

Kirsch Feature의 압축크기에 따른 인쇄체 숫자 인식에서의 성능 비교

김성우*, 최선아*, 차의영*
*부산대학교 신경회로망 및 실세계 응용 연구실

Performance Comparison of Various Kirsch Feature for Printed Numeral Recognition

Seong-Woo Kim*, Sun-Ah Choi*, Eui-Young Cha*
*Pusan Nat'l University Neural Network and Real World Applications Lab
E-mail : swkim, kiddy, eycha@harmony.cs.pusan.ac.kr

요 약

OCR 시스템에서 feature는 인식성능에 상당히 중요한 역할을 한다. gradient feature는 현재까지 개발되어진 여러 가지 feature들 중에서 폭넓게 사용되고 있는 것 중의 하나이다. 본 논문에서는 변형이 심한 인쇄체 숫자를 실험대상으로 하고, Kirsch mask를 이용한 방향성을 가지는 edge를 추출하여 신경망의 입력벡터로 사용할 때 압축의 크기에 따른 인식성능의 차이를 비교하고, 최적의 벡터크기를 제안한다.

1. 서론

Optical Character Recognition 시스템의 주요한 목표는 높은 인식률을 유지하면서 오인식을 최소화하는 것이다. OCR 시스템의 성능을 향상시키기 위해서 크게 세 가지 분야의 연구가 수행되어져 왔다. 첫째, 분류기(classifier)에 관한 연구이다. backpropagation, self-organizing map과 같은 신경회로망에 관한 연구와 Hidden Markov Model, Support Vector Machine와 같은 통계적 학습이론에 기반한 분류기에 관한 연구로 구분할 수 있다. 둘째, 복합분류기(multiple classifier)에 관한 연구이다. 이것은 단일 분류기의 인식 결과보다 복합분류기의 인식결과가 우수하다는 사실에 기반하여, 다수의 분류기를 조합하는데 관련된 연구이다. 마지막으로 각 클래스의 유일성을 나타내는 feature의 개발과 상호보완성을 가지는 feature의 조합에 관한 연구이다. 실제로, OCR 시스템의 성능은 사용되는 feature에 상당히 많이 좌우한다고 할 수 있다 [1].

본 논문은 feature에 관한 연구로서, Kirsch mask에 의하여 방향성을 갖는 edge 영상의 압축크기에 따른 변형이 심한 인쇄체 숫자 인식의 성능을 비교하고 실험을 통한 최적의 크기를 선정한다.

본 논문의 구성은 다음과 같다.

2장에서 특징 추출에 관하여 설명하고, 3장에서 본 논문에서 사용한 분류기인 신경회로망에 대해서 설명 하겠으며, 4장에서는 크기에 따른 인식률 비교에 대한 실험환경에 대해서, 5장에서는 실험결과를 분석하고, 6장에서는 결론 및 향후 연구과제를 제시한다.

2. 특징 추출

1) 배경

필기체 숫자와 인쇄체 숫자는 모두 획의 구성으로 이루어져 있다. 획이란 1차원 구조를 2차원 평면에 매핑해 놓은 것이다. 따라서, 특정 위치에 특정 방향의 획의 존재 유무가 각 숫자를 인식하는데 중요한 특징이 된다.

그러나, 이러한 획의 존재 유무뿐만 아니라 계산의 신속함과 하드웨어의 한계 등을 극복하기 위해서, 1차 미분으로 edge를 검출한 영상을 이용하여 특정 위치에서의 특정 방향의 획을 감지한다.

1차 미분을 이용한 edge 검출기들 중에서 8 이웃을 모두 고려하여 보다 정확한 edge의 검출이 가능한 Kirsch edge 검출기를 사용한다[2].

2) Kirsch Feature Vector

Kirsch mask를 이용한 edge 검출의 알고리즘은 다음과 같다.

A_0	A_1	A_2
A_7	(i,j)	A_3
A_6	A_5	A_4

$$G(i,j) = \max \left\{ 1, \max_{k=0}^7 [|5S_k - 3T_k|] \right\}$$

$$S_k = A_k + A_{k+1} + A_{k+2}$$

$$T_k = A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7}$$

$G(i,j)$ 는 픽셀 (i,j) 에서의 기울기값이고, A의 첨자는 modulo 8 연산을 구한 값이다[3].

수평(H), 수직(V), 오른쪽 대각선(R), 왼쪽 대각선(L) 방향성을 가지는 에지성분을 추출하기 위하여 다음 식을 적용한다[2].

$$G(i,j)_H = \max(|5S_0 - 3T_0|, |5S_4 - 3T_4|)$$

$$G(i,j)_V = \max(|5S_2 - 3T_2|, |5S_6 - 3T_6|)$$

$$G(i,j)_R = \max(|5S_1 - 3T_1|, |5S_5 - 3T_5|)$$

$$G(i,j)_L = \max(|5S_3 - 3T_3|, |5S_7 - 3T_7|)$$

이렇게 하여 그림 1과 같이 4 방향의 Kirsch feature를 추출한다.



[그림 1] Kirsch mask를 이용한 4 방향 추출

3) Mesh feature를 이용한 Kirsch feature 압축

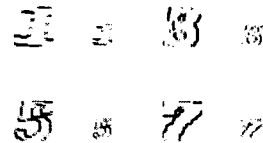
인식기의 성능을 최대화하기 위하여 feature 내부의 중요한 정보의 손실이 없이, 의미가 없는 정보의 최소화를 위하여 압축과정이 필요하다.

일반적으로 Mesh feature는 구현의 간단함과 뛰어난 성능으로 문자 인식에 흔히 사용되는 특징이다. 또한 이러한 Mesh feature는 다른 특징 추출 방법에 비해 영상 내의 관심 물체의 픽셀 개수, 즉 점밀도를 이용한 비선형 형태 정규화 방법에 더 적합하다[4].

크기 정규화된 영상을 $n \times n$ 부분 영역으로 분할한 다음, 각 부분 영역의 물체의 픽셀 개수를 계산하여 $n \times n$ 개의 feature vector를 추출한다.

이렇게 Kirsch feature에서 Mesh feature를 추출함으로써 vector의 크기를 압축할 수 있다.

그림 2는 32×32 의 Kirsch feature를 추출한 영상과 이 영상 크기의 $\frac{1}{4}$ 인 16×16 압축영상을 보여준다.



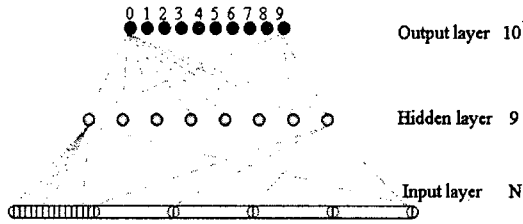
[그림 2] Mesh feature를 이용한 Kirsch feature 압축

3. 분류기

Kirsch feature는 입력 패턴의 모양에 기초한 structural feature가 아니고, 입력 패턴의 변형에 기초한 numerical feature이다. 그리고, numerical feature는 주로 신경회로망에 의해서 분류되어 입력패턴을 인식하는 것이 일반적이다[1].

따라서, 본 논문에서는 기존의 문자인식에 많이 사용되었던 Multi-Layer Perceptron 네트워크 구조에서 Backpropagation 알고리즘을 이용하여 학습한다.

네트워크의 입력층은 다양한 크기로 압축된 4 방향의 Kirsch feature vector와 전역적인 특성을 고려하기 위하여 다양한 크기로 압축된 원영상의 Mesh feature vector를 사용하고[2], 은닉층은 9개의 노드, 출력층은 10개의 노드로 구성되어 있고 각 노드는 각 숫자를 나타내는 클래스이다.

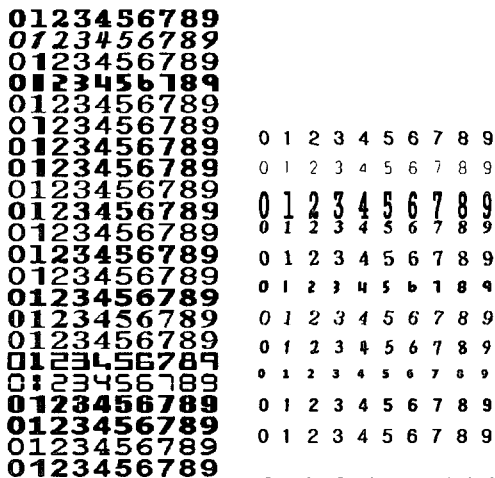


[그림 3] 네트워크 구조

그림 3에서 알 수 있듯이, 각 층의 노드는 다음 층의 노드와 fully connected 되어 있고, Input layer의 노드 개수는 Kirsch feature의 압축 크기와 입력벡터의 조합에 따라 가변적이다.

4. 실험 환경

본 논문의 실험에 사용된 영상은 22개의 1)한글 영문 조합 글꼴로 총 220개의 이진 숫자 영상을 학습데이터로 하였으며, 입력 데이터의 다양한 입력 변형을 포함할 수 있도록 학습 데이터에 대하여 다양하게 2)변형한 총 4840개의 변형 데이터에 대한 테스트를 수행한다.



[그림 4] 학습 데이터

[그림 5] 테스트 데이터

또한 다양한 종류의 글꼴에 대한 강인하고 신뢰도 높은 OCR 시스템의 구현을 위하여 학습은 단지 기본 220개의 데이터로만 수행하고, 테스트는 변형된 4840개의 데이터로 실험한다.

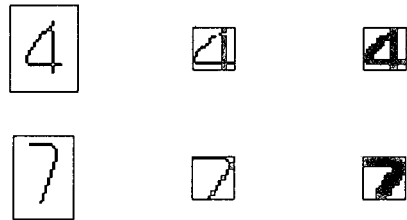
1) 실험데이터의 전처리

숫자 영상 데이터로 입력패턴을 생성하기 위한 전처리 과정으로서 다양한 크기의 입력 이미지에 대한 정규화는 문자 영상에 대한 효율적인 패턴벡터를 생성하는데 중요한 역할을 한다.

원영상은 28×33 크기이고, 영상의 중심에 각 숫자들이 놓여져 있다.

우선, 원영상에서 숫자의 외점사각영역을 분리하여, 그 부분 영역을 16×16으로 크기정규화를 수행한다.

그리고, 크기정규화된 영상에 대하여 모폴로지 연산으로 숫자 획의 굵기를 일정하게 조정된 획 정규화 방법에 의하여 숫자 영상을 개선한다[5].



(a) 원영상 (b) 크기정규화 (c) 획정규화

[그림 6] 전처리 과정

2) 입력벡터 생성

전처리가 완료된 영상을 이용하여 4 방향의 Kirsch feature를 추출한 후, 압축의 과정에서 Mesh feature를 추출하게 되는데, 여기서 16×16의 원영상을 $n \times n$ 의 부분영역으로 Mesh feature를 추출한다고 한다면, vector의 범위는 $0 \sim \frac{16}{n}$ 이 되므로, 이것을 부동소수점 0~1의 범위로 scaling하여 입력벡터로 사용한다.

- 1) 굴림, 궁서, 돋움, 명조, 바탕, 산세리프, 샘물, 시스템, 신명조, 안상수, 중고딕, 필기, 휴먼옛체, Arial, Courier, Fixdays, Impact, LED, Orbit-B, Symbol, Times New Roman, Verdana
- 2) 위치이동(4), 경사왜곡(4), 세션화(1), 중간값 필터링(1), 좌우 5°, 10° 회전(4), 모폴로지 확장, 가로확대축소(2), 세로확대축소(2), 가로세로확대축소(2) : 22종 * 12(21가지 변형+원형) * 10 = 4840개

5. 실험 결과 및 분석

문자인식에 있어서 Kirsch feature의 우수성을 검증하기 위하여 첫째, 전처리 결과 영상의 픽셀값을 입력벡터로 구성하여 테스트하고, 둘째, 전처리 결과 영상에서 4×4, 8×8 크기의 Mesh feature를 입력벡터로 하여 테스트하고, 셋째, 4 방향 Kirsch feature를 추출

[표 1] 입력벡터에 따른 인식률

입력벡터	압축크기	220개 테스트		4840개 테스트		Rank
		오인식 개수	인식률	오인식 개수	인식률	
G	8×8	0/220	100%	402/4840	91.69%	6
G	16×16	0/220	100%	402/4840	91.69%	6
M	4×4	1/220	99.55%	419/4840	91.34%	8
M	8×8	0/220	100%	213/4840	95.60%	3
HVRL	4×4×4	0/220	100%	241/4840	95.02%	4
HVRL	4×8×8	1/220	99.55%	314/4840	93.51%	5
HVRLG	5×4×4	0/220	100%	92/4840	98.10%	1
HVRLG	5×8×8	0/220	100%	140/4840	97.11%	2

G : normalized Images

M : Mesh feature vector for normalized Images

HVRL : Mesh feature vector of Kirsch feature for normalized Images

6. 결론 및 향후 과제

한 영상을 4×4, 8×8 크기의 Mesh feature로 입력벡터를 구성하여 테스트하고, 마지막으로 두 번째 입력벡터와 세 번째 입력벡터를 조합하여 지역적 정보와 전역적 정보를 동시에 포함하는 입력벡터를 구성하여 테스트한다.

표 1의 실험 결과를 살펴보면, 첫째, 입력벡터의 조합 측면에서 분석을 할 때, 지역적인 정보로만 구성된 HVRL 입력벡터가 전역적인 정보로만 구성된 G 입력벡터나, M 입력벡터보다 성능이 우수하다는 것을 알 수 있다. 그리고, 전역 정보와 지역 정보를 모두 포함하고 있는 HVRLG 입력벡터가 HVRL 입력벡터보다 훨씬 우수한 성능을 보인다.

둘째, 입력벡터의 압축크기 측면에서 분석을 하면, HVRL과 HVRLG 입력벡터의 경우에 16×16 영상을 원래 크기의 1/4인 8×8로 압축하는 것보다, 1/16인 4×4로 압축할 때 더 우수한 성능을 보인다는 것을 알 수 있다.

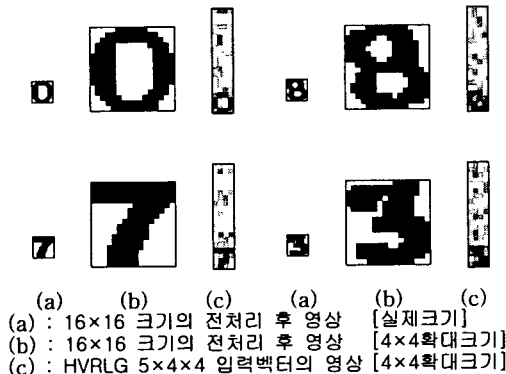
본 논문에서는 4 방향 Kirsch feature의 Mesh feature를 이용한 압축 정도와 입력벡터의 조합에 대한 성능을 인체체 숫자 인식에 적용하여 비교하였다.

가장 우수한 성능을 나타내는 입력벡터의 조합은 HVRLG로서 지역적 정보와 전역적 정보를 모두 포함하는 입력벡터이고, 16×16 영상을 4×4로 압축했을 때가 최적의 압축 크기인 것을 실험을 통하여 확인하였다.

향후 연구과제로는 HVRLG 5×4×4 입력벡터를 효과적으로 인식할 수 있는 여러 가지 분류기의 성능을 비교 분석하는 것이다.

[참고문헌]

- [1] Il-Seok Oh, Ching Y. Suen, "Distance features for neural network-based recognition of handwritten characters", IJDAR, Vol 1, pp. 73-88, 1998
- [2] Seong-Whan Lee, "Off-Line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network", IEEE transactions of Pattern Analysis and Machine Intelligence, Vol. 18, No. 6, pp. 648-652, 1996
- [3] W. K. Pratt, Digital Image Processing, New York : Wiley, 1978
- [4] 김두식, 이성환, "계층적 신경망 분류기를 이용한 다양한 언어, 활자체 및 크기의 대용량 문자 인식", 한국정보과학회논문지, 제 25권, 제 5호, pp. 792-801, 1998
- [5] 김도현, 강동구, 강민경, 차의영, "문자 인식을 위한 효율적인 획 정규화", 한국정보처리학회 추계학술대회, 상권, 2001



[그림 7] HVRLG 5×4×4 입력벡터의 영상화