M. E. Munich, P. Perona, "Continuous Dynamic Time Warping for Translation Invariant Curve Alignment with Applications to Signature Verification", (1999), Available at: <http://citeseer.nj.nec.com/munich99continuous.htm>.
- [6] M. Perizeau and R. Plamondon, "A comparative analysis of regional correlation, dynamic time warping and skeletal tree matching for signature verification", IEEE T-PAMI, Vol.12, No.7, pp.710-717, 1990.
- [7] C. Quek, R.W. Zhou, "Antiforgery: a novel pseudo-outer product based fuzzy neural network driver signature verification system", Pattern Recognition, Vol.23, pp.1795-1816, 2002.
- [8] D. Lettman and S. George , "On-line handwritten signature verification using wavelets and back-propagation neural networks", Proc. of ICDAR '01, Seattle, pp.596-598, 2001.
- [9] M. Fuentes, S. Garci-Salicetti, B. Dorizzi, "On line Signature Verification: Fusion of a Hidden Markov Model and a Neural Network via a Support Machine", Proc. of IWFHR-8, Canada, pp.253-258, 2002.
- [10] Marc Fuentes, Sonia Garcia-Salicetti, Bernadette Dorizzi, "On-Line Signature Verification: Fusion of a Hidden Markov Model and a Neural Network via a Support Vector Machine", Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), pp.253, August 2002.
- [11] Mohammad M. Shafiei, Hamid R. Rabiee, "A New On-Line Signature Verification Algorithm Using Variable Length Segmentation and Hidden Markov Models", Seventh International Conference on Document Analysis and Recognition Volume I, pp.443, August 2003.
- [12] H. S. Yoon, J. Y. Lee, H. S. Yang, "An On-Line Signature Verification System Using Hidden Markov Model in Polar Space", Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), pp.329, August 2002.
- [13] 김진환, 조혁규, 차의영, "고성능 동적 서명인증시스템 구현", 한국해양정보통신학회 논문지 제9권 4호, 2005.7.

저자소개



김 진 환(Jin-Whan Kim)

1989년 부산대학교 전산통계학과
(이학사)

1992년 연세대학교 컴퓨터과학과
(이학석사)

2006년 부산대학교 전자계산학과 (이학박사)

2003년 ~ 현재 : 영산대학교 컴퓨터공학과 조교수

2001년 ~ 현재 : (주)마이그룹 대표이사

※ 관심분야: 관심분야: 동적 서명인증, 필기체 문자인식, 다중 바이오인식, 유비쿼터스 컴퓨팅



조 재 현(Jae-Hyun Cho)

1998년 부산대학교 전자계산학과
(이학박사)

2001년 3월~현재 부산가톨릭대학교
컴퓨터공학과 부교수

2005년~현재 해양정보통신학회 논문지 편집위원

※ 관심분야: 신경회로망, 문자인식, 인간시각시스템