

Improvement of Historical-Hanja Recognition Using a Nonlinear Transform of Contour Directional Feature Vectors

Min Soo Kim¹⁾ and Jin Hyung Kim²⁾

Abstract

In Korea, OCR-based techniques have been developed for digital library construction of historical documents. In this paper, we propose the nonlinear transform of contour directional feature(CDF) vectors using logit and power transforms with skewness criterion to enhance the discriminant power. Experiments were conducted using samples from Seung-jung-won diaries(Diaries of King's Secretaries). Our results show that proposed method outperforms the others like Box-Cox transform in this database.

Keywords : feature extraction, character recognition, digital library

1. 서론

중국문자인식(Chinese character recognition)에 대한 연구는 수십 년 동안 이루어져 왔다(Kimura et al(1997), Kato et al(1999)). 중국문자는 중국뿐만 아니라, 우리나라와 일본, 베트남 등에서 사용되고 있다 - 중국에서는 한지(Hanji), 일본은 간지(Kanji) 그리고 우리나라에서는 한자(Hanja)라고 한다. 최근 들어, 필기체 문자인식의 분야에서 구조적 분석 접근방법(Yamamoto et al(1982))보다는 특징추출을 통한 통계적 패턴매칭(pattern matching)방법이 선호되고 있다. 그러므로, 현재의 광학문자인식(OCR)기술에 있어서 전체적인 시스템의 정확도는 추출된 특징벡터의 판별능력과 분류기의 일반적인 수행력에 달려 있다고 할 수 있다.

필기체 한자는 복잡한 구조와 많은 모양의 변이를 가지고 있으며, 사용되는 한자의 종류도 많을 뿐만 아니라 비슷한 형태가 너무나 많이 존재한다. 이러한 이유로, 픽셀(pixel)값에 기반한 매칭 방법들은 잘 수행되지 못한다. 그러므로, 모양 정규화(shape normalization)와 같은 전처리(preprocessing)와 한자분류를 위한 적절한 특징을 뽑을 수 있는 특징추출(feature extraction)이 필요하다.

Kim et al(2004a), Kim et al(2004b) 등에서 조선시대의 한자들로 쓰여진 고문서들의 디지털 라이브러리(digital library) 구축을 위한 광학문자인식의 연구를 찾아볼 수 있다. 이 연구들에서 디지털 라이브러리 구축의 전체적인 구성뿐만 아니라, 특징추출방법, 문자분할(segmentation) 방법, 분류 방법 등이 개발되고 있으며, 고문서 디지털화 작업에 정부기관들의 적극적인 투자가 이루어

1) Postdoctoral Fellow, CS Div, Dept. of EECS, KAIST, Daejeon, 305-701, Korea.
E-mail : mskim@ai.kaist.ac.kr

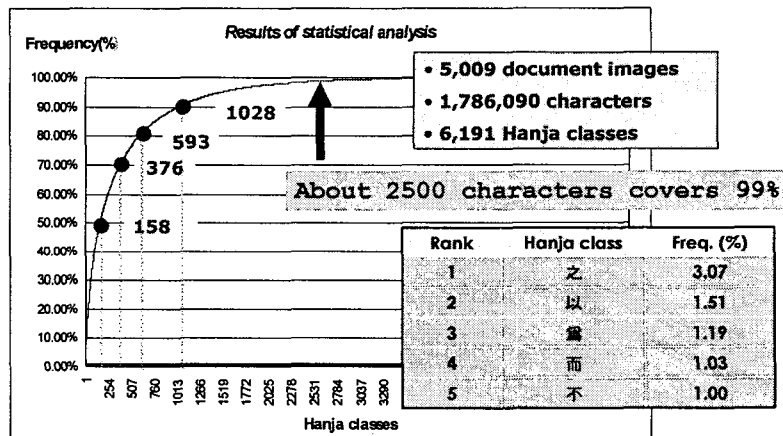
2) Professor, CS Div, Dept of EECS, KAIST, Daejeon, 305-701, Korea.

지고 있음을 알 수 있다. 고문서에 등장하는 한자는 현재는 사용되지 않는 글자가 상당수 있을 뿐만 아니라 붓으로 쓰여진 경우가 많아서 <그림1>과 같은 번짐 현상이 많고 필기자마다 같은 글자라도 필기습성에 따라 다르게 쓸 수 있다.



<그림1>번짐 현상이 있는 글자들

그런데, <그림2>와 같이 조선시대 고문서인 ‘승정원일기’ 제 29책에 등장하는 한자의 모형(종류)은 모두 6,191자이며, 이중 약 1000 글자모형이 90%, 약 2500 글자모형이 전체의 99%를 차지하고 있고 특히, ‘之’자의 빈도는 전체의 3.07%이다. 즉, 글자의 빈도수에 상당한 차이가 있기 때문에 심지어 ‘之’가 수만 번 등장할 때 한두 번 등장하는 글자모형도 있다는 것을 말한다. 본 연구와 같이 목적이 고문서 디지털화인 경우, 광학문자 인식시스템에서는 문서전체에서 몇 번 등장하지 못하는 글자에 대해 인식을 잘 하는 것보다는 자주 등장하는 글자들에 대해서 인식을 잘 하도록 분류기를 구성하는 것도 중요하다.



<그림2> 빈도분석을 통한 한자모형의 수

한자에는 글자는 다르지만, 같이 사용되는 글자도 있다. 즉, 다음의 세 글자(体, 體, 體)는 모두 몸 ‘체’자이다. 이와 같은 경우, 같은 몸 ‘체’자라고 쓰일지라도 모형을 각각 다르게 만들어서 훈련 집단(training set)을 얻어야 한다.

위와 같은 고문서 광학문자인식에서 특징벡터의 추출은 대단히 중요한 부분이다. 실제적으로 문자인식 기술의 가장 큰 부분은 특징추출이라고 해도 과언이 아니다. 본 연구에서는 고문서 한자인식의 특징 추출 방법으로 사용되는 윤곽선 방향 특징(contour directional feature) 추출방법의 특징벡터를 로짓 변환(logit transform)과 멱 변환(power transform)을 통해 선형판별분석(linear discriminant analysis)의 분류에 적합하도록 개선하는 방법을 제안하고자 한다.

2장에서는 비선형 형태정규화 방법과 윤곽선 방향 특징추출방법을, 3장에서는 제안하는 윤곽선 방향 특징벡터의 비선형 변환방법을 설명한다. 그리고 4장에서는 실험결과와 결론을 서술한다.

2. 비선형 정규화 및 특징추출방법

문자분할작업이 처리된 문서에 대해서 본 연구의 고문서 한자인식 시스템은 먼저, 비선형 형태 정규화 단계를 거친다. 그리고 윤곽선 방향 특징추출이 수행되고 다음으로 선형판별분석을 통한 분류단계가 처리되고 있다.

2.1 비선형 형태 정규화방법

필기체 한자에는 위치, 크기, 방향, 기울기 등에 있어 많은 변이가 존재한다. 정규화는 크게 선형정규화와 비선형 정규화로 나뉜다. 수학적 계산이 간단한 선형 정규화방법이 흔히 사용되지만, 형태가 불규칙하고 지역적으로 변이가 심한 필기 문자분류에는 충분한 성능을 발휘하지 못하는 단점이 있다. 결국, 이러한 제약 사항을 개선하기 위해 비선형 정규화 방법이 보편화되고 있다.

다양한 비선형 형태 정규화 방법들이 제안되고 있는데, 이것은 특징투영과 특징밀도 균일화 과정으로 볼 수 있다. 특징투영 과정은 입력 영상의 각 점에서 각 특징을 수평축 또는 수직축으로 투영하여 특징투영 히스토그램을 구한다. 특징밀도 균일화과정은 특징투영 히스토그램을 재표본화 함으로써 입력영상의 특징밀도를 균일화하는 것이다. 비선형 형태 정규화 방법은 점 밀도를 이용하는 방법과 교차횟수, 획 간격, 그리고 내접원에 바탕을 둔 획 밀도를 이용하는 방법들이 있다 (Lee et al(1994)). 비선형 형태 정규화 방법들의 정량적 평가에 의하면, 윤곽선 방향 특징을 사용하는 경우에 획 간격을 이용하는 방법이 우수한 인식 성능을 보이는 것으로 알려져 있다.

이 방법은 Tsukumo와 Tanaka(1988)가 처음 제안한 방법으로 입력영상에서의 획 간격을 특징으로 사용한다. 본 논문에서 이 방법이 사용되었으므로 간략한 설명을 한다. 획 간격을 이용하는 방법의 특징투영함수는 입력 영상의 점(i, j)에서의 수평축 특징 함수와 수직축 특징 함수를 사용하였다. 특징함수 $F_H(i, j)$ 와 $F_V(i, j)$ 는 각각 이웃하는 두 획사이의 간격에 대한 역수로 정의된다.

$$F_H(i, j) = \frac{1}{h(i, j)}, \quad F_V(i, j) = \frac{1}{v(i, j)}$$

이 때, $h(i, j)$ 는 점(i, j)의 오른쪽과 왼쪽에 이웃하는 두 획 사이의 수평간격을 나타내며, $v(i, j)$ 는 점(i, j)의 위와 아래에 이웃하는 두 획 사이의 수직간격을 나타낸다. 입력영상 $I \times J$ 에 대하여, 특징함수 $F_H(i, j)$ 와 $F_V(i, j)$ 를 수평축과 수직축으로 투영하는 특징투영함수는 다음 식과 같다.

$$H(i) = \sum_{j=1}^J F_H(i, j), \quad V(j) = \sum_{i=1}^I F_V(i, j)$$

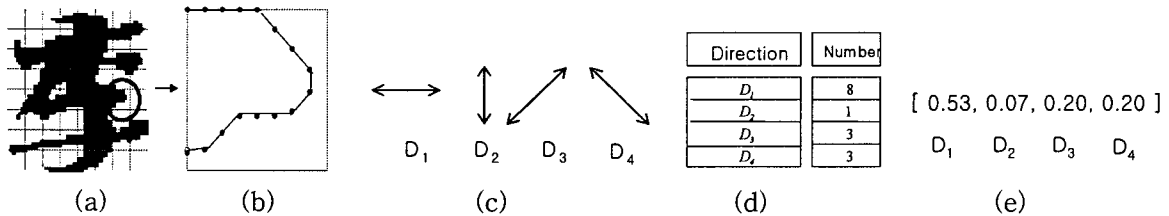
특징 밀도 균일화에서는 특징투영 히스토그램을 재표본화 함으로써 영상의 특징 밀도를 균일화하는 단계로서, 정규화 영상의 좌표(m, n)은 다음과 같이 계산할 수 있다. 이 때, 특징투영 함수를 누적하는 것으로 정의되는 뒤틀림 함수 $\Phi(i)$ 와 $\Psi(j)$ 를 제안하고, 이 함수를 이용하여 $M \times N$ 크기의 정규화 영상의 좌표를 얻을 수 있다.

$$m = \Phi(i) = A \sum_{k=1}^i H(k) + i_0, \quad n = \Psi(j) = B \sum_{l=1}^j V(l) + j_0$$

여기서, A와 B는 각각 $A = \frac{M}{\sum_{k=1}^I H(k)}$, $B = \frac{N}{\sum_{l=1}^J V(l)}$ 를 나타낸다.

2.2. 윤곽선 방향 특징(contour directional feature) 추출방법

<그림3>는 $N \times N$ 격자의 특징을 사용하는 후보선택방법으로, 임의의 블록으로부터 추출한 방향성에 대한 설명이다. 본 연구에서는 8×8 격자를 이용하였다. 즉, 각 격자별로 4개의 특징이 계산되므로, 전체는 256차원의 특징 값이 계산된다.



(a) 입력영상의 블록 (b)원이 가리키는 확대된 블록 (c)윤곽선의 4방향성분(수평,수직,오름대각선,내림대각선) (d) 블록내의 방향성분의 개수 (e) 계산된 방향성분의 비율(윤곽선방향 특징)

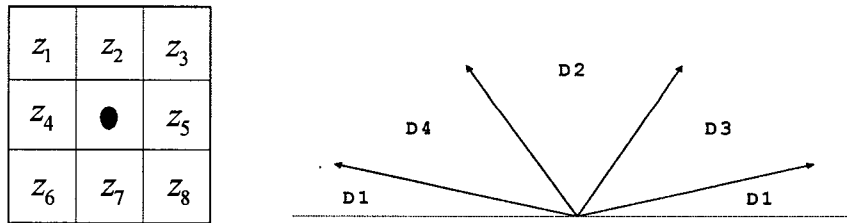
<그림3> 윤곽선 방향 특징추출방법 및 절차

입력영상이 비선형 정규화에 의해 $N \times N$ 블록으로 나누어지면, 각 블록에 대해 4차원 방향 벡터를 추출한다(Kimura 와 Shridhar(1991)). 검은 화소의 각도는 다음 <그림4>(a)에 있는 3×3 창(window)의 8-방향 이웃 화소에 의해 결정된다. 검은 화소(x, y)의 각도를 α 라 하면, 식은 다음과 같다.

$$\alpha = \tan^{-1}(G_x/G_y),$$

$$G_x = (z_6 + z_7 + z_8) - (z_1 + z_2 + z_3),$$

$$G_y = (z_3 + z_5 + z_8) - (z_1 + z_4 + z_6)$$

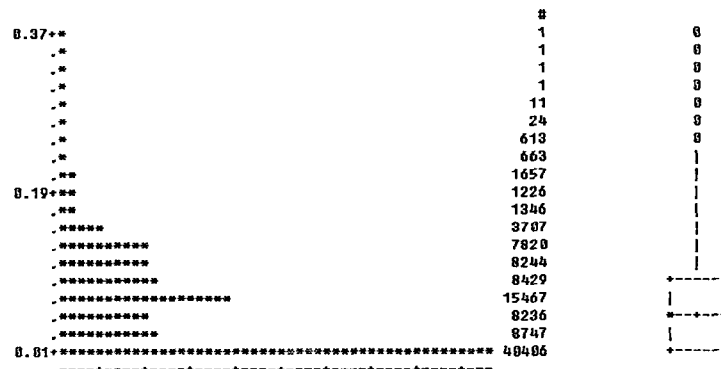


(a) 검은 화소의 8-이웃 (b) 각도의 4-그룹
 <그림4> 검은 화소의 8-방향 이웃 화소와 각도의 4-그룹

검은 화소의 각도는 각각 수평, 수직, 그리고 두 개의 대각선 성분에 대응하는 네 개의 그룹으로 나누어진다(<그림4>(b)). 따라서, 각 블록에 대한 4차원 방향벡터는 네 그룹에 속하는 블록내의 검은 화소의 개수로부터 추출한다.

3. 제안하는 특징벡터의 변환방법

2장의 설명으로부터, 입력된 고문서 한자 영상은 문자분할이 되고 난 후 윤곽선 방향 특징벡터가 추출됨을 알 수 있었다. 그런데, 윤곽선 방향 특징추출법에 의해 추출된 특징 값들의 분포를 살펴보면, 대부분이 다음 <그림5>와 유사한 형태를 취한다(적어도, 정규분포보다는 지수분포에 가까운 형태임을 알 수 있다.). 그런데, <그림5>는 전체 256개의 특징벡터 중 임의의 한 벡터(표본의 수, n=106,600)의 분포지만, 256개 대부분의 특징벡터들이 이와 유사한 형태의 분포를 하고 있다.



<그림5> 256개의 특징벡터 중 임의의 한 벡터의 히스토그램과 상자그림 (n=106,600)

특징 값들이 오른쪽으로 긴 꼬리를 갖는 왜도(skewness)가 양수(+)인 분포를 따르고 있음을 알 수 있다. 특히, 윤곽선 방향 특징추출법의 특성에 의해, 격자된 공간에 검은 픽셀이 없다면 그 픽셀의 4가지 특징 값들은 모두 0이므로 0의 빈도가 굉장히 많음을 알 수 있다.

그런데, 본 연구에서 사용되고 있는 고문서 한자 인식시스템은 선형판별함수기반 분류기이고, 선형판별함수에 의한 분류기를 사용할 때 다변량 정규분포 하에서 최적임을 알고 있다. 본 연구에서는 가능하면 특징벡터들의 값을 정규분포에 근사하게 할 수 있다면 인식률을 높일 수 있을 것으로 사려되어 변수변환방법을 고려하게 되었다.

주어진 자료를 정규분포에 근사하게 바꿔주는 변수변환 방법으로 우선 고려할 수 있는 것은 널리 알려진 Box-Cox변환 방법이다.

3.1 Box-Cox 변환방법

이 방법은 Box와 Cox(1964)에 의해 만들어진 변수변환방법이다. 그들은 다음과 같은 멱 변환의 모수족(parametric family)를 고려했다.

모든, $X > 0$ 에 대해서,

$$X^{(\lambda)} = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln X, & \lambda = 0 \end{cases}$$

그리고, $X^{\lambda_0} \sim N(\mu, \sigma^2)$ 라는 가정 하에서 로그 가능도 함수(log-likelihood function)를 원래의

관측값들에 의해 표현하면 다음과 같다.

$$l(\theta | x_n) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n [x_i^{(\lambda)} - \mu]^2 + (\lambda - 1) \sum_{i=1}^n \ln x_i$$

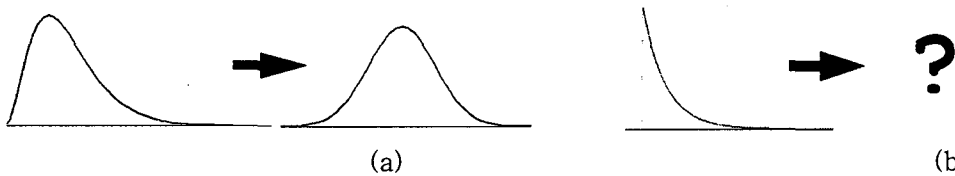
여기서, $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\lambda})$ 에서 $\hat{\mu}$ 와 $\hat{\sigma}^2$ 에 대해서 최대가능도추정량(maximum likelihood estimator)로 각각 $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i^{(\lambda)} = \overline{x^{(\lambda)}}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i^{(\lambda)} - \overline{x^{(\lambda)}})^2$ 대체하고 상수(constant)항을 배제하면,

Box-Cox 변환은 아래의 식을 최대로 해주는 적절한 λ 를 찾는 문제라고 할 수 있다.

$$l(\lambda) = -\frac{n}{2} \left[\frac{1}{n} \sum_{i=1}^n (x_i^{(\lambda)} - \overline{x^{(\lambda)}})^2 \right] + (\lambda - 1) \sum_{i=1}^n \ln x_i$$

3.2 제안하는 변환방법

고문서 한자 이미지로부터 윤곽선 방향 특징추출방법에 의해 추출된 특징벡터들은 마치 지수분포(exponential distribution)처럼 왼쪽 끝에 가장 높은 봉우리를 갖고 오른쪽으로 긴 꼬리를 갖는 모양이므로(<그림5>참조), 근본적으로 정규분포에 유사한 모양을 갖기가 굉장히 어렵다.



<그림6> Box-Cox에 의해 변환 용이한 분포(a)와 변환하기 어려운 형태의 분포(b)

<그림6>에서 보는 바와 같이 (a)형태의 분포에 대해서는 Box-Cox 변환의 수행력이 뛰어나겠지만, (b)형태에 대해서는 잘 작동하지 않을 수도 있다. 본 연구에서는 특징벡터들이 대부분 (b)의 형태와 유사하므로(마치 지수분포), 이와 같은 분포에 대해서, 그리고 궁극적인 목적이 정규분포와 유사하게 만들기 위함이라기보다는 판별력을 높이기 위함이기 때문에 딱 변환의 기준을 왜도를 0으로 하도록 하고자 한다. 본 연구에서는 정규분포의 가능도함수를 최대로 하는 Box-Cox 방법보다는 가능하면 왜도를 0에 가깝게 갈도록 딱 변환을 취하는 방법을 고려한다. 제안하는 변환방법은 윤곽선 방향 특징벡터의 특성을 고려하여 먼저, 로짓 변환(logit transform)을 시행한다 (Agresti(2002) 참조). 윤곽선 방향 특징벡터는 그 특성상 항상 0과 1사이의 실수 값을 갖게 되어 있다. 로짓 변환을 통해, $0 < x < 1$ 의 값을 $-\infty < \ln \frac{X}{1-X} < \infty$ 가 되도록 한 후, Box-Cox 변환처럼 딱 변환의 파라미터(parameter)를 추정한다. 하지만, 그 기준은 Box-Cox 변환과 달리, 가능하면 왜도가 0이 되도록 하는 파라미터를 추정하고자 한다. 제안하는 변환방법을 식으로 표현하면 아래와 같다.

$$X^{(\theta)} = \begin{cases} \left(\ln \frac{X}{1-X} \right)^\theta, & \theta \neq 0 \\ \ln \left(\ln \frac{X}{1-X} \right), & \theta = 0 \end{cases}$$

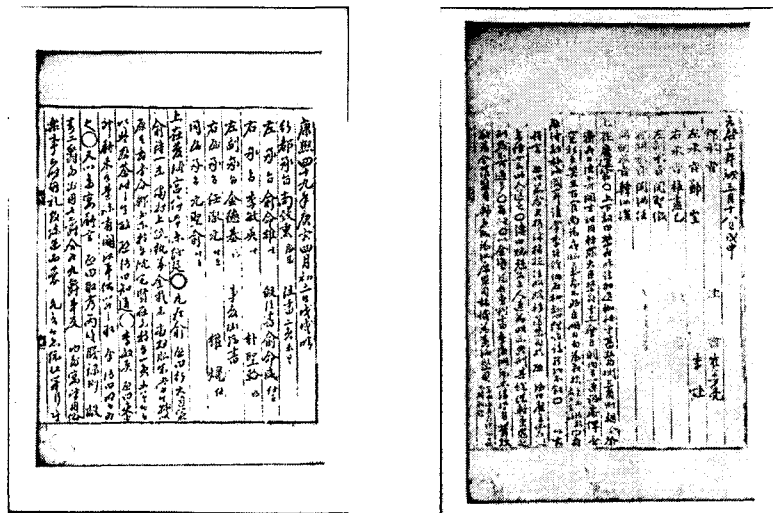
$X^{(\theta)}$ 를 위와 같이 정의할 때 $K(\theta)$ 를 최소로 하는 θ 를 추정한다. 이때 본 연구에서는 추정방법으로 격자탐색(grid search)을 이용하였다.

$$K(\theta) = \left| \frac{\frac{1}{n} \sum_{i=1}^n (X_i^{(\theta)} - \overline{X^{(\theta)}})^3}{S^3} \right|, \quad \text{이 때 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^{(\theta)} - \overline{X^{(\theta)}})^2$$

여기서 특징벡터를 정규분포에 근사하게 만들고자 하는 목적이 선형판별분석 인식기의 성능을 높이기 위함이라는 것을 생각하면, 정규분포의 다른 특성보다도 왜도를 0으로 만들 수 있는 기준이 합리적이라고 여겨진다.

4. 실험결과 및 결론

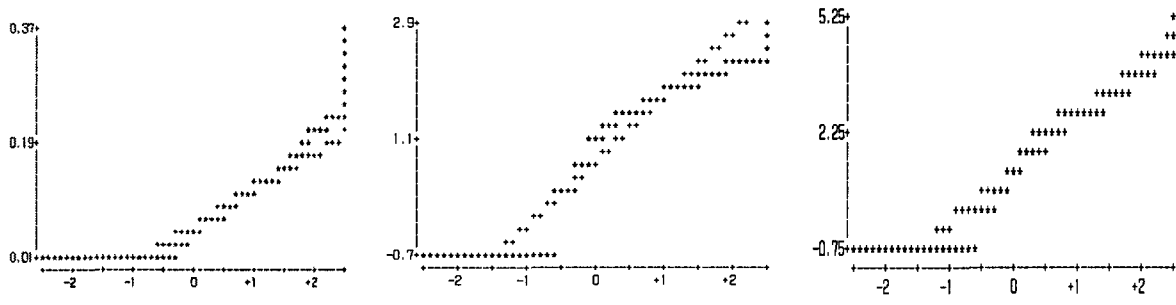
실험에 쓰인 데이터는 한국학 고문서인 승정원 일기 29책의 문서영상을 사용하였다<그림7>. 문서 내에 존재하는 한자의 수는 약 150만 자에 달하며 5,600개 이상의 모형을 가지고 있다. 학습 데이터는 문서 내에서 빈번히 사용되는 글자 1,066개 모형을 사용했으며, 이것은 문서 내 전체 한자 빈도수의 약 90%를 차지하고 있다(참조<그림2>). 학습 데이터 수는 모형 당 100자씩 즉, 총 106,600개의 글자 이미지를 이용하여 학습시켰다. 테스트 데이터 역시, 모형 당 100자씩, 총 106,600개를 사용하였다.



<그림7> 승정원일기의 예

<그림8>의 정규 확률그림에서, Box-Cox 변환 후 특징벡터나 제안된 변환 후 특징벡터 모두, 원래의 특징벡터의 정규 확률그림에 비해 훨씬 정규분포 모양에 근사함을 알 수 있다. 이것은 특징벡터의 값들이 0과 1사이의 실수 값인 점을 고려하여 제안된 변환에서 로짓 변환에 의해 어느 정도 유사 정규성을 갖도록 하는 역할을 수행한 것으로 사려된다. 그런데, Box-Cox 변환 후 특징

벡터의 왜도는 -0.3543 이고, 제안된 변환 후 특징벡터의 왜도는 0.0027이다. 제안된 변환방법은 기준이 왜도의 절대 값이 0에 근사하게 되도록 하는 것이었으므로 사실 당연한 결과이다.



(a) 원래의 특징벡터 (왜도=0.8988) (b) Box-Cox 변환 후 특징벡터 (왜도=-0.3543) (c) 제안된 변환 후 특징벡터 (왜도=0.0027)

<그림8> 각각의 특징벡터들에 대한 정규확률그림

본 연구에서는 고문서 낱자인식을 위해 분류기로서 선형판별분석에 기반한 방법을 사용하고 있다. 고문서 한자분류의 특성상 서포트 벡터머신(Support Vector Machine)이나 신경망 분류기 등 보다도 통계적 함수를 이용한 분류기를 사용하는 것이 보다 간단하다. 물론, 목적에 따라 서포트 벡터머신 등을 사용할 수도 있으나, 고문서에 등장하는 한자 모형(클래스)의 수는 수천 개이며, 현재는 사용하지 않는 한자 모형들도 상당수 포함되어 있어서, 서포트 벡터머신과 같은 이진분류기를 사용하게 되면 수천 개의 분류기를 구성해야할 뿐만 아니라, 수많은 파라미터의 조정이 요구된다. 또한 신경망 분류기, 역시 은닉 층의 개수라든지, 노드의 수 등의 조정해야 할 파라미터가 너무 많다. 실제적으로, 선형판별함수를 이용하는 분류기를 사용하여 비교적 간단하게 구축하고 부족한 인식률은 언어모델의 이용과 같은 후처리를 통해 보완할 수 있다.

<표1> 특징벡터별 선형판별분석의 인식률 결과

특징벡터	인식률(%)
Original CDF	91.11 (baseline)
Box-Cox 변환된 CDF	91.23
Logit 변환된 CDF	91.52
Logit 과 Box-Cox 변환된 CDF	91.45
제안된 변환된 CDF	91.89

<표1>의 실험결과에서, 제안된 방법을 이용하여 변환된 윤곽선 방향 특징벡터를 사용한 선형판별분석의 인식률이 가장 높게 나왔음을 알 수 있다. 표본의 수가 106,600개이므로 표준오차가 약 0.00087 이므로, 변환되지 않는 특징벡터(original CDF)에 대해서는 z-검정에 의해 $\alpha = 0.001$ 에서 유의한 차이가 있음을 알 수 있다. 또한, 나머지 변환된 특징벡터에 대해서도 $\alpha = 0.05$ 에서 유의한 차이가 있음을 알 수 있다. 물론, 이것은 승정원 일기라는 고문서에 대한 실험에 한정하여

나온 결과이므로, 다른 한자나 한글에 대해서도 유의한 차이가 있을 거라고 말할 수는 없을지 모른다. 그러나, 본 연구로부터 특징추출방법으로 얻어진 특징벡터를 다양한 변환 등을 통해 인식기 성능을 높일 수 있는 방법의 연구개발이 필요할 것으로 사려 된다. 또한, 요즘의 오프라인 문자인식에서는 예전처럼 전처리, 문자분할, 특징추출, 분류 등이 별도의 작업으로 여겨지지 않고 결합되는 형태를 많이 취한다. 심지어 이미지를 분할하는 작업에서도 분류기의 정보와 문맥정보를 이용하기도 한다. 이와 같은 맥락에서 본 연구는 추출된 특징벡터로부터 로짓 변환과 떡 변환을 이용하여 분류기의 판별 수행력을 높이도록 하는 것이었다고 할 수 있다.

참고문헌

- [1] Agresti, A.(1996). An Introduction to Categorical Data Analysis. Wiley & Sons, New York.
- [2] Box, G. E. P. and Cox, D. R.(1964). An Analysis of Transformations. *Journal of the Royal Statistical Society(B)*, Vol.26, No. 2, 211-252.
- [3] Kato, N., Suzuki, M., Omachi, S., Aso, H., and Nemoto, Y.(1999). A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 3, 258-262.
- [4] Kim, M. S., Cho, K. T., Kwag, H. K. and Kim, J. H.(2004). Segmentation of Handwritten Characters for Digitalizing Korean Historical Documents. *Lecture Notes in Computer Science*, Vol. 3163, 114-124.
- [5] Kim, M. S., Jang, M. D., Choi, H. I., Rhee, T. H. and Kim, J. H.(2004). Digitalizing Scheme of Handwritten Hanja Historical Documents. *In Proceedings of Document Image Analysis for Libraries(DIAL2004)*, Palo Alto, California, 321-327.
- [6] Kimura, F. and Shridhar, M.(1991). Handwritten Numerical Recognition based on Multiple Algorithm, *Pattern Recognition*, Vol.24, No.10, 969-983.
- [7] Kimura, F., Wakabayashi, T., Tsuruoka, S. and Miyake, Y.(1997). Improvement of handwritten Japanese character recognition using weighted direction code histogram. *Pattern Recognition*, Vol. 30, No. 8, 1329-1337.
- [8] Lee, S. W. and Park, J. S.(1994). Nonlinear Shape Normalization Methods for the Recognition of Large-set Handwritten Characters, *Pattern Recognition*, Vol.27, No. 7, 895-902.
- [9] Tsukumo, J. and Tanaka, H.(1988). Classification of Handprinted Chinese Characters using Nonlinear Normalization Methods. *In proceedings of 9th International Conference on Pattern Recognition*, 168-171.
- [10] Yamamoto, K. and Rosenfeld, A.(1982). Recognition of handwritten Kanji character by a relaxation method. *In Proceedings of the 6th International Conference on Pattern Recognition*, 395-398.