

# 웨이브렛 특징 추출을 이용한 숫자인식의 최적화

## Optimization Numeral Recognition Using Wavelet Feature Based Neural Network.

황성욱, 임인빈, 박태윤, 최재호  
전북대학교 컴퓨터공학과 미디어통신실(coolwind76@orgio.net)

Sung-Wook Hwang, In-Bin Yim, Tae-Yoon Park, Jae-Ho Choi

### 요약

본 논문에서는, 웨이브렛 변환과 잡음 섞인 숫자 영상에 대한 최적화 인식 훈련기법을 사용한 다계층 신경망을 제안하고, 이 시스템을 아라비아숫자 인식에 적용한다. 웨이브렛 변환을 이용해 원 영상 정보의 중요한 부분은 최대한 보존하면서 입력벡터의 크기를 줄임으로써 신경망의 노드 수와 학습 수렴시간이 줄어들도록 하였고, 최적화 인식 훈련기법은 데이터의 잡음을 점차적으로 높여가면서 훈련벡터에 적용, 인식률의 변화에 대해 살펴보았다.

잡음이 섞인 숫자 영상의 인식율을 높이기 위해 원 영상에 0, 10, 20, 30, 40, 50dB의 잡음을 섞은 영상을 훈련에 함께 사용하였다. 테스트 영상에 잡음이 30~50dB정도 섞였을 경우에는 원 영상만을 훈련에 이용했을 때와 잡음이 섞인 영상을 이용하여 훈련시켰을 경우에 인식율의 차이가 별로 없지만, 0~20dB정도 섞인 영상을 테스트에 사용할 때에는 0, 10, 20, 30, 40, 50dB의 잡음이 있는 영상을 훈련에 사용했을 때가 원 영상만을 훈련에 이용했을 경우에 비해 인식율이 9% 향상된다.

### Abstract

In this Paper, propose for MLP(multilayer perceptron) neural network that uses optimization recognition training scheme for the wavelet transform and the numeral image add to noise, and apply this system in Numeral Recognition. As important part of original image information preserves maximum using the wavelet transform, node number of neural network and the learning convergence time did size of input vector so that decrease. Apply in training vector, examine about change of the recognition rate as optimization recognition training scheme raises noise of data gradually.

We used original image and original image added 0,10,20,30,40,50dB noise for the increase of numeral recognition rate. In case of test image added 30~50dB, numeral recognition rate between the original image and image added noise for training is a little. But, in case of test image added 0~20dB noise, the image added 0, 10, 20, 30, 40, 50dB noise is used training. Then numeral recognition rate improved 9 percent.

**Key words** : Neural Network, Numeric Recognition, Wavelet, Noise Image, PSNR

## 1. 서론

컴퓨터를 이용해 숫자를 인식하는 문제는 60년 동안 중요한 토픽이었다. 그럼에도 불구하고 컴퓨터를 이용해 숫자를 인식하는 데에는 많은 문제점이 존재한다. 통신이 발달한 요즘 팩스, 이메일, 핸드폰 등 다양한 통신장비를 이용한 문서 전송이 이뤄지고 있다. 그 중 팩스를 이용해 문서를 전송할 경우 원 문서에 많은 잡음이 섞여서 전송되는 특징을 보인다. 이렇게 잡음이 많이 섞인 경우의 인식률은 급격히 감소하게 된다. 본 논문에서는

원 영상과 원 영상에 20, 30, 40, 50 dB의 잡음이 섞인 영상을 신경망의 훈련에 입력 값으로 넣어줌 이러한 인식률 감소를 줄일 수 있도록 하였다.

또한 특징추출 방법으로는 최근 들어 많이 사용되고 있는 웨이브렛 변환 방법을 사용하였다. 웨이브렛 변환을 하게 되면 저대역의 계수에는 문자의 기본적인 모양 정보가, 고대역 부분의 계수에는 나머지 경계부분(수평, 수직, 대각 방향)의 정보가 실리게 된다.[1] 이러한 계수들을 이용하여 인식시스템에서 사용할 특징벡터 값을 만드는데, 신경망의 학습시 계산량은 각 층의 노드 숫자

가 증가할 수 록, 급격한 증가를 보이기 때문에 특징벡터의 크기를 줄이기 위해 두 번의 웨이브렛 변환 과정 중 중요한 정보가 실린 부분을 선택하였다.

## II. 특징 추출

### 1. Haar 웨이브렛

숫자 영상은 본질적으로 선으로 그려진 형태를 띤다.[2] 따라서 선 마디의 부분적인 검출과 선구조의 전체적인 검출을 하는 것이 숫자 영상의 경우에 알맞은 특징 추출방법이라 할 수 있다.

웨이브렛은 팽창과 이동에 의해 하나의 기본 함수  $W(x)$ 를 만들 수 있다. Haar 웨이브렛은 팽창함수(*scaling function*)  $\phi(x)$ 로부터 만들어 지는데, 그 표현은 식 (1)과 같이 정의된다.

$$W_H = c_1\phi(2x) - c_0\phi(2x-1) \quad (1)$$

$$= \phi(2x) - \phi(2x-1)$$

$$= \begin{cases} 1 & \text{if } 0 < x < 1/2 \\ -1 & \text{if } 1/2 < x < 1 \\ 0 & \text{if otherwise} \end{cases}$$

여기에서  $\phi(x)$ 는 상자함수(*box function*)로 식 (2)와 같이 정의된다.

$$\phi(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{if otherwise} \end{cases} \quad (2)$$

Haar 웨이브렛 변환을 한 번 하게 되면 그림 1처럼 4개의 웨이브렛 계수로 분해된다.

### 2. 특징벡터

Harr 웨이브렛 변환은 빠른 계산 속도로 선 마디의 부분적 검출과 선구조의 전체적인 검출에 적합하다.[3] 따라서 본 논문에서는 숫자 영상에서 특징추출을 위해 Harr 웨이브렛 변환을 사용한다.

16×16 화소로 정규화 된 숫자 영상에 웨이브렛 변환을 한 번 적용하면 8×8 화소를 가진 네 개의 영상  $\{A^{j-1}, B^{j-1}, C^{j-1}, D^{j-1}\}$ 가 만들어진다. 따라서 벡터의 크기는 4×8×8 화소의 크기를 갖는다. 저대역(*lowpass*)영역  $\{A^{j-1}\}$ 에 영상의 중요한 정보가 많이 실리고, 그 외의  $\{B^{j-1}, C^{j-1}, D^{j-1}\}$ 영역에는 영상의 수직, 수평, 대각선 정보가 실린다. 웨이브렛 변환의 이러한 특징을 근거로 효율적인 특징 추출을 하기 위하여 4가지의 특징추출 방법을 적용, 비교해 보았다.

각각의 방법에 대해 살펴보면,

- 1) 분할된 4개의 영상  $\{A^{j-1}, B^{j-1}, C^{j-1}, D^{j-1}\}$ 을 모두 입력벡터로 결정하는 경우와 숫자 영상을 웨이브렛 변환

할 경우,

- 2) 숫자영상의 가장자리 부분에 0이 몰리는 경향을 이용하여 네 개의 영상 중  $\{A^{j-1}\}$ 만 그대로 특징벡터로 이용하고,  $\{B^{j-1}, C^{j-1}\}$ 는 오른쪽 두 열을 제외한 나머지 6열의 값만을,  $\{D^{j-1}\}$ 영상은 양쪽 가장자리 1열씩을 제외한 가운데 6열의 값만을 특징벡터로 이용하는 경우,
- 3)  $\{A^{j-1}, B^{j-1}, C^{j-1}\}$ 영상의 값과  $\{D^{j-1}\}$ 영상을 한 번 더 웨이브렛 변환을 거친 후  $\{K^{j-2}, L^{j-2}, M^{j-2}\}$ 부분을 입력벡터로 결정 한 경우,
- 4) 마지막으로  $\{A^{j-1}\}$ 부분과  $\{E^{j-2}, F^{j-2}, G^{j-2}\}$ ,  $\{H^{j-2}, I^{j-2}, J^{j-2}\}$ ,  $\{K^{j-2}, L^{j-2}, M^{j-2}\}$ 부분을 입력 벡터로 결정 하는 경우이다.

이러한 시도는 영상의 특징은 최대한 보존하면서 입력 벡터의 크기를 줄임으로써, 신경망의 학습 수렴시간을 줄이기 위한 것이다.

또한 신경망의 인식능력을 증가시키기 위해 특징벡터의 계수 값들을 [0, 1] 사이로 정량화 한다.[1]

### 3. 잡음 첨가

숫자 영상 화질의 평가 척도로 PSNR(Peak Signal-to-Noise Ratio)을 사용하였다.[4] 본 논문에서 사용한 영상은 16×16크기의 이진데이터를 가지므로 PSNR은 식 3와 같다.

$$PSNR = 10 \log_{10} \frac{1}{e} \quad (3)$$

여기에서  $e$ 는 MSE(mean squared error)로서, 원 영상에 대한 비교 영상의 오류 파워에 해당한다.

$$e = \frac{1}{N_1 N_2} \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} |L_o(i, j) - L_c(i, j)|^2 \quad (4)$$

여기에서  $N_1, N_2$ 는 각각 영상 파일의 수평, 수직방향 화소 수이다.

본 논문에서는 원 숫자 영상에서 추출된 특징벡터 값과 원 숫자 영상에 0, 10, 20, 30, 40, 50dB의 잡음을 섞은 영상으로부터 추출된 특징 벡터 값을 모두 훈련벡터로 사용하였을 경우 인식률의 변화에 대해 살펴보았다.

### 4. 분류기

분류기는 3개의 층으로 이뤄진 다계층 신경망의 구조 [5]를 가진다. 입력층은 특징벡터의 크기인 208개의 노드로 구성되고 출력층은 숫자 영상 0~9에 대응되는 노드 10개로 구성된다. 은닉층의 노드 수는 4개로 구성된다. 각 층의 노드는 다음 층의 노드와 전체연결 된다. 연결초기값은 -0.01~0.01 사이에 분포하고 평균값이 0을 갖도록 하고, 역전파학습 알고리즘[6]을 이용하여 학습을 진행한다. 다계층 신경망의 구조는 그림 1과 같다.

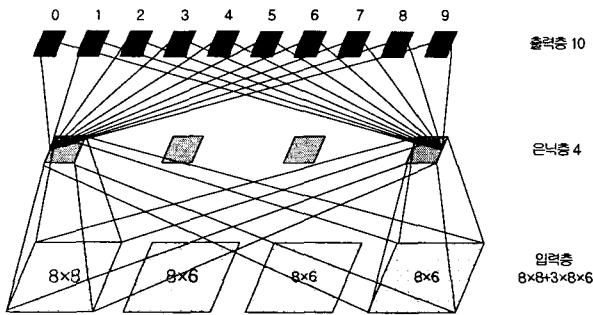


그림 1. 다계층 신경망의 구조

III. 실험결과 및 비교

실험에 쓰인 숫자는 워드프로세서에 있는 12개의 글씨체, 0~9까지 총 120가지를 사용했다. 효율적인 특징 벡터를 결정하기 위해 4가지 방법을 이용하여 실험하였다.

- 1) 실험 A는 한번의 웨이브렛 변환 후 분할된 4개의 영상  $\{A^{j-1}, B^{j-1}, C^{j-1}, D^{j-1}\}$  모두 다 입력벡터로 결정, 총  $4 \times 8 \times 8$  크기의 영상을 이용하고,
- 2) 실험 B는 숫자영상의 가장자리에 0값이 몰린다는 특징을 이용하여  $\{A^{j-1}\}$ 는 그대로 이용하고  $\{B^{j-1}, C^{j-1}\}$ 는 오른쪽 두 열을 제외한 나머지 6열의 값만을,  $\{D^{j-1}\}$  영상은 양쪽 가장자리 1열씩을 제외한 가운데 6열의 값을 특징벡터로 이용하는 경우, 총  $8 \times 8 + 3 \times 8 \times 6$  크기의 영상을 이용하고,
- 3) 실험 C는  $\{A^{j-1}, B^{j-1}, C^{j-1}\}$  부분과,  $\{D^{j-1}\}$ 만 웨이브렛 변환을 한 번 더 거친 후  $\{K^{j-2}, L^{j-2}, M^{j-2}\}$  부분을 입력 벡터로 결정,  $3 \times 8 \times 8 + 3 \times 4 \times 4$  크기의 영상을 이용하고,
- 4) 실험 D는  $\{A^{j-1}\}$  부분과  $\{E^{j-2}, F^{j-2}, G^{j-2}\}$ ,  $\{H^{j-2}, I^{j-2}, J^{j-2}\}$ ,  $\{K^{j-2}, L^{j-2}, M^{j-2}\}$  부분을 입력 벡터로 결정,  $8 \times 8 + 9 \times 4 \times 4$  크기의 영상을 입력벡터로 사용한 경우이다.

평균	실험A	실험B	실험C	실험D
인식율(%)	98.33	98.33	93.33	92.5
오인식율(%)	0.83	0.83	4.16	5.00
거부율(%)	0.83	0.83	2.50	2.50
훈련시간(초)	937.685	937.685	918.320	601.525

-시뮬레이션 시스템: P4-1.5, 256MB  
표 1. 입력벡터별 인식율 및 훈련시간 비교

표 1처럼 실험 A와 B는 같은 인식율을 보이고, 실험 C, D, E는 상대적으로 인식율이 낮은 것을 볼 수 있다. 따라서 본 논문에서는 가장 좋은 인식율과 상대적으로 짧은 훈련시간을 갖는 실험 B에서 사용한 특징벡터로 결정하였다.

잡음이 섞인 숫자 영상을 신경망의 훈련에 사용하기 위해 각각의  $16 \times 16$  크기의 이미지에 20, 30, 40, 50dB의

잡음을 더했다. 잡음의 정도에 따른 최적의 인식률을 찾기 위해 다섯 가지의 실험을 실시하였다. 실험 1에서는 원 숫자 영상만을 훈련에 사용한 것이고, 실험 2는 원 숫자 영상과 원 영상에 50dB의 잡음이 섞인 영상을, 실험 3은 실험 2에 사용한 훈련 데이터에 40dB의 잡음이 섞인 영상을, 실험 4는 30dB의 잡음이 섞인 영상을, 실험 5는 20dB의 잡음이 섞인 영상을, 실험 6은 10dB의 잡음이 섞인 영상을, 실험 7은 0dB의 잡음이 섞인 영상을 훈련에 사용한 것이다. 테스트에 사용한 숫자는 훈련에 사용된 원 영상 120가지에 0, 10, 20, 30, 40, 50dB의 잡음을 누적시킨 3,600가지를 사용하였다.

평균	실험1	실험2	실험3	실험4	실험5	실험6	실험7
인식율(%)	80.59	74.4	80.95	77.14	79.88	80.71	89.52
오인식율(%)	12.73	17.5	10.83	12.97	9.88	8.21	0.83
거부율(%)	6.66	8.09	8.21	9.88	10.23	11.07	9.64

표 2. 평균 인식율

표 2에 보여 지듯 실험 7의 경우, 즉 원 영상과 50dB와 0, 10, 20, 30, 40, 50dB의 잡음이 섞인 숫자 영상을 훈련으로 사용했을 때 인식율이 가장 뛰어나게 나타난다. 그림 2는 각 실험별 잡음 영상에 대한 인식율을 나타낸다. 각 실험에서 30dB~50dB 사이의 잡음이 섞인 영상에 대해서는 인식율이 90% 이상을 보이지만, PSNR이 20dB이하 부터는 인식율이 급격히 감소함을 볼 수 있다. 각각의 실험에서의 잡음별 인식율은 그림 3~9과 같다

IV. 결론

본 논문에서는 잡음 섞인 숫자 영상의 인식율을 높이기 위해 신경망의 훈련에 잡음 섞인 영상을 입력 데이터로 결정하는 것을 제안하였다. 표 2에서처럼 0, 10, 20, 30, 40, 50dB의 잡음이 섞인 영상을 훈련데이터로 결정 한 경우가 가장 좋은 인식율을 보였다. 이는 훈련 데이터에 0dB의 잡음이 섞인 영상을 넣어 줌으로써 0dB의 잡음이 섞인 영상을 인식하는데 나은 성능을 보이기 때문이다.

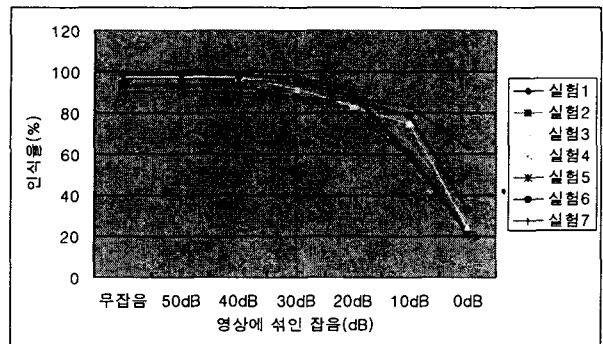


그림 2. 실험별 잡음 영상에 대한 인식율

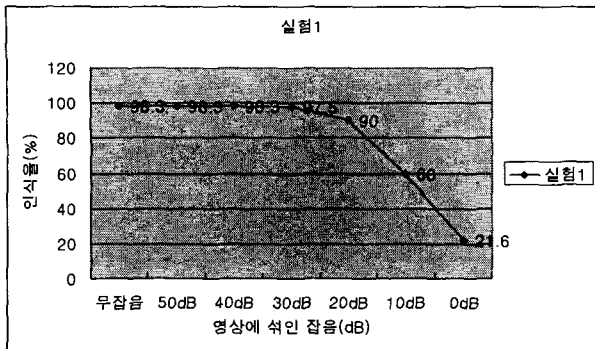


그림 3. 실험1의 경우 잡음 영상에 대한 인식률

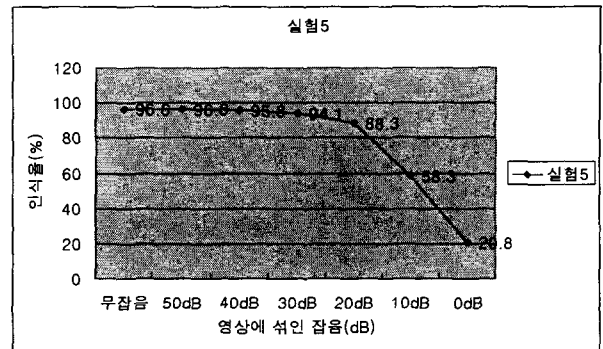


그림 7. 실험5의 경우 잡음 영상에 대한 인식률

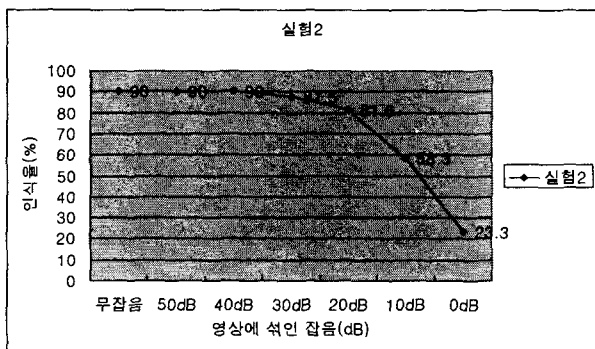


그림 4. 실험2의 경우 잡음 영상에 대한 인식률

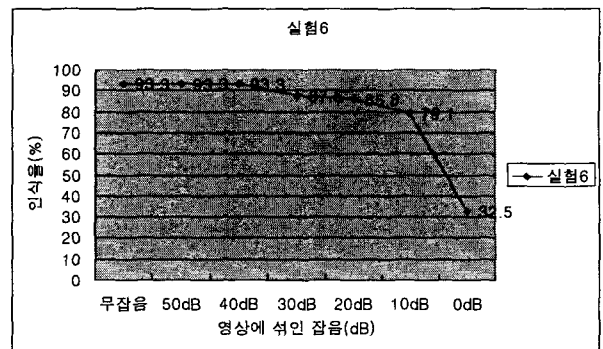


그림 8. 실험6의 경우 잡음 영상에 대한 인식률

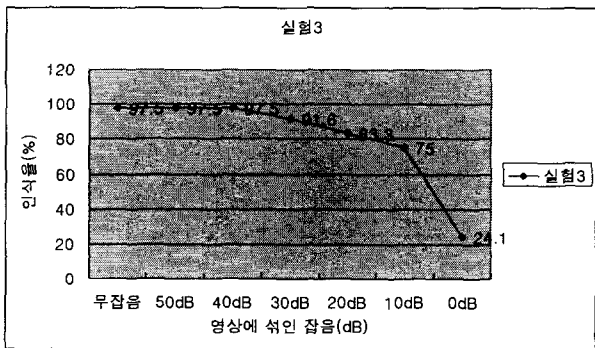


그림 5. 실험3의 경우 잡음 영상에 대한 인식률

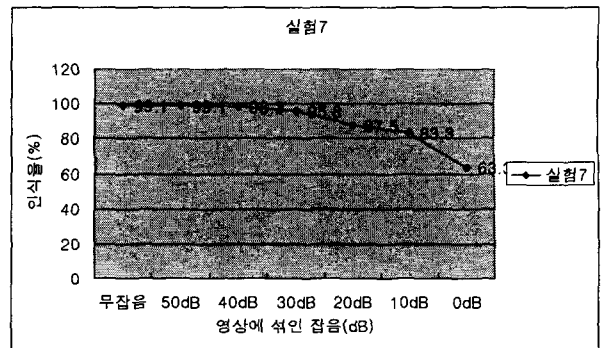


그림 9. 실험7의 경우 잡음 영상에 대한 인식률

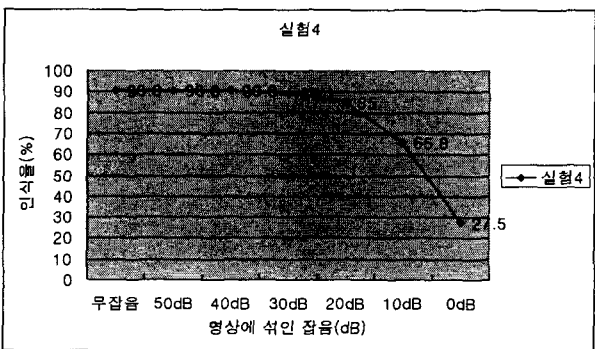


그림 6. 실험4의 경우 잡음 영상에 대한 인식률

### V. 참고문헌

- [1] Suzete E. N. Correia, João M. de Cavalho, "Optimizing the Recognition Rates of Unconstrained Handwritten Numerals Using Biorthogonal Spline Wavelets," IEEE conference, Pattern Recognition, Vol.2, pp 251254, 2000
- [2] Seong-Whan Lee, Young-Joon Kim, "Multiresolution Recognition of Handwritten Numerals with Wavelet Transform and Multilayer Cluster Neural Network," IEEE conference, Document Analysis and Recognition, Vol.2, pp1010-1013, 1995
- [3] Y.Y Tang, L.H Yang, *World Scientific*, "Wavelet Theory and Its Application to Pattern Recognition", 2000
- [4] 정미숙, "효율적인 웨이브렛 팩킷 알고리즘을 이용한 영상부호화에 관한 연구", 全北大學校, 1984
- [5] 김희승, 생능출판사, 영상인식 -영상처리, 컴퓨터비전, 패턴인식, 신경망- pp355-360, 1998
- [6] 김문현, 생능출판사, "인공지능", pp245-254, 2001