

서명검증에서 특징값을 고려한 판단 경계 설정에 관한 연구

이 흥 열, 김 재 희

연세대학교 전기·컴퓨터공학과 Intelligent Vision 연구실
전화 : (02) 361 - 2869

Determination of Decision Boundary Using Feature Values in the Signature Verification

Heung Yeol Lee, Jaihie Kim

Dept. of Electrical and Computer Engineering, Yonsei university

E-mail : ofelixo@unitel.co.kr

Abstract

Usually, more reference signatures result in better performance in signature verification. However, registering many signatures may be a tedious work for users, so algorithms that use less signatures for the registration without increasing error rate is needed.

In this paper, we find the features such as pen-down duration, the number of locally minimum velocity points, and the number of locally maximum curvature points. Then we find the relationship between these features and the optimal decision boundary. We apply this relationship in deciding threshold for signature verification.

Experimental results show that the method using three reference signatures has almost same error rate as algorithms with many references.

1. 서 론

최근 해킹등 적대적인 의도로 이루어지는 외부로부터의 침입에 대해서 컴퓨터에 저장된 중요한 데이터의 보안과 사용자의 확인 및 관리의 문제가 대두되고 있

다. 이러한 문제 해결을 위해 개인의 서명을 통하여 자동으로 본인 여부를 확인하는 연구가 계속 진행되고 있다[1][2]. 이는 서명을 통한 인증(identification) 방법이 도장이나 ID카드와 같이 분실 할 위험이 없고, 또 지문인식처럼 사용자에게 불필요한 불편감을 유발시킬 우려도 없다는 강점을 가지고 있기 때문이다.

온라인 서명 검증의 접근 방법은 크게 함수적 접근 방법과 매개변수적 접근 방법으로 나누어진다. 함수적 접근 방법은 입력 신호를 시간에 따른 특징함수로 나타낸 후 서명의 시작에서 끝까지 모든 함수 값을 비교함으로써 서명의 진위를 판별하고[3], 매개변수적 접근 방법은 서명으로부터 서명의 특징이 되는 매개변수를 추출, 이를 매개변수 공간에 대응시킴으로써 등록된 서명과 대표서명과의 거리를 구하여 진위 여부를 판별하게 된다[4][5].

기존 논문에서의 서명검증 방법은 다수의 참조서명으로부터 임계치를 결정, 서명검증을 수행하기 때문에 실제로 사용자가 서명을 여러 번 반복하여 등록시켜야 하는 번거로움이 있다. 그리고 서명의 변화도 정보를 이용하여 임계치를 결정하기 때문에 임계치를 설정할 때 서명의 역동적인 신호값이나 국부적 변화를 반영하지 못하며, 따라서 참조서명의 개수가 줄어드는 경우에는 적절한 판단 임계치의 결정이 어렵게 된다.

그래서 본 논문에서는 적은 참조서명으로도 낮은 오류율을 나타내도록 함수적 접근 방법을 응용하여 임계치를 설정하는 방법에 대하여 연구하였다. 먼저 참조서명에 대한 최적의 임계치를 실험적으로 구하고, 이 임계치와 관련이 있는 서명의 특징에는 어떤 것이

있는지, 그리고 서명자의 서명 특징이 그 서명자의 최적의 임계치와 선형적인 관계가 있는지를 조사하고 관련이 있는 특징들의 수치를 이용하여 가변적으로 임계치를 결정하였다.

2. 매개변수적 방법에서 다중 군집화를 이용한 서명 검증

매개변수적 접근방법에서, 복수의 프로토타입을 이용하는 서명검증방법에서는 서명 데이터를 여러 개의 군으로 군집화하는 과정이 필요하다. 서명을 군집화하는 방법으로는 K-means 알고리즘이나 Forge 알고리즘과 같이 군의 개수가 미리 정해진 상태에서 군집화(supervised clustering)를 수행하는 방법과 ISODATA와 같이 군의 개수를 데이터 분포의 특성에 따라 시스템이 결정하는 방법(unsupervised clustering)이 있다. 서명데이터의 경우, 각 서명자마다 데이터의 분포가 달라지기 때문에 K-means 알고리즘이나 Forge 알고리즘과 같은 방법으로는 서명 데이터들의 이상적인 군집화가 불가능하다.

2.1 ISODATA방법의 문제점

ISODATA의 방법을 사용하여 서명의 군집화를 수행하고 서명의 특징값에 대한 변화도를 이용하여 임계치를 결정하는 경우, 참조 서명의 개수가 줄어들어 인하여 서명 데이터에 대한 절대적인 정보량이 감소하므로 서명의 변화도 정보가 서명자의 서명 특성을 충분히 반영시키기 어렵다[6]. 그래서 이러한 방법을 사용하였을 때 서명 데이터의 개수가 3개로 줄어들 경우의 오류율은 급격하게 높아지게 된다. 따라서 참조서명의 개수가 적을 경우, 다이내믹 프로그래밍 기법(DP Matching)[7]과 같이 서명의 역동적인 신호값이나 국부적 변화 등을 고려할 수 있는 방법이 필요하다.

3. 함수적 방법에서 서명의 특징정보를 이용한 임계치 설정 방법

함수적 접근 방법에서 쓰일 수 있는 특징으로는 속도, 속도, 압력, 가속도 특징이 있으며, 그 중에서 속도 특징이 가장 우수하므로[3] 본 논문에서는 속도 특징에 대해서 다이내믹 프로그래밍을 수행하였다. 속도 특징을 얻는 방법은 다음과 같다. 태블릿으로부터 입력받은 신호가 n개의 점으로 구성되었다고 할 때, 서명 신호는 식 (1)과 같이 표현된다.

$$Signal(i) = (x_i, y_i, t_i, p_i), i = 1, 2, \dots, n \quad (1)$$

이 때, x_i 는 서명의 i 번째 x좌표이고 y_i 는 서명의 i 번째 y좌표이며, t_i 는 $i-1$ 번째 서명 신호가 들어온 후 i 번째 서명신호가 들어올 때까지 걸린 시간, p_i 는 i 번째 서명 신호의 압력을 나타낸다. 이러한 n개의 서명 신호들로부터 i 번째 점에서의 속도 특징 $v(i)$ 는 식 (2)와 식 (3)으로 구한다.

$$v(i) = (v_x, v_y) = (0, 0) \quad i = 1 \quad (2)$$

$$v(i) = (v_x, v_y) = \left(\frac{x_i - x_{i-1}}{t_i}, \frac{y_i - y_{i-1}}{t_i} \right) \quad i = 2, 3, \dots, n \quad (3)$$

위와 같은 속도 특징을 이용, 서명의 상이도를 사용하여 임계치를 결정하는 함수적 접근방법 역시 참조서명의 개수가 줄어드는 경우에는 오류율이 급격하게 증가하게 된다. 그래서 본 논문에서는 서명의 특징값을 이용하여 임계치를 설정하는 방법을 제시하였다.

3.1 최적의 임계치와 서명 특징과의 관계 도출

우선 각 서명자들로부터 얻어낸 3개의 참조서명 각각을 서명의 프로토타입으로 정의하고, 전체의 검증 서명 데이터에 대해서 최적의 임계치를 실험적으로 조사하였다. 여기에서 구해진 최적의 임계치와 서명데이터로부터 얻을 수 있는 76개의 특징값 각각을 2차원 평면에 맵핑(mapping)시키고, 이들 관계 중에서 선형적인 관련이 있는 특징에는 어떤 것들이 존재하는지를 조사하였다.

3.2 선형근사화를 통한 최적의 임계치와 서명 특징간 관계의 일반화

위의 과정을 통해서 임계치와 특징값 사이에서의 선형적인 관계를 가지는 특징을 찾은 후 이 특징값과 최적의 임계치 사이의 선형근사화를 수행하고, 그럼으로써 발생하는 최적의 임계치와의 오차를 최소화하기 위해 본 논문에서는 수치해석에서 사용하는 최소자승법(Least Square Approximation)을 사용하였다.

3.3 임계치 설정의 최적화 방법

앞에서 구한 최적의 임계치와 선형적인 관계를 보이는 특징이 여러 개 존재할 경우 이들 특징을 하나씩 고려하여 서명 검증을 수행하는 경우보다 이들의 관계를 모두 고려하였을 때 보다 낮은 오류율을 얻을 수 있다.

본 논문에서는 n개의 특징에 대해서, 최적의 임계치와 특징에 대해 일차방정식으로 표현되는 선형 근사화된 관계식의 기울기와 Y절편에 대해서 scaling factor의 개념을 이용하여 최적의 해를 찾는 방법을 사용하였다.[8]

i번째 서명의 특징값을 x_i 라고 하고, i번째의 최적의 임계치를 θ_i 라고 하였을 때, x_i 와 θ_i 사이의 선형근사화된 관계식은 다음과 같다.

$$\theta_i = A_i x_i + B_i \quad (4)$$

식(4)와 같이 선형근사화된 관계식의 기울기와 Y절편을 각각 A_i , B_i 라고하고, i번째 특징의 기울기 조건을 사용하였을 경우 올바르게 검증될 확률과 Y절편의 조건을 사용하였을 때 올바르게 검증될 확률을 각각 $P(X/A_i)$, $P(X/B_i)$ 라고 하면 최적화된 기울기 A_{op} 와 y절편 B_{op} 는 식(5)와 식(6)처럼 표현된다.

$$A_{op} = \sum_{i=1}^n (A_i P(X/A_i)^a)^r \quad (5)$$

$$B_{op} = \sum_{i=1}^n (B_i P(X/B_i)^a)^r \quad (6)$$

물론 임계치 θ_i 는 기울기 A_i 와 Y절편 B_i 가 모두 존재하여야만 결정되므로 식(6)의 관계가 성립한다.

$$P(X/A_i) = P(X/B_i) \quad (7)$$

위의 확률 값을 판단 임계치를 θ_i 로 설정하였을 경우의 검증률이다.

식(5)와 식(6)에서 a와 r을 변화시키면서 가장 최소의 오류를 나타내는 선형화된 임계치 식의 기울기와 Y절편을 실험적으로 결정하게 된다. 그리고 식(5)와 식(6)에서 구해진 기울기와 Y절편으로부터 최적화된 임계치 θ_{op} 는 식 (8)에 의해 결정된다.

$$\theta_{op} = A_{op}x + B_{op} \quad (8)$$

4. 실험 결과

ISODATA를 이용한 서명 검증 방법에서는 참조서명의 개수가 많을 경우 기존의 접근 방법에 비해 우수한 성능을 나타내고 있지만, 참조서명의 개수가 적은 경우에는 크게 개선되지 않은 결과를 보이고 있다. 기존의 함수적 방법에서 참조 서명 데이터의 수를 3개로 제한하였을 경우의 오류율은 표 1에서 보는 바와 같이 10%이상으로 매우 높다.

표 1. 함수적 접근 방법에서 참조서명의 개수에 따른 오류율의 변화

참조서명의 개수	3개	6개	9개	12개	20개
오류율	10.48%	6.98%	5.71%	4.98%	4.79%

76개의 특징에 대해 임계치와의 관계 여부를 조사해본 결과, 임계치와 관계가 있는 특징으로는 최소 속력점의 개수와 최대 곡률점의 개수, 그리고 펜이 눌린 시간입을 확인하였다. 그림 4부터 그림 6은 최적의 임계치와 각 특징값과의 관계이며 그림내의 선분은 서명 데이터들을 선형 근사화시킨 결과이다.

그림 7은 식 (5)와 식(6)의 a와 r을 동시에 변화시키면서 실험을 수행한 결과이다. a = 4, r = 1.2 에서 최소의 오류율을 보이고 있으며, 그때의 평균 오류율은 5.18%이다. 표 2의 경우와 비교하였을 때, 특징을 하나씩만 적용시킨 경우의 최소 오류율(5.34%)보다 scaling factor를 이용하여 세 가지의 특징을 모두 고려한 경우 보다 높은 검증율을 보이고 있다.

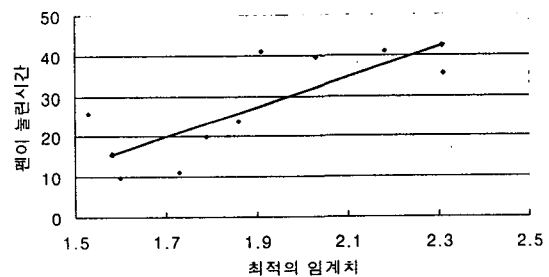


그림 4. 펜이 눌린 시간과 최적의 임계치와의 관계

5. 결 론

기존의 방법에서는 서명의 평균 변화도나 상이도를 사용하여 임계치를 결정하므로 참조 서명데이터의 개수가 많은 경우에는 높은 검증률을 얻을 수 있지만, 서명데이터의 개수가 적어질 경우 오류율이 급격하게 높아지게 된다. 본 논문에서 제안하는 방법을 사용하여 사용자마다 다른 임계치를 설정함으로써, 적은 수의 참조서명 데이터만으로도 높은 검증률을 나타낼 수 있었다.

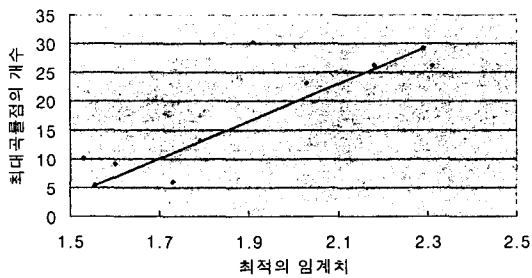


그림 5. 최대곡률점의 개수와 최적의 임계치와의 관계

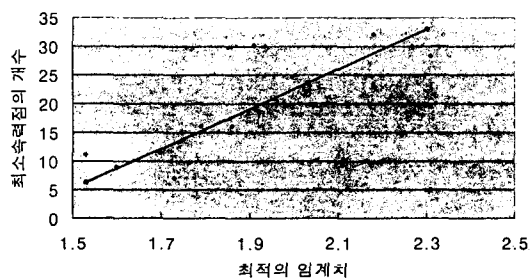


그림 6. 최소속력점의 개수와 최적의 임계치와의 관계

표 2. θ_{min} 을 선형화 시킨 경우의 오류율

특징	임계치와 특징과의 관계식	오류율
펜이 눌린 시간	$\theta = 0.024 x_1 + 0.743$	6.14%
최대 곡률점의 개수	$\theta = 0.027 x_2 + 0.897$	5.34%
최소 속력점의 개수	$\theta = 0.025 x_3 + 0.884$	5.62%

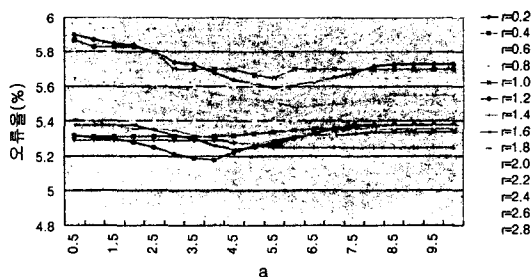


그림 7. a와 r을 변화시킨 경우 검증률의 변화

참고 문헌

- [1] R. Plamondon and Guy Lorette, "Automatic Signature and Writer Identification - The State Of Art", Pattern Recognition, Vol. 22, No. 2, pp. 107-131, 1989
- [2] E. Howbrook "Signature Recognition for Security Systems", Proc. of the Conference on Microprocessors in Automation and Communication, university of Kent, England, pp. 377-384, 1978
- [3] 유재룡, "온라인 서명 검증을 위한 특징의 비교 분석 및 적용에 관한 연구", 연세대학교 대학원 석사학위 논문 1994
- [4] 박명수, "온라인 서명 검증에서 특징 집합에 대한 각 서명별 가중치 설정 방법", 연세대학교 대학원 석사학위 논문 1994
- [5] G.Lorette and R.Plamondon,"Dynamic Approaches to Handwritten Signature Verification", in R. Plamondon and C. G. Leedham(eds), Computer Processing of Handwriting, World Scientific Pub, pp. 21-47, 1990
- [6] Donald H. Foley, "Consideration of Sample and Feature Size", IEEE Trans. on Information Theory, Vol. IT-18, No. 5, pp. 618-626, September, 1972
- [7] T. C. Hu, "Dynamic Programming Combinatorial Algorithms", Addison - Wesley Publishing Company, pp. 107 - 173, 1992
- [8] Earl gose, Richard Johnsonbaugh and Steve Jost, "Pattern Recognition and Image Analysis", Prentice Hall Inc, 1996
- [9] Jaihie Kim, J.R. Yu, and S. H. Kim, "Learning of prototype and decision boundaries for a verification problem having only positive samples", Pattern Recognition Letters, Vol. 17, pp. 691-697, 1996.