

구조적 접근방식의 온라인 자동 서명 검증 기법

김성훈*

A Structural Approach to On-line Signature Verification

Seong-Hoon Kim *

요약

이 논문에서는 온라인 서명 검증에 대한 새로운 접근 방법으로, 서명을 필기의 기본 구성 요소들로 표현하는 구조적인 접근 방법을 제시한다. 속력의 국부 극소점을 분할점으로 하여 얻어지는 부분을 원시 패턴으로 정의하였고, 서브패턴으로서 방향 변화에 의해 얻어지는 단순 회전형 성분, 첨형 성분, 종형 성분의 세 가지 종류를 정의하여 서명을 구조적으로 표현하였다. 그리고 서브패턴을 기본 단위로 한 동적 계획 정합 알고리즘을 제시하였다. 또한, 학습 샘플로부터 국부적인 부분에 대한 변화도와 복잡도를 추출하여 참조 패턴의 학습과 진위 판단 임계치를 설정하였다. 실험을 통하여, 함수적 접근과 매개변수적 접근 그리고 제안된 구조적 접근 방법을 동일 조건에서 검증률, 처리시간, 메모리의 측면에서 비교하였다. 제안된 방법에서는, 서명의 국부적인 가중치를 적용하고 복잡도를 판단임계치의 설정에 사용함으로써 평균 오류율이 14.2%에서 4.05%로 검증률이 크게 향상되었다. 또한, 시간과 메모리의 측면에서 함수적 접근 방법에 비해 효과적이면서도 함수적 접근 방법의 최고 성능에 근접한 검증률을 얻을 수 있었다.

Abstract

In this paper, a new structural approach to on-line signature verification is presented. A primitive pattern is defined as a part segmented by a local minimal position of speed. And a structural description of signature is composed of subpatterns which are defined as such forms as rotation shape, cusp shape and bell shape, acquired by composition of the primitives regarding the directional changes. As the matching method to find identical parts between two signatures, a modified DP(dynamic programming) matching algorithm is presented. And also, variation and complexity of local parts are computed from the training samples, and reference model and decision boundary are derived from these. Error rate, execution time and memory usage are compared among the functional approach, the parametric approach and the proposed structural approach. It is found that the average error rate can be reduced from 14.2% to 4.05% when the local parts of a signature are weighted and the complexity is used as a factor of decision threshold. Though the error rate is similar to that of functional approaches, time consumption and memory usage of the proposed structural approach are shown to be very effective.

▶ Keyword : 구조적 서명검증(structural signature verification), 국부 모양(local shape), 필기표현(handwriting representation), 서명 복잡도(signature complexity), 국부적 가중치(local weight)

• 제1저자 : 김성훈
• 접수일 : 2005.08.31, 심사완료일 : 2005.09.23
* 영동대학교 컴퓨터공학과 교수

1. 서론

그동안의 오랜 연구에도 불구하고 자동 서명 검증의 문제는 실용적인 측면에서 만족할 만한 성과를 거두지 못하고 있다. [1] 자동 서명 검증의 문제를 어렵게 만드는 주된 첫 번째 요인으로 동일인의 서명 내에서도 자체적인 약간의 변화를 가진다는 점이다. 이로 인해 자동 서명 검증은 개인의 변화를 최대한 흡수하는 동시에 가짜 서명은 정확하게 색출해내야 하는 요구사항을 만족시켜야 한다. 두 번째로는 일반적인 패턴인식 문제의 학습과정에서 사용되는 상대적인 (negative) 학습 샘플 데이터 집합을 서명 검증의 문제에서는 사전에 확보할 수 없다는 것이다. 이것은 학습과정에서 포지티브 샘플 데이터 집합만을 이용해야 한다는 제약 때문에 서명검증 문제에서의 판단 임계치 설정을 어렵게 만든다. 세 번째 문제로는 학습 샘플 데이터의 개수가 충분히 많이 제공되지 못하는 제약을 가진다. 이로 인해 평균치나 분산치와 같은 통계적인 속성값의 정확성을 확보하기 어렵다는 한계를 갖는다. 네 번째로, 자동 서명 검증 문제를 생체 보안 응용분야에 활용하기 위해서는 저장 데이터의 일정 크기로의 제한과 서명의 재복원 방식이 가능한 저장 방식이어야 한다는 것이다.

지금까지의 서명검증에 대한 접근은 크게 두가지로 함수적 접근(functional approach)와 매개변수적 접근(parametric approach)로 나누어진다[2]. 서명 검증 연구의 초기단계에서 주로 행하여진 함수적 접근은 입력된 신호를 시간에 대한 특징함수로 나타내고 서명의 시작에서부터 끝까지 모든 함수값을 비교함으로써 입력된 서명의 진위를 판별한다[3]. 이에 반해 매개변수적 접근은 서명으로부터 서명의 특징이 될 매개변수를 추출하여 이를 매개변수 공간에 대응시켜 미리 등록된 서명을 대표하는 중심값과의 거리를 계산하여 진위여부를 판별한다[4].

두 접근방법의 이러한 한계는 서명을 구성하는 국부적인 형태를 고려하지 않고 단순히 특징함수 또는 특징 집합의 평균치 벡터로 표현하여 서명에 존재하는 다양한 정보를 끌어내어 사용하지 못하였기 때문에 앞서 언급한 서명의 자체적인 변화성 흡수문제와 학습과정에서의 제약성을 극복하지 못하는 한계를 지닌다. 즉, 서명에 존재하는 정보로는 서명

을 구성하는 기본적인 형태의 조합, 국부적인 부분에서의 변형 형태와 복잡한 정도, 검증에 유용한 정도, 등의 다양한 정보를 활용하지 못하는 것이다[5].

이 논문에서는 서명 검증에 대한 새로운 접근방법으로 서명으로부터 다양한 정보를 끌어내어 사용하기 위해 서명의 표현에 있어서 서명의 구성요소들로 나타내고, 서명검증에서 국부적으로 중요한 부분과 유용한 특징을 선택적으로 사용하는 구조적인 접근방법을 제안하고자 한다.

II. 서명검증의 구조적 접근방법

2.1 구조적 서명 검증 방법의 특성

구조적인 서명 검증은 기존의 서명 검증 방법에 비해 다음과 같은 특성을 갖는다.

1. 서명을 구조적으로 표현한다(그림 1 참조). 서명을 최하위 단위의 기본 요소들로 분할하여 이들로로부터 서명을 재구성하면서 서명의 특징을 비교한다. 그러므로 비교시간이 감소된다. 또한 동일 부분의 매핑이 정확해지며, 모양의 분석이 가능하다. 그리고 국부적인 부분에서의 강조가 가능하다.

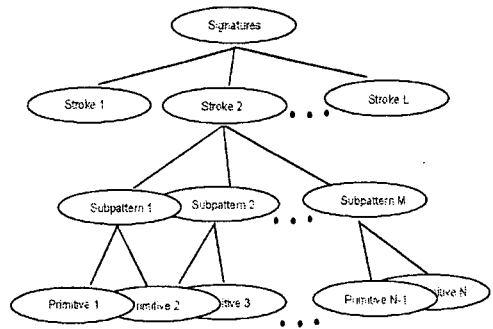


그림 1. 서명의 구조적 표현
Fig. 1 Structural representation of signature

2. 서명의 구성 성분들에 대해 국부적으로 서명 검증에 중요한 부분과 유용한 특징을 선택적으로 사용할 수 있도록 참조 서명의 모델을 형성한다.

즉, 서명을 구성하는 세그먼트에 대해 진서명내의 변화가 심한 부분은 비교 과정에서 덜 중요하게 참여시키고 변화가 거의 없는 안정된 부분은 강조하여 비교되도록 한다. 또한 세그먼트가 단순하여 다른 사람이 모조하기 쉬운 부분에서는 중요도를 약화시키고, 반대로 서명이 복잡하여 상대방이 모조하기 어려운 부분에서는 강조하여 비교될 수 있도록 한다(그림 2 참조).

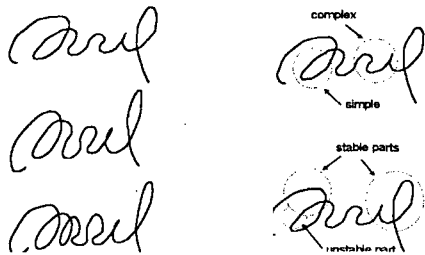


그림 2. 서명의 국부적 변화성과 복잡성의 예
Fig. 2 An example of local complexity and variation of signature

3. 최종 판단 임계치의 결정을 서명의 전체 변화도와 복잡도에 의해 결정함으로써 검증 오류율을 안정시킬 수 있다.

서명 검증 과정 중의 최종 진위 판단 단계에서 사용되는 판단 임계치는 서명 검증의 성능을 좌우하는 중요한 요소가 된다. 지금까지의 서명 검증 방법에서는 판단 임계치의 설정을 모조서명의 샘플을 이용한 검증 오류율의 추정 실험을 통하여 각 사람별로 판단 임계치를 설정하거나(4), 아니면 진서명들간의 비교를 통한 차이값의 변화의 정도에 따라 판단 임계치를 설정하였다(6)(7). 모조서명 샘플들에 의한 검증률 추정 실험은 실제로 검증 시스템을 할용시 서명 등록 자료부터 많은 갯수의 모조 서명 샘플을 확보해야 하므로 비실용적이다. 또한, 진서명들간의 변화 정도에 따라 임계치를 설정하는 방법은 모조 서명을 사용하지 않는다는 장점을 가지지만, 이 방법은 근본적으로 유형-I 에러율(진서명 거부 에러율)을 예측할 뿐이지 유형-II 에러율(모조서명 인정 에러율)은 고려하지 못한다는 점에서 바람직하지 못하다.

하지만, 진서명의 여러가지 속성들에 의해 서명의 복잡성을 산출할 수 있으며, 복잡한 서명은 일반적으로 상대방이 모조하기 어려운 정도를 나타낸다고 가정할 수 있다. 즉 모조가 어려운 서명의 경우 모조서명과의 차이값의 분포는 모조가 쉬운 서명에 비해 상대적으로 큰 차이값을 나타내게 된다. 서명에 대해 사람의 모조 능력의 수학적인 모델링은 필기 모델의 가설을 바탕으로 하여 Brault와 Plamondon에 의해 연구된 바 있다(8). 이 논문에서는 이러한 관점에서 임계치의 결정에서 변화도 뿐만 아니라 서명의 복잡도와와의 상관 관계를 구하여 활용하였다.

2.2 구조적 서명 검증 시스템의 구조

이와 같은 특성을 지닌 서명 검증 시스템의 구성을 (그림 3)에 보이고 있다.

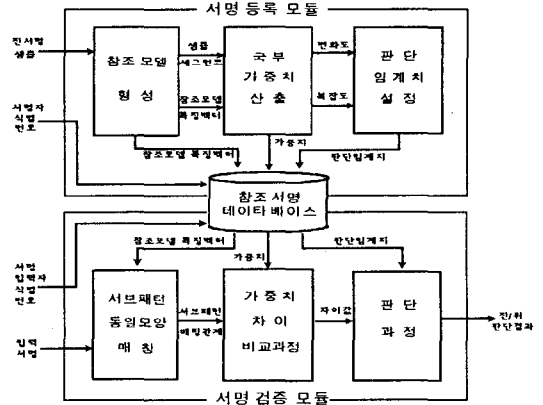


그림 3. 서명검증시스템의 구성도
Fig. 3 Signature Verification System Architecture

서명 검증 시스템은 등록 모듈과 검증 모듈로 구성된다. 서명 등록 모듈은 몇개의 진서명 샘플을 입력받아서 참조 모델 형성, 국부 가중치 산출, 그리고 판단 임계치의 과정을 거쳐 참조 서명에 대한 정보를 생성하고 이를 데이터베이스에 저장한다. 서명 검증 모듈은 검증해야 될 서명을 입력받아 데이터베이스에 저장된 참조 모델과 비교하는 서브패턴 동일 모양 매칭, 가중치 차이 비교과정, 그리고 판단 과정을 거쳐서 입력 서명의 진위를 판단한다.

III. 서명의 구조적 표현

일반적으로 점의 좌표들만으로 구성되는 온라인 필기 데이터를 문자 또는 기호, 서명 등의 의미를 가진 패턴으로 인식하기 위해서는, 인식 대상인 패턴을 적절한 크기의 서브 패턴(subpattern)으로 나누고, 이들 서브 패턴을 다시 최소 단위의 원시패턴(primitive pattern)들로 나누어 표현한다. 가령, 한글 문자의 경우 하나의 문자는 초성, 중성, 종성과 같은 자소 단위의 중간 패턴으로 구성되며, 자소 단위의 중간 패턴은 몇개의 획들로 구성된다. 인식 방법에 따라, 획 자체를 인식의 최소 단위인 원시 패턴으로 설정하는 경우[9]도 있고, 획을 더 작은 부분으로 분할한 부분획을 인식의 최소 단위인 원시 패턴으로 설정할 수도 있다. 이처럼 패턴 인식에서 패턴(pattern)을 더 작은 단위의 서브패턴으로 축소시켜 표현하려는 이유는 주어진 패턴이 복잡하여 인식하려는 클래스의 갯수가 많게 되는 경우, 그만큼 더 많은 갯수의 특징들이 사용되어야 하고, 이로 인해 인식기의 설계가 복잡해지기 때문이기도 하다[10].

3.1 프리미티브의 정의

획을 보다 작은 단위의 부분획으로 표현하려는 연구는 Higgins[11], Guy Lorette[12] 등이 있다. 그러나, 이러한 방법들은 문자와 문자 또는 자소와 자소간의 연결이 하나의 획내에서 발생한 경우, 이를 분리하여 자소 또는 문자를 인식하기 위한 목적으로 고안된 것이다. 서명 검증에서 획을 부분획으로 분할하려는 목적은 문자 인식의 경우와는 달리, 획내에서 서명 검증에 중요한 부분을 찾아서 다른 부분과 차별하여 사용하고자 하는 것이다.

서명검증을 위하여 서명의 획을 부분획으로 분할하는데 있어 만족되어야 할 사항을 다음과 같이 설정한다

- 1) 분할되는 점이 서명의 위치이동(translation), 크기(size), 회전(rotation)의 변화에 불변이어야 한다.
- 2) 동일한 서명의 샘플들간에 국부적인 변화가 있더라도, 변화가 없는 동일부분에 대해서는 분할점의 위치가 서로 일치하는 분할이 일어나야 한다.

- 3) 분할되어 얻어지는 조각들의 모양은 일정 갯수의 원시 패턴으로 정의가 가능하여야 한다.

이러한 분할점을 얻기 위해 여러 가지 방법들이 제시되어 활용이 가능하겠으나[참고문헌 추가바람], 여기에서는 최소 속력점(minimal speed point)을 분할점으로 하여 얻어지는 부분획을 원시 패턴으로 사용하였다. 국부 극소 속력점을 분할점의 기준으로 얻어지는 원시 패턴의 유형은 기본적으로 (그림 4)와 같이 5가지로 정의한다.

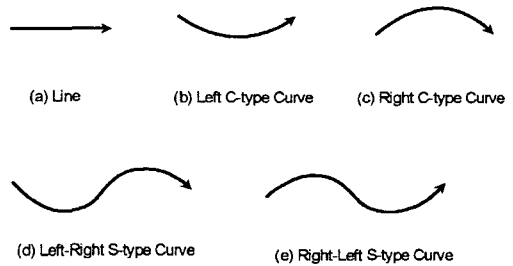


그림 4. 5가지의 기본 원시 패턴
Fig. 4 Five primitives of segmented stroke

3.2 서브패턴의 형성 및 구분

필기 곡선의 모양은 원시 패턴의 연결 모양에 따라 결정된다. 이전의 원시 패턴의 유형과 다음의 원시 패턴의 유형, 그리고 연결되는 회전 각도가 국부 모양을 결정하는데 중요한 요소가 된다.

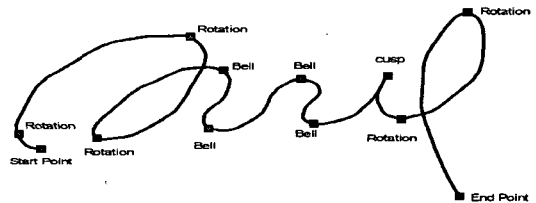


그림 5. 서명에서의 연결 유형의 예
Fig. 5 An example of connection types

서명에서의 원시 패턴들간의 연결 모양을 보면 (그림 5)에서 보인 것처럼 단순회전형(rotate), 첨형(cusp), 종형(bell) 중의 하나로 나타나는 것을 알 수 있다. 이 서명은 'rotate - rotate - rotate - bell - bell - bell - bell - cusp - rotate - rotate'의 연결 순서로 서명이 이루어진다.

3.3 서브 패턴의 정합 방법

패턴 인식에서 연속된 일련의 스트링을 비교하기 위한 방법으로 동적 계획(dynamic programming) 정합 방법이 주로 사용된다. 동적 계획 정합은 두 스트링간에 추가 또는 삭제 등의 변형을 비선형적인 매핑에 의해 흡수하여 두 스트링간의 최소 차이값과 이때의 매핑 관계를 찾아준다.[13]

비교해야 될 패턴 A, B가 다음과 같은 스트링으로 나타내어진다고 하자.

$$A : a_1, a_2, a_3, \dots, a_I \dots\dots\dots (3-1)$$

$$B : b_1, b_2, b_3, \dots, b_J \dots\dots\dots (3-2)$$

이 때, I, J 는 각 패턴의 구성요소의 갯수이다.

두 패턴을 비교하기 위해서는 각 구성요소들간에 적절한 대응 관계(mapping)가 주어져야 하는데, 이를 워핑 함수 F 라 하고 다음과 같이 표시한다.

$$F : \alpha(1), \alpha(2), \alpha(3), \dots, \alpha(k), \dots, \alpha(K) \dots\dots\dots (3-3)$$

이 때, $\alpha(k) = (i, j)$ 와 j 는 패턴의 구성요소

워핑 함수 F 의 임의의 구성 요소의 차이는 다음과 같이 표시한다.

$$\alpha(k) = \alpha(i, j) = \|a_i - b_j\| \dots\dots\dots (3-4)$$

워핑 함수에 의해 패턴 A, B를 매칭시켜 구해지는 전체적인 차이는 가중치 거리 공식에 의해 다음과 같이 표현된다.

$$D(A, B) = \frac{Min}{F} \left[\frac{\sum_{k=1}^K \alpha(k)}{K} \right] \dots\dots\dots (3-5)$$

전체 차이 계산식은 중간 과정까지의 차이 계산식 $g(\alpha(k))$ 에 의해 귀납적인 표현이 가능하다.

$$k = 1, g(\alpha(k)) = g(1, 1) = d(\alpha(1))$$

$$k \geq 2, g(\alpha(k)) = g(i, j)$$

$$= \underset{\alpha(k-1)}{Min} \left[g(\alpha(k-1)) + d(\alpha(k)) \right] \dots\dots\dots (3-6)$$

이 때, $\alpha(k-1)$ 은 $\{(i, j-1), (j-1, i), (i, j)\}$ 중의 하나이다.

전체 차이값은 다음과 같이 계산된다.

$$D(A, B) = \frac{1}{K} g(\alpha(K)) \dots\dots\dots (3-7)$$

동적 계획 정합을 서브패턴 단위로 나타내어진 필기 곡선에 대해 적용하고자 하는 경우, 추가 및 삭제의 변형 처리와 함께 병합(merge) 연산이 필요하다. 원시 패턴의 곡선이 굴곡이 큰 경우에는 중간 부분이 두개로 나뉘어지는 경우가 종종 발생한다(그림 6 참조). 그러므로 두개의 원시 패턴이 하나의 원시 패턴으로 합쳐져서 비교가 이루어지는 경우 보다 작은 차이값을 얻을 수 있다.

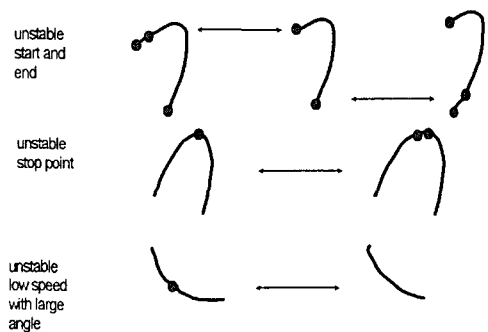


그림 6. 원시 패턴의 분리 변형
Fig. 6 Deformed separation of primitives

추가와 삭제의 연산과 함께 병합 연산이 가능한 조건에서의 중간 워핑 경로는 (그림 7)에서 보인 것과 같이 바뀐다.

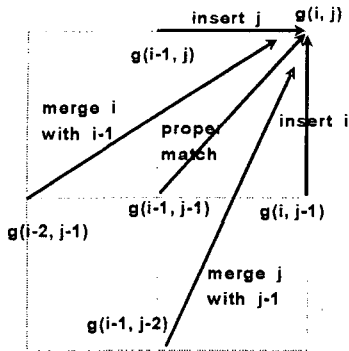


그림 7. 병합연산이 추가된 중간 워핑 경로
Fig. 7 Warping path inserted with merging

그리고 식 (3-7)에서의 중간 정합 과정까지의 누적차이 값의 계산식은 다음과 같이 정의된다.

$$g(c(k)) = \text{Min} \left[\begin{array}{l} g(i-1, j-1) + d(i, j), \\ g(i, j-1) + d(mul, j), \\ g(i-1, j) + d(i, mul), \\ g(i-1, j-2) + d(i, m(j, j-1)), \\ g(i-2, j-1) + d(m(i, i-1), j) \end{array} \right]$$

..... (3-8)

여기서, $m(i, i-1)$ 또는 $m(j, j-1)$ 은 두 서브패턴을 하나의 서브패턴으로 병합하는 연산이다.

IV. 참조 모델 데이터 생성과 국부적 가중치

서명을 등록하는 과정은 서명자의 식별 번호와 M개의 진서명 샘플(대개, M은 20개 내외)을 입력받아 진서명에 대한 참조 데이터를 만드는 과정이다.

4.1 참조 모델 형성

이 논문에서는 서명을 서브패턴 단위의 세그먼트로 구성하므로 서명에 대한 특징은 각 세그먼트에 대한 특징 벡터들로 구성된다. 샘플들로부터 추출된 특징 벡터들로 표현된 세그먼트들의 나열(sequence)을 참조 모델이라 부른다. 참조 모델에 포함되는 각 세그먼트들은 샘플들로부터 모양이 가장 안정되게 나타난 부분들만을 추출하여 이들로부터 특징 벡터를 얻도록 한다. 각 세그먼트들에 대한 중요도를 반영하기 위해 국부적 가중치를 산출한다. 이 과정에서 국부적인 변화도와 복잡도를 사용한다. 그러므로 서명에 대한 참조 모델은 각 세그먼트의 특징 벡터와 국부적인 가중치, 그리고 판단임계치가 된다.

M개의 진서명 샘플들을 입력받아 참조 모델을 형성하는 과정은 (그림 8)과 같이 4단계의 과정을 거쳐서, 서명 샘플들에 대해 가장 안정된 세그먼트들의 특징 벡터를 구한다.

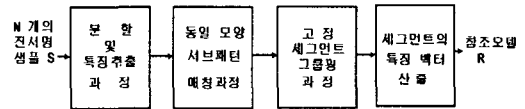


그림 8. 참조모델 형성과정
Fig. 8 Generation process of reference model

4.1.1 분할 및 특징 추출 과정

서명 샘플들의 집합을 S라 하자

$$S = \{S_1, S_2, S_3, \dots, S_m, \dots, S_M\} \dots\dots\dots (4-1)$$

여기서, M은 샘플의 갯수이다. 서명 샘플 S_m 이 분할 과정을 거쳐 서브패턴 단위로 구성된 경우, 다음과 같이 표현한다.

$$S_m = [s_1^m, s_2^m, s_3^m, \dots, s_{N_m}^m, \dots, s_{N_m}^m] \dots\dots\dots (4-2)$$

여기서, N_m 은 서명 샘플 S_m 의 세그먼트의 개수이다. s_n^m 은 m번째 샘플의 n번째 세그먼트를 나타내는 특징 벡터이다.

$$s_n^m = [\text{특징1, 특징2, \dots, 특징k, \dots, 특징K}]$$

$$= [s_{n,1}^m, s_{n,2}^m, \dots, s_{n,k}^m, \dots, s_{n,K}^m]$$

..... (4-3)

여기서, K는 특징의 개수이다.

각각의 세그먼트를 묘사하는 특징들은 매개변수적인 방법(6)에서 유사한 종류의 속성들로 구성시켰으며, 여기서는 <표 1>과 같이 간단한 특징집합을 선택하였다. 특징1~5는 각 세그먼트의 모양을 나타내는 특징이고, 특징5~9는 각 세그먼트의 동적인 속성을 반영하는 특징이며, 특징10~14는 각 세그먼트에서의 압력의 변화를 반영하는 특징들이다.

표 1. 세그먼트의 특징들
Table 1 Feature set of a segment

특징들	설 명
특징 1	세그먼트의 폭
특징 2	길이
특징 3	첫번째 피크의 폭
특징 4	두번째 피크의 폭
특징 5	두 원시 패턴의 연결 각도
특징 6	세그먼트의 끝기 시간
특징 7	최대 속력
특징 8	평균 속력
특징 9	최대속력 - 평균 속력
특징 10	최대 압력
특징 11	최소 압력
특징 12	평균 압력
특징 13	최대 압력 - 최소 압력
특징 14	최대 압력 -평균 압력

4.1.2 동일 모양의 서브패턴 매칭 과정

M개의 서명 샘플들에 대해 임의의 기준 샘플을 지정하여 다른 샘플들과 서브 패턴 단위의 동적 계획 정합을 행하고, 이 과정을 기준 샘플을 다른 샘플로 바꾸어 반복 수행하여, 동적 계획 정합의 차이값의 평균이 최소가 되는 기준 샘플을 찾는다. 즉, 기준 샘플은 식 (4-4)를 만족하는 S_i 로 지정한다.

$$\min_{S_i \in S} \left[\sum_{S_j \in S} D(S_i, S_j) \right] \dots\dots\dots (4-4)$$

여기서, $D(S_i, S_j)$ 는 서브패턴 단위의 동적 계획 정합의 차이 계산값이다.

4.1.3 고정 세그먼트 그룹핑 과정

이 과정은 선택된 기준 샘플의 각 세그먼트들에 대해, 매칭 과정의 결과로 얻어진 워핑 함수로부터 매핑이 이루어진 다른 샘플들에서의 세그먼트를 그룹핑하는 과정이다. 이때, 그룹핑되는 세그먼트는 워핑함수가 proper matching이 이루어진 경우의 세그먼트만을 선택한다. 즉, 기준 서명의 세그먼트에 매핑된 다른 샘플의 세그먼트가 null point 이거나 병합된 세그먼트가 아닌 단일 세그먼트의 경우만을 선택한다.

$$G_n = \{g_1^n, g_2^n, g_3^n, \dots, g_{I_n}^n\}, \quad i \leq n \leq N, \dots\dots\dots (4-5)$$

여기서, I_n 은 n번째 세그먼트와 proper matching된 샘플 세그먼트의 개수이다.

(그림 9)에서 동일 모양의 샘플 세그먼트를 그룹핑한 결과의 한 예를 보이고 있다.

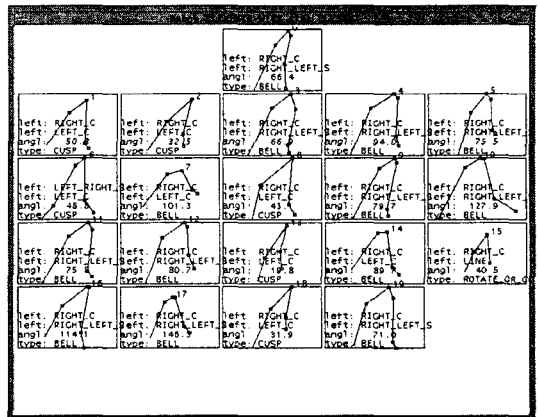


그림 9. 동일 모양의 서브패턴을 그룹핑한 결과
Fig. 9 An example of grouping the same type of subpatterns

4.1.4 세그먼트의 특징 벡터의 산출

참조 모델 R 의 각 세그먼트들의 특징 벡터는 그룹핑되는 고정 세그먼트들로부터 계산된다. 즉, 동일한 모양을 가지는 샘플 세그먼트들의 특징 벡터들로부터 특징 공간상의 중심값과 표준편차를 계산하여 참조 모델 R 의 n 번째 세그먼트의 특징 벡터 r_n 을 할당한다.

$$\begin{aligned} r_n &= \langle m_n, \sigma_n \rangle, \quad 1 \leq n \leq N_r \\ m_n &= [m_1^n, m_2^n, \dots, m_k^n, \dots, m_K^n] \dots (4-6) \\ \sigma_n &= [\sigma_1^n, \sigma_2^n, \dots, \sigma_k^n, \dots, \sigma_K^n] \end{aligned}$$

여기서,

$$m_k^n = \frac{1}{I_n} \sum_{i=1}^{I_n} g_{i,k}^n \dots (4-7)$$

$$\sigma_k^n = \sqrt{\frac{1}{I_n} \sum_{i=1}^{I_n} (g_{i,k}^n - m_k^n)^2} \dots (4-8)$$

이며, $g_{i,k}^n$ 는 n 번째 세그먼트에 그룹핑된 고정 세그먼트 집합 G_n 에서의 i 번째 샘플 세그먼트 g_i^n 의 k 번째 특징값을 의미한다.

4.2 국부적 가중치 산출

세그먼트에 대한 가중치는 국부적인 변화도와 복잡도에 의해 산출된다. 국부적인 변화도는 다음과 같이 정의된다.

$$\begin{aligned} V &= [v_1, v_2, \dots, v_n, \dots, v_N] \\ v_n &= \frac{1}{M} \sum_{m=1}^M \alpha r_{n, s_m^m} \dots (4-9) \end{aligned}$$

여기서, $\alpha r_{n, s_m^m} = \sqrt{\sum_{k=1}^K \left(\frac{s_{n,k}^m - m_k^m}{\sigma_k^m} \right)^2} \dots (4-10)$

$\alpha r_{n, s_m^m}$ 은 R 의 n 번째 세그먼트의 특징벡터와 m 번째 샘플의 n 번째 세그먼트의 특징벡터와의 가중치 거리(weighted

Euclidean distance)이다. 그러므로, n 번째 세그먼트의 변화도 v_n 는 참조 모델의 세그먼트와 샘플들의 세그먼트간의 평균 차이값을 의미한다.

서명을 구성하는 세그먼트들에 대해 진서명내의 변화가 심한 부분에서는 비교 과정에서 덜 중요하게 참여시키고 변화가 거의 없는 안정된 부분은 강조하여 비교되도록 한다. 즉, 국부적인 가중치는 변화도에 대해서는 역(negative)의 관계이므로, 국부적 변화도에 의한 가중치의 효과는 다음과 같은 식에 의해 산출되며, 값의 범위는 0~1 사이의 값을 가지도록 변환한다.

$$\begin{aligned} W^v &= [w_1^v, w_2^v, \dots, w_n^v, \dots, w_{N_r}^v] \\ w_n^v &= \frac{1}{1 + \exp\left(\frac{m_v - v_n}{\sigma_v}\right)} \dots (4-11) \end{aligned}$$

여기서 m_v 은 v_n 의 평균, σ_v 은 v_n 의 표준편차이다.

국부적인 복잡도는 Plamondon의 정의[8]에서 서브패턴 유형의 반복에 의한 감쇄효과를 추가하여 다음과 같이 설정한다.

$$\begin{aligned} C &= [c_1, c_2, \dots, c_n, \dots, c_M] \\ c_n &= \left\{ 1 + 0.9 \frac{\sum_{j=1}^{L_{\max}(n)} \theta(n, j)}{L_{\min}(n) \frac{\pi}{2}} \right\} \\ &= \frac{|L(n+1) - L(n)|}{L(n+1) + L(n)} \frac{L(n)}{T(n)} * K^{r(n)} \dots (4-12) \end{aligned}$$

단, $r(n)$ 은 세그먼트 n 의 같은 모양 반복횟수이며, K 는 반복시의 difficulty의 감쇄효과 상수(0.89)이다.

서명을 구성하는 세그먼트가 단순하여 다른 사람이 모조하기 쉬운 부분에서는 중요도를 약화시키고, 반대로 서명이 복잡하여 상대방이 모조하기 어려운 부분에서는 강조하여 비교될 수 있도록 한다. 즉, 국부적인 가중치는 복잡도에 대해서는 순(positive) 관계이므로, 국부적 복잡도에 의한 가중치의 효과는 다음 식에 의해 산출되며, 값의 범위는 0~1 사이의 값을 가지도록 변환한다.

$$W^c = [w_1^c, w_2^c, \dots, w_n^c, \dots, w_{N_r}^c]$$

$$w_n^c = \frac{1}{1 + \exp\left(\frac{c_n - m_c}{\sigma_c}\right)} \dots\dots\dots (4-13)$$

여기서 m_c 은 c_n 의 평균, σ_c 은 c_n 의 표준편차이다.

변화도와 복잡도를 동시에 고려한 국부적 가중치는 다음과 같이 두 요소에 의해 결정되는 국부적 가중치의 기하 평균으로 설정한다.

$$W = [w_1, w_2, \dots, w_n, \dots, w_{N_r}]$$

$$w_n = \sqrt{w_n^v * w_n^c} \dots\dots\dots (4-14)$$

V. 판단임계치 설정과 판단 과정

5.1 판단 임계치의 설정

검증 문제에서 서명의 진·위 판정을 위해서는 적합한 참조 모델의 설정은 물론 이에 적절한 판단 임계치를 설정해야 한다. 서명 상호간의 상이도는 서명별로 상이도의 축에서 나타나는 위치와 이 값의 출현 범위가 달라진다.

진서명이 비교적 안정적인 때, 서명간의 상이도는 대체적으로 상당히 작고 이 값들의 범위도 좁다. 하지만, 불안정한 서명인 경우는 서명간의 상이도는 대체적으로 상당히 크고 이 값들이 나타나는 범위도 크다. 이 때, 샘플들 간의 상이도는 대체적으로 크거나 작다는 것은 상이도의 평균이 크거나 작다는 것으로 해석할 수 있다. 또한, 상이도의 범위는 편차로 표현될 수 있다. 일반적으로 서명의 판단임계치는 이러한 변화를 수용할 수 있을 만큼 크게 설정하여야 함과 동시에 모조 서명이 진 패턴으로 오인되지 않도록 너무 크지 않게 설정하여야 한다. 이러한 두가지 조건이 모두 만족되려면, 진 서명간의 차이와 진 서명과 모조 서명들간의 차이에 대한 사전 지식이 있어야 한다. 그러나, 실제적으로 모든 진 서명의 샘플을 충분히 많이 얻을 수 없고, 더욱

이 모조 서명이 어떠한 형태로 어느 정도의 정교함으로 발생할 지 모조 서명의 샘플을 통하여 알 수가 없다.

그러므로, 판단 임계치를 얻는 과정은 진 서명의 일부 즉, 주어진 학습 샘플으로부터 서명의 변화정도를 나타내는 요소를 이용하여 판단 임계치를 설정하는 방법이 지금까지 사용되어 왔다[6][7]. 변화도만에 의한 임계치 설정은 다음과 같다.

$$\theta = c \cdot \mu \text{ (여기서 } c \text{는 상수)} \dots\dots\dots (5-1)$$

위 식에서 c 값은 모든 서명들에 고정된 상수로 사용한 다. 샘플 집합으로부터 진 서명의 변화도는 상이도의 평균으로 정의된다.

$$\mu = \frac{1}{M} \sum_{m=1}^M \text{Diss}(R, S_m) \dots\dots\dots (5-2)$$

이때, $\text{Diss}(R, S_m)$ 는 참조 모델 R 과 서명 샘플 S_m 과의 전체 상이도로서 다음식에 의해 결정된다.

$$\text{Diss}(R, S_m) = \frac{\sum_{n=1}^{N_r} \alpha(r_n, s_n^m) \cdot w_n}{\sum_{n=1}^{N_r} w_n} \dots\dots\dots (5-3)$$

이 방법에 의해 판단 임계치를 설정하는 경우, 모든 사람에게 동일한 임계치를 사용하는 경우보다는 검증 오류율을 상당히 줄일 수 있다. 하지만, 이 방법은 유형-I 오류의 측면에서 검증 오류율을 낮춘 것이지, 모조 서명을 진 서명으로 오인하는 유형-II 오류는 고려하지 않은 것이다.

유형 II 오류를 고려한 임계치의 설정이 되기 위해서는 진 서명의 자체의 변화도뿐만 아니라 진 서명에 대한 모조 서명의 상이도의 분포가 어떻게 나타날지를 예측해야 한다. 그러나 주어진 서명에 대한 가능한 모든 모조 서명의 샘플을 얻는 것은 실제 응용의 경우 어려운 일이다. 다만, 모조 서명의 분포가 어떻게 나타날 지에 대한 예측으로 진서명의 복잡도를 예측하여 사용할 수 있다. 즉, 서명이 복잡하여 모

조하기 어려운 경우에는 모조 서명의 분포가 진 서명의 분포에서 보다 큰 값의 범위에서 나타날 것이며, 반대로 서명이 단순하여 모조가 쉬운 경우, 모조 서명의 분포가 진 서명의 분포에 가깝게 나타날 것이라고 예측하는 것이다.

서명의 복잡도는 앞서 정의된 국부적 변화도의 합으로 다음과 같이 계산된다.

$$C_{total} = \sum_{n=1}^{N_r} C_n \dots\dots\dots (5-4)$$

기존의 방법을 사용하여 모든 사람에게 고정된 상수값으로 사용하였던 상수 c 값을 변화시킴에 따라 최소 검증 오류율을 나타내는 c의 크기와 복잡도 C_{total} 의 상관 관계는 선형적인 관계로 모델링하고, 다음과 같은 식으로 표현된다. 상수 p와 q는 서명샘플에 대해 c와 C_{total} 의 분포도를 이용하여 최소 자승법으로 계산된다.

$$c = p \cdot C_{total} + q \dots\dots\dots (5-5)$$

이때, c와 C_{total} 의 관계를 이용하여 판단 임계치를 설정하는 산출식은 다음과 같다.

$$\theta = c' \cdot (p \cdot C_{total} + q) \cdot \mu$$

(여기서, c' 은 상수) $\dots\dots\dots (5-6)$

5.2 서명 검증 모듈

입력된 서명 T는 먼저 분할 과정을 거쳐 서브패턴 단위의 세그먼트로 표시된다.

$$T = [t_1, t_2, t_3, \dots, t_n, \dots, t_{N_r}] \dots\dots\dots (5-7)$$

여기서, N_r 은 입력 서명 샘플 T의 세그먼트의 개수, t_n 은 n번째 세그먼트를 나타내는 특징 벡터이다.

$$t_n = [t_{n1}, t_{n2}, t_{n3}, \dots, t_{nk}, \dots, t_{nk}] \dots\dots\dots (5-8)$$

여기서, K는 특징의 개수이다.

입력된 서명 T와 참조 모델 R과의 동일한 형태를 갖는 세그먼트들끼리의 매핑관계를 얻기 위해 세그먼트 단위의 동적 계획 정합을 행한다. 즉, 모양의 특징에 근거하여 동일 모양의 세그먼트들끼리의 대응관계를 얻어낸다.

비교과정에서는 다음 식에 의해 동일한 모양을 갖는 세그먼트들끼리 국부적 가중치를 적용하여 참조 모델과 입력 서명간의 가중치 거리를 구한다.

$$Diss(R, T) = \frac{\sum_{n=1}^{N_r} d(r_n, t_n) \cdot w_n}{\sum_{n=1}^{N_r} w_n} \dots\dots\dots (5-9)$$

판단 과정에서는 비교 차이값이 저장된 임계치보다 작으면 진서명으로, 그렇지 않으면 모조서명으로 판정한다.

if $Diss(R, T) \leq \theta$ then T is True $\dots\dots\dots (5-10)$
 else T is Forgery

VI. 실험결과

6.1 실험 데이터

실험을 위한 샘플 서명은 9사람의 서명을 대상으로 하였다. 이 때, 한사람당 120개의 진서명과 다른 사람들이 이를 위조한 100개의 모조서명으로 검증용 실험을 행하였다. 실험 방법은 한사람의 서명 내에서 진서명 집합으로부터 20개의 샘플을 추출하여 참조 모델을 구하는데 사용하였고, 이를 제외한 나머지 중에서 진서명 오인율을 구하기 위하여 100개의 진서명을 사용하였다. 또한, 그 사람의 모조서명 집합에서 모조서명 오인율을 구하기 위하여 100개의 모조서명을 한다.

6.2 실험 결과

다른 방법들과의 비교를 위하여, 함수적 접근방법(7)과 매개변수적인 방법(6)을 구현하여 동일한 실험 데이터에 의해 <표 2>와 같은 비교실험 결과를 얻었다.

결과에서 보면, 검증을 측면에서 가장 우수한 서명 검증 방법은 동적 계획 정합을 이용한 함수적 접근 방법이며, 특징 함수로 속도 함수를 사용하고, 군집화에 의한 다수의 참조서명을 가지도록 하는 경우에 평균 오류율 3.37%의 우수한 성능을 보인다.

구조적인 접근의 기본 방법 만에 의한 검증 오류율은 가장 좋지 않은 결과를 나타내고 있는데, 이것은 아직까지 구조적 접근에서 사용된 세그먼트 단위의 특징집합이 함수적

접근에서의 필기 신호의 샘플링 시간 단위의 속도 함수만큼 세밀하게 서명을 표현하고 있지 못함에 기인한다. 구조적 접근의 검증률 향상을 위해서는 세그먼트 단위의 특징집합에 대한 연구가 계속적으로 이루어질 필요가 있음을 시사한다. 하지만, 구조적인 접근에서는 서명을 세그먼트 단위로 기술하여 세그먼트로부터 국부적인 복잡도나 변화도 등의 정보를 부가적으로 끌어내는 것이 가능하다.

표 2. 서명 검증 방법들의 결과 비교
Table 2 Comparison result with other signature verification techniques

검증방법		비교항목	평균 오류율 (%)	처리시간		저장 정보량 (byte)	검증률의 안정성 (유형 II 오류율)			
				등록과정 (sec)	검증과정 (msec)		유형 I 오류 5%	유형 I 오류 10%	유형 I 오류 15%	유형 I 오류 20%
함수적 접근	$\theta = c \cdot \mu$ 임계치설정과 다수의 참조서명 허용		3.37	32.5	150	6K	3.2	1.6	0.6	0.2
매개 변수적 접근	23개의 특징에 개인별 가중치 적용		4.28	0.0216	0.11	0.6K	3.0	2.5	2.0	1.0
구조적 접근	기본 방법		14.2	1.49	7.40	1.5K	35.0	21.5	13.2	7.4
	국부적 가중치를 적용한 경우		9.67	1.50	7.48	1.6K	23.0	10.2	5.4	3.2
	복잡도에 의한 임계치를 설정한 경우		4.05	1.50	7.48	1.6K	3.0	1.2	0.5	0.0

〈표 2〉의 실험결과에서 보듯이, 국부적 가중치와 복잡도에 의한 판단 임계치를 사용함으로써 검증률이 향상됨을 알 수 있다. 이 방법의 실험 결과에서 최소의 검증오류율 4.05%는 함수적 접근에서의 3.37%보다는 성능이 우수하지는 못하지만, 처리 시간과 저장 정보량의 측면에서 실용적인 가능성을 보이고 있다. 특히, 구조적인 접근에서 주목할 것은 유형 II 오류율의 안정성을 보장 받을 수 있다는 것이다. 비교 도표에서 보는 것처럼 유형 I 오류율을 어느 정도 (5 ~ 20%) 감소할 때의 유형 II 오류율은 여타 다른 방법에 비해 매우 우수한 결과를 보이고 있다

VI. 결론

이 논문에서는 서명 검증 문제에 대해 기존의 접근방식은 차별되는 새로운 구조적 접근방식을 제시하였고, 필기의 분할과 표현, 이에 따른 동적 계획 정합 알고리즘, 등 필기 패턴의 구조적인 인식 방법을 다루었고, 서명 검증의 문제에 적용하기 위해 서명의 등록과정과 판단 과정의 구체적인 서명 검증 방법을 제시하였다. 그리고, 실험을 통하여 기존의 방법인 함수적 방식과 매개변수적인 방법을 동일 조건에서 비교 실험하여, 검증률, 처리시간, 정보저장량, 서명 검증률의 안정성, 등의 다각적인 측면에서 비교를 행하였다.

제안된 방법은 검증률의 측면에서는 아직도 함수적 방법보다 약간 차이로 우수하지 못하였으나, 국부적인 가중치나 복잡도에 의한 국부적인 가중치의 적용이 가능하게 되었으며, 보다 정확한 판단 임계치의 설정이 가능함을 보였다. 그리고, 서명 검증의 실제적인 활용을 고려하는 경우에 더 유용하게 사용될 수 있다는 것을 알 수 있었다.

향후의 연구로는 지문인식이나 얼굴인식, 등의 다른 매체의 생체인증 방법과 서명검증을 혼합하여 사용자의 신뢰성과 인증의 정확도를 동시에 확보하는 다중 생체인증을 연구하는 것이 필요할 것으로 본다.[14][15]

참고문헌

[1] R. Plamondon and S.N. Srihari, "On-Line and Off-line Handwriting Recognition: A Comprehensive Survey," IEEE Trans. on PAMI, Vol. 22, No.1, pp.63-84, 2000.

[2] R. Plamondon and G. Lorette, "Automatic signature verification and writer authentication - The state of art," Pattern Recognition, Vol. 22, No.2, pp.107-131, 1989.

[3] Sukhan Lee and Jack C. Pan, "Offline Tracing and Representation of Signatures," IEEE Trans. on S.M.C., Vol.22, No.4, pp.755-771, July/Aug. 1992.

[4] H. D. Crane and J. S. Ostrem, "Automatic Signature Verification using a three-axis-force-sensitive Pen," IEEE Trans. on Systems Man and Cybernetics, SMC-13, No. 3, pp.329-337, 1993.

[5] 김성훈, 장경식, 유재룡, 김재희, "On-line Signature Verification을 위한 특징선택에 관한 연구," 제 2 회 인공지능, 신경망 및 퍼지시스템 종합학술대회 논문집, pp.200-209, 1992년 10월.

[6] Seong Hoon Kim, Myeong Soo Park, Jaihie Kim, "Applying Personalized Weights to a Feature Set for On-line Signature Verification," Proc. of the 3rd International Conference on Document Analysis and Recognition, pp.882-885, Montreal Canada, Aug 1995.

[7] Jaihie Kim, Jae Ryong Yu and Seong Hoon Kim, "Learning of Prototypes and Decision Boundaries

for a Verification Problem having only Positive Samples," Pattern Recognition Letters, Vol. 17, pp.691-697, 1996.

[8] Jean-Jules Brault and Réjean Plamondon, "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery," IEEE Trans. on S.M.C., Vol. 23, No. 2, pp.400-413, Mar/Apr 1993.

[9] 전병환, 김성훈, 김재희, "퍼지 결정 트리를 이용한 온라인 필기 문자의 계층적 인식," 대한 전자공학회 논문지, 제31권 B편 제3호, pp.132-140, 1994년 3월.

[10] King Sun Fu, Syntactic Pattern Recognition and Application, Prentice-Hall Inc., p.1-5, 1982.

[11] C. A. Higgins and P. E. Bramall, "Pen-based Computers and Cursive Script Recognition," Proc. of InfoScience '93, Korea Information Science Society, pp.113-120, 1993.

[12] S. Bercu and Guy Lorette, "On-line Handwritten Word Recognition: An Approach based on Hidden Markov Models," Proc. of IWFHR '93, pp.385-390, 1993.

[13] 유재룡, 박명수, 김성훈, 김재희, "온라인 서명검증을 위한 전처리 및 Dynamic Programming 알고리즘의 결정요소들의 최적화," 제3회 인공지능, 신경망 및 퍼지시스템 종합학술대회/전시회 논문집, pp.459-463, 1993년 10월 18일 - 20일.

[14] 신강호, 채호병, 정연만, "순환 고조파 필터를 이용한 회전불변 지문 인식에 관한 연구," 한국컴퓨터정보학회 논문집, 8권 3호, pp.94-99, 2003년 9월.

[15] 조민환, "체인 코드를 이용한 훼손된 지문의 인식에 관한 연구," 한국컴퓨터정보학회 논문집, 8권 3호, pp.63-68, 2003년 12월.

저자소개



김성훈

1988년 2월 서강대학교 공학사
 1990년 2월 연세대학교 공학석사
 1996년 2월 연세대학교 공학석사
 1996년 3월~현재 영동대학교 컴퓨터
 공학과 부교수

<관심분야> 생체인식, 콘텐츠 기술