

論文96-33B-9-10

# 은닉 마르코프 모델과 신경회로망을 이용한 정면 얼굴인식 (Frontal View Face Recognition Using the Hidden Markov Model and Neural Networks)

尹康植\*, 咸永國\*, 朴來弘\*\*

(Kang-Sik Yoon, Young Kuk Ham, and Rae-Hong Park)

## 요 약

본 논문에서는 은닉 마르코프 모델과 신경회로망 (HMM-NN)을 이용한 얼굴인식 알고리즘을 제안한다. 전처리 단계에서 locally adaptive threshold (LAT)로 얼굴의 에지를 찾아내고, 얼굴의 일반적인 지식에 기초하여 특징을 추출한 후, 이 특징들을 이용하여 데이터베이스를 구축한다. 학습단계에서는 forward-backward 알고리즘을 이용하여 각 사람의 HMM 파라미터를 만든다. 인식단계에서는 HMM에 의해 계산된 확률값을 NN의 입력 데이터로 사용한다. 컴퓨터 시뮬레이션 결과 제안한 HMM-NN 알고리즘이 기존의 얼굴인식 알고리즘보다 높은 인식률을 보였다.

## Abstract

In this paper, we propose a face recognition algorithm using the hidden Markov model and neural networks (HMM-NN). In the preprocessing stage, we find edges of a face using the locally adaptive threshold (LAT) scheme and extract features based on generic knowledge of a face, then construct a database with extracted features. In the training stage, we generate HMM parameters for each person by using the forward-backward algorithm. In the recognition stage, we apply probability values calculated by the HMM to subsequent neural networks (NN) as input data. Computer simulation shows that the proposed HMM-NN algorithm gives higher recognition rate compared with conventional face recognition algorithms.

## I. 서 론

얼굴은 인간에게 있어 독특한 특징이다. 심지어 일란성 쌍둥이의 얼굴도 서로 조금은 다르다는 것을 관찰할 수 있다. 얼굴은 상대방의 신원과 감정을 나타내고 전달하는 데 중요한 역할을 하므로 사회생활에 있어 얼굴인식의 중요성이 점차 증가하고 있다. 인간은 별 어려움없이 어떤 장면에서도 얼굴을 찾아 그 얼굴 주인공의 신원까지 알아낼 수 있다. 이러한 사람의 능력은 안경, 머리모양의 변화와 같은 외부적인 요인과 얼굴표현, 노쇠와 같은 얼굴자체의 변화에도 영향을 적게

받는다. 그러나, 이러한 능력을 갖춘 자동인식 시스템을 구축하기는 매우 어렵고 또한, 사람이 사용하는 지식이 너무 방대하여 지식의 일부분만을 이용하여 인식하는 방법이 연구되고 있다.<sup>[1]</sup> 이를 위해 풀어야 할 문제로 인식문제, 얼굴표정의 분석문제, 얼굴특징에 기초를 둔 분류문제 등이 있다.

얼굴인식 시스템에는 두가지 입력영상을 고려할 수 있는데, 이에 따라 정면 얼굴영상인식<sup>[2-8]</sup>과 측면 얼굴영상인식<sup>[9,10]</sup>이 있고, 두가지 방법을 혼합하여 사용하는 방법<sup>[11,12]</sup> 등이 있다. 정면 얼굴영상이 측면 얼굴영상보다 더 많은 정보를 포함하고 있기 때문에 주로 정면 얼굴영상을 인식 시스템의 입력으로 하여 많은 연구가 이루어졌다. 정면 얼굴을 이용하여 얼굴을 인식하는 기존의 방법에는 먼저 눈의 위치를 알아낸 후 얼

\* 準會員, \*\* 正會員, 西江大學校 電子工學科

(Dept. of Electronic Eng., Sogang Univ.)

接受日字:1995年10月27日, 수정완료일:1996年8月28日

굴에 대한 기초지식을 이용하여 각 구성성분들 사이의 위치관계를 이용하는 방법,<sup>[6]</sup> 에지영상에서 정합 영역 추출후 2차 통계 특징을 이용하여 정합하는 방법<sup>[8]</sup>이 있고, 최근에는 상관 계수를 이용하여 얼굴의 대칭축을 찾아 얼굴의 각 부분의 위치를 찾는 방법,<sup>[13]</sup> 움직이는 얼굴에서 에지 화소 카운팅 방법으로 얼굴의 각 부분의 위치를 찾는 방법,<sup>[14]</sup> 얼굴의 각 부분을 deformable templates로 찾은 뒤 특징값을 추출하는 방법<sup>[15]</sup>이 있다.

상관 계수와 에지 화소 카운팅방법은 얼굴이 한쪽으로 치우쳐 있거나 눈이나 이마부분이 머리로 감추어져 있으면 특징을 찾기 어렵고, 2차 통계 특징을 이용한 정합 방법의 경우는 조명에 의한 에지의 변화에 민감하고 취득된 영상의 기울기 보정이 어려운 단점이 있다. 그리고 얼굴에서 각 구성성분을 찾기 위해 눈의 위치를 단지 눈의 깜빡임으로 찾는데 모든 입력영상에 대해 이런 제한조건을 두는 것은 무리이다.

따라서 본 논문에서는 얼굴의 각 구성성분들 사이의 위치관계를 사용하여 특징들을 추출한 후 인식을 시도하였다.

본 논문에서는 안경을 쓴 사람의 얼굴영상인 경우 안경영상이라 정의하였고, 정면 얼굴의 각 구성 성분들 사이의 위치관계를 얼굴에 대한 기초지식을 사용하여 특징들을 추출한 후 인식하였다. 여기서 기초지식이라 함은 얼굴영상은 좌우 대칭이고, 눈썹 아래에 눈이 있고 그 사이에 코가 존재하며, 코 아래에 입술이 위치한다는 것 등이다.

이러한 위치관계와 얼굴의 형태에 관련된 특징점들을 사용할 경우 특징을 많이 사용하여 많은 사람을 구분할 수 있다는 장점은 있으나, 얼굴과 배경을 분리하는 단계에서 얼굴의 형태가 조명과 각도에 의해 왜곡이 일어날 가능성이 크고 안경영상인 경우 조명에 매우 민감하므로 인식률이 낮아진다.

본 논문에서는 얼굴의 각 구성성분들인 눈, 코, 그리고 입을 지식에 기초한 접근 방법을 이용하여 자동추출하고 HMM-NN을 이용한 얼굴인식 방법을 제안하였다.

본 논문의 II장에서는 특징들의 올바른 추출을 위한 전처리에 대해 기술하였고, III장에서는 HMM-NN을 이용한 제안한 얼굴인식 방법에 대하여 논하였으며, IV장에서는 실험결과 및 분석, V장에서는 결론에 대해서술하였다.

## II. 전처리

에지 연산자로 임계값을 갖는 4 방향 Sobel 에지연산자를 사용하였고, 올바른 특징추출을 위하여 팽창(dilation)과 수축(erosion) 방법에 의해 얼굴영상에서 에지의 끊어짐을 방지했고, 지식에 기초하여 특징을 추출했다.

### 1. 에지추출

에지추출 연산자는 4 방향 Sobel 연산자를 사용하였고, 임계값은 Robinson이 제안한 locally adaptive threshold (LAT)를 사용하여 자동으로 결정하였다.<sup>[16]</sup>

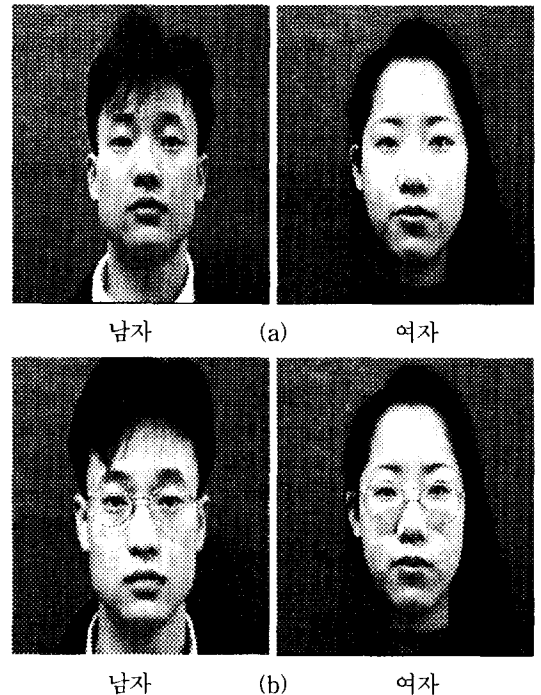


그림 1. 실험에 사용한 얼굴영상에  
(a) 얼굴영상 (b) 안경영상

Fig. 1. Examples of frontal view face images used in experiments.

(a) face images, (b) images in glasses.

그림 1은 실험에 사용된 영상들중 얼굴영상과 안경영상의 한 예를 보였다. 영상을 취득할 때 여자의 경우 한쪽 귀나 양쪽 귀가 머리카락으로 가리워지는 경우가 많으나 머리카락이 눈부위나 이마를 가리는 경우를 제외하곤 비교적 얼굴 각 부분의 위치를 찾는 데 문제가 없어서 실험영상에 귀가 가리워지는 사람도 포함하였다. 그리고 LAT 알고리즘을 이용해서 얻어진 에지영

상은 얼굴윤곽선이나 각 부위의 주변의 에지가 끊어짐을 볼 수 있는데 보다 좋은 특징추출을 위해 형태학상의 팽창으로 에지의 끊어짐을 어느 정도 보완하였다.

2. 얼굴과 배경의 분리

본 논문에서는 원영상을 이진화한 후 지식에 기초한 에지 추적 방법에 의하여 얼굴과 배경 및 어깨를 분리하였다. 또한 에지의 끊어짐이 적어야 얼굴과 배경 및 어깨를 잘 분리할 수 있으므로 이진 영상에 대한 형태학상의 팽창 (morphological dilation) [17] 하였다.  $W$  를  $3 \times 3$  창,  $F$  를 얼굴의 이진영상, 그리고  $f$  를 에지점이라 할 때, 이진화된 영상에 structure element라 정의되는  $3 \times 3$  창을 움직이며 에지점이 하나라도 포함되면  $3 \times 3$  창의 모든 부분을 에지점으로 결정한다. 즉, 팽창연산을  $\oplus$ 로 표시하면,

$$W \oplus F = \bigcup_{f \in F} (W + f) \quad (1)$$

로 형태학상의 팽창연산을 정의할 수 있으며, 그리고, 눈, 코, 그리고 입의 끊어진 에지를 연결하기 위하여

$$\begin{aligned} C(F, G) &= [F \oplus (-G)] \ominus G \\ &= e [d(G, -F), -F] \end{aligned} \quad (2)$$

와 같은 연산을 사용하였다. 여기서  $\ominus$ 는 수축연산을 나타내며  $F$ 는 이진 얼굴 영상이고  $G$ 는  $1 \times 3$  창이다.  $e$ 는 수축을 의미하고  $d$ 는 팽창을 의미한다. 그림 2은 형태학상의 팽창으로 구한 영상이다. 그림 2에서 보듯이 에지의 끊어짐이 많이 보완됐으며 지식에 기초한 특징추출을 통해 얼굴의 각 부위의 위치를 찾아낼 수 있다.

3. 지식에 기초한 특징추출

본 논문에서는 기존의 논문 [2,18-20]에서와 같은 방법으로 지식을 표현했고, 지식에 기초한 특징 추출 방법으로 눈 위에서 코 위까지의 거리 (a), 코 위에서 입 위까지의 거리 (b), 코 중앙에서 왼쪽 눈의 오른쪽 중앙까지의 대각선의 거리 (c), 코 중앙에서 오른쪽 눈의 왼쪽 중앙까지의 대각선의 거리 (d), 입 중앙에서 좌측 얼굴 윤곽선까지의 거리 (e) 등 5개의 특징값을 추출하였다. 5개의 특징값을 미간 사이의 거리로 정규화하여 영상 크기의 변화에 대처하였다. 그리고 안경영상의 경우 두꺼운 테나 빨테 안경을 제외한 금속테를 사용했는데 이것은 에지영상에서 눈부분의 안경에 의한 에지가 되도록 적어야 눈의 위치를 찾는 데 용이하기 때문

이다. 그림 2에 특징 추출 결과를 보였다.

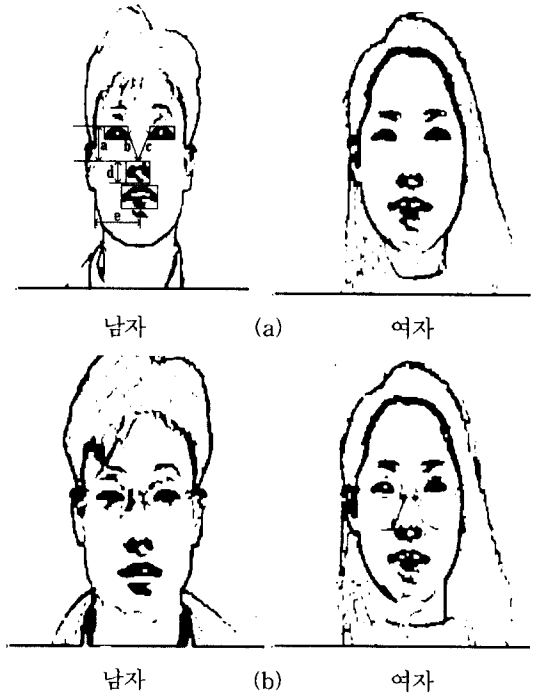


그림 2. 형태학상의 팽창후의 추출된 에지 (a) 얼굴영상 (b) 안경영상

Fig. 2. Extracted edges using morphological dilation. (a) face images, (b) images in glasses.

III. HMM-NN을 이용한 제안한 얼굴인식

1. HMM [21] 알고리즘

HMM은 신호를 doubly stochastic process로 나타낼 수 있는 stochastic 모델이다. 즉, 관찰열이라 불리는 특징의 연속열  $O_1, O_2, \dots, O_T$ 와 감추어진 (hidden) 부분인 상태열  $s_1, s_2, \dots, s_T$ 로 나타낼 수 있다. HMM은 천이 (transition)에 의해 서로 연결된 상태들의 모임으로서 각 천이에는 2가지의 확률이 관련되어 있다. 하나는 현재의 천이가 이루어질 천이 확률이고, 다른 하나는 천이가 이루어졌을 때 유한개의 관측 대상으로부터 각 출력 심벌을 내는 조건부 확률을 규정하는 출력 확률이다. HMM을 나타내기 위해서 먼저 기호들을 정의하자. 상태수, 관측 심벌수는 각각  $N, M$ 으로, 상태 집합, 심벌 집합, 관측열의 길이를 각각  $Q = \{q_1, q_2, \dots, q_M\}$ ,  $V = \{v_1, v_2, \dots, v_M\}$ ,  $t = 1, 2, \dots, T$ 으로 표기한다. 이로부터 HMM을 구성하

는 3가지 파라미터  $A$ ,  $B$ ,  $\pi$ 를 설명하면 다음과 같다.  $t$ 번째 관측 심볼열이 상태  $q_t$ 에 있고,  $t+1$  번째 관측 심볼열이 상태  $q_j$ 를 선택할 상태 천이 확률 행렬  $A$ 는

$$A = \{a_{ij}\}, a_{ij} = \Pr(q_j \text{ at } t+1, q_i \text{ at } t) \quad (3)$$

로,  $t$ 번째 관측 심볼열이  $q_i$  상태에서 심볼  $v_k$ 를 선택할 관측 심볼 확률 벡터  $B$ 는

$$B = \{b_j(k)\}, b_j(k) = \Pr(v_k \text{ at } t, q_j \text{ at } t) \quad (4)$$

로, 초기 상태에서 상태  $q_i$ 를 선택할 초기 상태 확률 벡터  $\pi$ 는

$$\pi = \{\pi_i\}, \pi_i = \Pr(q_i \text{ at } t = 1) \quad (5)$$

로 주어진다. 관측열은  $O = O_1, O_2, \dots, O_T$ 로 주어진다. 이상의 정의를 이용하면 HMM은  $(A, B, \pi)$ 로 표시할 수 있다. 이 모델에서 모델화에 영향을 크게 미치는 중요도는 관찰 심볼 확률 벡터  $B$ 가 가장 크며, 다음은 상태 천이 확률 행렬  $A$ , 그리고 초기 상태 확률 벡터  $\pi$ 의 순이다.

본 연구에서는 얼굴 특징 데이터의 개수가 관찰열의 길이 5가 되고 그리고 특징의 순서를 변화시키면서 인식률을 계산했는데 II. 3.에서 언급한 특징의 순서일 때 가장 인식률이 높았다. 상태수도 3에서 10까지 변화시키면서 실험했고, 비교적 계산량이 적으므로 left-to-right 모델<sup>[22]</sup>을 사용하였다.

#### 1) HMM 학습 알고리즘

학습 알고리즘으로는 forward-backward 알고리즘<sup>[22]</sup>과 Baum-Welch 재추정 (reestimation) 알고리즘<sup>[22]</sup>이 사용된다. 전자는 모델 파라미터  $\lambda = (A, B, \pi)$ 가 주어졌을 때 관측열  $O$ 의 확률, 즉  $P(O|\lambda)$ 를 계산하는 알고리즘이고, 후자는 초기 파라미터들로부터  $\Pr(O|\lambda)$ 를 최대로 하는  $\lambda = (A, B, \pi)$ 를 계산하는 알고리즘이다. 현재 모델을  $\lambda = (A, B, \pi)$ 라 하고 다시 추정된 모델을  $\lambda' = (A', B', \pi')$ 라 하면 새로운 모델 파라미터는 다음과 같이 구할 수 있다.

$$\pi'_i = \lambda_i(i), \quad 1 \leq i \leq N \quad (6)$$

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{T-1} \alpha_t(i) \beta_t(i)} \quad (7)$$

$$b'_j(k) = \frac{\sum_{t=1, O_t=V_k}^T \alpha_t(j) \beta_t(j)}{\sum_{t=1}^T \alpha_t(j) \beta_t(j)} \quad (8)$$

여기서  $a'_{ij}$ 과  $b'_j(k)$ 는 각각 재추정된 천이 확률과 관측 심볼 확률이고,  $\lambda_i(i)$ 는 시간  $t = 1$ 일 때 상태  $q_i$ 에 있을 확률, 그리고  $V_k$ 는 관측열 중  $k$ 번째 관측 심볼을 가리킨다.

#### 2) HMM 인식 알고리즘

인식 알고리즘으로는 forward 알고리즘, backward 알고리즘 혹은 Viterbi 알고리즘<sup>[22]</sup> 등을 사용할 수 있다. 본 연구에서는 left-to-right 모델에 적합한 Viterbi 알고리즘을 사용하였다. Viterbi 알고리즘은 가장 확률이 높은 하나의 경로만을 찾기 때문에 계산량이 적으며, 자기 자신의 상태와 상위의 상태로만 천이가 가능한 left-to-right 모델에 적합한 알고리즘이다. 즉 어느 시점에서부터 임의의 노드에 도달하는 확률은 직전 노드에서의 확률에다 현재의 노드로 천이할 때의 추가 확률을 합하여 계산하고, 이런 계산을 반복함으로써 어떤 시점의 현재 노드까지 오는 경로중 최적인 것 하나만을 기억한다.

#### 2. HMM-NN을 이용한 제한한 얼굴인식

HMM을 얼굴인식 알고리즘으로 사용하기 위해 총 600장의 영상중 300장은 학습시, 나머지 300장은 인식 대상으로 사용하였다. 실험에서 사용한 HMM은 left-to-right 모델로 가정하여 시간축에 대해서 작은 번호의 상태에서 큰 번호로의 상태로만 천이가 가능하고, 2개 이상의 상태를 한번에 천이가 불가능하므로 이것을 고려하여 상태수가 4일 경우의 초기 조건을

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (9)$$

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \quad (10)$$

와 같이 나타내었다. Left-to-right 모델인 경우 반드시 가장 큰 번호의 상태에서 끝나므로  $a_{44} = 1.0$ 이어야 하며 본 실험에서는 자기 자신으로의 상태 천이의 경우 0.2, 나머지 큰 번호의 상태로의 천이 확률값은 0.8로 실험적으로 결정하였다.

한편 기존의 알고리즘들중 인식률이 높은 NN<sup>[23]</sup>은 데이터상의 오류와 표현하기가 쉽지않는 애매모호한

데이터를 처리하는 분류기로서 본 실험에서는 II. 3에서 설명한 바와 같이 각 사람에서 추출한 5개의 특징값을 가지고 입력 데이터와 인식 데이터를 만든 후 인식하였다. 그리고 학습함과 판정함은 입력 데이터의 특성을 고려하여 모두 0.3을 사용하였고, 최적의 중간층 노드의 수를 찾기 위해 중간층과 출력층 사이의 가중치값과 전체 학습패턴에 대한 각 중간층 노드의 출력값의 평균을 가지고 불필요한 노드를 찾아 제거하는 방법<sup>[24]</sup>을 사용하여 결정하였다. 그리고 HMM에서 오인식된 데이터를 처리하고 얼굴에서 추출한 데이터가 NN으로 분류하기 쉬운 static 데이터 패턴이므로 HMM의 다음 단계에 NN을 결합해서 인식을 시도하였다.

HMM-NN<sup>[25]</sup>은 음성인식과 문자인식에서 dynamic time warping 조건하에서도 우수한 모델링을 할 수 있는 HMM<sup>[22,26,27]</sup>을 전단계, 그리고 정적 데이터 패턴에 대해 우수한 분류능력이 있는 NN을 후단계에 놓고 인식을 시도하였다. 즉 각 사람에서 추출한 5개의 특징값을 입력으로 사용한 후의 HMM에서의 인식은 Viterbi 알고리즘을 이용하여 가장 큰 확률값을 가지는 사람을 인식의 결과로 보기 때문에 만약에 여러 사람이 서로 비슷한 확률값을 가지면 오인식률이 높아지기 때문에 인식단에서 Viterbi 알고리즘으로 최종 시간축의 각 상태에 대한 조건부 확률값을 구하고 적절한 scaling을 취한 다음 NN의 입력으로 사용한다. 특히, NN의 입력 데이터는 HMM에서 계산된 확률값이고 인식될 사람의 확률이 다른 사람의 확률값보다  $10^3 \sim 10^5$  배 정도 크기 때문에 인식될 사람마다 NN의 입력 데이터가 clustering되는 장점이 있으므로 NN의 수렴 속도도 약 10초(MIPS R4000 100MHz) 이내로 빠르고 인식률도 높아진다. 특히, 후단계인 NN에서는 빠른 학습시간과 보다 정확한 인식을 위해 최적의 중간층 노드의 수를 NN에서와 같이 결정하였다. 그림 3(a)는 전체적 실험 과정을 나타냈는데 HMM에서 인식할 한 사람에 대해 학습된 각 사람의 HMM 파라미터와 Viterbi 알고리즘으로 구한 확률값을 NN의 입력단으로 사용함을 보여준다. 또한, HMM의 전단계에 NN을 결합한 NN-HMM<sup>[28,29]</sup>은 HMM의 후단계에 NN을 결합한 형태와 인식률을 비교함으로써 NN이 미치는 영향을 살펴보았다. 이 경우의 NN은 HMM의 입력으로 사용하는 일반적인 유클리디안 거리에 의한 vector quantization (VQ)를 대신해 NN을 사용해서 HMM

의 입력 데이터를 만든다. 이것은 음성과 문자의 경우 말하는 사람, 글 쓰는 사람마다 같은 사람이라도 처리해야 할 데이터가 무척 많기 때문에 isolated나 speaker dependent 데이터와 같은 경우 상대적으로 데이터가 적어서 NN을 VQ로 사용하는 것이 가능하다.

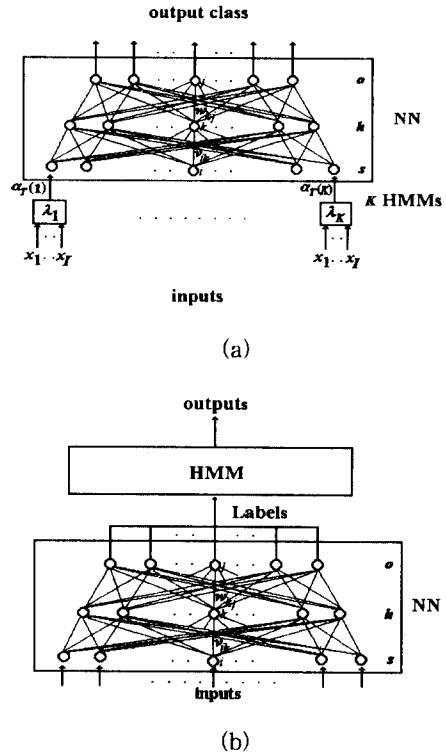


그림 3. HMM-NN과 NN-HMM의 블록도  
(a) HMM-NN (b) NN-HMM

Fig. 3. Block diagrams of the HMM-NN and NN-HMM.

(a) HMM-NN, (b) NN-HMM.

그림 3(b)를 보면 얼굴에서 추출한 5개의 데이터를 NN의 입력으로 사용한 후 TOP-N<sup>[28,29]</sup> 방법을 사용하여 가장 큰 값을 가지는 출력층의 노드에 해당되는 번호와 나머지 노드들에는 0을 레이블링한다. 그리고 가장 큰 값을 연속으로 2개, 3개를 취해서 각각 레이블링하면 인식률이 가장 큰 값을 취했을 때 인식률이 더 증가하는데 이것은 NN의 출력층에서 인식되는 사람이 여러명 존재할 수 있다는 것을 보여주고 이 값들을 포함해서 HMM의 입력데이터로 사용해서 오인식률을 줄이는 역할을 한다. HMM-NN에 비해서 NN-HMM에서의 NN의 수렴속도는 약 1시간 정도로 느리는 단점이 있다.

#### IV. 실험 결과 및 분석

본 논문에서는 남자 23명과 여자 7명 총 30명에 대해 각각 20장씩 취득하고, 사람별로 특징들을 추출한 후 다시 데이터베이스화해서 총 600장에 대한 특징을 사람별로 데이터베이스화했다. 이 중에는 안경영상 60장이 포함되어 있다. 이렇게 추출한 특징값들은 사람을 명확히 구분할 수 있는 올바른 특징만을 추출하여야 하므로 다음과 같은 제약조건을 두었다. 첫째, 얼굴부분이 취득된 영상의 가운데 부분에 있어야 한다. 둘째, 영상 취득시 눈은 뜬 상태로 입은 다물어진 상태로 취득되어야 한다. 셋째, 얼굴부분이 기울어지거나 회전되지 말아야 한다. 넷째, 얼굴부분이 그림자 등으로 가려져서는 안된다. 다섯째, 수염, 머리카락 등이 얼굴을 가려서는 안된다. 따라서, 본 논문에서는 위의 제약조건들 하에서 정규화된 눈과 입의 크기와 서로간의 위치관계에 대한 특징값 5개를 사용하여 얼굴인식 시스템을 구성하였다. 이 때 사용한 인식 알고리즘은 기존의 방법으로 널리 사용된 유클리디안 거리법,<sup>[19]</sup> 삼각형<sup>[6]</sup>과 사다리꼴 멤버십 함수<sup>[20]</sup>를 이용한 인식, NN,<sup>[21]</sup> 뉴로-퍼지 알고리즘,<sup>[19]</sup> 그리고 HMM-NN을 이용해 제안한 알고리즘들을 사용했다. 영상취득시 임의의 시간간격을 주고 영상을 취득하였으며 각각의 특징값은 미간사이의 거리값으로 정규화한 5개의 특징값으로 데이터베이스를 구축했다.

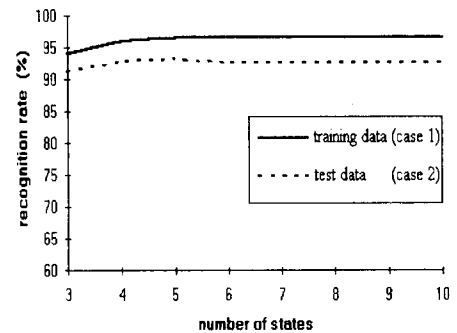
그리고 각 인식을 실험에서는 데이터베이스로 300장의 영상을, 나머지 300장은 인식대상으로 인식률을 구하는 데 사용했다.

기존의 인식방법으로 많이 사용되는 방법은 유클리디안 거리<sup>[9]</sup>에 의한 방법으로 이에 대한 정규화 방법과 인식률은 기존의 방법을 따랐고, 식 (11)로 정규화한 경우에 인식률이 가장 높았고 이 결과는 기존의 논문<sup>[18,19]</sup>에서의 결과와 비슷했다.

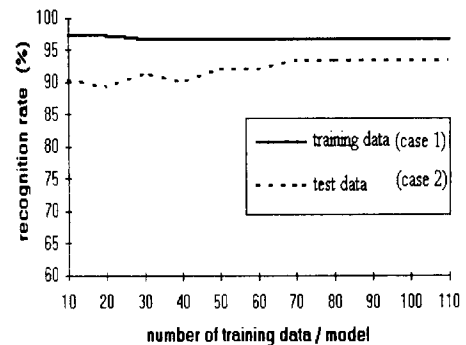
$$\sigma_i^2 = \frac{1}{K} \sum_{k=1}^K \frac{1}{L} \sum_{l=1}^L (r_{ikl} - r_{ik})^2, \quad 1 \leq i \leq 5 \quad (11)$$

여기서,  $\sigma_i$ 는  $i$ 번째 특징의 정규화 값이고,  $r_i$ 는 데이터베이스의 특징값,  $K$ 은 인식할 사람수,  $L$ 은 데이터베이스의 수를 각각 나타낸다. 그리고 퍼지 멤버십 함수<sup>[19]</sup>를 이용하여 인식하는 경우, 언어변수의 수와 근접한 언어변수 간의 min-max값에 따라 인식결과가 다르게 나타나고 min-max값이 0.8일 때 삼각형과 사

다리꼴형 멤버십 함수를 이용한 경우 인식률이 각각 85.7%, 86.3%로 나타났다. 일부 영상 중에는 각각의 특징값이 서로 차이가 작기 때문에 이런 특징값은 유클리디안 거리에 의한 인식률을 떨어뜨리는 원인이 되므로 애매모호한 특징값을 처리하기 위해 퍼지 멤버십 함수를 이용하면 본 실험과 같이 유클리디안 거리에 의한 것보다 인식률이 약 1.3% 높게 나왔다. 그리고 NN<sup>[21]</sup>을 이용한 인식에 있어서는 기존의 오류 역전과 알고리즘의 단점인 학습시간과 국부 최소값을 극복하기 위해 학습률과 관성항을 학습시 가중치의 변화와 함께 변화시키는 방법을 이용했고, 학습률과 관성항 값은 실험적으로 결정했으며 인식률은 90.3%이었다. 그리고 뉴로-퍼지 알고리즘<sup>[19]</sup>인 경우 입력단에 특징값 5개를 15개의 퍼지값으로 퍼지화한 후, NN의 입력으로 사용하였으며, 인식률은 91.3%이었다.



(a)



(b)

그림 4. HMM의 인식률

(a) 상태수에 따른 인식률 (b) 학습데이터 수에 따른 인식률

Fig. 4. Recognition rate of the HMM algorithm. (a) Recognition rate as a function of the number of states, (b) Recognition rate as a function of the number of training data.

그림 4(a)는 HMM의 상태수에 따른 인식률의 변화를 보여주는데, HMM에 있어서 테스트패턴과 학습패턴인 경우 상태수가 5일 때 인식률이 각각 93.3%, 96.7%로 가장 높았으며 상태수를 더 증가시켜도 인식률은 일정함을 실험적으로 관찰할 수 있었고, HMM-NN의 입력데이터는 HMM의 인식률이 상태수가 5일 때 가장 높으므로 상태수를 5로 고정시키고 Viterbi 알고리즘을 사용해서 NN의 입력데이터를 만든 후 실험했다. 그리고 학습시 학습패턴의 수에 따른 인식률의 변화는 그림 4(b)에서와 같이 각 사람에 해당되는 학습패턴의 수를 처음 10에서 시작해서 같은 학습패턴의 수를 계속해서 10만큼 붙여나가면서 학습패턴의 수를 110일 때까지 테스트패턴과 학습패턴에 대한 인식률을 구했으며 테스트패턴의 경우 인식률은 서서히 증가하면서 학습패턴의 수가 70개일 때부터 93.3%로 일정하게 되었고, 학습패턴의 경우 인식률은 처음 학습패턴의 수가 10, 20일 때 97.3%로 높았다가 점점 그 수가 증가하면서 96.7%로 감소하면서 일정하게 됨을 볼 수 있는데 이는 HMM 알고리즘이 학습시 학습패턴의 수가 인식률에 어느 정도 영향을 미치나, 학습패턴의 수를 계속 증가시키면 인식률은 더 이상 학습패턴의 수에 영향을 받지않는 성질이 있음을 말해준다. 따라서 HMM-NN의 실험시 전단계인 HMM은 학습패턴의 수를 70, 상태수를 5로 고정시키고 NN의 입력데이터를 만들었다. 그리고 HMM의 오인식률은 테스트패턴에 대해 약 6%정도 즉, 18명정도가 오인식된 사람인데 실제로 HMM에 의한 Viterbi 알고리즘으로 이 영상을 찾아보면 주로 안경영상에서 안경테 주위의 조명의 밝기에 의해 눈위치가 제대로 찾아지지 못하는 경우에 해당되고, 이 오인식된 영상에서 추출한 데이터를 처리하기 위해 HMM의 후단계에 NN을 사용하였다.

NN-HMM의 경우에도 그림 4에 나타난 결과와 비슷한 결과를 보였고 최종 인식률은 95.7%로 나타났다. 따라서, HMM에서 인식률을 높이려고 전단계 혹은 후단계에 NN을 결합해서 실험했는데, 후단계에 NN을 결합한 HMM-NN은 HMM만 사용했을 때보다 인식률이 크게 개선되었다. 전단계에 NN을 결합한 NN-HMM은 NN의 출력층에서 가장 큰 값을 가지는 노드 1개에 대해 레이블링하는 것보다 연속적인 큰 값 2, 3개에 대해 레이블링하면 NN에서 오인식될 사람을 고려하여 HMM의 입력 데이터를 만드는 것과 같아서 인식률이 3개에 대해 레이블링한 것이 가장 높았고, 3

개에 대해 레이블링한 것의 인식률이 2개의 레이블링보다 인식률이 크게 개선되지 않은 것은 대부분이 NN의 출력층에서 2명의 사람에서 추출한 특징값들이 서로 비슷하고 출력값도 다른 노드에 비해서 매우 크므로 레이블링을 3개로 하더라도 인식률이 크게 증가하지 않았다.

그리고 HMM-NN에서 후단계인 NN의 최적의 중간층 노드의 수는 기존의 방법<sup>[18,19]</sup>과는 다르게 최소한의 노드 수를 가지고 실험했다. 즉, 출력층의 노드의 수가 실험대상의 사람수와 같으므로 만약 인식할 사람 수가  $K$ 이면 최소한  $\log_2 K$ 개<sup>[30]</sup> 이상되었을 때 수렴 속도가 빠르므로 본 실험에서는 30명에 대한 인식을 고려하면 최소한 5개 이상의 중간층 노드의 수를 가져야 하기 때문에 처음에 노드의 수 5개를 가지고 인식률을 계산하면서 중간층의 노드의 수를 증가시켜 나갔다. 특히 NN의 입력데이터인 HMM의 출력값은 인식할 사람과 나머지 다른 사람들과 비교해 보면 매우 큰 차이가 나므로 입력데이터가 마치 two-class 문제나 XOR 문제에 적절한 입력데이터와 비슷하므로 NN만으로 실험한 경우 28개보다 비교적 적은 10개 이상의 중간층 노드의 수를 가질 때 인식률이 더 이상 변하지 않아서 최적의 중간층 노드의 수를 10으로 결정했고, 입력패턴특성상 쉽게 입력패턴을 선형 분리할 수 있으므로 NN만의 실험보다 수렴속도와 인식률이 크게 증가했다.

표 1. 기존의 인식 알고리즘과 제안한 인식 알고리즘의 인식결과 비교

Table 1. Comparison of recognition rate of the conventional and proposed algorithms.

인식방법		인식률 (%)
유클리디안 거리법 <sup>[9]</sup>		85.0
퍼지 알고리즘	삼각형 멤버쉽함수 <sup>[6]</sup>	85.7
	사다리꼴형 멤버쉽함수 <sup>[20]</sup>	86.3
NN 알고리즘 <sup>[2]</sup>		90.3
뉴로-퍼지 알고리즘 <sup>[18,19]</sup>		91.3
HMM 알고리즘 <sup>[21]</sup>		93.3
NN-HMM 알고리즘		95.7
HMM-NN 알고리즘		100.0

실험결과, 표 1에서 보듯이 제안한 HMM-NN 알고리즘이 인식률이 가장 높음을 알 수 있었다. 이것은 많

은 실험 영상 중에서 영상 취득 조건에 맞는 것만을 가지고 실험한 것에 그 원인이 있다. 따라서 개선해야 될 점은 영상 취득 조건을 좀 더 완화시킨 후의 얻은 영상으로 인식률을 높이는 것이다. 특히 안경영상의 경우 정확한 특징값을 추출할 수 있는 방법과 영상취득 시 취득조건에 거의 영향을 받지않은 보다 효율적인 특징에 대한 연구가 필요하다.

## V. 결 론

본 논문에서는 얼굴의 중요 특징인 눈, 코, 그리고 입의 위치에 따라 5개의 특징을 이용한HMM-NN 얼굴인식 알고리즘을 제안하였다. 전처리 부분에서는 올바른 특징추출을 위해서 배경과 얼굴부분을 정확히 분리하였으며, 특징추출 단계에서는 얼굴에 대한 지식을 이용하여 눈, 코, 그리고 입의 위치를 찾았다. 이러한 위치관계를 이용하여 5개의 특징을 추출하여 인식단으로 넘겨주었다. 인식단에서는 우수한 모델링을 할 수 있는 HMM을 전단계, 그리고 정적 데이터 패턴에 대해 우수한 분류능력이 있는 NN을 후단계에 결합함으로써 HMM의 오인식된 결과를 NN에서 효과적으로 처리하여 전체적인 인식률을 높일 수 있었다.

앞으로의 연구과제로 본 논문에서 사용한 5개의 특징 외에 영상취득시 주위의 환경 변화에 덜 민감한 특징의 추출과 좀 더 효율적인 인식 알고리즘에 대한 연구가 필요하다.

## 참 고 문 헌

- [1] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expression: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, Jan. 1992.
- [2] Y. K. Ham, S. Y. Lee, and R.-H. Park, "Knowledge-based face recognition using neural networks," in *Proc. Int. Joint Conf. Neural Networks*, pp. 2125-2128, Nagoya, Japan, Oct. 1993.
- [3] 이상영, 함영국, 박래홍, "지식에 기초한 특징추출과 역전파 알고리즘에 의한 얼굴인식," 전자공학회논문지-B, 제 31-B권, 제 7호, pp. 119-128, 1994년 7월
- [4] M. D. Kelly, "Edge detection in pictures by computer using planning," in *Machine Intelligence*, B. Meltzer and D. Michie, Eds., Edinburgh University Press, Edinburgh, vol. 6, pp. 397-409, 1971.
- [5] I. Craw, H. Ellis, and J. R. Lishman, "Automatic extraction of face features," *Pattern Recognition Letters*, vol. 5, no. 2, pp. 183-187, Feb. 1987.
- [6] K. M. Lim and K. W. Oh, "A face recognition system using fuzzy logic," in *Proc. First Korean-Japan Joint Conf. Computer Vision*, pp. 506-509, Seoul, Korea, Oct. 1991.
- [7] 박용근, 김영일, "ART와 다층 퍼셉트론을 이용한 화상인식 시스템의 성능 분석," 전자공학 회계종합학술대회 논문집, 제 15권, 제 1호, pp. 492-496, 부산, 1992년 6월
- [8] K. Matsuda, T. Kageyama, and T. Aibara, "Application of the second order statistics for the recognition of human faces," in *Proc. Int. Conf. Image Processing*, pp. 543-546, Singapore, Sep. 1992.
- [9] L. D. Harmon, S. C. Kuo, P. F. Ramig, and U. Raudkiv, "Identification of human face profiles by computer," *Pattern Recognition*, vol. 10, no. 2, pp. 301-312, Mar. 1978.
- [10] T. Aibara, K. Ohue, and Y. Oshita, "Human face profile recognition by a P-Fourier descriptor," *Optical Engineering*, vol. 32, no. 4, pp. 861-863, Apr. 1993.
- [11] H. Agawa, G. Xu, Y. Nagoshima, and F. Kishino, "Image analysis for face modeling and facial image reconstruction," in *Proc. SPIE Visual Communication and Image Processing 90*, vol. 1360, pp. 1184-1197, Lausanne, Switzerland, Oct. 1990.
- [12] H. S. I. Horace and M. C. N. James, "Human face recognition using Dempster-Shafer theory," in *Proc. Int. Conf. Image Processing '94*, vol. II, pp. 292-295, Austin, Texas, Nov. 1994.
- [13] A. J. Colmenarez and T. S. Huang, "Frontal view face detection," in *Proc. SPIE Visual Communications and Image*



- Processing '95*, vol. 2501, pp. 90-98, Taipei, Taiwan, May 1995.
- [14] L. C. D. Silva, K. Aizawa, and M. Hatori, "Detection and tracking of facial features." in *Proc. SPIE Visual Communications and Image Processing '95*, vol. 2501, pp. 1161-1172, Taipei, Taiwan, May 1995.
- [15] X. Xie, R. Sudhakar, and H. Zhuang, "On improving eye feature extraction using deformable templates," *Pattern Recognition*, vol. 27, no. 6, pp. 791-799, Jan. 1994.
- [16] G. S. Robinson, "Edge detection by compass gradient masks," *Computer Graphics Image Processing*, vol. 6, no. 5, pp. 492-501, Oct. 1977.
- [17] C. R. Giardina and E. R. Dougherty, *Morphological Methods in Image and Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, pp. 1-35, 1988.
- [18] 이상영, 함영국, 박래홍, "뉴러 퍼지 알고리즘을 이용한 얼굴인식," 전자공학회논문지-B, 제 32-B권, 제 1호, pp. 50-63, 1995년 1월
- [19] S. Y. Lee, Y. K. Ham, and R.-H. Park, "Recognition of human front faces using knowledge-based feature extraction and neuro-fuzzy algorithm," in press, *Pattern Recognition*, vol. 29, no. 11, Nov. 1996.
- [20] Y. K. Ham, S. Y. Lee, and R.-H. Park, "Fuzzy-based recognition of human front faces using the trapezoidal membership function," in *Proc. FUZZ-IEEE/IFES '95*, vol. IV, pp. 1799-1806, Yokohama, Japan, Mar. 1995.
- [21] 윤강식, 함영국, 박래홍, "은닉 마르코프 모델을 이용한 정면 얼굴인식," 1995년도 대한전자공학 회 회로 및 시스템 연구회/전력전자 연구회 합동 학술발표회 논문집, pp. 95-98, 1995년 10월
- [22] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 267-293, Feb. 1989.
- [23] D. J. Burr, "Experiment on neural net recognition of spoken and written text," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-36, no. 7, pp. 1162-1168, July 1988.
- [24] 송현경, 이영진, "역전파 학습시 중간층의 역할 분석과 노드 수 축소 방법," 전자공학회논문지-B, 제 28-B권, 제 3호, pp. 65-78, 1991년 3월
- [25] R. P. Lippmann and E. Singer, "Hybrid neural-network/HMM approaches to word-spotting," in *Int. Conf. Acoustics, Speech, Signal Processing '93*, vol. I, pp. 565-568, Minneapolis, USA, Apr. 1993.
- [26] A. Kundu, Y. He, and P. Bahl, "Recognition of handwritten word: First and second order hidden Markov model based approach," *Pattern Recognition*, vol. 22, no. 3, pp. 283-297, Aug. 1989.
- [27] W. S. Kim and R.-H. Park, "Off-line recognition of handwritten Korean and alphanumeric characters using hidden Markov models," in press, *Pattern Recognition*, 1995.
- [28] P. L. Cerf, W. Ma, and D. V. Compennolle, "Multilayer perceptrons as labelers for hidden Markov models," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 1, pp. 185-193, Jan. 1994.
- [29] G. Rigoll, "Maximum mutual information neural networks for hybrid connectionist-HMM speech recognition systems," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 1, pp. 175-184, Jan. 1994.
- [30] 심재형, 이종혁, 윤태훈, 김재창, 이양성, "신경망을 이용한 모음의 학습 및 인식 방법," 전자 공학 회논문지, 제 27권, 제 11호, pp. 144-150, 1990년 11월

— 저 자 소 개 —



尹 康 植(準會員)  
 1995년 서강대학교 전자공학과 졸업(공학사). 현재 서강대학교 전자공학과 대학원 석사과정 재학중. 주관심분야는 패턴 인식과 컴퓨터 비전



威 永 國(準會員)  
 1990년 2월 서강대학교 전자공학과 졸업(공학사). 1992년 2월 서강대학교 대학원 전자공학과 졸업(공학석사). 1996년 서강대학교 대학원 전자공학과 졸업(공학박사). 현재 국방정보체계연구소 선임연구원. 주관심분야는 패턴 인식과 컴퓨터 비전



朴 來 弘(正會員)  
 1976년 서울대학교 전자공학과 졸업(공학사). 1979년 서울대학교 대학원 전자공학과 졸업(공학석사). 1982년 Stanford Univ. 졸업(공학석사). 1984년 Stanford Univ. 졸업(공학박사). 현재 서강대학교 전자공학과 교수. 주관심분야는 컴퓨터 비전, 영상부호화, 패턴 인식 등임