
필적 및 서명에 대한 Off-line 자동분석시스템

The Off-line Verification System of Signature of Handwrite

김세훈, Seihoon Kim, 하정요, Jeungyo Ha, 김계영 Gyeyoung Kim, 최형일 Hyungil Choi

*숭실대학교 미디어학과, **컴퓨터비전연구실

요약 필적 감정은 개인의 고유한 필적 개성을 이용하여 임의의 두 필기 문장 또는 텍스트가 동일인에 의해 작성되었는지를 판별하는 기술로 유서대필 및 보안수사, 서명의 검증, 범죄 수사 등에 활용되어지고 있다. 이러한 작업은 감정 전문가의 판단기준에 의해 필적의 유사성을 판별하기 때문에 객관성 결여 및 과도한 소요 시간, 과도한 처리 비용의 문제를 내포하게 된다. 이러한 문제를 해결하여 판별의 객관성과 업무의 신속한 처리를 가능하게 하기 본 논문에서는 컴퓨터를 통한 패턴 분석을 적용하여 두 필적의 유사성을 판별하는 방법을 본 논문에서는 제안한다. 이를 위하여 본 논문은 학습단계와 자동분석단계로 나뉘며, 학습단계에서는 입력된 문서영상에서 필적의 영역을 추출한 후, 특징을 추출하고 DTW연산을 통하여 학습을 한다. 자동분석단계에서는 대조할 문서영상에서의 특징을 추출하고 입력된 문서영상과 대조할 문서영상간의 마할라노비스 거리(Mahalanobis Distance)를 구하여 서명 및 필적에 대한 유사도를 도출한다. 실험은 4명의 필적을 이용하여 비교하였으며, 우수한 결과를 보였다.

핵심어: 필적감정, 서명감정, Off-line, Handwrite, Signature, Verification, DTW, Mahalanobis Distance

1. 서론

필적감정(Handwrite verification)은 개인의 고유한 필적 개성을 이용하여 임의의 두 필기 문장 또는 텍스트가 동일인에 의해 작성되었는지를 판별하는 기술이다. 현재 이러한 감정 작업은 국가에서 인정한 감정 기관에서 이루어지고 있다. 이러한 곳에서 이루어지는 감정 업무는 유서대필 및 보안수사, 서명검증, 범죄수사 등에 활용되어지고 있는 분야가 있다. 필적 감정 전문가에 의해 필적 및 서명의 유사성을 몇 가지 특징(길이와 두께, 각도, 필압 등)을 기준으로 판별하게 되는데 여기에는 몇 가지 문제점이 있다. 여러 문제점 중 대표적인 세 가지를 들어 설명하면, 첫째로는, 사람이 쓴 필적을 사람이 감정함으로써 감정시의 감정 전문가의 상황이나 부차적인 이유로 객관성이 결여될 수 있다. 둘째로, 존각을 다루는 상황의 경우 감정을 하여 결과를 도출해 내는 데까지의 수행 시간의 지연으로 범죄 증명, 또는 기타 민, 형사상의 소에서 증거 자료로 제출을 할 수 없는 경우가 발생할 수 있다. 마지막으로 기존의 필적감정 시 필적 감정 전문가를 통해 감정이 이루어 졌는데, 이러한 전문가의 수가 적고 또한 양성 하는 데에 많은 어려움이 있기 때문에 많은 비용의 지출이 발생하게 된다. 따라서 위의 문제점에 대한

해결 방안과 업무의 자동화를 위해 컴퓨터를 이용한 패턴분석을 통해 문서의 유사성을 분석 하고자 한다.

필적 및 서명의 자동분석 시스템에는 크게 On-line분석 시스템과 Off-line분석시스템으로 2가지 종류로 나눌 수 있다. On-line시스템은 디지털이거나 타블렛 등 기타 입력도구를 이용하여 필적 및 서명을 입력하여 분석하는 시스템으로 필적 및 서명에 대해 입력되는 순서에 대하여 각 문자 대한 중심 간의 거리를 이용하여 유사도를 도출 내는 시스템[1]등 입력되는 순서와 관련이 있는 특징을 주로 이용하여 패턴인식 작업을 수행하여 유사도를 도출하고 있다. 이와는 다르게 Off-line분석시스템은 필적이나 서명을 컴퓨터의 입력장치인 스캐너를 통하여 그림 및 사진 파일로 저장하여 분석하는 시스템으로써 필적이나 서명에 대하여 입력 순서를 부여하기 어려우며, 그 필적이나 서명에 대하여 다른 특징을 찾아내어 분석을 하는 시스템의 형태를 갖추고 있다. 그 예로써 서명감정에 관하여 구조적인 특징의 유사도를 이용한 방법[2]과, HMM 분류기를 이용하여 감정하는 방법[3], HMM분류기와 SVM분류기를 각각 이용하여 감정 하는 방법[4], DTW를 이용하여 Dynamic Programming을 통한 서명의 유사도를 판단하는 방법[5], DTW를 이용하여 그 유

사도를 Mahalanobis Distance를 통하여 비유사도를 측정하는 방법[6] 등이 있다.

앞서 설명된 방법들의 경우 필적의 감정보다는 서명의 감정을 위주로 연구되고 있어, 서명과 같이 길이가 짧은 패턴에 효과적으로 수행을 하고 있으며, 더욱이 한글 필적인식에 대한 자료는 많았으나, 한글 필적감정의 경우에는 자료를 거의 찾아볼 수 없을 정도 연구된 자료가 부족했다. 그렇기 때문에 제안하는 시스템을 통하여 한글, 영어, 등의 필적에 범용성을 가지게 될 것으로 보이며, 서명과 같이 짧은 패턴과 동시에 필적과 같은 긴 패턴에 대하여 처리를 각각으로 수행할 수 있으므로 필적과 서명을 한 가지 방법으로 동시에 수행할 수 있으므로 그 기대효과가 매우 클 것으로 고려된다.

제안하는 시스템의 주요 프로세스는 문자의 영역 추출, 특징추출, 특징의 학습, 학습된 특징의 분석, 대조본의 특징추출, 두 특징간의 유사도 검출, 분석결과 도출 등 이다. 따라서 본 논문에서는 필적이나 서명의 영역 추출하는 방법, 추출된 특징의 학습방법, 학습된 특징의 분석 방법, 두 특징간의 비교 방법에 관하여 설명한다.

본 논문의 구성은 제 2절에서는 필적 자동분석시스템의 전체적인 개요를 기술하고, 제 3절에서는 학습 단계에서 필적 영상의 획득과 필적의 영역을 추출하고, 특징을 획득하는 과정에 대한 방법 및 특징을 분석하는 단계로 기술하고 있으며, 제 4절에서는 필적 자동분석 단계인 대조본에 대한 특징 추출과 분석, 그리고 두 가지 필적에 대하여 비교하는 과정에 대하여 기술되고 있다. 제 5절에서는 실험 및 결과를 보이고, 마지막으로 제 6절에서는 결론 및 향후 연구에 관하여 기술한다.

2. 자동분석시스템의 개요

본 논문에서 제안하는 시스템의 개요는 그림1과 같이 학습 단계와 필적 자동 분석 단계로 나뉘게 된다. 학습 단계에서는 먼저 필적 영상을 입력하면 그 문서에서 필적 영역을 추출하게 되는데, 이때 자소 단위와 단어 단위의 영역을 추출을 행하게 되고, 이때의 결과를 이용하여 각 영역의 크기와 거리를 바탕으로 통합하는 과정을 수행하게 된다. 그 추출된 영역 내에서 필적의 영역의 특징 점을 추출하게 되는데 이때 프로젝션을 통해 그 영역 내에 존재하는 필적의 프로파일 결과를 합성한다. 그 후 필적 영역의 특징을 학습하게 되는데 특징간의 변위를 측정하고, 학습된 특징의 분석을 수행함으로써 학습 단계가 완료되고, 그 후 대조본에 대하여 학습 단계의 과정과 같은 과정을 거쳐 특징을 추출하고 그 추출된 특징을 이용하여 학습된 자료의 특징간의 거리 정보를 추출하여 유사도를 측정하게 되면 분석결과가 나오

게 된다.

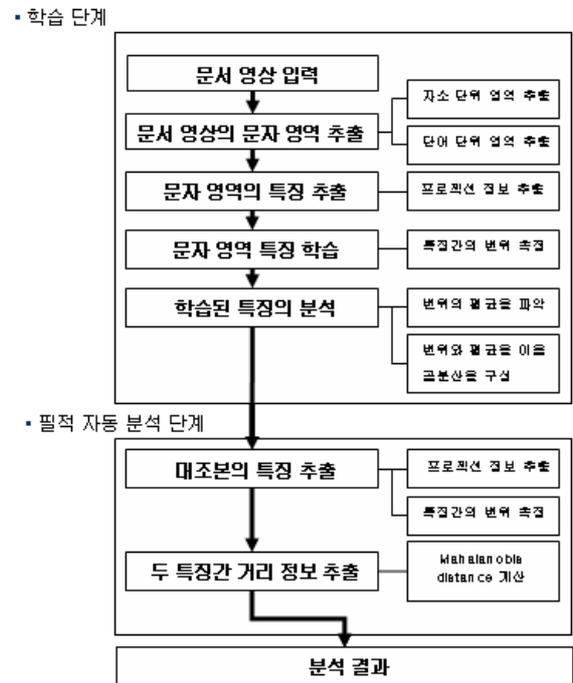


그림 1. (시스템 개요도)

3. 학습 단계

3.1 문서 영상 입력

본 시스템은 Off-line 자동분석시스템으로서 필적의 입력을 컴퓨터의 입력장치인 스캐너를 이용하여 B/W스캔 방법으로 입력을 받게 된다.

3.2 문서 영상의 문자 영역 추출

필적 영역과 배경 영역을 구분하기위해 레이블링 알고리즘을 이용한 결과를 바탕으로 자소 단위의 영역인식을 하게 된다. 이렇게 레이블링 된 결과를 바탕으로 최소포함사각형을 구성하고, 먼저 각각의 최소 포함 사각형의 위치를 이용하여 겹치는 영역을 제거하게 된다. 겹치는 영역을 제거하게 된 후 최소 포함 사각형간의 무게 중심 간의 거리를 이용하여 일정 범위 안에 있는 경우를 조사하여 문자 단위의 영역을 추출하게 된다. 위에서 추출한 결과를 바탕으로 필적이나 서명에 관한 최종 특징 추출 영역을 결정하게 된다.

3.3 문자 영역의 특징 추출

문서 영상의 특징을 추출하기위해 각각의 영역별로 가로 줄에 대한 프로젝션을 이용하여 각각의 문자의 영역에 대하여 특징을 추출한다. 이렇게 추출된 특징을 이용하여, 필기 문자열에 대한 학습과 비교분석과정을 수행한다.

3.4 문자 영역의 특징 학습

위에서 추출된 특징을 바탕으로 다른 문자열과 비교를 위해서, 한사람의 문자열에 대한 학습의 과정을 거치게 된다. 학습 과정을 바탕으로 개인에 대한 필적의 유형을 파악하고, 비교, 분석이 가능하게 되어 진다.

이 과정에서는 DTW(Dynamic Time Warping)알고리즘 [8]이 사용 되었다. DTW는 기준이 되는 특징의 패턴과 입력된 특징의 패턴간의 유사도를 동적 프로그래밍(Dynamic programming)을 이용해 구하는 방법이다.

이전 단계까지의 여러 장의 학습 결과를 DTW에 넣게 되는데 DTW는 서로 다른 길이의 특징 벡터의 비교 연산이 용의하기 때문에 학습단계에서 서로 다른 길이의 문자열의 특징에 대한 비교 연산이 가능하게 되는 결과가 그림2와 같은 결과를 가진다.

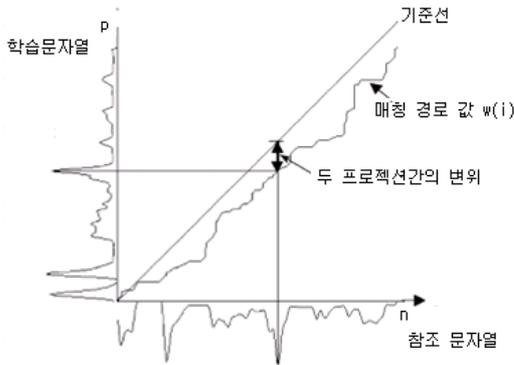


그림 2. (참조문자열과 학습문자열의 DTW의 예)

두 문자 영역의 투영된 데이터 특징을 각각 $Q=q_1, q_2, \dots, q_m$, $C=c_1, c_2, \dots, c_n$ 라고 했을 때, 이들 사이의 DTW 매칭 경로의 값을 식(1)과 같이 정의 할 수 있다.

$$d(w_k) = d(q_k, c_k) = |[Q(q_k) - \mu_Q] - [C(c_k) - \mu_C]| \quad (1)$$

이렇게 얻어진 매칭경로를 최소화 하는 경로 값을 식(2)와 같이 표현한다.

$$DTW(Q, C) = \min \frac{1}{K} \sqrt{\sum_{k=1}^K W_k} \quad (2)$$

$$W_k = \{w_1, w_2, \dots, w_k\}$$

이때의 누적거리는 식(3)으로 구하게 된다.

$$\gamma(i, j) = d(q_i, c_j) + \min \gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1) \quad (3)$$

DTW는 두 패턴간의 누적거리를 최적화하는 (m,n)평면의 최적경로 $m=w(n)$ 을 찾는 방법이다. 이 방법을 이용하여 학습 문자열중의 대표 패턴을 하나 정하여 참조 문자열로 지정을 한 후 이것과 학습 문자열간의 패턴을 학습하고, 참조 문자열과 학습 문자열간의 최적경로(w_n)를 계산하여, 특징으로서 사용하게 된다.

3.5 학습된 특징의 분석

3.4절에서 DTW를 통하여 추출된 경로를 이용하여 학습된 특징 분석 단계의 파라미터로 활용하게 되는데, 식(4)와 같이 표현하게 된다.

$$V_k = [w_k(1) - G_1, w_k(2) - G_2, \dots, w_k(L_1) - G_n]^T \quad (4)$$

여기서 각각 G, k, t 는 식(5)와 같다.

$$G_k = \frac{p}{n} \quad (5)$$

$$k = 1, \dots, N-1$$

$$t = 1, \dots, n$$

마지막으로 학습된 특징의 분석을 위하여 상관관계에 대한 정보를 분석에 사용하기 위해 공분산 계산을 하게 되는데, 이때 학습 파라미터들의 평균은 식(6)으로 표현가능하며,

$$\mu = \frac{1}{N-1} \sum_{k=1}^{N-1} V_k \quad (6)$$

파라미터와 평균의 차를 이용한 공분산은 식(7)과 같다.

$$\Sigma = \frac{1}{N-1} \sum_{k=1}^{N-1} (V_k - \mu)(V_k - \mu)^T \quad (7)$$

과 같이 계산이 되어 진다. 위와 같은 계산을 입력되는 학습 문자열에 대하여 3.4절과 3.5절을 반복하여 수행하여 학습 문자열의 특징을 학습하게 된다.

4. 필적 자동 분석 단계

4.1 분석 문자열의 특징 추출

3절에서의 학습 단계의 작업을 분석 문자열에도 수행하여 DTW를 통한 1차원 특징 벡터를 추출하게 된다. 학습과정과 마찬가지로 같은 과정으로 추출하게 되는데 이때의 차이점은 학습 단계에서는 특징을 누적 시켰지만, 여기서는 분석을 하기위하여 하나의 분석 문자열의 결과만을 가지게 된다.

4.2 두 특징간의 거리정보 추출

학습된 문서 영상들의 DTW를 이용한 특징 벡터와 분석할 문자열의 특징 벡터를 비교하게 되는 과정으로써, 비교, 분석 후 그 유사도를 도출해 내게 된다.

마할라노비스 거리(Mahalanobis Distance) [9]가 비교하는 방법으로써 사용되어지게 되는데, 다변량 벡터에 대하여 효과적으로 거리를 측정할 수 있는 방법으로써 유사도를 판단하는데 사용하였다.

분석 문자열의 특징을 $F(i) = w(i) - \mu$ 라고 하면 두 특징간의 거리정보는 식(8)과 같이 표현할 수 있다.

$$d = (F - \mu)^T \Sigma^{-1} (F - \mu) \quad (8)$$

DTW를 통한 마할라노비스 거리는 결과 값이 작게 나와 이를 보완하기위해 식(9)의 과정을 수행하여 거리의 값을 보완하였다.

$$S = 1 - \frac{d}{n} \quad (9)$$

5. 실험 및 결과

본 논문에서 기술한 방법은 문서영상에서 글자의 영역을 찾고 그 글자영역에서 추출된 특징을 바탕으로 학습시키고, 비교할 문자열을 입력하여 비교하여 유사도를 나타내는 방법이다. 학습된 양과 문자열에 따라 실험결과가 차이를 보였으며 실험에 사용된 환경은 Pentium4 3.0GHz, RAM512 시스템을 사용하였고 5명의 인원의 글씨 2가지 종류로 분류하여 각각 5회, 10회, 15회, 20회씩 학습을 시켰으며, 각각

의 인원을 서로 모두 비교해가며 결과를 도출하였다.

(단위 : %)

	A	B	C	D	E
A	99	85	79	70	82
B	0	96	0	0	0
C	100	100	100	100	100
D	100	100	100	100	100
E	100	100	100	100	100

표 1. (5회 학습 결과)

(단위 : %)

	A	B	C	D	E
A	100	88	89	89	80
B	86	99	85	87	79
C	77	78	93	58	0
D	100	78	88	99	65
E	82	83	88	84	98

표 2. (10회 학습 결과)

(단위 : %)

	A	B	C	D	E
A	100	0	100	100	100
B	95	100	97	92	85
C	85	99	99	96	0
D	91	93	95	100	92
E	95	91	92	90	100

표 3. (15회 학습 결과)

(단위 : %)

	A	B	C	D	E
A	100	100	100	100	100
B	100	100	100	100	100
C	100	100	100	100	100
D	100	100	100	100	100
E	100	100	100	100	100

표 4. (20회 학습 결과)

6. 결론 및 향후 연구 방향

필적 및 서명 검증은 개인 간의 고유한 필적 개성을 통해 임의의 두 필기 문장 또는 텍스트가 동일인에 의해 작성되었는지를 판별하는 기술이다. 보통은 감정관 이를 행하여 수행하지만, 주관적이라는 단점이 있고, 소모 시간과 비용의 문제를 가지게 된다. 본 논문은 이러한 문제점을 컴퓨터를 이용한 패턴 인식을 통해 객관적이고 효율적인 판단을 하게 하는 목적을 가지고 있다. 컴퓨터를 이용한 패턴인식을 통한 필적 및 서명 검증은 현재 On-line과 Off-line, 이 2가지의 방법이 있으며, 본 논문은 Off-line검증 방식에 대하여 기술하고 있다.

Off-line방식을 통하여 효율적이고 객관적으로 필적 및 서명에 대하여 최소 포함 사각형을 이용한 문자 영역의 추출 후, DTW를 통한 학습이라는 방법을 통해 개인의 고유한 필적 개성을 분석하고 있으며, DTW로 나온 결과를 마할라노비스 거리(Mahalanobis Distance)를 통하여 유사도를 도출하는 방식을 쓰고 있다.

실험 및 결과를 통하여 프로그램의 우수성을 보여주고 있으나 몇 가지 문제점이 발생하였다. 그 문제점은 학습된 양과 문자열에 따라 실험 결과가 차이를 보이게 되는데, 이에 따른 이유를 분석하면 다음과 같다.

첫째, 학습된 문서의 양이 적당하여야 한다. 본 연구에서 기술한 방법대로 처리를 하였을 경우 실험에 대하여 학습의 횟수가 너무 적을 경우 문자의 패턴을 제대로 분석하지 못함으로써 유사도 특징에 실패하게 되는 결과를 가지게 되었고, 이와는 반대로 너무 많을 경우는 학습되는 패턴이 너무 많아 학습 결과가 모해해 짐으로써, 다른 패턴들과의 구분이 불가능 했다. 둘째로는 기울어지거나 왜곡된 문자열이 학습 문자열에 포함되었을 경우 유사도 판별에 문제가 생기게 된다. 이러한 문자열들이 학습될 경우 그 문자열의 영향을 받아 학습 데이터 전반에 영향을 미치게 되어 유사도 판별 시, 잘못된 결과가 도출되는 경우가 발생하게 되어 진다.

결론적으로 위에서 언급한 문제점이 포함된 경우를 제외하고 인식에 대한 실험을 수행한 결과 95%의 높은 변별력을 보여주었으며, 글씨체에 따라 다르지만 결과가 좋을 경우 100%의 변별력을 보여주었다.

향후 연구는 문자가 기울어지거나 왜곡 되었을 때 이를 보정하거나 제외하여 학습 패턴을 보완하는 방법과 다른 특징들의 복합적인 적용에 관한 연구를 추가하게 되면 더욱더 감정에 대한 결과가 향상될 것이라 사료된다.

Acknowledgement

본 연구는 대검찰청 과학수사 1팀 문서감정실의 지원으로 이루어졌습니다.

참고문헌

- [1] Basabi Chakraborty, Goutam Chakraborty "A new feature extraction technique for on-line recognition of handwritten alphanumeric characters" Information Sciences 148, pp. 55-70, 2002
- [2] Kai Huang, Hong Yan "Off-line signature verification using structural feature correspondence" Pattern Recognition vol. 35, pp.2467-2477, 2002

- [3] Alan McCabe "Hidden Markov Modelling with Simple Directional Features for Effective and Efficient Handwrite Verificaion"
- [4] Edson J.R. Justino, Flavio Bortolozzi, Robert Sabourin " A comparison of SVM and HMM classifiers in the off-line signature verification" Pattern Recognition Letters vol. 26 pp. 1377-1485, 2005
- [5] MARC PARIZEAU, REJEAN PLAMONDON "A Comparative Analysis of Regional Correlation, Dynamic Time Warping, and Skeletal Tree Matching for Signature Verification" IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE vol. 12 no7 1990
- [6] B. Fang, C. H. Leung, Y. Y. Tang, K. W. Tse, P. C. K. Kwork, Y. K. Wong "Off-line signature verification by the tracking of feature and stroke positions" Pattern Recognition vol. 36 pp. 91-101, 2003
- [7] 박희주, 김진호, 오광식 "새로운 자소분리 기법을 이용한 필기체 한글 인식 시스템" 한국퍼지 및 지능 시스템학회 논문지 Vol. 5, No3, pp.101-110, 1995
- [8] MARC PARIZEAU, REJEAN PLAMONDON "A Comparative Analysis of Regional Correlation, Dynamic Time Warping, and Skeletal Tree Matching for Signature Verification" IEEE Transactions on pattern analysis and machine intelligence Vol. 12, No7, pp.710-717, 1990
- [9] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart "Tutorial The Mahalanobis distance" Chemometrics and Intelligent Laboratory Systems Vol.50 pp.1-18, 2000