

동적 프로그래밍을 이용한 오프라인 환경의 문서에 대한 필적 분석 방법

(A Verification Method for Handwritten text in Off-line
Environment Using Dynamic Programming)

김 세 훈[†] 김 계 영^{**} 최 형 일^{***}
(Se-hoon Kim) (Gye-young Kim) (Hyung-il Choi)

요 약 필적 감정은 개인의 필적 개성을 이용하여 임의의 두 필기 문장 또는 텍스트가 동일인에 의해 작성되었는지를 판별하는 기술이다. 본 논문은 패턴 인식 기술을 사용하여 효과적으로 필적을 분석하고 판별하는 오프-라인 환경에서의 검증 방법을 제안한다. 본 논문에서 연구된 방법의 핵심 절차는 문자 영역 추출, 문서의 구조적 특징을 반영하는 특징의 추출, DTW(Dynamic Time Warping) 알고리즘과 주성분 분석을 이용한 특징 분석이다. 실험 결과는 제안하는 방법의 우수한 성능을 보여준다.

키워드 : 필적검증, 한글, 동적 시간 정합, 동적 프로그래밍, 주성분 분석, 마할라노비스 거리

Abstract Handwriting verification is a technique of distinguishing the same person's handwriting specimen from imitations with any two or more texts using one's handwriting individuality. This paper suggests an effective verification method for the handwritten signature or text on the off-line environment using pattern recognition technology. The core processes of the method which has been researched in this paper are extraction of letter area, extraction of features employing structural characteristics of handwritten text, feature analysis employing DTW(Dynamic Time Warping) algorithm and PCA(Principal Component Analysis). The experimental results show a superior performance of the suggested method.

Key words : Handwriting Verification, Hangul, DTW, Dynamic Programming, PCA, Mahalanobis Distance

1. 서 론

필적학에서 필적은 글씨체로서 그 사람의 성격, 버릇,

외모, 체형까지 반영한다고 정의한다. 이는 개인에 대한 필적 개성을 나타내며, 그 필적 개성을 분석하여 개인 식별의 연구와 응용이 가능하다. 서명 또한 행위자 자신의 동일성을 표시하기 위해 사용되어지는 것으로 개인 별의 특징을 가지게 되며, 이는 필적과 마찬가지로 개인 식별 연구와 응용이 가능 해진다. 이러한 개인 식별 연구와 응용 작업을 필적(서명)감정이라 한다.

필적(서명)감정은 2가지 이상의 필적(서명) 사이에서 개인 필적의 항상성과 회소성 존재를 식별하여, 이들 필자가 같은 사람인가의 여부를 판단하는 일이다. 필적감정의 검사방법은 대체로 자료의 필적이 자연 필인가, 작의 필인가, 서자기술은 어느 정도인가, 위필의 의혹은 없는가, 대조자료는 적절한가 등을 검사하고, 이어 자세하게 자획구성, 자획형태, 필순, 배자, 필압, 필세, 오자, 옴용 등을 식별한다. 이러한 감정 업무는 유서대필 및 보안수사, 서명검증, 범죄수사 등에 활용되어지고 있고, 현재 이러한 감정 작업은 국가에서 인정한 감정 기관 및 일부 사립감정원에서 이루어지고 있으며, 감정 전문

· 본 논문은 숭실대학교 교내연구비 지원 및 2006년 정부(교육인적자원부)의 재원으로 한국 학술진흥재단의 지원을 받아 수행된 연구비 지원으로 수행되었음(KRF-2006-005-J03801)

[†] 학생회원 : 숭실대학교 미디어학과
krhoonse@hotmail.com

^{**} 종신회원 : 숭실대학교 컴퓨터학부 교수
gykim11@ssu.ac.kr

^{***} 종신회원 : 숭실대학교 글로벌미디어학부 교수
hic@ssu.ac.kr
(Corresponding author임)

논문접수 : 2009년 5월 18일

심사완료 : 2009년 9월 30일

Copyright©2009 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제36권 제12호(2009.12)

가의 판단기준에 의하여 필적(서명)의 유사성을 판별하게 된다. 하지만 감정 전문가에 의해 필적의 유사성을 판단하기 때문에 객관성 결여 및 과도한 소요 시간과 처리 비용의 문제를 가지게 된다. 이를 해결하기 위해 주관적인 요소를 적게 하고, 과학성을 부여하려고 하는 시도에서 컴퓨터를 이용한 분석방법들이 연구되고 있고 그 방법을 통하여 기존 수작업의 문제를 하나하나 줄여나가고 있다.

컴퓨터를 이용한 필적의 분석방법에는 크게 On-line 환경에서의 분석 방법과 Off-line 환경에서의 분석 방법 2가지 종류로 나눌 수 있다. On-line 환경에서의 분석 방법은 디지털타이저나 타블렛 등 기타 입력장치를 이용하고, 필적 및 서명을 입력하여 분석하는 방법이다. 이 방법은 필적 및 서명에 대해 입력되는 순서에 대하여 각 문자에 대한 중심 간의 거리를 이용하여 유사도를 도출 내는 방법[1] 등으로 입력되는 순서와 관련이 있는 특징을 주로 이용하여 패턴인식 작업을 수행하여 유사도를 도출하고 있다. 이와는 다르게 Off-line 환경에서의 분석 방법은 필적이나 서명을 컴퓨터의 입력장치인 스캐너를 통하여 그림 및 사진 파일로 저장하여 분석하는 방법으로써 필적이나 서명에 대하여 입력 순서를 부여하기 어려우며, 그 필적이나 서명에 대하여 다른 특징을 찾아내어 분석을 하는 형태를 갖추고 있다. Off-line 환경에서의 분석 방법에는 기존에 개발된 방법인 구조적인 특징의 유사도를 이용한 방법[2]과 HMM(Hidden Markov Model)분류기를 이용하여 감정하는 방법[3], HMM분류기와 SVM(Support Vector Machine)분류기를 각각 이용하여 감정 하는 방법[4], DTW(Dynamic Time Warping)알고리즘을 이용한 다이내믹 프로그래밍을 통한 서명의 유사도를 판단하는 방법[5], DTW 알고리즘을 이용하여 그 유사도를 마할라노비스 거리 알고리즘을 통하여 비유사도를 측정하는 방법[6,7] 등이 있다.

본 논문은 Off-line 환경 하에서 한글 필적(서명)에 대한 특징 추출 및 분석에 대한 방법을 제안한다. 기존 연구의 문제점인 한글 비교 분석이 가능하게 하기 위해 한글 필적에 대한 문자 영역 추출 문제를 해결하고자 각 문자 영역 추출 방법을 연구하고, 추출된 문자 영역에서의 효과적인 특징 추출방법 및 기존의 짧은 패턴에서 효과적으로 수행하던 방법들을 다양한 길이를 가진 개인별 특징에 대해 비교 분석하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 제안하는 시스템의 개요를 설명한다. 3절에서는 문서영상에서의 문자영역 추출 방법과 문자영역에서의 특징추출 방법을 제안하며, 4절에서는 다른 필적과의 비교 분석을 위한 PCA 알고리즘과 DTW 알고리즘에 대하여 설명하고 그

알고리즘을 이용하여 비선형자료 분석방법에 대하여 제안을 하고, 5절에서는 제안하는 방식을 이용한 실험을, 6장에서는 이러한 실험 결과를 토대로 결론을 제시한다.

2. 문자의 영역 및 특징 추출

2.1 문자영역추출

문자의 영역추출은 입력 받은 영상에서 처리의 기준이 되는 문자의 단위 및 범위를 나눌 수 있는 중요한 과정이며, 자소단위 및 필적(서명)의 크기와 위치 정보를 구할 수 있게 되고, 문자의 전체적인 특징 정보[8]를 나타낸다.

일반적으로, 문자 영역추출은 다음의 그림 1과 같이 3가지 단계를 거쳐 이루어진다.

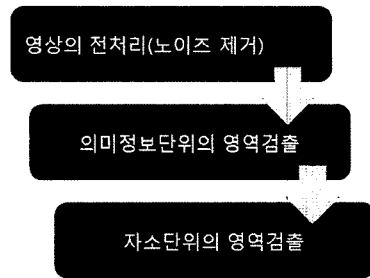


그림 1 문자 영역 추출 과정

영상의 전처리 과정에서는 Off-line으로 입력 받은 영상을 아래의 그림 2의 상단 그림과 같이 전처리과정을 거쳐 노이즈 성분을 제거한 뒤, 문서 영상의 문자부분을 추출하게 된다.

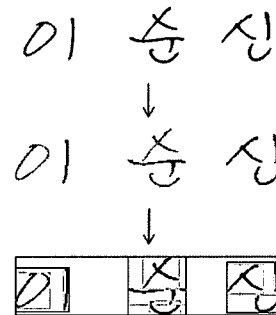


그림 2 문자 영역추출

다음의 과정으로, 의미정보단위의 영역검출 과정에서는 그림 2의 중간 그림과 같이 레이블링 알고리즘을 이용하여, 영상의 연결성분을 검출하여, 각 연결 성분을 구분 짓는다. 마지막 과정으로 의미정보단위의 영역검출이 이루어지는데, 의미정보단위의 영역검출은 그림 2 하

단의 그림과 같이 녹색 사각형으로 표시된 자소단위의 영역검출과, 빨간색 사각형으로 표시된 문자단위의 영역검출, 파란색 사각형으로 표시된 문단단위의 영역검출을 의미한다.

자소단위의 영역검출 과정은 최소단위의 연결영역성분을 검출하고, 각 레이블링된 성분에 대하여 연결요소 검사과정을 거쳐 그림 2와 같이 최소포합사각형을 추출한다. 이 추출된 최소포합사각형을 자소단위의 영역으로 정의하고 사용한다.

문자단위의 영역검출은 자소단위의 검출된 영역에 대하여 다음과 같은 2가지의 규칙을 이용하여 결정한다. 첫째, 문자가 포함하고 있는 정보는 세로로 2개, 가로로 2개 이상의 성분을 포함하지 않는다. 둘째, 글자가 쓰여질 때에는 일정한 거리를 기준으로 문자와 단어가 구분된다. 이 2가지 조건을 이용하여 문자단위영역을 추출한다.

먼저, 문자단위영역을 추출하기 위하여, 최소포합사각형들간의 거리정보를 구해 비교하는 과정이 필요하다.

i 개의 개수를 가진 최소포합사각형(M_i)에 대하여 거리(d_i)는 다음의 식 (1)과 같이 계산한다.

$$d_i = \frac{\sqrt{(M_i^{C_x} - M_{i+1}^{C_x})^2 + (M_i^{C_y} - M_{i+1}^{C_y})^2}}{-\frac{1}{2} \sqrt{(M_i^w)^2 + (M_i^h)^2} + \sqrt{(M_{i+1}^w)^2 + (M_{i+1}^h)^2}} \quad (1)$$

w 는 최소포합사각형의 넓이, h 는 최소포합사각형의 높이, i 는 최소포합사각형의 인덱스, C 는 최소포합사각형의 중심점(C_x, C_y)를 의미한다.

계산된 거리를 이용하여 문자영역을 추출하기 위하여, 각 최소포합사각형들(M_i)에 대하여, 다음의 식 (2)를 이용하여, 임계 거리보다 작은 사각형을 통합하여 새로운 문자단위영역($newM_i$)을 정의한다.

$$where(d_i < threshold) \quad (2)$$

$$new M_i^{C_x} = \frac{1}{2} \{ (M_i^{C_x} + M_{i+1}^{C_x}) - |M_i^{C_x} - M_{i+1}^{C_x}| \}$$

$$new M_i^{C_y} = \frac{1}{2} \{ (M_i^{C_y} + M_{i+1}^{C_y}) - |M_i^{C_y} - M_{i+1}^{C_y}| \}$$

$$new M_i^w = (M_i^w + M_{i+1}^w) + |M_i^w - M_{i+1}^w|$$

$$new M_i^h = (M_i^h + M_{i+1}^h) + |M_i^h - M_{i+1}^h|$$

2.2 문자영역의 특징추출

본 논문에서 제안하는 특징 추출 방법은 문서 영상의 각 문자의 획에 대하여 중심점과의 거리정보를 이용하는 방법이다.

기준에 제안된 방법 중 하나인 투영 방법을 이용한 특징 추출 방법은 각 문자에 대하여 일정 범위의 검은 색의 영역의 분포를 계산하여 문자마다의 고유의 특징을 추출한다. 다음의 그림 3과 같이 추출된 결과를 볼 수 있다.

하지만, 예를 들어 ‘이’와 같은 영역에서는 동그란 부

이 순 신



그림 3 투영의 방법

분 ‘o’이 있는 부분에서도 두께나 분포 및 모양에 따라 분포가 틀리며, 다른 문자인 ‘ㅣ’영역에서도 o과 다르게 분포를 나타내므로, 각 문자에 대한 고유의 특성을 추출이 가능하다. 하지만, 문자영상의 크기나 방향에 변화에 따라 다른 결과를 보이게 되는 문제점이 있다.

본 논문에서 제안하는 특징추출방법은 다음의 그림 4와 같다. 먼저, 필적(서명) 영상에 대하여 각각의 문자영역내의 문자 영상을 세선화알고리즘을 이용하여 한글문자의 구조적특징[9]에 따라 문자 내의 획 정보를 추출, 추출된 획 정보를 이용하여 단점, 연결점, 끝꼭짐, 분기점, 교차점등의 점에 대한 정보[10]를 구한다.

그러나, 획을 이용하여 비교를 하기 위하여 추출되는 특징은 크기와 모양 방향이 일치해야한다는 기준을 가지게 된다. 그러나 크기는 문자열 전체에서 일관성만 있으면 문제가 되지 않으므로 본 방법을 통하여 모양과 방향정보를 사용하여 지역적인 특징정보를 산출하고, 문자 크기에 독립적이고 모양과 방향에 따라 다른 평가가 가능한 특징추출을 수행 하고자 한다.

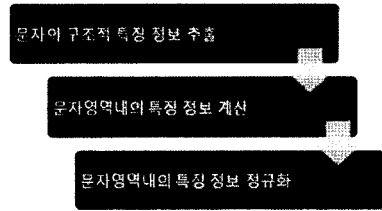


그림 4 특징 추출 과정

앞서 거론된 바와 같이, 크기에 독립적이고 모양과 방향에 따라 다른 평가가 가능하기 위하여, 문자 단위영역에서 중심을 구한 뒤 그 중심과 각 획소까지의 거리를 최소포합원의 반경을 이용하여 정규화 하여 그 각 획소들을 추적하여 특징을 다음의 그림 5와 같이 추출한다.

점에 대한 정보 중, 단점들 간의 거리 정보에서 최소포합사각형의 (x_{min}, y_{min})과 (x_{max}, y_{max})를 계산하고, 그 중심(x_c, y_c)과 단점들 간의 거리정보를 이용하여 최대포합원의 반경 R 을 계산한다. 획소들의 위치를 추적하기 위하여 시작점을 검출하고, 한글 및 여러 문자의 필기체의 특징상 왼쪽에서 오른쪽으로 쓰이는 점에 착

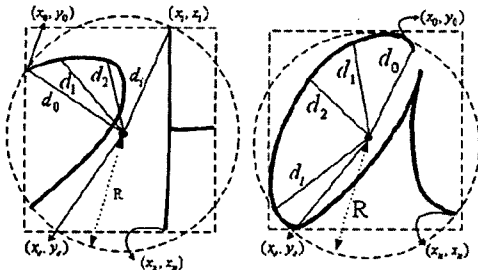


그림 5 중심점과의 거리정보를 이용한 특징 추출

안하여 좌측 상단에 각각 가중치를 주어 시작점을 찾는다[11,12].

그 다음단계로, 시작점을 선정 후 추출된 점들을 후보군으로 분류하여 다음 추적 대상을 반복적으로 탐색하여, 점들의 다음 시작점을 결정하게 되며, 전체 획에 대한 추적이 가능하게 된다.

최소포함사각형(M_i)의 개수를 i 라 할 때, 시작점이 결정된 뒤 시작점으로부터 추적되는 흑화소의 위치를 $C = \{(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)\}$ 라 할 때 최대포함원의 반경과 최소포함원의 반경을 이용하여 정규화된 거리 $L = \{l_0, l_1, \dots, l_n\}$ 를 식 (3)과 같이 산출하게 된다.

$$l_i = \sum_{x=0}^n \sum_{y=0}^n \frac{\sqrt{(x_n - x_c)^2 + (y_n - y_c)^2}}{R_i} \quad (3)$$

이렇게 하여 추출된 정규화된 거리 L 을 이용하여 추출된 특징을 확별로 관리하여 특징으로 이용한다.

3. 비선형적 자료 형태의 비교분석

추출된 특징들의 형태는 일정한 개수로 정해지지 않은 비선형적인 자료의 형태를 가지게 된다. DTW(Dynamic Time Warping)알고리즘[13]은 기준이 되는 특징의 패턴과 입력된 특징의 패턴간의 유사도를 동적 프로그래밍(Dynamic Programming)을 이용해 구해내는 방법이다. 이 방법을 통해 서로 다른 특징 길이를 가지고 있는 특징 벡터들 간의 비교가 가능하므로, 비선형적인 대응관계로부터 유사도 산출이 가능하다. 비선형적인 자료의 형태를 DTW알고리즘의 정합을 이용 다음의 그림 6과 같은 형태로 DTW알고리즘을 이용하여 두 특징 벡터간의 정합을 수행한다.

참조 문자열은 문자의 특징을 비교해 내기위한 기준 문자열로서, 그 사람의 대표적인 필적 특징을 의미한다. 학습 문자열은 대표적인 필적에, 사람이 문자를 쓰면서 변화하는 특징을 학습하기 위한 문자열이다.

두 문자의 영역의 투영된 데이터 특징을 각각 $Q = \{q_0, q_1, \dots, q_n\}$, $T = \{t_0, t_1, \dots, t_p\}$ 라고 했을 때, 이들 사이의 i 개의 DTW매칭 경로(w_i)의 값($d(w_i)$)를 식 (4)

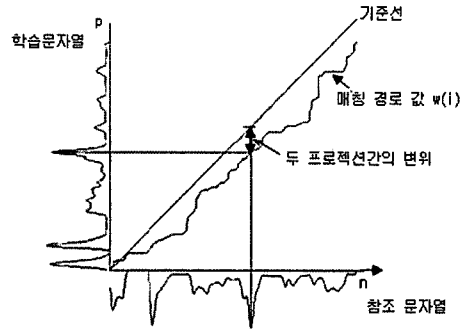


그림 6 비선형적 특징의 정합

과 같이 정의할 수 있다.

$$d(w_i) = d(Q_n, T_p) = ||Q(q_n) - \mu_n + [T(t_p) - \mu_p]|| \quad (4)$$

μ 는 데이터 특징의 평균을 나타낸다.

투영된 두 데이터 특징의 경로($d(w_i)$)를 최소화하는 경로 값을 식 (5)와 같이 표현한다.

$$DTW(Q, C) = \min \frac{1}{i} \sqrt{\sum_1^i W_i} \quad (5)$$

$$W_i = \{w_1, w_2, \dots, w_i\}$$

기준선을 바탕으로 특징데이터를 경로를 최소화 하는 값의 결과인 식 (5)의 w_i 값 이용하여, 식 (6)의 과정을 거쳐 k 개의 길이를 가진 특징벡터(f)로서 저장한다.

$$f_k = [w_i(1) - G_1, w_i(2) - G_2, \dots, w_i(k) - G_k] \quad (6)$$

이때의 G 는 다음과 같다.

$$G_n = \frac{p}{n} k \quad (7)$$

식 (6)의 과정으로 생성된 특징 벡터들을 다음의 그림 7과 같이 한사람의 필적(서명)에 대하여 특징행렬을 생성하게 된다.

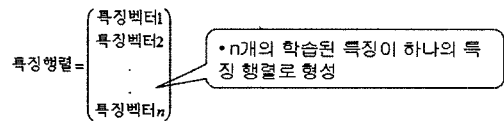


그림 7 특징 행렬의 형성

PCA(Principle Component Analysis)알고리즘[14]은 고차원 데이터로부터 데이터의 구조를 밝히거나, 데이터의 차원을 낮추는데 많이 이용되는 다변량 통계 분석법이다.

n 개의 특징 행렬 $X = \{x_1, x_2, \dots, x_m\}^T$ 의 각 벡터가 k 개의 개수를 가진다면, 학습된 특징벡터는 $x_m = \{x_m(1), x_m(2), \dots, x_m(k)\}$ 가 되며, 각 특징 벡터의 평균 벡터(Ψ_m)는 다음의 식 (8)과 같이 구할 수 있다.

$$\Psi_m = \frac{1}{k} \left[\sum_1^k \chi_m(k) \right] \quad (8)$$

특징행렬의 특징벡터(Φ)와 평균 벡터(Ψ) 사이의 차는 식 (9)와 같이 구한다.

$$\Phi_m = \chi_m(k) - \Psi_m, \quad k=1, \dots, k \quad (9)$$

다음의 식 (10)을 이용하여 공분산 행렬(C)을 구한다.

$$C = \frac{1}{m} \left[\sum_1^m \Phi_m \Phi_m^T \right] \quad (10)$$

행렬(C)의 고유값(λ)과 고유벡터(e)는 식 (11)의 관계를 이용하여 구한다.

$$C_e = \lambda_e \quad (e \neq 0) \quad (11)$$

계산된 고유 벡터들의 행렬을 A 라 하고, 다음의 식 (12)와 같이 데이터 벡터 χ 의 주성분 값(ψ)를 구한다. 가장 높은 고유값을 가지고 있는 고유벡터를 이용하여 주성분을 형성하고 공분산 행렬로부터 고유벡터를 얻은 후, 고유값에 따라 정렬함으로 성분 선택과 특징벡터(ψ) 형성과정을 거친다.

$$\psi = A(\chi - \Psi) \quad (12)$$

최종적인 유사도를 도출하기 위하여 위와 같은 학습의 과정을 거친후 비교할 비교 문자열과의 효과적인 비교를 위해, 본 논문은 마할라노비스 거리(Mahalanobis Distance)알고리즘을 이용한다.

마할라노비스 거리(Mahalanobis Distance)알고리즘은 군집분석에서 가장 많이 사용되는 거리개념으로서, 변수들 사이의 표준편차와 상관관계를 고려하여 만들어진 거리로서, 두 지점의 단순한 거리뿐만 아니라, 변수의 특성을 나타내는 표준편차와 상관계수가 함께 고려된다는 특징을 가지고 있다.

비교할 문자열에 대하여 2절에서 제안된 특징 추출 방법을 이용하여 추출한다. 학습 문자열의 특징에 대한 공분산(C) 및 고유 벡터(e)를 이용하여 비교문자열과의 정합과정을 거쳐 변위를 측정하여 비교할 문자열의 특징정보(F)로 저장한다.

분석할 문자열의 특징(F), 학습된 특징의 공분산(C)와 학습된 특징벡터값(ψ)를 이용하여 산출한 분석된 특징간의 거리정보(MD)는 식 (13)과 같이 표현할 수 있다.

$$MD = (F - \psi)^T C^{-1} (F - \psi) \quad (13)$$

이렇게 산출된 결과는 정규화 되지 않아 특징의 길이에 따라 유사도가 서로 다르게 산출 될 뿐 아니라, 거리가 클수록 적은 값을 가지는 비유사도이다. 식 (13)과 같이 정규화된 유사도(S)로 재 산출한다.

$$S = 1 - MD \quad (14)$$

4. 실험 및 결과

4.1 실험 환경

본 실험을 위해 사용한 컴퓨터는 Intel Pentium Core 2 Duo의 CPU와 1GB의 RAM을 사용하였고, 운영체제는 Microsoft의 Windows XP Home edition을 사용하였으며, 사용언어는 Microsoft의 Visual C++ 8.0을 사용하여 제작하였다.

4.2 실험 및 결과 평가

본 논문에서 제안된 추출방법 및 다양한 길이를 가진 개인별 특징에 대해 비교 분석하는 방법에 대하여 실험 하였다.

실험 결과를 내보이기위해 실험군을 선정하여, 각 실험군별로 100개의 필적을 제공 받아, 본인 필적과 대리 필적을 구분, A,B,C,D,E 5그룹으로 구성 하여, 각각 그룹에 대하여 5, 10, 15 개 종류의 입력된 필적을 기준으로 학습의 횟수로 구분하여 실험을 진행하였다.

A, B, C, D, E 각 그룹에 대하여, 실험결과를 다음의 표 1~3으로 정리하였고, 표 내부의 음영 표시된 영역은 같은 필적을 비교하였을 때를 표시하고, 투영과 제안된 방법에 대한 비교를 수행하기 위하여 2개의 항목으로 따로 결과를 나타내었다.

기존에 연구된 투영을 통한 방법[8]과 본 논문에서 제안한 특징 추출과 비교분석방법의 두 가지 특징을 비교하여 결과를 다음의 표 1~3과 같이 표시하였으며, 실험 결과를 통하여 제안하는 특징방법의 우수성과 제안하는 분석 방법에 대하여 다른 방법과의 비교를 통하여 차이를 보이고 있다.

기존의 연구 방법인 투영을 통한 방법은 입력을 받은 영상에 대하여, 문자열에 대한 특징을 추출할 때 영상을 기준열에 대하여 투영방법을 이용하여 특징을 추출하고, DTW알고리즘을 이용하여 비교 분석한 방법이다.

표 1 5회 학습된 결과 단위: %

5회	A		B		C		D		E	
	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법
A	92	83	94	21	94	18	90	15	75	27
B	40	21	64	64	60	65	0	48	1	60
C	92	45	31	40	87	63	88	60	62	61
D	92	0	92	0	93	0	90	74	93	0
E	92	0	92	18	92	17	93	10	91	75

표 2 10회 학습된 결과 단위: %

10회	A		B		C		D		E	
	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법
A	90	75	91	11	79	23	95	9	78	28
B	31	25	63	46	63	53	29	73	26	76
C	54	3	24	71	81	45	79	43	37	40
D	72	74	88	70	90	86	86	60	85	71
E	89	5	88	15	88	8	79	11	72	63

표 3 15회 학습된 결과 단위: %

15회	A		B		C		D		E	
	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법	투영	제안된 방법
A	88	69	89	7	90	0	93	26	73	5
B	23	18	54	38	66	48	23	65	25	71
C	16	0	16	62	78	32	55	18	28	17
D	84	65	70	63	89	85	83	53	82	68
E	86	81	57	77	85	78	74	78	84	53

위의 표에서 보는 바와 같이 각각 5회 학습 시(표 2)에 가장 변별력 있게 구별하는 모습을 보였다. 하지만 실험 및 결과를 통하여 제안하는 분석 방법의 몇 가지 문제점이 발생하였다. 그 문제점을 분석해보면, 표 3에서 보이듯이 학습된 문자열의 양이 많을 경우 학습되는 문자열의 상의한 상태를 가지게 되어 여러 학습된 문자열들의 패턴이 모호해져 오히려 비유사도가 증가함을 알 수가 있었다. 실제로 실험 결과(표 3)의 C, D의 학습 패턴들에서 발생하였으며, 5회의 학습에서 좋은 결과를 보이던 결과들이 10회 이상의 학습 결과에서 가장

비유사도가 높은 문자열로 분류됨을 알 수 있었다.

위의 그림 8과 그림 9는 기존 방법과 제안된 방법의 효과적인 비교를 위하여 그룹 A를 기준으로 결과를 나타내었다. 세로축은 정확도를 표현하며, 가로축은 학습된 횟수를 나타내었다. 그림 8의 경우, 투영의 방법과 비교하였을 때 제안한 방법이 더 좋은 결과를 나타냄을 볼 수 있었고, 그림 9의 경우, 기존에 제안된 방법과 비교하여 더 좋은 결과가 나타남을 볼 수 있었다.

5. 결론

필적은 글씨체로서 그 사람의 성격, 버릇, 외모, 체형까지 반영한다고 정의한다. 이는 개인에 대한 필적 개성을 나타내며, 그 필적 개성을 분석하여 개인 식별의 연구와 응용이 가능하다. 서명 또한 행위자 자신의 동일성을 표시하기 위해 사용되어지는 것으로 개인별의 특징을 가지게 되며, 이는 필적과 마찬가지로 개인 식별 연구와 응용이 가능해진다. 이러한 연구 분야를 일반적으로 감정 판 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 논문에서 제안된 방식은 다음과 같다. Off-line 검증 방식을 이용하여, 필적 및 서명에 대하여 스캐너 및 기타 입력장치를 통하여 입력을 받은 후, 입력받은 필적에 대하여 최소 포함 사각형을 계산하는 연산을 이용한 문자 영역의 추출 후, 필적의 궤적을 추적하는 방식의 특징 추출과정을 거쳐, Dynamic Programming 방법을 이용한 nonlinear-PCA 알고리즘을 통한 학습방법을 이

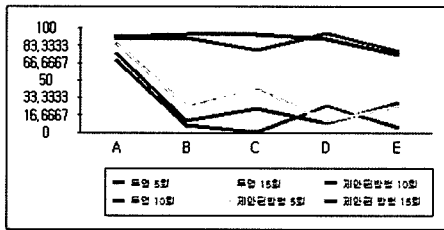


그림 8 A 기준으로 투영된 특징과 제안된 특징추출방법의 결과의 비교

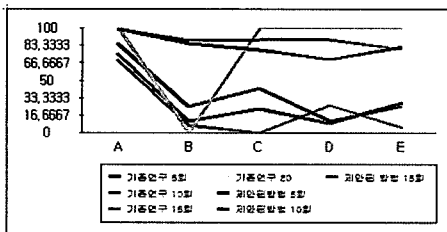


그림 9 A 기준으로 기존 연구 방법의 결과와 제안된 방법의 결과의 비교

용하여 개인의 고유한 필적 개성을 분석하고 있으며, PCA를 통해 도출된 결과를 마할라노비스 거리(Mahalanobis Distance) 알고리즘을 통하여 유사도를 도출하는 방식을 제안한다.

실험 및 결과를 통하여 제안하는 방법에 대하여 기존의 방법보다 제안된 방법이 더 높은 변별력을 가짐을 알 수 있다. 제안된 특징 추출 방법으로 인하여 기존 투영방법으로 시행하던 방법상의 두께 정보보다 개인의 필순에 따라 유사도 도출이 정확도 있게 문자열간의 유사도를 도출해낼 수 있음을 실험을 통해 알 수 있었고, 5회 정도의 짧은 학습을 통하여 좋은 결과를 보였음을 알 수 있다.

참 고 문 헌

- [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 Verification," *Accepted in Proceedings of the Sixth Pacific Rim International Conference on Artificial Intelligence, (PRICAI 2000)*, Melbourne, 2000.
- [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 no.7, 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] S. H. Kim, K. s. Oh, H. I. Choi, "Off-line Verification System of the Handwrite Signature or Text, Using a Dynamic Programming," *Computational Science and Its Applications - ICCSA2007*, vol.4705, pp.1014-1023.
- [8] H. J. Park, J. H. Kim, K. S. Oh, "Handprinted Korean Characters Recognition System by Using New jaso Decomposition Method," *Journal of fuzzy logic and intelligent systems*, vol.5, no.3, pp.101-110, 1995.
- [9] D. G. You, G. H. Kim, "Slant Correction of Handwritten Strings using Properties of Korean Characters," *Journal of KIISE : Software and applications*, vol.30, no.2, pp.93-102, 2003.
- [10] W. H. Jo, "Introduction of Character Recognition, and Development of New Recognition-System," *Proceedings of the Spring Conference, Korean Statistical Society*, pp.68-74, 1992.
- [11] S. S. Yoon, H. R. Byeon, I. B. Lee, "Feature extraction motivated by human information processing method and application to handwriter character recognition," *Korean Journal of Cognitive Science*, vol.9, pp.1-11, 1998.
- [12] G. Y. Kim, "A Quality Evaluation System of a Handwriting String by Global and Local Features," *Journal of Korean Society for Internet Information Korean Society for Internet Information Bimonthly*, vol.5, no.6, pp.121-128, 2004.
- [13] Eamonn J. Keogh, Michael j. Pazzani, "Derivative Dynamic Time Warping," *In First SIAM International Conference on Data Mining (SDM'2001)*, 2001.
- [14] Lindsay I Smith, "A tutorial on Principal Components Analysis," *February 26*, 2002. Page 2. Chapter 1. Introduction, 2002.



김 세 훈

2006년 한국교육개발원(공학사). 2008년 숭실대학교대학원 미디어학과(공학석사) 2008년~현재 숭실대학교 일반대학원 미디어학과 박사과정. 관심분야는 영상처리, 컴퓨터비전, 패턴인식, 증강현실 등임



김 계 영

1990년 숭실대학교 전자계산학과(공학사) 1992년 숭실대학교대학원 컴퓨터학과(공학석사). 1996년 숭실대학교대학원 컴퓨터학과(공학박사). 1996년~1997년 한국전자통신연구원(Post Doc.). 1997년~2001년 한국전력공사 전력연구원(선임연구원) 2001년~현재 숭실대학교 컴퓨터학부(부교수). 관심분야는 컴퓨터비전, 형태인식, 생체인식, 증강현실, 영상 및 신호처리 등임



최 형 일

1979년 연세대학교 전자공학과(공학사) 1982년 미시간대학교 전산공학과(공학석사). 1987년 미시간대학교 전산공학과(공학박사). 1995년~1997년 퍼지 및 지능시스템학회 이사. 1996년~1998년 정보과학회 컴퓨터비전 및 패턴인식 연구회 위원장. 1997년 IBM Waston Lab 방문연구원. 2005년~2006년 한국정보과학회 이사. 1987년~현재 숭실대학교 글로벌미디어학부(교수). 관심분야는 컴퓨터비전, 패턴인식, 퍼지이론, 비디오검색, 인터페이스 에이전트 등임