

KL 변환과 신경망을 이용한 개인 얼굴 식별

論 文

48A-1-10

Human Face Identification using KL Transform and Neural Networks

金容柱*, 池承桓**, 柳在炯***, 金貞煥***, 朴玟用§

(Yong-Joo Kim* · Seung-Hwan Ji** · Jae-Hyung Yoo*** · Jung-Hwan Kim*** · Mignon Park§)

Abstract - Machine recognition of faces from still and video images is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neural networks. In addition, human face identification has numerous applications such as human interface based systems and real-time video systems of surveillance and security. In this paper, we propose an algorithm that can identify a particular individual face. We consider human face identification system in color space, which hasn't often considered in conventional methods. In order to make the algorithm insensitive to luminance, we convert the conventional RGB coordinates into normalized CIE coordinates. The normalized-CIE-based facial images are KL-transformed. The transformed data are used as the input of multi-layered neural network and the network are trained using error-backpropagation methods. Finally, we verify the system performance of the proposed algorithm by experiments.

Key Words : 얼굴 인식, 얼굴 식별, KL 변환, 역전파 신경망, CIE 좌표계

1. 서 론

정지 영상이나 동영상으로부터 다수의 얼굴을 인식하거나 한 개인의 얼굴을 식별하는 기술은 영상 처리나 패턴 인식, 컴퓨터 비전과 신경망 등과 같은 다양한 분야에서 활발히 연구되어지고 있으며, 상업적으로나 법적으로 수많은 응용 사례를 가지고 있다. 이러한 얼굴 인식 기술(Face Recognition Technology, FRT)은 여권, 신용 카드, 주민 등록증, 운전 면허증, 그리고 범인의 얼굴 사진 등과 같은 제한된 형태의 정지 영상을 사용하는 방법에서부터 동영상 감시 시스템 등과 같은 실시간 인식에 이르기까지 그 종류가 다양하다.

얼굴 영상 인식 기술은 일반적으로 주어진 배경의 정지 영상이나 동영상에 대하여 입력 영상 내의 한 명 이상의 사람이 주어진 데이터베이스 내에 존재하는지 확인하는 것으로 정의될 수 있으며, 탐색 영역을 좁히기 위해서 인종, 나이, 성별 등과 같은 부수적인 정보들이 사용될 수도 있다. 일반적으로 얼굴 인식의 문제는 대부분 영상으로부터의 얼굴 영역의 분리, 얼굴 특징점 추출, 그리고 분류 과정의 세 부분으로 나누어 처리되어진다[1].

인간과 기계의 얼굴 인식에 관한 수많은 연구들은 과거 20년이 넘는 기간 동안 심리학, 신경 과학, 그리고 공학 분야에서 폭넓게 진행되어 왔다. 그러나 몇몇 예외적인 경우를 제외하고는 얼굴 영상의 인식 기술은 심리학이나 신경생리학과는 별개로 발전하였다. 얼굴 인식 기술의 초기 단계인 1970년 중반 동안에는 입력 얼굴로부터 특징점을 추출하여 저장되어 있는 특징점과 비교하는 전형적인 패턴 분류 방법이 주류를 이루었다. 하지만 많은 제약점과 낮은 성능으로 인해 1980년 대에서는 거의 연구가 되지 않았었다. 그러나, 휴먼 인터페이스 중심의 연구가 활발해졌던 1990년 대에 이르러서는 이 얼굴 인식 기술이 매우 중요한 이슈로 성장하였다. 특히 최근 몇 년 동안 정면 얼굴 영상을 사용하는 인식 방법에 있어서 얼굴 영역의 분리와 눈, 코, 입 등과 같은 특징을 추출하는 부분에 많은 진전이 있었고, 영상의 통계적 처리 방법이나 퍼지나 신경망 등과 같은 여러 지능 알고리즘에 의한 접근 방법 등이 개발되었다[2][3].

정지 영상을 사용한 얼굴 인식 방법은 몇 가지 장단점을 가지고 있다. 예를 들어 여러 범 죄자의 사진 중에서 특정인의 얼굴을 찾아내는 문제를 다룰 경우에는 사진 촬영시 가해지는 여러 가지 제약 조건들에 의해서 얼굴의 분리가 다소 쉬워질 수 있으나, 공공 장소나 사무실과 같은 복잡한 배경에 인물이 들어 있는 사진이나 정지 영상에서는 얼굴의 분리가 어렵다[1]. 한편 감시 카메라로부터 얻어진 동영상에서는 사람의 움직임을 단서로 해서 보다 쉽게 얼굴을 분리해낼 수 있다[2]. 본 논문은 위에서 언급한 얼굴 인식의 세 가지 중 두 번째와 세 번째에 초점을 맞춘다. 배경이 복잡한 경우, 첫 번째의 얼굴 영역의 분리 과정은 매우 어려운 과정이므로 이 과정에 실패할 경우 그 이후 실패가 무의미하므로 본 논문에서는 연구 대상에서 제외한다. 또 본 논문

* 正 會 員 : LG電子 Multimedia연구소 研究員

** 正 會 員 : 未來産業 MRC 研究員

*** 正 會 員 : 延世大 工大 電子工學科 博士課程

§ 正 會 員 : 延世大 工大 電子工學科 教授 · 工博

接受日字 : 1997년 8월 18일

最終完了 : 1998년 11월 26일

에서는 얼굴 인식의 기초가 되는 한 개인의 얼굴을 식별하는 경우에 대해서만 다룬다. 대부분의 개인 얼굴 식별 알고리즘은 다수의 얼굴을 인식하는 알고리즘으로 확장 가능하므로 본 논문의 연구 목표는 개인 식별의 문제로 국한한다.

본 논문에서는 컬러 영상을 얼굴 식별 시스템의 입력으로 사용한다. 그림 1에 제안된 얼굴 식별 알고리즘을 나타내었다. RGB(red-green-blue) 좌표계로 표현되는 입력 영상을 정규화된 CIE(Commission Internationale de L'Eclairage 1931, 1964) 좌표계[4][5]로 변환하여 밝기의 변화에 따른 영상의 변화를 제거하였다. 또한 컬러 영상을 사용하는 경우에는 흑백 영상에 비해 3배의 정보를 가지므로 얼굴 영상을 데이터베이스화하기 위해 사용되는 기억 장소도 그만큼 필요하게 되며, 식별 알고리즘의 실행에 사용되는 시간도 길어진다. 이와 같은 문제를 해결하기 위해 본 논문에서는 KL 변환을 사용하여 얼굴 영상 정보의 차원을 줄인다. 기존의 KL 변환을 이용한 얼굴 식별 알고리즘에서는 KL 변환된 벡터의 유클리드 거리를 비교하여 얼굴을 식별하였다[2]. KL 변환된 영상 벡터는 2차원 영상을 1차원의 벡터로 변환한 것이므로 영상의 한 쪽 방향(벡터화 되면서 서로 연결되는 방향)의 정보는 남아있지만 다른 한 쪽 방향의 정보를 잃어버리게 된다. 그러므로 유클리드 거리를 사용하는 경우 한 쪽 방향의 정보의 연관성에 대해서 비교할 수 없다. 본 논문에서는 학습된 역전과 신경망을 이용한 비선형 거리 함수를 이용하여 얼굴 영상을 비교한다. 거리 함수가 신경망에 의해 학습되므로 유클리드 거리를 이용하는 경우, 비교에 포함되지 않는 벡터상의 서로 떨어진 원소들 끼리의 관계도 비교에 영향을 미치게 되어 식별 성능을 높일 수 있다.

본 논문은 정규화된 컬러 영상과 KL 변환 및 신경망을 이용한 얼굴 식별 알고리즘을 제안하고, 그 성능을 RGB 영상을 사용한 경우와 비교한다. 본 논문의 2 장에서는 정규화된 컬러 좌표계를 이용한 컬러 얼굴 영상의 처리에 대해 설명하고, 3장에서는 KL(Karhunen-Loève) 변환에 대해서 설명하며, 4 장에서는 신경망을 이용한 개인 얼굴의 학습과 식별 시스템에 대해서 기술한다. 5 장에서는 실험을 통해 알고리즘의 성능을 확인하며, 6 장에서는 결론을 내리기로 한다.

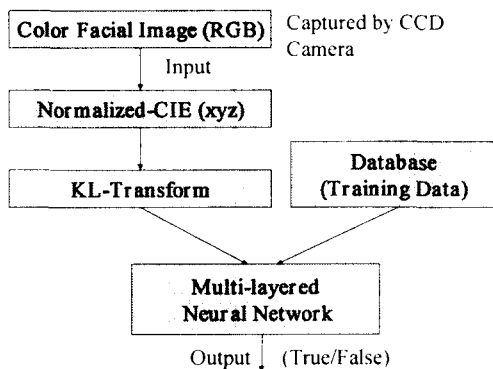


그림 1 얼굴 식별 알고리즘.
Fig. 1 The face identification algorithm.

2. 컬러 얼굴 영상의 처리

본 논문에서 제안한 알고리즘에서는 RGB 좌표로 표현되는 얼굴 영상을 정규화된 CIE 좌표계변환하여 얻어진 색상(chromaticity) 성분값을 벡터 성분으로 이용한다. CIE 좌표계는 얼굴의 인식과 식별에 RGB보다 좋은 성능을 보이며, 정규화된 CIE 좌표계[4][5]에서는 휘도(luminance) 성분만을 제거할 수 있기 때문에 조명의 영향에 크게 좌우되지 않는 매우 강한 성능을 기대할 수가 있다.

세 가지 성분을 이용하여 모든 색을 재구성할 수 있다는 것은 오래 전부터 알려져 온 사실이다. 이 중에서도 RGB 좌표계는 빨강, 녹색, 파랑의 세가지 색으로 모든 색을 표현하는 좌표로 색 좌표계에서 가장 많이 사용되어지고 있다. 하지만 색으로부터 휘도 성분을 얻어내는 경우에는 RGB 좌표계에서는 직접 구할 수가 없는 반면, CIE 좌표계에서는 휘도 성분만을 하나의 주성분으로 취하고 있기 때문에 휘도 값을 직접 얻어낼 수가 있다. RGB 좌표계로 되어 있는 입력 영상은 다음과 같은 행렬 변환식을 이용하면 CIE 좌표계로 변환할 수 있다[4].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5943 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

색깔은 크게 색상과 휘도의 두 성분으로 나누어질 수가 있는데 CIE 좌표계의 Y 값은 휘도 성분만을 나타낸다. 이 휘도 성분은 조명에 매우 민감하기 때문에 이 성분을 제거하면 조명의 변화에 대하여 안정적인 식별 성능을 얻을 수 있다. X, Z 값에도 휘도 성분이 섞여 있으므로 다음과 같은 정규화 과정을 통해서 휘도 성분을 제거할 수 있다[4].

$$D = X + Y + Z \quad (2)$$

$$x = \frac{X}{D} \quad y = \frac{Y}{D} \quad z = \frac{Z}{D} \quad (3)$$

여기서 $x + y + z = 1$ 이므로 x, y 두 값만 가지고 색상 성분을 표현할 수 있다.

3. Karhunen-Loève 변환을 이용한 벡터 차원 감소

호텔링 변환(Hotelling Transform), 혹은 주성분 분석 기법(Principal Component Analysis, PCA)라고도 불리는 KL 변환은 벡터 표현의 통계적인 특징에 기반을 둔다. KL 변환은 데이터의 통계적인 처리에 의해 입력 데이터들의 중요한 성분을 구해내고 결과적으로 중요하지 않은 성분은 제거하여 데이터의 압축 효과를 얻을 수 있는 선형 알고리즘이다[6]. 본 논문에서는 얼굴 영상 데이터베이스와 얼굴 식별 시스템에 사용되는 데이터의 양을 줄이기 위해 KL 변환을 이용한다. KL 변환을 이용하여 정보량을 감소시키면 얼굴 정보를 저장하기 위한 기억 공간이 감소하며 얼굴 식별에 사용되는 신경망의 학습 시간도 감소한다.

입력 얼굴의 2차원 영상은 $(N \times N)$ 다음과 같은 N^2 차

원 벡터 \mathbf{x} 로 변환될 수 있다:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N^2} \end{bmatrix} \quad (4)$$

여기서 x_i 는 벡터 \mathbf{x} 의 i 번째 원소이다.

\mathbf{x} 의 모집단의 공분산 행렬(covariance matrix)을 \mathbf{C}_x 라 하고, \mathbf{C}_x 의 고유벡터(eigenvector)와 고유값(eigenvalue)를 $\mathbf{e}_i, \lambda_i (i=1, 2, \dots, N^2)$ 라 하자. 이 때 고유벡터와 고유값은 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N^2}$ 을 만족하는 순서로 정렬되어 있다고 가정한다. \mathbf{C}_x 의 고유벡터가 각각 행벡터를 구성하는 변환 행렬 \mathbf{A} 는 다음과 같이 나타낼 수 있다:

$$\mathbf{A} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1N^2} \\ e_{21} & e_{22} & \dots & e_{2N^2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N^2 1} & e_{N^2 2} & \dots & e_{N^2 N^2} \end{bmatrix} \quad (5)$$

여기서 e_{ij} 는 i 번째 고유벡터의 j 번째 원소이다. 이 때 KL 변환식은 다음과 같이 표현된다:

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x) \quad (6)$$

여기서 \mathbf{m}_x 은 \mathbf{x} 의 모집단의 평균 벡터이다.

\mathbf{C}_x 는 실대칭 행렬(real symmetric matrix)이므로 항상 직교 정규벡터를 구할 수 있으며, $\mathbf{A}^{-1} = \mathbf{A}^T$ 이 된다[8]. 그러므로 \mathbf{x} 는 변환된 벡터 \mathbf{y} 로부터 다음 식에 의해 복원될 수 있다.

$$\mathbf{x} = \mathbf{A}^T \mathbf{y} + \mathbf{m}_x. \quad (7)$$

\mathbf{C}_x 의 모든 고유벡터로 이루어진 \mathbf{A} 대신 고유벡터 중 다른 것보다 큰 고유값을 갖는 K 개의 고유벡터 ($\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K$)로 구성된 행렬 \mathbf{A}_K 를 고려해 보자. \mathbf{A}_K 를 \mathbf{A} 대신 KL 변환 행렬로 사용하는 경우, 변환 벡터 \mathbf{y} 는 다음 식에 의해 K 차원 벡터가 된다:

$$\mathbf{y} = \mathbf{A}_K(\mathbf{x} - \mathbf{m}_x) \quad (8)$$

또한 역변환 식 (7)은 다음과 같은 근사식으로 표현된다:

$$\hat{\mathbf{x}} = \mathbf{A}_K^T \mathbf{y} + \mathbf{m}_x \quad (9)$$

큰 고유값에 해당하는 고유벡터는 원래의 데이터를 표현

하는 데 상대적으로 크게 기여하게 되며, 기여도가 충분히 큰 벡터만으로 구성된 변환 행렬 \mathbf{A}_K 를 이용하는 것에 의해 KL 변환으로 데이터의 크기를 줄이게 된다. 정규화된 CIE 좌표계에서 표현되는 컬러 영상의 한 픽셀은 x 와 y 의 두 가지 요소로 표현 가능하므로, 얼굴 벡터의 원소 수는 픽셀 수의 두 배가 된다. 그림 2에서와 같이 $N \times N$ 픽셀로 구성된 2차원 영상은 $2N^2$ 차원의 벡터로 변환되며, 변환된 CIE 얼굴 벡터는 K 차원으로 압축된다.

원래의 데이터 \mathbf{x} 와 K 개의 고유벡터만을 사용하는 경우 얻어지는 근사값 $\hat{\mathbf{x}}$ 사이의 평균 제곱 에러(mean square error) R 은 다음 식과 같이 나타낼 수 있다:

$$R = \sum_{j=1}^{N^2} \lambda_j - \sum_{j=1}^K \lambda_j = \sum_{j=K+1}^{N^2} \lambda_j. \quad (10)$$

정규화된 평균 제곱 에러(normalized mean square error) \bar{R} 은 다음 식과 같이 모든 고유값의 합에 대한 평균 제곱 에러 R 의 비로 표현된다:

$$\bar{R} = \frac{R}{\sum_{j=1}^{N^2} \lambda_j} = \frac{\sum_{j=1}^{N^2} \lambda_j - \sum_{j=1}^K \lambda_j}{\sum_{j=1}^{N^2} \lambda_j} \quad (11)$$

위 식에서 $K=n$ 인 경우 \bar{R} 은 0이 되며, 이것은 모든 고유벡터를 사용하는 경우에 오차가 0이 됨을 의미한다. 또한 K 가 감소함에 따라 \bar{R} 은 증가하며, 정규화된 평균 제곱 에러의 최대값은 1이 된다. 이와 같이 \bar{R} 값은 KL 변환에 의한 근사화의 정보 보존율을 나타낸다고 생각할 수 있으므로, 본 논문에서는 이 값을 얼굴 벡터의 압축의 성능을 평가고 효용성을 입증하는 데 사용한다.

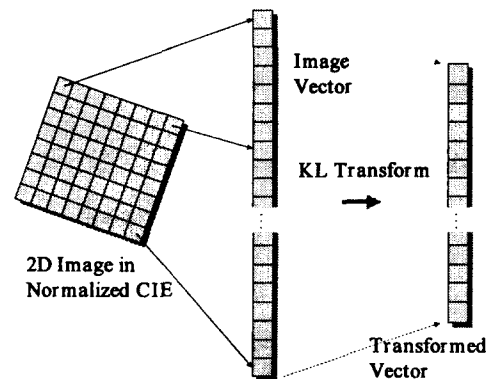


그림 2 2차원 영상의 벡터화와 KL 변환에 의한 벡터의 차원 감소.

Fig. 2 Vector representation of a 2D image and dimension reduction using the KL transform.

4. 얼굴 식별 시스템에 사용된 신경망

입력된 영상에서 얼굴을 포함한 사각 영역은 그림 2에서와 같이 일련의 한 벡터로 만들어진다. R, G, B 3개의 정보를 다룰 경우 한 픽셀은 벡터상에서 R, G, B의 연속된 3성분을 차지하며, 따라서 벡터의 총 성분 수는 얼굴을 이루는 영상 데이터 총 픽셀 수의 3배가 된다. 정규화된 CIE의 경우에는 x, y 의 연속된 두 성분을 차지하며 결과적으로 벡터의 총 성분 수는 얼굴을 이루는 영상 데이터 총 픽셀 수의 2배가 된다. 입력 영상으로부터 얻어진 벡터는 KL 변환에 의해 K 차원으로 압축된다.

KL 변환을 이용한 기존의 얼굴 식별 방법[2]에서는 주로 유클리드 거리를 이용하여 비교/식별 한다. 이 경우 얼굴 벡터의 생성시 픽셀의 좌우 관계는 직접적으로 비교에 반영되지만, 픽셀 상하 간의 상호 연관 관계는 상대적으로 반영 정도가 크게 떨어지므로 결과적으로 식별에 필요한 정보 일부를 활용하지 못하는 문제가 있다. 그러므로 본 논문에서는 얼굴 식별을 위한 거리 함수로 그림 3의 구조를 갖는 다층 역전파 신경망(backpropagation neural network)[7]을 이용한다. 다층 역전파 신경망은 비선형 함수를 학습시킬 수 있으므로 이를 이용하면 학습 과정에 의하여 이러한 잃어버린 픽셀 상하 간의 연관성을 다시 부각시킬 수 있다. 신경망을 사용하는 또 다른 장점으로, 같은 사람의 얼굴이라도 다양하게 입력될 수 있는 얼굴 영상을 학습에 의해 하나의 시스템으로 처리할 수 있다는 점이다. 유클리드 거리를 사용하는 경우에는 동일 인물의 다른 입력을 인식하려면 실험에 의해 인식 한계선을 다시 조정해야만 한다. 하지만 이 방법은 서로 다른 모습을 하고 있는 개인 얼굴 영상들이 특징 벡터 공간 안에 고르게 분포하고 있어야만 한다는 단점을 가진다. 만약 불규칙한 비선형적인 경계를 갖고 있는 경우에는 이 방법으로는 만족할만한 결과를 얻을 수 없다. 그림 3과 같은 다층 신경망은 비선형적인 경계선으로 나눌 수 있기 때문에 여러가지 다른 모습을 하고 있는 개인의 얼굴 영상도 학습에 의해 인식할 수 있다.

역전파 신경망은 교사 신호가 있는 학습(supervised learning) 알고리즘으로, 식별하고자 하는 인물의 추출된 얼굴 영역에 대하여 KL 변환된 입력에 대해서는 정(正, TRUE)의 출력력을, 또 다른 사람의 얼굴과 얼굴이 아닌 영역의 입력에는 부(否, FALSE)의 출력력을 주어 학습시킨다. 부의 출력력을 내는 데이터의 학습은 특정 개인 얼굴의 식별을 위한 경계면을 보다 정확하게 학습시키는 데 기여한다. 이부의 데이터는 다른 사람들의 얼굴 영상, 얼굴과 모양이 비슷한 영상, 또 얼굴과 색이 같은 영상, 그리고 얼굴과 모습이 전혀 다른 영상 등으로 구성하였다. 정, 부의 입력 영상은 모두 그림 2와 같이 벡터화 되어 KL 변환에 의해 K 차원으로 압축된 후 역전파 알고리즘의 입력 데이터로 사용된다.

본 논문에서는 신경망의 출력으로 입력된 영상이 학습된 얼굴인가 아닌가를 판단하는 기준으로 사용하며, 이상적인 식별자인 경우 '얼굴이다'(正, TRUE)와 '아니다'(否, FALSE)의 두 가지가 출력되어야 하지만, 역전파 신경망의 출력은 미분 가능한 함수로 구성되므로, 실제로는 0과 1사이

의 값을 갖도록 구성된다. 그러므로 얼굴 입력에 대해서는 1에 가까운 값이, 얼굴이 아닌 입력에 대해서는 0에 가까운 값이 출력되도록 학습된다.

본 논문에서 사용된 다층 역전파 시신경망은 입력층(input layer)과 1개의 은닉층(hidden layer), 출력층(output layer)으로 구성되며, 입력층의 뉴런 수는 20개, 은닉층의 뉴런 수는 30개, 출력 뉴런은 1개이다. 은닉층의 개수가 늘어날수록 약간 성능은 증가하지만 학습 시간은 계산량의 증가로 인하여 기하급수적으로 늘어난다. 실험 결과 하나의 은닉층을 사용하는 경우와 둘 이상의 은닉층을 사용할 때의 성능 차이가 크지 않았으므로 하나의 은닉층을 사용하였다. 은닉층의 뉴런(노드) 수를 증가시키며 실험한 결과 30개까지는 성능이 계속적으로 상승하지만 그 이상의 뉴런에 대해서는 성능 유지 또는 오히려 성능 하락의 역효과를 보였다. 가중치 수정시 사용되어지는 이득 값인 μ 은 0.08을 사용하였으며, 이 값 역시 실험에 의해 결정하였다. 입력층의 뉴런 수는 KL 변환을 통해 변환된 벡터의 성분의 개수와 같으며, 출력층은 목표 얼굴이 맞는가(TRUE)/틀린가(FALSE)이므로 1개의 뉴런을 사용한다.

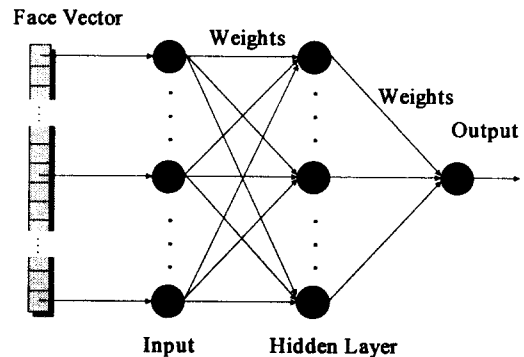


그림 3 얼굴 벡터 학습과 식별에 사용되는 다층 신경망의 구조.

Fig. 3 Multi-layered neural network for facial vector training and identification.

5. 얼굴 식별 시스템과 실험 결과

본 논문의 개인 얼굴 식별 시스템 구현을 위한 비전 시스템은 크게 컬러 디지털 CCD 카메라, 영상 데이터 캡처 보드, 알고리즘이 부여된 소프트웨어로 구성되어 있다. 소프트웨어에는 신경망 역전파 학습 알고리즘, 벡터의 KL 변환 처리 등으로 구성된다. 개인 얼굴 식별 시스템에서는 CCD 카메라로부터 개인 얼굴 영상을 획득해 학습 데이터를 생성한다. 실험은 카메라로부터 영상을 획득하여 얼굴 부분만을 추출, 기억되어 있는 얼굴과 비교하여 식별하는 것으로 한다. 개인의 얼굴을 학습시키기 위해 얼굴의 영상을 다수개 획득하여 얼굴 부분만 일정한 크기로 잘라내어 벡터화한 후 KL 변환시켜 역전파 신경망을 이용하여 얼굴 영상을 학습시킨다. 실험을 위한 입력 영상으로는 학습시 사용한 데이터 영상과 별도로 추출한 테스트 영상을 사용하였다.

본 논문에서의 실험에 사용된 원 영상은 CCD 카메라 입력으로부터 320×240 픽셀의 크기로 캡처(capture)된 24비트 RGB 영상으로, 이 중 추출된 얼굴 영상의 해상도는 40×40이며, 크기 2×2의 모자이크화 한 20×20의 영상을 시스템의 입력으로 사용한다. 실험에 사용된 시스템은 133MHz 펜티엄 CPU와 16M 바이트의 RAM을 사용하는 PC상에 구현되었다.

그림 4는 식별하고자 하는 사람의 얼굴 영상으로부터 추출된 입력 데이터의 예이다. 이와 같이 추출된 데이터는 신경망의 학습에 사용되는 정의 입력이 되는데, 본 실험에서는 100개의 추출된 영상 데이터를 이용한다. 또한 보다 정확한 학습 효과를 얻기 위해 부의 출력을 갖는 데이터의 경우 대해서도 100개의 데이터 영상을 추출하여 학습에 사용하였다. 이 100개의 데이터 영상에는 10명의 다른 사람의 얼굴 영상 7개씩, 총 70개의 얼굴 영상과 사람 얼굴과 색이 비슷한 물체 영상, 사람 얼굴과 모양이 비슷한 물체 영상 등을 포함한 얼굴이 아닌 영역의 영상 30개로 구성되어 있다. 이 영상의 예는 그림 5에 나타내었다. 따라서 학습에 사용된 총 데이터 수는 200개가 된다.

RGB 입력 영상에서는 하나의 픽셀이 R, G, B 세 개의 값을 가지므로 KL 변환의 입력으로 사용될 벡터 성분의 개수는 1200개(20×20×3)가 되며, 정규된 CIE 좌표계로 변환하는 경우에는 x , y 두 값만을 가지므로 데이터 벡터의 성분 개수는 800개(20×20×2)가 된다. 하나의 성분은 메모리에서 1 바이트(8 비트)의 크기를 가지며 0~255 사이의 값을 가진다.

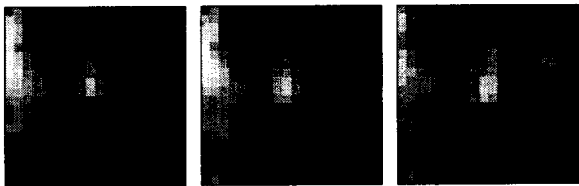


그림 4 신경망의 학습에 사용된 정(TRUE)의 데이터의 예.
Fig. 4 Examples of "true" face vectors in the neural network learning data.



그림 5 신경망의 학습에 사용된 부(FALSE)의 데이터의 예.
Fig. 5 Examples of "false" face vectors in the neural network learning data.

KL 변환은 입력 영상의 벡터 공간을 식별하고자 하는 목표 얼굴의 고유 공간으로 매핑시키는 역할을 한다. 이 고유 공간은 목표 얼굴에 대한 중요한 방향으로의 성분 값 분포를 나타낸다. 이 과정은 특정 개인의 얼굴의 고유 공간을

구하는 것이 목적이므로 정의 출력을 내는 100개의 영상에 대해서 변환 행렬을 구한다. 그림 4와 같은 영상 100 개에 대한 평균 벡터 \mathbf{m}_x 와 공분산 행렬 \mathbf{C}_x 를 구하고 공분산 행렬 \mathbf{C}_x 의 고유벡터와 고유값을 구한다. RGB 경우는 입력 영상의 벡터 차원은 1200×1이므로, 벡터 \mathbf{m}_x 는 1200×1, 행렬 \mathbf{C}_x 는 1200×1200의 차원을 갖고 따라서 고유값은 모두 1200개가 되고, 각각의 고유값에는 1200×1 차원의 고유 벡터를 가진다. 정규화된 CIE로 처리하는 경우는 \mathbf{m}_x ,

\mathbf{C}_x 각각 800×1, 800×800의 차원을 가지며, 고유값은 800개, 고유벡터는 800×1크기로 800개가 구해진다. KL 변환 행렬을 구하는데 걸리는 시간은 RGB인 경우 약 40분, CIE인 경우 약 15분 정도로, CIE의 경우 계산 시간이 적게 걸렸다. 이것은 CIE의 경우 행렬의 크기가 800×800이므로 RGB의 1200×1200인 경우보다 계산량이 적기 때문이다.

데이터를 압축시키려면 이 고유값 중에서 큰 값을 갖는 것을 몇 개 선택하고 그 고유값에 해당하는 고유벡터를 이용한다. 고유값이 크다는 것은 그만큼 입력 영상의 중요한 방향 성분임을 의미하는 것이다. 본 논문의 얼굴 식별 실험에서는 RGB, CIE 모두의 경우 20개의 고유값을 사용한다 ($K=20$). 표 1은 정규화된 CIE에서의 추출된 20개의 고유값들이다. 표에서 처음 20개의 고유값이 비교적 큰 수임을 알 수 있다. 결과적으로 20개의 고유벡터를 열 벡터로 취하는 행렬이 KL 변환의 변환 행렬이 되며 변환 행렬은 정규화된 CIE의 경우에는 800×20 차원이 된다. KL 변환의

\mathbf{A}_k 가 800×20 차원을 가지므로 입력 벡터 \mathbf{x} 가 들어오면 미리 구한 목표 개인 얼굴 영상의 평균 벡터 \mathbf{m}_x 와 변환 행렬 \mathbf{A}_k 로부터 20×1의 차원을 갖는 벡터 \mathbf{y} 를 구할 수가 있다. 이 벡터 \mathbf{y} 는 벡터 \mathbf{x} 의 공분산 행렬의 고유 벡터로의 방향값을 나타내며 20차원의 고유 공간에 분포한다. K 와 고유값을 값을 식 (11)에 대입하면 KL 변환에서의 정규화된 평균 제곱 에러를 계산할 수 있다. 주성분 수 K 값을 선택하는 기준은 식 (11)의 정규화된 에러 \bar{R} 의 크기가 작으면서 동시에 신경망의 입력으로 적절한 값이어야 한다. \bar{R} 은 K 값이 클수록 작아지지만, K 가 크면 사용되는 벡터의 차원이 커져 데이터 저장 공간이 많이 필요해지며 신경망의 입력의 갯수가 많아진다. 이 두 가지 관점에서의 적절한 균형(tradeoff)을 필요로 한다. 그림 6은 총 고유값 수 중에서 KL 변환을 위해 사용한 주요 고유값 수 K 값에 따른 \bar{R} 의 값을 나타낸 것이다. 실험에서 사용된 값인 $K = 20$ 인 경우, 정규화된 CIE에서 $\bar{R} = 0.1871$ 이며 이는 원래 데이터의 $1/\bar{R} = 0.813(81.3\%)$ 의 정보를 포함한다는 것을 의미한다. 그림 6에서 이 $K = 20$ 일 때의 \bar{R} 의 값을 '*' 표로 표시해 놓았다.

KL 변환을 통해 변환된 벡터의 20개 성분이 그대로 신경망의 입력이 되므로 역전파 신경망의 입력층의 뉴런 수는 20개가 되며 출력층은 목표 얼굴이 맞는가(TRUE)/틀린가(FALSE)이므로 1개의 뉴런을 사용한다. 이와 같이 구성된

역전과 신경망을 사용하여 출력이 정(TRUE)인 데이터 100개, 부(FALSE)인 데이터 100개, 합쳐서 200개의 데이터를 학습시켰다. 그림 7은 역전과 신경망 학습 오차를 나타낸 것이다. 총 200개의 학습 데이터들의 한차례의 학습 과정을 1세대라고 할 때 학습 데이터가 8000세대 정도 진행하면 역전과 학습 오차가 더 이상 크게는 줄어들지 않아 10000세대까지만 학습시켰다. 200개의 데이터로 신경망을 학습시키는 시간은 약 25분 정도이다. 신경망의 학습에 걸리는 시간은 RGB와 정규화된 CIE의 경우 계산량의 차이가 없으므로 같게 된다.

표 1 큰 수치를 갖는 목표 영상 데이터의 20개의 고유값(정규화된 CIE).

Table 1 The largest 20 eigenvalue(normalized CIE coordinate system).

No.	Eigenvalues				
5	760.6	283.7	265.8	175.4	130.8
10	112.8	94.8	81.1	67.9	57.8
15	52.1	43.2	39.2	36.7	35.4
20	29.5	25.1	21.9	20.8	19.3

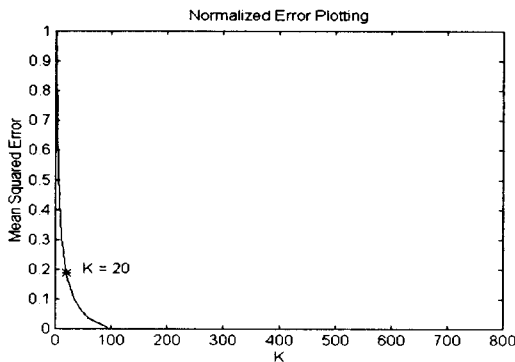


그림 6 K값에 따른 KL 변환의 정규화된 평균 제곱 에러.

Fig. 6 The normalized errors of the KL-transformed face vector.

신경망 학습 알고리즘은 시그모이드(sigmoid) 함수를 이용한 역전과 알고리즘[7]을 사용하였으며 오차 값은 오차 제곱의 평균값을 나타낸 것이다. 또 학습시 입력 데이터의 선별 방식은 순차식이 아닌 임의의 순서로 하였다. 가중치 수정시 사용되어지는 이득 값인 μ 은 0.08을 사용하였다. 미리 추출된 학습 데이터를 이용하여 신경망을 학습시킨 후, 학습에 사용된 데이터와 테스트 데이터를 사용하여 식별 성능을 확인하였다. 실험에 사용된 테스트 데이터를 구성하는 영상도 학습 데이터 영상과 마찬가지로 100개의 정의 출력을 갖는 목표 얼굴 영상과, 다른 10명의 얼굴 영상을 각각 8개씩 추출하여 만든 80개의 영상, 그리고 얼굴 그림이나 사람의 피부 내부 영역, 또는 이와 비슷한 색과 모양을 지닌 영상 20개로 구성한다. 조명의 영향을 알아보기 위해 이 200개의 테스트 영상은 10가지의 다른 조명 조건에서 추출되었다.

개인 얼굴 식별 실험에서의 식별 결과는 다음과 같은

네가지 종류로 구분되어질 수 있다. 먼저 식별이 성공하였을 때는 두가지 경우가 존재하는데, 식별하고자 하는 얼굴을 입력하였을 때 정의 출력을 주는 경우와 식별하고자 하는 얼굴이 아닌 입력에 대하여 부의 출력을 주는 경우가 있다. 본 논문에서는 전자의 경우인 전체 정의 데이터에 대한 정의 출력(성공)의 비율을 r_{TT} 라 나타내고 후자의 경우인 전체 부의 데이터에 대한 부의 출력(성공)의 비율을 r_{FF} 라고 정의한다. 또 식별이 실패한 경우에도 두가지 종류가 있는데, 정의 데이터에 대하여 부의 출력을 주는 경우와 부의 데이터에 대하여 정의 출력을 주는 경우로 나눌 수 있다. 앞 경우와 마찬가지로 본 논문에서는 전자의 경우를 r_{TF} , 후자의 경우를 r_{FT} 로 정의한다. r_{FT} 의 비율은 외부 사람들의 출력을 제한하는 감시·보안 시스템에서는 매우 중요한 성능 평가 요소이다.

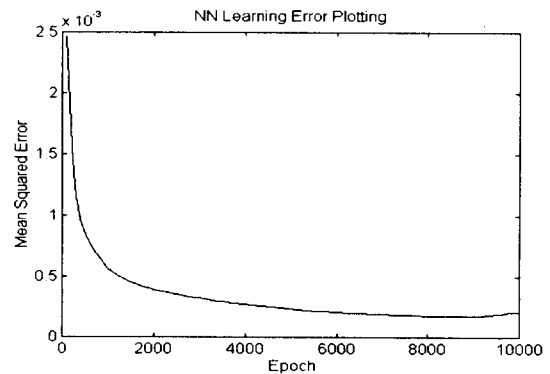


그림 7 역전과 신경망의 학습 과정에서의 에러 변화.

Fig. 7 The error graph of the backpropagation neural network in the face learning.

개인 얼굴 식별 실험 결과는 표 2와 3에 나타내었다. 표 2는 RGB 값을 입력으로 사용했을 때의 결과이고, 표 3은 정규화된 CIE 값을 입력으로 사용했을 때의 결과이다. 표의 결과는 다섯 번의 얼굴 식별 실험 후에 구한 평균 식별 결과이다. RGB 좌표와 정규화된 CIE 좌표 모두에 있어서 대체로 안정적인 결과를 얻을 수 있었는데, 특히 학습 데이터에 대한 실험(Training)에서는 거의 100%에 가까운 식별률을 보였다. 그리고 RGB 좌표를 그대로 다루는 것보다는 CIE 좌표로 변환하여 정규화 처리한 것이 전반적으로 나은 결과를 나타내도 확인할 수 있었다. 정규된 CIE 좌표계를 사용하는 경우에는 조명의 영향을 없앨 수 있어 RGB보다는 매우 안정적인 결과가 나왔다고 생각된다. 특히 정규화된 CIE에서는 학습 데이터의 경우에는 100세대 정도만 학습시켜도 100%의 식별률을 나타내었고 임의의 테스트 데이터에 대해서도 90%에 가까운 식별률을 보여 실제 식별 시스템으로의 적용 가능성을 확인할 수 있었으며, 데이터 양도 2/3으로 줄어 KL 변환 행렬을 구하는데 걸리는 시간과 자료 저장에 필요한 공간을 줄일 수 있었다. 이러한 얼굴 식별률은 기존 논문과 수치적으로는 비슷하지만 다양한 조명 하에서 얻은 실험 결과이므로 매우 가치가 있다고 생각된다. 또 결과를 보면 이 중에서 실제 사람의 얼굴이 아닌 영

상(다른 부위의 피부, 사물, 벽 등)의 입력에 대해서는 RGB 나 정규된 CIE 모두에 있어서 γ_{FF} 가 매우 높아 얼굴 검출 문제로의 적용 또한 상당히 좋은 성능을 나타낼 수 있다는 것을 확인하였다.

실제 입력 데이터에 대하여 식별 시스템의 결과를 얻을 때는 식별하고자 하는 사람에 대하여 미리 추출된 데이터로 계산된 KL 변환 행렬과 이에 의해 학습된 신경망의 가중치를 이용한 계산만을 사용한다. 즉, 입력 영상을 정규화된 CIE 좌표계로 변환한 뒤 이를 벡터화하여 KL 변환 행렬을 곱해 차원을 감소시킨 고유 공간의 벡터를 구하고, 이 고유 공간의 벡터를 신경망에 입력하여 신경망의 출력을 비교하는 것으로 식별이 이루어지는 것이다. 본 실험에서 200개의 데이터 모두에 대해 식별 결과를 얻는 시간은 약 5초로 측정되었다.

표 2 얼굴 식별 시스템의 식별 결과(RGB).

Table 2 The identification result of the RGB faces.

Data	Training	Test
TRUE	200(100)	200(100)
Success	197(99)	164(84)
γ_{TT}	99/100	84/100
γ_{TF}	1/100	16/100
γ_{FF}	98/100	80/100
γ_{FT}	2/100	20/100
Performance	98.5%	82%

표 3 얼굴 식별 시스템의 식별 결과(정규화된 CIE).

Table 3 The identification result of the normalized CIE faces.

Data	Training	Test
TRUE	200(100)	200(100)
Success	200(100)	179(94)
γ_{TT}	100/100	94/100
γ_{TF}	0/100	6/100
γ_{FF}	100/100	85/100
γ_{FT}	0/100	15/100
Performance	100%	89.5%

5. 결 론

본 논문에서는 CCD 카메라로부터 받은 영상으로부터 얼굴 영역을 추출하고 추출된 얼굴 영역으로부터 한 개인의 얼굴을 식별하는 알고리즘을 제안하였다. 본 논문에서는 지금까지는 잘 시도되지 않았던 컬러 공간에서의 접근을 시도하였으며, RGB 좌표와 휘도치 성분을 제거한 정규화된 CIE 좌표에 대하여 결과를 비교하였다. 휘도 성분은 조명의 영향에 크게 좌우되기 때문에 이를 최대한 줄이기 위해 정규화된 CIE 좌표계에 대하여 변환된 입력 벡터를 이용하였다. 또 과도한 입력 벡터의 데이터 양을 줄이기 위해서 주 성분 분석 기법인 KL 변환을 사용하였다. KL 변환에 의해 구해진 특징 벡터는 얼굴 영상의 중요한 방향으로의 성분을

의미하며 기억하고자 하는 얼굴의 고유 공간에 분포한다. 이 특징 벡터의 경계를 정하고 식별하는 방식으로는 기존에 많이 사용하던 유클리드 거리를 이용하지 않고 비선형적인 특성을 갖는 다층 역전파 신경망을 사용하였다.

본 논문에서는 실험을 통해 입력 벡터의 RGB 좌표값과 CIE 좌표 값을 이용한 경우에 대하여 제안된 개인 식별 시스템의 성능을 확인하였다. 실험 데이터는 학습 데이터와 테스트 데이터로 나누어 실험하였고, 그 결과 상당히 만족스러운 식별률을 확인할 수 있었다. 특히 정규화된 CIE 좌표 값을 이용하는 방법이 RGB 좌표 값을 이용하는 경우에서보다 안정적인 식별 결과를 나타내는 것도 볼 수 있었다. 이것은 정규화된 CIE 좌표를 사용한 방법이 조명에 대한 영향을 최대한 감소시켰기 때문이라 생각된다.

감사의 글

본 연구는 한국과학재단 특정기초연구비(96-10-02-15-01-3) 지원에 의하여 이루어진 연구로서, 지원에 감사를 드립니다.

참 고 문 헌

- [1] Chellappa, Wilson, and Sirohey, "Prolog to Human and Machine Recognition of Faces : A Survey," Proceedings of IEEE, Vol. 83, No. 5, pp.704-740, 1995.
- [2] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, pp.71-86, 1991.
- [3] Shigeru Akamatsu, Tsutomu Sasaki, and Hideo Fukamachi, "An Accurate and Robust Face Identification Scheme," Proceedings of IEEE, pp.217-220, 1995.
- [4] M.A.Sid-Ahmed, "Image Processing: Theory, Algorithms & Architectures," McGraw-Hill International Editions, pp.277-375.
- [5] Markus Groß, "Visual Computing," Springer-Verlag, pp.76-87, 1994.
- [6] Rafael C. Gonzalez and Paul Wintz, "Digital Image Processing," Addison-Wesley Publishing Company, 2nd Edition, pp.122-125, 1987.
- [7] Simon Haykin, "Neural Networks : A Comprehensive Foundation," Prentice Hall, pp. 363-370.
- [8] Gilbert Strang, "Linear Algebra and Its Applications 3rd Edition," Harcourt Brace Jovanovich International Editions, pp.294-297.
- [9] Dana H. Ballard and Christopher M. Brown, "Computer Vision," Prentice Hall, pp17-148.
- [10] Kazuhiko Kawamura, Sugato Bagchi, Moenes Iskarous, and Magued Bishay, "Intelligent Robotic Systems in Service of the Disabled," IEEE Transactions on Rehabilitation Engineering Vol. 3

No.1, 1995.

- [11] Henry A. Rowley, Schumeet Baluja, and T. Kanade, "Human Face Detection in Visual Scenes," Technical Report CMU-CS-95-158, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1995.
- [12] Kah-Kay Sung and T. Poggio, "Learning Human Face Detection in Cluttered Scenes," Computer Analysis of Images and Patterns, pp.432-439, 1995.



박민용(朴 玟 用)

1973년 연세대학교 전자공학과 졸업. 1977년 동대학원 졸업(석사). 1982년 동경대 대학원 졸업(박사). 1982년~현재 연세대 전자공학과 교수. 주관심분야: 퍼지 시스템, 의용 전자 공학 등.

저 자 소 개



김용주(金 容 柱)

1995년 연세대학교 전자공학과 졸업. 1997년 동대학원 전자공학과 졸업(석사). 1997년~현재 LG 전자 멀티미디어 연구소.



지승환(池 承 桓)

1969년 9월생. 1992년 연세대학교 전자공학과 졸업. 1994년 동 대학원 전자공학과 졸업(석사). 1998년 동 대학원 전자공학과 졸업(박사). 1998년~현재 미래산업(주) 미래연구센터.



유재형(柳 在 炯)

1974년 1월 28일 생. 1997년 국가대 대표학과 졸업. 1997년~현재 동대학교 대표학과 석사과정. 1959년 11월 18일 서울 생. 1978년~1983년 연세대학교 전자공학과(학사). 1983년~1985년 연세대학교 전자공학과(석사). 1986년~현재 한국통신 통신망연구소 교환망관리연구실장(선임연구원). 1995년~현재 연세대학교 전자공학과 박사과정.



김정환(金 貞 煥)

1961년생. 1984년 충남대학교 전자공학과 졸업. 1984년~1986년 삼보컴퓨터. 1986년~1996년 현대고등학교 교사. 1996년 연세대학교 전자계산학과 공학 석사 취득. 1996년~현재 동대학원 전자공학과 박사과정. 1997년~현재 정화인쇄(주) 부설연구소 이사.