

회전 및 이동 영상을 이용하는 모듈 구조 신경망 기반 필기체 숫자 인식

임 길 택[†] · 남 윤 석^{**} · 진 성 일^{***}

요 약

본 논문에서는 필기체 숫자 인식을 위한 방법으로서 회전과 이동 영상을 이용하는 모듈 구조의 신경망 기반 인식 방법을 제안한다. 전체 숫자 패턴 공간은 서로 중첩되는 여러 개의 작은 군집으로 나뉘어지는데, 이 군집들 상에서 전문가망으로 동작하는 MLP 신경망들이 구현된다. 그리고 군집마다 구현된 MLP들을 결합하기 위하여 게이팅망을 또한 MLP로 구현하였다. 학습 단계에서는 입력 숫자 영상에 회전과 이동의 디터밍 연산을 가함으로써 얻은 새로운 숫자 영상들을 이용하여 학습한다. 인식 단계에서는 입력 영상만을 인식하지 않고 입력 영상과 회전 및 이동된 영상들을 이용하여 인식한다. 이 과정에서 여러 개의 인식기 출력값들이 계산되는데, 이들을 다양한 방법을 이용하여 결합함으로써 최종 인식을 한다. 실험 결과 제안한 필기체 숫자 인식 방법이 우수한 인식 성능을 보임을 알 수 있었다.

Handwritten Numeral Recognition Based on Modular Neural Networks Utilizing Rotated and Translated Images

Kil-Taek Lim[†] · Yun-Seok Nam^{**} · Sung-Il Chien^{***}

ABSTRACT

In this paper, we propose a modular neural network based classification method for handwritten numerals utilizing rotated and translated images of an input image. The whole numeral pattern space is divided into smaller spaces which overlap each other and form multiple clusters. On these multiple clusters, multiple multilayer perceptrons (MLP) neural networks, specialized in those clusters, are constructed. Thus, each MLP acts as an expert network on the corresponding cluster. An MLP is also used as a gating network functioning as a mediator among the multiple MLPs. In the learning phase, an input numeral image is dithered by two geometric operations of translation and rotation so that new numeral images similar to original one are generated. In the recognition phase, we utilize not only input numeral image, but also newly generated images through the rotation and the translation of the original image. Thus, multiple output values for those generated images were combined to make class decision by various combination methods. The experimental results confirm the validity of the proposed method.

1. 서 론

일반적으로 인식 대상 패턴들이 매우 복잡한 결정

경계를 이루는 경우 간단한 구조로는 높은 인식률을 얻기가 매우 힘들다. 따라서 이러한 복잡한 인식문제들을 해결하기 위해서 최근의 연구들은 다중 특징 및 다중 인식기를 이용하는 복잡한 구조를 가지는 인식시스템에 대한 연구로 진행되어왔다[1-7]. 이러한 다중 인식기에 대한 연구는 크게 다음과 같이 나눌 수 있

† 정 회 원 한국전자통신연구원 우정기술연구부 선임연구원
** 정 회 원 한국전자통신연구원 우정기술연구부 우정자등회팀장
*** 정 회 원 경북대학교 진지진기공학부 교수
논문집수 2000년 2월 15일, 심사완료 2000년 5월 29일

다 첫째, 다중 특징을 이용하여 각 특징별로 인식기를 구현한 후 인식기들을 결합하여 인식률을 향상시킨다. 이 경우 각 인식기는 각 특징에 대해서 전문가망으로 동작하는데 서로 다른 패턴공간에서 구현된 인식기를 통합하여 인식한다[8-10]. 둘째, 다중 인식기는 서로 다른 알고리즘이나 구조로 구현되어 결합된다[11, 12]. 셋째, divide and conquer의 방법을 적용하여 대분류 후 상세분류를 통해 인식을 하는데 이 경우 패턴공간을 보다 작은 공간으로 나누고 각 인식기들은 나누어진 특정한 패턴공간내에서 동작하게된다[13-15].

문자 인식에 대한 연구는 작성된 문자의 형태에 따라서 필기체와 인쇄체 인식, 그리고 인식 대상 클래스수에 따라서도 인식의 난이도가 다르므로 인식 대상에 따라 다양한 방법이 적용되어왔다. Kimura [12] 등은 필기체 숫자 인식에서 두 가지 알고리즘을 이용하여 인식기를 구현하였는데, 사용하는 특징에 따라 통계적 인식기와 트리 인식기를 구현하여 결합하였다. 다중 특징을 이용하는 연구로 Shrihari 등[9]은 다섯 종류의 특징에 대해 각각의 신경망 인식기를 구현하고 결합하였다. Cao 등[13]은 다중특징과 다단계인식 방법을 이용하여 필기체 숫자 인식 시스템을 구현하였다. 인식 방법은 인식 단계별로 서로 다른 특징을 이용하여 대분류 후 상세분류 과정을 거쳐 인식을 하는 형태이다. Suen 등 [16]은 다중 전문가 시스템을 제안하였는데 4개의 알고리즘을 이용하여 각각의 전문가 시스템을 구현하고 이들을 결합하여 높은 인식률을 얻었다. Togawa 등[17]은 3단계로된 계층적 구조의 인식 방법으로 다중 LVQ (learning vector quantization)를 이용하여 간지 인식 시스템을 구현하였다. 이 방법에서는 각 LVQ망에 대해 유사 패턴간의 분별력을 높이기 위해 각 군집마다 서로 다른 수용 영역을 가지게 한 것이 특징이다. Itawa 등 [18]은 하나의 스템망 (stem network)과 다수의 브랜치망 (branch network)으로 이루어진 CombNet-II을 제안하였다. 스템망은 입력 패턴과 각 군집 대표벡터간의 유사도에 따라 대분류를 하고 상세분류기로서 MLP (multilayer perceptrons)를 이용하는데, 스템망의 출력이 브랜치망의 출력과 결합되어 인식이 된다.

본 논문에서는 게이팅망과 모듈화된 MLP 구조를 가지며 입력 숫자 영상에 회진과 이동 연산의 디더링 (dithering) 방법을 이용하는 필기체 숫자인식 방법을 제안한다. 필기체 숫자 인식기의 구현을 위해 전체 패

턴 공간을 서로 중첩하는 보다 작은 공간으로 나누고 다중 군집을 구성한다. 초기 다중 군집의 구성은 SOM (self organizing map) 망을 이용하였는데 각 군집의 반경에 따라 서로 중첩하게 하여 새로운 군집을 만들었다. 구성된 각 군집에 대해서 전문가망으로 동작하는 MLP를 구현하고 이들을 결합시키기 위해 게이팅 기능의 MLP를 구현한다. 학습단계에서 각 군집을 담당하는 전문가 MLP들과 게이팅망은 각각 분리되어 학습이 된다. 이 단계에서 각 학습 문자들에 대해 회진과 이동의 두 가지 디더링 과정을 수행함으로써 확장된 학습문자 집합이 생성된다. 인식 단계에서는 입력된 숫자 영상에 대해 게이팅망이 각 군집 MLP들의 출력값을 가중함으로써 각 클래스에 대한 신뢰도값이 출력된다. 이 단계에서는 기존의 방법들과 달리 입력 영상만을 인식하지 않고 입력 영상과 그것의 회진 및 이동된 영상들을 이용하여 인식한다. 이 과정에서 각 영상들 마다 인식기의 출력값들이 서로 틀리며, 인식 결과도 다르게 나타날 수 있다. 본 논문에서는 이러한 서로 다른 여러 개의 인식기 출력값들을 다양한 방법으로 결합함으로써 최종 인식을 하였다. 인식실험은 무제한 필기체 숫자에 대해서 이루어졌는데 제안한 인식시스템이 매우 높은 인식률을 보였다.

2. 군집의 생성과 확장

2.1 SOM망을 이용한 초기 군집의 생성

대규모의 문자인식에 대한 전략 중 하나로 'divide and conquer' 방법이 많이 적용되어 왔다[15]. 이러한 전략을 위해서는 군집화 과정이 필요하며 K-means [19], SOM [20] 등의 알고리즘이 많이 이용된다. 본 논문에서는 학습 패턴들에 대해 SOM 알고리즘을 이용하여 초기 군집을 구성하였다. 학습 후의 SOM망 출력노드들은 학습 문자집합의 확률 분포를 나타내게되고 군집의 대표 벡터로 사용될 수 있다. N_T 개로 이루어진 학습 샘플 집합을 $S = \{X_k, k=1, 2, \dots, N_T\}$ 이라 할 때 이 집합은 M 개의 분리된 부분집합 $S_i, i=1, 2, \dots, M$ 으로 나누어진다. 이때 부분집합 S_i 를 군집 i 라 한다. 입력문자들을 군집에 레이블링하기 위해서는 출력노드와의 유사도를 계산하는데 군집 i 의 대표벡터를 SOM의 i 번째 출력노드에 해당하는 w_i 벡터로 할 때 다음과 같이 계산된다.

$$d_i(X_k) = \|X_k - w_i\| = \sqrt{\sum_{n=1}^N (x_n - w_{in})^2} \quad (1)$$

이 식에서 N 은 패턴의 차원이며 w_n 은 군집 대표벡터의 n 번째 성분이다. 그러므로 가장 매칭이 잘되는 출력노드는 다음과 같이 결정된다.

$$d_m(X_k) = \min_i d_i(X_k), i=1, 2, \dots, M \quad (2)$$

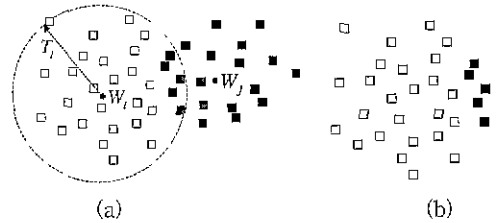
위 식을 계산하면 입력 패턴벡터 X_k 는 군집 m 에 속하게 된다. S 의 모든 패턴벡터에 대해 레이블링 과정을 거치면 S 는 M 개의 분리된 집합으로 나누어진다.

2.2 군집의 확장

계층적이며 모듈 구조를 가지는 일반적인 문자인식기는 대분류 후 상세분류 과정을 거쳐 인식을 하게 된다. 이 경우 대분류 단계에서 일어나는 오류는 상세분류 단계까지 전파가 된다. 실제 S_i 에 속하는 패턴이 S_i 상에서 학습된 인식기에 입력될 때 인식결과는 완전히 잘못될 것이다. 그러므로 이러한 오류를 감소시키기 위해서는 군집화 단계에서 많은 주의가 요구된다. 일반적인 군집을 형성하는 방법은 샘플 패턴공간을 서로 중첩되지 않게 즉, 군집원들이 중복되지 않게 분리되게 나누게 된다. 그러나 본 논문에서는 군집들이 위치하는 샘플 패턴공간을 서로 중첩되게 나누는 방법으로 군집을 확장하였다. 즉, 어떤 특정 군집 S_i 의 영향권 내에 들어올 수 있는 다른 군집의 패턴들을 군집 S_i 의 원소로 함으로써 군집 S_i 를 확장하는 것이다. SOM의 각 출력노드에 의해 정의될 수 있는 샘플 집합 S 는 반경 T_i 에 따라 식 (3)과 같이 S_i 로 확장된다.

$$S_i = \{X_k \mid \|X_k - W_i\| \leq T_i, k=1, 2, \dots, N_T\} \quad (3)$$

여기서, $T_i = \max \|X_k - W_i\|$, for $X_k \in S_i$ 이다. 이와 같은 과정을 그림 1에 개념적으로 나타내었다 (그림 1)의 (a)에서 흰 사각형으로 표현된 패턴은 W_i 에 가깝고 검은 사각형으로 표현된 패턴은 W_i 에 가깝게 위치한다. 따라서 군집 S_i 는 흰 사각형 패턴들을 포함하고 S_j 는 검은 사각형 패턴을 포함한다 (그림 1)의 (a)를 보면 S_i 에 있는 5개의 패턴들이 T_i 에 의해 정의되는 흰 사각형 패턴 영역에 위치하므로 (그림 1)의 (b)와 같이 S_i 에 새롭게 추가된다.



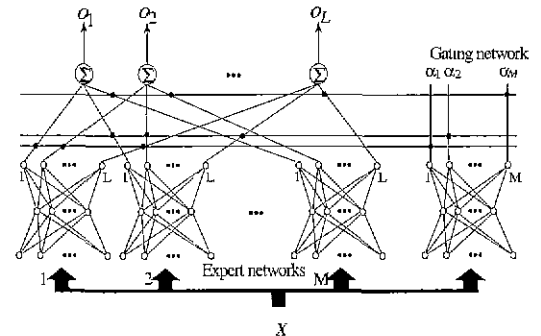
(그림 1) S_i 와 확장된 S_i . (a) S_i : 흰 사각형들 S_j , 검은 사각형들 (b) S_i .

3. MLP 모듈 구조의 문자인식기

본 논문에서는 확장된 각 군집마다 MLP 신경망을 적용하는 모듈구조의 숫자 인식기로 무제한 필기체 숫자 인식을 한다. M 개의 확장된 특정 군집에 대해서 M 개의 MLP들이 학습되어 해당 군집에 대해서 상세분류기로서 동작한다. 즉, 각 MLP는 해당 군집에 대한 전문가망으로 동작한다. 그러므로 입력이 된 숫자가 해당 MLP이외의 MLP에 의해 인식된다면 그 인식 결과는 잘못될 가능성이 매우 많다. 따라서, 적절한 MLP의 선택을 해주는 MLP 선택기, 즉 게이팅망(gating network)이 필요하다. 이를 위해서 본 논문에서는 MLP를 이용하여 게이팅망을 구현한다. 이때 게이팅망의 출력노드의 개수는 전체 군집 전문가 MLP의 수와 같아야 한다. 이러한 동작을 하는 인식기의 전체 구조를 (그림 2)에 나타내었다. (그림 2)의 인식기에서 어떤 특정 클래스 i 에 대한 인식기의 출력값은

$$o_i = \sum_{k=1}^M a^k f_i^k \quad (4)$$

와 같다. 여기서 o_i 는 클래스 i 에 대한 출력값이고,



(그림 2) 제한된 인식기의 구조

α^k 는 k 번째 군집에 대한 게이팅망의 출력값으로써 입력 패턴이 k 번째 군집에 속할 가능성을 나타낸다. f_i^k 는 k 번째 MLP의 클래스 i 에 대한 출력값을 나타낸다. (그림 2)의 인식기의 결정규칙은 o_m 이 식 (5)와 같이 계산될 때 입력패턴의 클래스는 m 으로 결정된다

$$o_m = \max o_i, i=1, 2, \dots, L \quad (5)$$

4. 입력 숫자 영상의 디더링

필기체 숫자는 필기 형태와 도구, 그리고 환경 등에 따라 매우 많은 변형을 보인다. 이러한 무 제약 필기체 숫자에 대한 완전한 인식을 위해서 고려될 수 있는 것 중 하나는 변형되어 나타날 수 있는 모든 숫자 형태를 확보하는 것이다. 그러나 실제의 경우에 있어서 이러한 대처방법은 불가능하므로 입력 문자에 대한 비선형 정규화 방법[21] 그리고 영상결함모델 방법 [22]등이 연구되었다. 본 논문에서는 필기체 숫자 패턴에 대해 비선형 정규화 방법은 적용하지 않고 회전과 이동 연산을 통한 디더링을 함으로써 확보되는 학습 숫자 집합을 확장하고자 하였다.

4.1 입력 영상의 회전

숫자 영상 $f(x, y)$ 는 $f'(x', y')$ 이 생성된다. θ 만큼 회전된 좌표는 회전되어서

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (6)$$

와 같이 얻어진다.

4.2 입력 영상의 이동

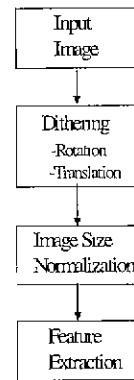
숫자 영상 $f(x, y)$ 는 $f'(x', y')$ 이 생성된다. 수평과 수직방향으로 각각 dx, dy 만큼 이동되는 좌표는

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \quad (7)$$

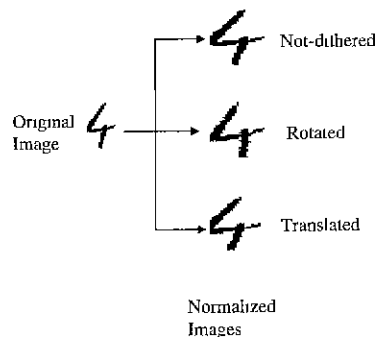
와 같이 얻어진다.

입력 숫자 영상은 특징벡터의 형태로 변환되어서 인식기에 입력되는데 특징추출과정은 (그림 3)과 같다. 디더링된 영상의 최소 외접사각형 내의 영상에 대해 크기 정규화를 한 후 특징이 추출된다. 영상의 좌표는

입력 영상의 최소 외접사각형의 좌상 꼭지점인 $(x=0, y=0)$ 가 원점이 된다. 이동의 경우에 디더링된 영상의 최소 외접사각형은 입력영상의 최소 외접사각형과 같다. 즉, x 또는 y 가 0보다 작거나, 입력영상의 최소외접사각형의 넓이나 높이보다 큰 영역에서는 특징이 추출되지 않고 무시된다. (그림 4)에 입력 문자 영상에 디더링을 가했을때의 예를 나타내었는데, 회전 영상은 원시 영상을 10° 회전한 후 크기 정규화를 한 것이고, 이동 영상은 원시 영상을 좌로 2화소, 즉, $dx=-2, dy=0$ 로 이동한 것이다. 이동 영상은 원시영상의 왼쪽 2화소들을 제거하고 오른쪽 2화소에는 흰화소로 채운 것이 된다. 실제의 경우에서는 접촉된 인접 문자들을 분리해낼 때 화소의 제거 현상과 회전 현상은 자주 일어나는 현상이다. 2가지 디더링 방법으로 생성된 새로운 문자 영상들은 실제의 경우에서도 동일한 클래스로 나타날 가능성이 있는 것들을 알 수 있다.



(그림 3) 특징 추출과정

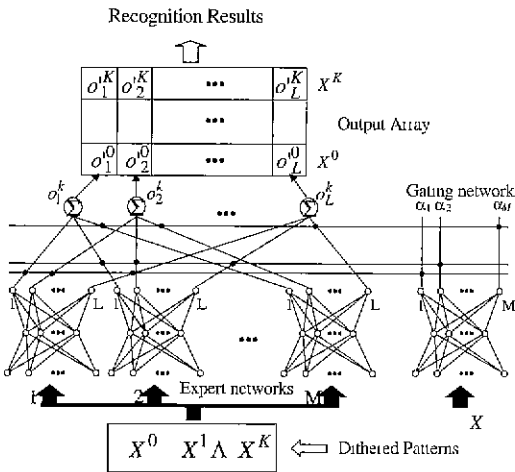


(그림 4) 영상 디더링의 예

5. 디더링 영상을 이용한 인식

구현된 인식기에 입력된 숫자가 학습에 사용된 숫자와 유사하다면 인식기는 높은 출력값을 나타낼 것이다. 그러나 학습에 사용된 숫자와 매우 다른 형태를 가지는 숫자들에 대해서는 인식성능이 저하될 가능성이 높게 된다. 본 논문에서는 이러한 문제점에 대처하기 위해, 인식단계에서 기존의 다른 방법과 다르게 입력 숫자만을 이용하지 않고, 입력 숫자 영상과 그것의 회전 및 이동된 영상을 이용함으로써 인식성능을 높이고자 한다

입력 패턴에 대해 K 개의 디더링 숫자들이 생성되어 인식기에 입력된다. 원시 입력 문자와 각각의 디더링 숫자들은 (그림 5)와 같이 인식기에 입력되어 그 출력값들이 출력 배열에 저장된다. 이러한 $K+1$ 개의 출력 벡터 값들을 이용하여 최종 인식 클래스를 결정하기 위해서 다양한 방법들을 이용한다.



(그림 5) 제안된 인식기의 인식 방법

인식기의 출력값들을 결합하는 각 방법들을 설명하기 위한 심볼들은 다음과 같다. 디더링 숫자 X^k 가 입력될 때 클래스 i 에서의 출력값은

$$o_i^k = \sum_{m=1}^M \alpha_{ij}^m(X^k) \quad (8)$$

와 같다. 여기서 M 은 전체 MLP의 수와 같다. 디더링된 각 문자들에 대한 각 MLP의 출력값들은 출력 배열에 변환되어 저장되는데 이들은 결합방법에 따라 결합된다.

o_i^k 의 변환된 값을 o_i^s 로 두고, $A = (1, 2, \dots, L)$ 을 L 개의 클래스 레이블의 집합으로 정의한다. 또한 입력 숫자 X 를 디더링한 후 생성된 모든 숫자들에 대한 출력값들을 결합한 후의 인식 클래스를 $D(X)$ 로 정의한다. 예를 들어 주어진 X 가 클래스 j 로 분류될 때 $D(X) = j$ 가 된다. <표 1>에 본 논문에서는 사용한 디더링 된 패턴들에 대한 인식기 출력값들을 결합하는 다양한 방법들을 나타내었다

<표 1> 인식기 출력값의 결합 방법 요약

Methods	o_i^k	Decision
method I-1	o_i^k	$D(X) = j,$ if $j = \arg \max_{i \in A} \sum_{k=0}^K o_i^k$
method I-2	$\frac{1}{s^k} o_i^k$	
method II-1	$(o_j^k - o_i^k) \delta_{ij}$	
method II-2	$\frac{1}{s^k} (o_j^k - o_i^k) \delta_{ij}$	
method II-3	$(\frac{1}{s^k} o_j^k + (1 - \frac{1}{s^k}) o_i^k) \delta_{ij}$	
method III-1	δ_{ij}	
method III-2	$o_j^k \delta_{ij}$	
method III-3	$\frac{1}{s^k} o_j^k \delta_{ij}$	

출력값들을 결합하는 가장 간단한 방법은 출력 배열에 있는 출력값들을 단순히 합하는 것이다(방법 I-1). 이 경우에는 변환값은

$$o_i^k = o_i^k \quad (9)$$

와 같다. 최종 인식 결과는 식 (10)과 같이 주어지는데 본 논문에서 사용하는 모든 방법에 대해 동일한 결정 함수가 이용된다.

$$D(X) = j, \text{ if } j = \arg \max_{i \in A} \sum_{k=0}^K o_i^k \quad (10)$$

즉, A 에 속하는 1에서 L 까지의 모든 클래스에 대한 출력값들 중 가장 높은 출력값을 가지는 클래스를 인식 결과로 취한다. 방법 I-2에서는 출력값 o_i^k 를 정규화하여 결합한다. 즉,

$$s^k o_i^k = o_i^k \quad (11)$$

가 된다. 여기서 $s^k = \sum_{i=1}^L o_i^k$ 로 k 번째 디더링된 패턴에

대한 스케일로서 $\frac{1}{s^k} \sum_i o_i^k = \sum_i o_i^k = 1$ 이 된다

$$o_i^k = \frac{1}{s^k} o_i^k \delta_{ij} \quad (19)$$

방법 II에서는 세 종류의 결합방법이 있는데 주어진 입력 패턴에 대한 인식 신뢰도지수에 따라 출력값들이 결합된다. k 번째 디더링 숫자에 대한 신뢰도를 rf^k 라 할 때 신뢰도는 방법 II-1에서

$$rf^k = o_j^k - o_j^k \quad (12)$$

와 같다. 여기서 $o_i^k = \max_{j \in A} o_j^k$ 으로 X^k 에 대한 L 개의 출력값 중 최대값이며 $o_j^k = \max_{j \in A_2} o_j^k$ 는 두 번째 큰 값으로 j^* 는 최대값의 인덱스이며 j^2 는 두 번째 최대값의 인덱스이다. 이때 $A_2 = A - \{j^*\}$ 이다. 방법 II-1에서의 변환값은

$$o_i^k = rf^k \delta_{ij} \quad (13)$$

으로 하였다. 이때 $i = j$ 이면 $\delta_{ij} = 1$ 이고, 그렇지 않으면 $\delta_{ij} = 0$ 이다. 이 방법에서는 주어진 디더링 문자에 대해 인식시스템 출력 최대값을 신뢰도지수로 바꾸고 나머지 값들은 0으로 두는 것이다. 방법 II-2에서는 방법 I-2와 같은 정규화 과정을 거친다. 즉,

$$o_i^k = \frac{1}{s^k} rf^k \delta_{ij} \quad (14)$$

이 된다. 방법 II-3에서는 Takahasi 등이 [23]에서 제시한 신뢰도 지수를 사용한다. 이 경우에

$$rf^k = \frac{1}{s^k} o_i^k + (1 - o_i^k / o_j^k), \quad (15)$$

$$o_i^k = rf^k \delta_{ij} \quad (16)$$

와 같이 하였다.

방법 III-1에서는 인식기의 출력값들을 결합하는 다른 방법으로 보우팅을 사용한다. 이를 위해서 변환값은 최대값의 경우에 1이 되고 그렇지 않은 경우에는 0이 된다. 그러므로 변환값은

$$o_i^k = \delta_{ij} \quad (17)$$

이 된다. 이 방법을 조금 변형한 것이 방법 III-2와 III-3으로 각각 식 (18)과 (19)에 따라 o_i^k 가 계산된다.

$$o_i^k = o_j^k \delta_{ij} \quad (18)$$

방법 III-2는 방법 III-1을 출력값으로 가중한 것이며 방법 III-3은 정규화하여 가중한 것이다.

6. 인식실험

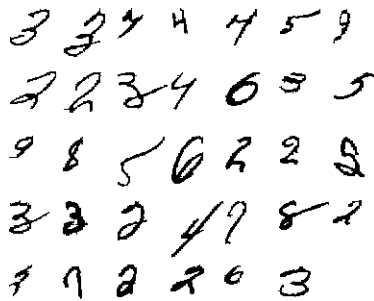
6.1 SOM을 이용한 군집 생성과 군집의 획정

본 논문의 실험에 사용한 데이터는 Concordia대학교의 필기체 숫자 데이터베이스이다. 이 데이터베이스는 0에서 9까지의 숫자가 각 600자씩 6000자로 구성되어있다. 이 중 상위 4000자는 학습에 사용하였고 하위 2000자는 테스트에 사용하였다. SOM을 이용하여 초기 군집을 만들 때 입력 특징으로는 144차원으로 된 숫자영상의 기술기 벡터[34]에서 추출된 54차원의 주성분벡터[25]로 하였다. 출력노드는 36개이며 0.85의 초기학습률로 100회 반복 학습하였다.

6.2 군집별 전문가 MLP의 학습

획정된 군집들에서 학습되는 MLP들은 모두 동일한 구조를 가진다. 입력 특징으로는 숫자 영상의 기술기 벡터에 대한 주성분 벡터를 사용하는데 군집 생성에서와 다르게 144차원으로 되어있다. 입력 노드의 개수는 144이며 출력 노드의 개수는 10개이며 은닉층의 노드 수는 30개이다. 목표벡터는 입력 패턴의 클래스에 해당하는 성분은 1로 두고 나머지 성분은 0로 하였다. 학습률은 0.1로 하였고 관성항은 0.7, 최대 학습 반복 횟수는 100회로 제한하였다.

입력 숫자 영상을 디더링할 때 θ 는 5와 10로 두어 4개의 새로운 숫자 영상을 만들어 내었다. θ 를 이렇게 둔 이유는 학습 숫자 영상들을 직접 관찰해보면 5° 이하로 회전시킬 경우 영상의 변형이 두드러지지 않으며 10° 보다 크게 회전시킬 경우 변형이 너무 크다고 판단되었기 때문이다. 이동의 경우는 dx 와 dy 를 2로 제한하였다. 이 경우 25개의 조합이 존재하게되는데 $\max(|dx|, |dy|)$ 가 2보다 작은 경우를 제외하였다. 이동 변이가 작을 때는 입력 영상의 변화가 작아 디더링의 효과가 거의 나타나지 않기 때문이다. 그러므로 이동의 경우는 16개의 새로운 조합이 가능하여 전체 20개의 디더링 영상이 생성될 수 있다. 실제 실험에 있어서 이들을 모두 저장한 다음에 학습을 하게 된다면 매우 큰 메모리가 필요하고 학



(a) 숫자 영상들

[3, 8]	[3, 2]	[7, 4]	[4, 8]	[7, 4]	[5, 1]	[9, 3]
[2, 8]	[2, 8]	[3, 4]	[4, 9]	[6, 0]	[3, 5]	[5, 3]
[9, 8]	[8, 6]	[5, 4]	[6, 0]	[2, 8]	[2, 8]	[2, 3]
[3, 2]	[3, 2]	[2, 8]	[4, 1]	[7, 9]	[8, 4]	[2, 8]
[3, 9]	[7, 9]	[2, 8]	[2, 8]	[6, 0]	[3, 8]	

(b) 혼동쌍 [올바른 클래스, 오인식된 클래스]

(그림 6) 인식 결과 오류 숫자

<표 4> 디더링의 효용성 테스트 실험 결과

Methods	Learning	Testing	Recognition rate	Remark
LN-TN	No dithering	No dithering	96.20	
LN-TD	No dithering	Dithering	96.30	
LD-TN	Dithering	No dithering	97.95	
LD-TD1	Dithering	Dithering	98.30	method I-2
LD-TD2	Dithering	Dithering	98.20	method II-2
LD-TD3	Dithered	Dithering	98.25	method II-3

본 논문에서 제안된 인식 방법은 입력 영상만을 이용하는 것이 아니라 회전과 이동의 디더링 과정을 통해 얻은 변형된 영상들도 이용하고 있다. 이러한 본 논문의 제안한 방법이 효용성이 있는가를 실험하여보았다. <표 4>는 디더링 방법의 효용성을 보여주는 것으로, 여기에서 'L'은 학습을 나타내며 'T'는 테스트, 즉, 인식을 나타낸다. 또한 'D'는 디더링, 'N'은 디더링을 하지 않았다는 것을 나타낸다. 인식단계에서 디더링을 하지 않은 경우들은 단순히 그림 2의 인식기의 출력값만으로 최종 인식을 하는 경우이다. 이 표를 살펴보면 학습 단계에서만 디더링을 행하는 LD-TN 방법이 인식 단계에서만 디더링을 하는 경우인 LN-TD 방법보다 우수하였다. 즉, 회전과 이동의 디더링 방법은 인식 단계에서 보다 학습 단계에서 디더링을 하는 것이 인식률이 보다 높음을 알 수 있다. 그러나, 학습과 인식 단계 모두에서 디더링을 한 경우인 LD-TD1, LD-TD2, LD-TD3의 인식률들이 가장 높음을 알 수 있다. 여기서 한가지 주목할 점은 LD-TN의 경우 97.95%의 인식률인데, <표 2>의 방법 III-2와 III-3에서 K=6인 경우의 인식률은 LD-TN보다 높지만, K=20인 경우의 인식률은 이보다 낮게 나왔다. <표 2>의 실험 결과에서도 알 수 있었지만, 서

로 유사한 디더링 영상을 너무 많이 이용하면 인식률이 도리어 떨어짐을 알 수 있다.

7. 결 론

이 논문에서는 필기체 숫자를 인식하기 위하여 회전과 이동 영상을 이용하는 모듈 구조를 가지는 신경망 기반 인식 방법을 제안하였다. 인식시스템의 구현을 위해 전체 패턴 공간을 서로 중첩하는 보다 작은 공간으로 나누고 다중 군집을 구성하였다. 초기 다중 군집의 구성은 SOM 망을 이용하였는데 각 군집의 반경에 따라 서로 중첩하게 하여 새로운 군집을 만들었다. 구성된 각 군집에 대해서 전문가망으로 동작하는 MLP를 구현하고 이들을 결합시키기 위해 게이팅 기능의 MLP를 구현하였다. 이때 각 MLP의 학습에 있어서는 학습 집합을 확장하고자 회전과 이동의 2가지 디더링 과정을 거쳐 생성된 패턴들을 포함하였다. 인식기의 동작은 입력된 문자에 대해 게이팅망이 각 군집 MLP들의 출력값을 가중함으로써 각 클래스에 대한 신뢰도 값을 출력한다. 본 논문의 인식 방법은 기존의 방법들과 달리 입력 영상만을 인식하지 않고 입력 영상에 대해서 회전과 이동의 두가지 디더링을 통하여 얻은 새로운 숫자 영상들을 인식하였다. 이때 여러 개의 인식기 출력값이 계산되는데 이들에 대해 다양한 방법들을 적용하여 결합함으로써 보다 높은 인식률을 얻고자 하였다. 인식실험 결과 제안한 여러 가지 결합 방법들이 높은 인식률을 보임을 알 수 있었다. 특히, 인식시스템의 디더링 효용성에 관한 실험에서는 학습 단계에서의 디더링이 인식단계에서 보다 더 효과적이며 두 단계 모두에 디더링을 적용한 경우가 가장 좋은 성능을 보

임을 알 수 있었다.

참 고 문 헌

- [1] A. Krzyzak, W. Dai, and C. Y. Suen, "Unconstrained handwritten character classification using modified backpropagation model," *Proc. of the First IWFHR*, Montreal, Canada, pp.155-166, 1990.
- [2] S. Cho, "Neural-network classifiers for recognizing totally unconstrained handwritten numerals," *IEEE Transactions on Neural Networks*, Vol.8, pp.43-53, 1997.
- [3] T. Akiyama and N. Hagita, "Automated entry system for printed documents," *Pattern Recognition*, Vol.23, pp.1141-1154, 1990.
- [4] J. H. Kim, K. Kim, and S. Chien, "Korean and English character recognition system using hierarchical classification neural network," *Proc. of ICSMC*, Vancouver, Vol.1, pp.759-764, Oct. 1995.
- [5] D. Lee and S. N. Srihari, "A theory of classifier combination : the neural network approach," *Proc. of the Third ICDAR*, Montreal, Canada, pp.42-45, 1995.
- [6] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16, pp.66-75, 1994.
- [7] Y. S. Huang and C. Y. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17, pp.90-94, 1995.
- [8] K. Miyahara and F. Yoda, "Printed Japanese character recognition based on multiple modified LVQ neural network," *ICDAR*, Tsukuba, pp.250-253, 1993.
- [9] S. N. Srihari, "Recognition of handwritten and machine printed text of postal address interpretation," *Pattern Recognition Letters*, Vol.14, pp.291-302, 1993.
- [10] Y. M. Baek, K. T. Lim, and S. I. Chien, "Off-line handwritten Hangul recognition based on multiple features and modular partially connected multi-layer perceptrons," *Proc. of IWFHR*, Taejon, pp.269-278, 1998.
- [11] L. Xu, A. Krzyzak, and C. Y. Suen, "Method of combining multiple classifiers and their application to handwritten numeral recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.22, pp.418-435, 1992.
- [12] F. Kimura and M. Shridhar, "Handwritten numeral recognition based on multiple algorithms," *Pattern Recognition*, Vol.24, No.10, pp.976-983, 1991.
- [13] J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," *Pattern Recognition*, Vol.28, No.2, pp.153-160, 1995.
- [14] A. Waibel, H. Sawai, and K. Shikano, "Modularity and scaling in large phonemic neural networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.37, pp.1888-1898, 1989.
- [15] B. S. Kang, S. I. Chien, and K. T. Lim, "Large scale pattern recognition system using hierarchical neural network and false alarming nodes," *Proc. of ICTAI*, Newport, California, pp.196-203, Nov. 1997.
- [16] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai, and L. Lam, "Computer recognition of unconstrained handwritten numerals," *Proc. of IEEE*, Vol.80, No.7, pp.1162-1180, 1992.
- [17] F. Togawa, T. Ueda, T. Aramaki, and A. Tanaka, "Receptive field neural network with shift tolerant capability for Kanji character recognition," *IJCNN*, pp.1490-1499, 1988.
- [18] A. Itawa, H. Kawajiri, and N. Suzumura, "Classification of handwritten digits by a large scale neural network 'CombNET-II'," *IJCNN*, Singapore, Vol.1, pp.1021-1026, Nov. 1991.
- [19] M. Nadler and E. Smith, *Pattern Recognition Engineering*, John Wiley & Sons, New York, 1993.
- [20] T. Kohonen, "The self-organizing map," *Proc. of IEEE*, Vol.78, No.9, pp.1464-1480, 1990.
- [21] S. Lee and J. Park, "Nonlinear shape normalization methods for the recognition of large-set handwritten characters," *Pattern Recognition*, Vol.27, No.7, pp.895-902, 1994.

[22] H. S. Baird, "Document image defect models," *IAPR Workshop on SSPR*, Murray Hill, NJ, pp.38-46, 1990.

[23] H. Takahashi, "A neural network OCR using geometrical and zonal pattern features," *Proc. of ICDAR*, Paris, Vol 2, pp.821-828, 1991.

[24] G. Srikanan, S. W. Lam, and S. N. Srihari. "Gradient-based contour encoding for character recognition," *Pattern Recognition*, Vol.29, pp.1147-1160, 1996.

[25] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison Wesley. pp.148-156, 1989.



임길택

e-mail : ktlim@etri.rc.kr
 1993년 경북대학교 전자공학과 (공학사)
 1995년 경북대학교 전자공학과 (공학석사)
 1999년 경북대학교 전자공학과 (공학박사)

1999년~현재 한국전자통신연구원 우정기술연구부
 선임연구원
 관심분야 : 패턴인식, 신경회로망, 영상처리 등



남윤석

e-mail : ysnam@etri.re.kr
 1984년 아주대학교 산업공학과 (학사)
 1989년 Polytechnic Univ (New York), Dept. of Industrial Engineering(공학석사)
 1992년 Polytechnic Univ.(New York), Dept of Industrial Engineering(공학박사)
 1993년~현재 한국전자통신연구원 우정기술연구부
 우정자동화팀장
 관심분야 : 소프트웨어 공학, 패턴인식 등



진성일

e-mail : sichien@ec.kyungpook.ac.kr
 1977년 서울대학교 전자공학과 (공학사)
 1981년 한국과학기술원 전자공학과 (공학석사)
 1988년 카네기멜론 대학교 전자공학과 (공학박사)

1988년~현재 경북대학교 전자전기공학부 교수
 관심분야 : 영상처리, 컴퓨터비전, 패턴인식 등