論文2000-37CI-3-4

고유특징과 다층 신경망을 이용한 얼굴 영상에서의 눈과 입 영역 자동 추출

(Automatic Extraction of Eye and Mouth Fields from Face Images using MultiLayer Perceptrons and Eigenfeatures)

柳 淵 植 *, 吳 世 泳 **

(Yeon-Sik Ryu and Se-Young Oh)

요 약

본 논문은 얼굴영상에서 눈과 입 부위를 추출하기 위한 알고리즘을 제안하였다. 첫째로, 눈과 입의 에지 이진 화소 집합의 고유 값 (Eigenvalue) 과 고유 벡터 (Eigenvector) 로 부터 추출한 정보들은 눈과 입을 찾기 위한 좋은 특징이 된다. 눈과 입 부위의 긍정적 샘플과 부정적 샘플로부터 추출한 고유 특징들로 다층 신경망을 학습하여 특정 영역이 눈과 입 부위 포함하는 정도를 나타내도록 하였다. 둘째로, 시스템의 강건성 확보를 위해 서로 다른 구조의 단일 MLP를 묶어서 그 결과를 이용하는 Ensemble network 구조를 사용하였다. 두 눈과 입에 각각 별도의 Ensemble network을 사용하였고, 각 Ensemble network내 MLP들의 출력이 최대가 되는 영역의 중심 좌표들을 평균하여 최종 위치를 결정하였다. 셋째로, 특징 정보 추출 검색 영역을 줄이기 위해 얼굴 영상 에지 정보와 눈과 입의 위치 관계를 이용해 눈과 입의 대략적인 영역을 추출하였다.

제안된 시스템은 적은 수의 정면 얼굴에서 추출한 고유 특징들로 학습된 Ensemble network을 사용하여 학습에 사용되지 않은 다른 사람들의 정면얼굴 뿐만 아니라 일정한 범위 내 자세 변화에서도 좋은 일반화 성능을 얻고 있으며, 작은 범위 내에서의 얼굴 크기 변화나 좌우 20° 이내의 자세 변화에 대해서도 신경망의 일반화 기능을 이용하여 강건한 결과를 얻고 있음을 확인하였다.

Abstract

This paper presents a novel algorithm for extraction of the eye and mouth fields (facial features) from 2D gray level face images. First of all, it has been found that Eigenfeatures, derived from the eigenvalues and the eigenvectors of the binary edge data set constructed from the eye and mouth fields are very good features to locate these fields. The Eigenfeatures, extracted from the positive and negative training samples for the facial features, are used to train a MultiLayer Perceptron(MLP) whose output indicates the degree to which a particular image window contains the eye or the mouth within itself. Second, to ensure robustness, the ensemble network consisting of multiple MLPs is used instead of a single MLP. The output of the ensemble network becomes the average of the multiple locations of the field each found by the constituent MLPs. Finally, in order to reduce the computation time, we extracted the coarse search region for eyes and mouth by using prior information on face images.

The advantages of the proposed approach includes that only a small number of frontal faces are sufficient to train the nets and furthermore, lends themselves to good generalization to non-frontal poses and even to other people's faces. It was also experimentally verified that the proposed algorithm is robust against slight variations of facial size and pose due to the generalization characteristics of neural networks.

* 正會員, LG電子㈜
(LG Electronics Inc.)
** 正會員, 浦港工科大學校 電子 컴퓨터工學部
(Electrical and Compute Engineering Division, POSTECCH)

※ 이 논문은 한국 학술진흥재단의 학술연구비(1999년)와 BK21 사업을 통하여 포항공과대학교 전자·컴퓨터 분야에 주어진 교육부의 재정 지원에 의해 연구되었습니다.
接受日字:1999年 9月 15日, 수정완료일 : 2000年 3月14日

Ⅰ. 서 론

사람의 눈과 입의 특징은 얼굴인식, 인증과 감정을 전달 하는 데 있어서 매우 중요한 역할을 한다. 이러한 관점에서 사람의 눈과 입의 자동추출 방법은 많은 응용범위를 갖고 있다. 이제까지의 일반적인 눈과 입의 자동추출 방법은 템플릿 패턴 정합 (Template Pattern Matching), 명암(Intensity), 얼굴의 기하학적인 정보를 이용하는 방법 등으로 구분할 수 있다[1]. 기하학적인 정보를 이용한 얼굴 인식 연구[2,3,4,5]에서는 반드시 얼굴의 특징점을 추출하여야 하며, 주로 에지 영상의 수직, 수평 특성을 이용하고 있다. 그러나 수직, 수평 방향의 에지 영상만 이용하는 경우 정면얼굴에 대해서는 어느 정도 대략적인 위치를 찾을 수 있으나 자세가 변하는 경우에는 정확하게 찾기 어려울 것으로 보인다.

템플릿 패턴 정합을 이용하는 경우에는 변경된 자세를 수용하기 위해서 많은 양의 학습 샘플을 필요로 하게 된다[1]. Beymer[6]는 얼굴인식 시스템에서 눈과 코의 하단을 찾기 위해 5 단계의 계층적인 처리를 수행하여 자세의 변화를 수용하고 있다. 각 단계에서는 얼굴 전체 영역에 대한 상호관계 계산과 여러 종류의 회전된 탬플릿을 이용한 계산 등을 위해 많은 사람의 다양한 자세에서 얻은 눈과 코의 템플릿 및 계산 시간을 필요로 하고 있다. Juel[7]은 영상에서 얼굴을 찾기 위한 정보로 눈, 코와 입을 찾기 위해 에지 강화 영상(Edge Enhancement Image)을 3개의 신경망에 입력하는 구조를 이용하고 있다. 영상공간에서 눈, 코, 입에 할당된 신경망의 출력과 얼굴의 기하학적인 관계를 이용하여 얼굴영역을 찾고 있다. 우리의 구조와 비슷하지만 이들은 신경망의 입력으로 에지 강화 영상을 이용함으로써 많은 수의 입력 뉴런을 필요로 하고 있으며, 자세의 변화 등에 적절히 대응하고 있지 못하다.

윤호섭[8]은 두 눈의 위치를 찾는 단계에서 2-pass 레이블링 방법, Hough 변환, 히스토그램 분석 등을 사용함으로써 많은 연산 시간을 필요로 한다. 최동선[9]은 눈의 위치를 추출하는 방법으로 이진 에지 영상을 라벨링하고 각 라벨링된 영역의 면적, 둘레, 원형도 등으로부터 얻은 유사도를 검사하여 눈을 찾고 있다. 그러나 자세가 변화하는 경우는 두 눈의 면적, 둘레, 원형도 등의 정보가 변화하므로 정확한 위치를 찾기 힘들 것으로 보인다.

우리는 많은 학샘플을 필요로 하지 않으면서도 자세의 변화와 명암 변화등에도 적절하게 대응하는 알고리즘을 제안한다. 이진 에지 영상을 이용하여 얼굴의 경계선을 구하고[3], 두 눈과 입의 기하학적인 특징들을 이용하여 각각을 위한 대략적인 영역을 결정한다. 대략적인 영역에 존재하는 화소들을 2차원 상에 존재하는 데이터 집합으로 간주한다. 이 데이터 집합의 분포는 고유 값(Eigenvalue)과 고유 벡터(Eigenvector), 무게중심 등으로 특징지을 수 있다. 이러한 특징들을 고유특징(Eigenfeature)이라 정의한다. 이것은 고 차원의 얼굴영상을 저 차원의 특징 벡터로 차원 줄이기(Dimension Reduction)을 수행하고, 또한 고 차원에서는 다양하게 표현되는 정보를 낮은 차원에서 대표성을 갖는 표현 즉, 이진 데이터의 분포 모양으로 나타내는 효과가 있다. 그러나 이진 데이터의 분포 모양도 획일적으로 동일하지 않으며, 또한 모든 분포 모양을 획득할 수는 없으므로 비 선형 투사기능을 갖는 MLP(MultiLayer Perceptron)을 이용하여 정규적인 패턴을 학습하고 정규적인 패턴에서 벗어나는 부분은 MLP의 일반화 기능을 이용하여 결과를 얻도록 하였다. 일반화 성능을 향상하기 위해 여러 개의 MLP 출력 결과를 합성하여 최종 영역을 결정하는 Ensemble network 구조를 이용하였다.

MLP학습은 눈과 입의 영역을 정확하게 포함하고 있는 긍정적 샘플(Positive Sample)들과 눈과 입을 정확하게 포함하고 있지 않은 부정적 샘플(Negative Sample)들에서 추출한 고유 특징들을 입력 벡터로 이용한다. 출력은 긍정적 샘플에 대해서는 '+1', 부정적 샘플에 대해서는 '-1' 이 되도록 학습하고, 실험에서 각 MLP 는 실험영역의 학습된 고유특징에 대한 유사도를 출력하도록 하여 가장 큰 유사도를 출력하는 영역의 중심좌표를 얻는데 이용된다. 각 부위마다 1개의 Ensemble network 가 구현되었고, Ensemble network 내의 각 MLP 출력으로부터 얻은 중심 좌표들을 합성하여 Ensemble network의 최종 좌표를 얻는다.

제 Ⅱ 장에서는 눈과 입의 대략적인 영역을 추출하는 방법에 대해 설명하고, 제 Ⅲ 장에서는 신경망의 학습방법과 대략적인 영역에서 눈과 입의 영역을 추출하는 방법에 대해 설명한다. 제 Ⅳ 장에서는 실험 및 결과에 대해 기술한다.

Ⅱ. 눈과 입의 대략적인 영역 추출

본 논문에서 대상으로 하는 흑백 얼굴 영상은 92×112 크기의 256 밝기 영상이다. 영상은 Cambridge 대학의 Olivetti연구실 데이터 베이스에서 얻었다. 입력 영상은 정면얼굴과 좌우 20° 이내에서 회전된 측면 얼굴을 포함하고 있다. 눈의 크기는 20×10, 입의 크기는 40×20 영역에 포함되며, 눈과 입은 영상의 중앙을 중심으로 분포하고 배경도 균일한 것으로 가정한다. 그림 1은 전체 시스템의 구성이다.

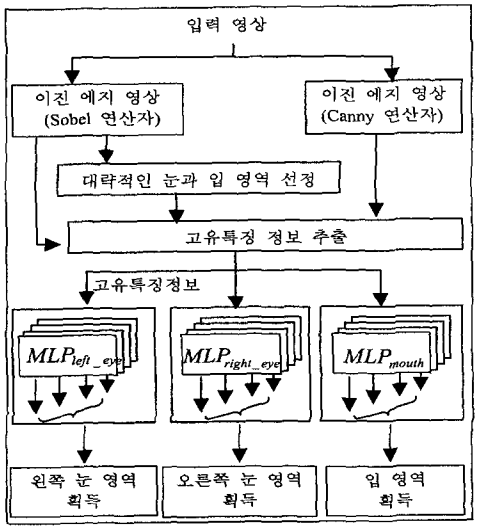


그림 1. 전체 시스템 구성도
Fig. 1. The system block diagram.

대략적인 눈과 입 부위의 설정은 전체 계산시간과 성능에 중요한 요소이므로 가급적 작은 영역을 찾는 것을 목표로 한다. 얼굴영상은 일정한 기준점으로 이동하거나 크기 등을 일정하게 만드는 등의 별도의 정규화(Normalization) 및 에지 강화 과정을 거치지 않는다. 사람의 눈과 입의 위치는 얼굴 크기, 자세등에 따라 다르게 나타나므로 획일적으로 그 위치를 결정할 수는 없다. 그러므로 먼저 눈과 입을 포함하는 영역을 다른 영역과 분리할 필요가 있다.

첫번째 단계로 Sobel 연산자를 적용하여 얻은 이진 영상 정보를 이용하여 두 눈과 입을 포함하는 영역을 설정하고 그 안에서 눈과 입의 기하학적인 관계를 이용하여 대략적인 눈과 입의 영역을 결정한다. 먼저 영상에서 눈과 입을 포함하는 영역을 찾기 위해 입력 영상의 수직, 수평 방향 에지 정보를 이용한다[6]. 입력영상을 $I(x,y)$라고 할 때, $I_{IV}(x,y)$, $I_{HE}(x,y)$는 $I(x,y)$에 Sobel의 수평 방향 마스크와 수직 방향 마스크를 적용하여 얻은 에지 영상을 적당한 값(Threshold) 으로 이진화한 것이다. H_v는 $I_{VE}(x,y)$의 i열의 수직방향 투영 합이고, $H_h(j)$는 $I_{HE}(x,y)$의 행의 수평 방향 투영 합이다(그림 2). 얻어진 $H_v(i)$, $H_h(j)$을 이용하여 대략적인 얼굴의 폭(width)과 상하(height)를 결정 짓는 윈도우의 경계선을 위한 x_L ,x_R, y_H, y_L을 얻는다(그림 2). y_L은 입 근처의 행(row)을 나타낸다. 그러나 이 위치는 얼굴 자세에 따라 입술의 중간이나 윗 부분 또는 아래 부분이 될 수 있으므로 (15 화소)을 사용하여 입술 전체가 포함되도록 y_L을 재 설정한다.

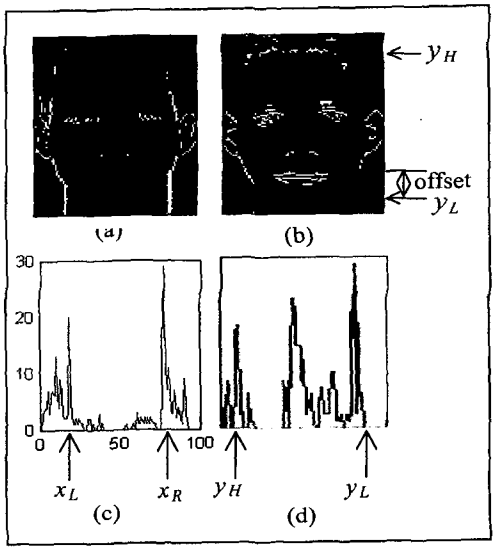


그림 2. 대략적인 얼굴 영역을 찾은 예
(a) $I_{VE}(I,J)$: 수직 에지 이진 영상, (b) $I_{HE}(i,j)$: 수평 에지 이진 영상 (c) $H_v(i)$: (a)의 수직방향으로의 합, (d) $H_h(j)$
Fig. 2. An example of determining the face boundary.
(a) $I_{VE}(I,J)$: Vertical edge dominance map, (b) $I_{HE}(i,j)$: Horizontal dege dominance map, (c) $H_v(i)$: The sum of vertical projection of (a), (d) $H_h(j)$: The sum of horizontal projection of (b)

두번째 단계로 얻어진 얼굴 영역에서 우리가 찾고자 하는 눈의 위치와 입의 위치를 찾기 위해 눈과 입의 기하학적인 관계로부터 다음의 가정을 설정한다 (그림 3 (a) 참조).

- 눈 부위(Eye_Region) 는 첫번째 단계에서 찾은 얼굴 영역에서 상단 영역에 존재한다. 이마의 높이와 얼굴 자세에 따라서 눈의 위치가 변화하므로 눈이 존재할 수 있는 영역을 그림 3(a) 상단 사각형과 같이 설정한다.
- 입 부위(Mouth_Region)는 첫번째 단계에서 찾은 얼굴 영역에서 하단 영역에 존재한다. 자세에 따라 입의 위치도 변화하므로 입이 존재 할 수 있는 영역을 그림 3(a) 하단 사각형과 같이 설정한다.

눈의 대략적인 위치를 좀더 작은 영역으로 줄이기 위해서는 눈썹과 눈을 구분해야 할 필요성이 있다. 이를 위해 $I_{VE}(x, y)$를 이용한다. 눈과 눈썹의 경우 수평 에지 성분이 주를 이루는 특징을 갖고 있다. 그러나 눈은 눈썹에 비해 수직 에지 성분도 많이 포함되어 있으므로 을 이용하여 눈의 대략적인 영역을 좀더 작게 설정할 수 있다. 즉, 눈의 대략적으로 설정한 상단 구간에 대하여 식 (1), (2)와 같이 행의 수평 방향 투영 합을 구하여 E_c를 구한다. 는 눈과 눈썹이 있는 대략적인 영역 중에서 수직 에지가 가장 강한 행을 의미 하므로 이 행 근처에 눈이 존재할 가능성이 높다

$$H_{hv}(j)=\sum_{i=1}^{V_r} I_{VE}(i, j),\quad V_r=\left(\frac{2}{3}\right)*(y_L-y_H) \tag{1}$$

$$E_c=\arg\ \underset{j}{Max}(H_{hv}(j)) \tag{2}$$

물론 이것 또한 자세의 영향으로 두 눈의 위치가 수평관계가 아닌 경우나 빛의 조건 등으로 인해 정확한 위치는 아니다. 이를 극복하기 위해 우리는 를 중심으로 상하 15 화소의 여유을 갖는 영역을 잠정적으로 눈 부위로 설정한다 (그림 3(b) 참조).

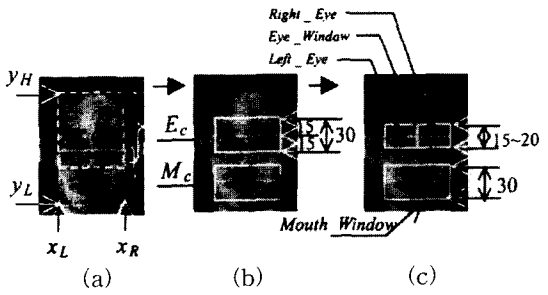


그림 3. 대략적인 영역의 설정
Fig. 3. The 3-stage process of determining the left eye, right eye and mouth regions.

셋째 단계로 이 영역에 대해서 다음의 가정을 적용하여 눈의 위치를 포함하는 대략적인 윈도우를 최종적으로 얻는다 (그림 3(c) 참조).

- 눈은 수평 에지 성분이 강한 특성을 갖고 있으므로 E_c를 중심으로 상하에 존재하는 두 영역에서 $H_h(j)$의 값이 월등히 큰 쪽에 눈이 존재 할 확률이 높다.

이러한 가정을 적용하여 자세의 변화에 따른 두 눈의 상하 위치를 고려하여 적당한 크기로 *Eye_Window*를 설정한다.

$$Eye\_Window=RECT(x_L, E_c-5, x_R, E_c+10) \tag{3}$$
$$\text{if}\ \sum_{j=E_c}^{E_c+15} H_h(j)\geq\alpha * \sum_{j=E_c-15}^{E_c} H_h(j)$$

$$Eye\_Window=RECT(x_L, E_c-10, x_R, E_c+5) \tag{4}$$
$$\text{if}\ \sum_{j=E_c}^{E_c+15} H_h(j)<\left(\frac{1}{\alpha}\right) * \sum_{j=E_c-15}^{E_c} H_h(j)$$

$$Eye\_Window=RECT(x_L, E_c-10, x_R, E_c+10) \tag{5}$$
$$Otherwise$$

여기서 RECT (x_1, y_1, x_2, y_2)는 점 (x_1, y_1), (x_1, y_2), (x_2, y_2), (x_2, y_1)들이 구성하는 직사각형이고, α 는 실험적으로 1.5를 사용하였다. 즉, 식 (2) 에서 얻은 E_c를 중심으로 상하 15 화소사이의 수평 에지 맵의 합 중에서 어느 한 쪽의 합이 다른 쪽의 α배 이상 큰 경우에는 합이 큰 쪽으로 *Eye_Window* 를 약간 크게 설정해주며, 그렇지 않은 경우에는 상하 동일영역의 크기로 설정한다. 이렇게 하는 이유는 두 눈이 평행한 경우와 경사진 경우를 구분하여 영역을 설정하기 위한 것이다. 이렇게 찾아진 영역의 크기는 $(x_R-x_L)\times$(15 또는 20)가 되고, 이 영역을 세로로 이등분 *Left_Eye, Right_Eye* 하여 영역을 설정한다(그림 3(c) 참조).

Mouth_Window 의 설정은 입의 수평 에지 특성이 매우 강한 것을 이용한다. 먼저 눈의 경우와 동일한 방법으로 입이 존재할 가능성이 있는 넓은 영역에 대하여 수평 에지 성분 합 $H_h(i)$이 가장 큰 행(M_c)를 찾는다 (그림 3(b) 참조). 물론 이것도 자세의 변화와 영상의 명암등에 따라 정확한 위치라고 판단할 수 는 없다. 그러므로 최종적으로 이 행을 중심으로 상하 15 화소 씩 영역을 두어 입이 존재할 가능성이 있는 영역으로 Mouth_Window 를 설정한다.

$Mouth\_Window = RECT(x_L, M_c - 15, x_R, M_c + 15)$ (6)

이것을 통하여 대략적인 Mouth_Window 의 크기는 $(x_R - x_L) \times 30$로 설정된다(그림 3(c) 참조).

Ⅲ. 신경망의 학습과 시험

다양한 신경망중에서도 MLP는 비 선형 투사 관계에 유용하며, 학습 패턴에 대한 일반화 기능을 갖고 있다. 사람의 눈과 입 부위의 영상은 자세에 따라 다양해도 에지 이진영상의 고유특징을 이용하면 적은 수의 학습 샘플로도 많은 사람들의 눈과 입 모양을 대표할 수 있으며, 신경망의 일반화 기능이 그러한 역할을 강화할 수 있을 것으로 판단된다.

얼굴 특징부위 추출기는 크게 왼쪽 눈, 오른쪽 눈, 입을 위한 Ensemble network 3개로 구성되었고, 각 Ensemble network은 여러 개의 서로 다른 구조를 갖는 MLP로 구성된다. Ensemble network은 한 개의 신경망 구조가 갖는 단점을 보완하기 위해 서로 다른 구조를 갖는 여러 개의 신경망의 결과를 합성하는 개념적인 network이다. Ensemble network을 이용하는 주된 이유는 동일한 입력 조건에 대해서도 신경망은 구조에 따라 서로 다른 결과를 얻을 수 있으며, 결과를 융합하여 다른 신경망의 부족한 부분을 보완하도록 하여 전체적으로 보다 우수한 결과를 얻을 수 있기 때문이다. 각각의 MLP는 동일한 학습 패턴으로 독립적으로 학습되며, Ensemble network은 독립적으로 학습된 MLP들로부터 가장 높은 유사도를 나타내는 영역의 중심 위치를 얻은 뒤 그 위치를 합성하여 최종적인 중심 위치를 얻는다. 여기서 신경망의 출력을 그대로 합성하는 것이 아니고, 각 신경망의 출력이 최대가 되는 영역의 중심을 합성하는 것이므로 개념적인 network이라 하였다. 이 때 각 MLP마다 가중치를 다르게 하여 합성할 수 도 있지만, 본 논문에서는 평균을 취함으로써 동일한 가중치를 적용하였다. 각 신경망의 학습방법은 Resilient Back Propagation 알고리즘[10]을 이용하였다.

신경망의 학습에 있어서 사람이 물체를 인식하는데 방법을 모델링 하였다. 즉

- 데이터 베이스에 저장된 물체와 정확하게 정합이 되는 것을 찾는 것-긍정적 샘플
- 정합이 발생하지 않는 집합을 만족하지 않는 것을 찾는 방법[11]-부정적 샘플

등이다. 신경망이 위의 두 가지 특성을 동시에 갖도록 하기 위해 긍정적 샘플과 부정적 샘플들로 학습이 이루어진다. 그림 7(a)는 학습의 긍정적 샘플(그림 8 (a), (c))과 부정적 샘플(그림 8(b), (d))를 얻는데 사용한 영상이고, 그림 7(b)는 시험 영상의 일부 예이다 (전체 시험 영상은 그림 12에 있음). 위 영상에서 학습을 위한 샘플은 눈의 경우 20×10의 크기로, 입의 경우 40x20의 크기로 추출되었다. 각 MLP의 출력은 선형출력특성을 갖는 뉴런 1개로 구현되었으며, 긍정적 샘플에 대해서는 '+1', 부정적 샘플에 대해서는 '-1' 값을 원하는 출력으로 하여 학습한다. 출력 값이 클수록 긍정적 샘플에 유사한 것으로 해석하며, 값이 작을수록 부정적 샘플에 유사한 것으로 정의한다. 궁극적으로 우리가 원하는 것은 일정한 영역에서 최고 유사값을 갖는 영역의 중심 좌표(u_c, v_c)를 얻는 것이다.

1. 고유 특징(Eigenfeature)의 추출

신경망의 학습에 있어서 입력과 출력의 표현(Representation)은 신경망의 성능에 매우 중요한 역할을 한다[12]. 영상을 그대로 입력으로 이용하는 것 보다는 고유특징을 추출하여 얻고자 하는 목적에 필요한 정보만을 추출해 이용하면 입력 벡터의 차원을 줄일 수 있고, 신경망의 학습에도 효과적이다.

입력 영상에서 얻어진 에지 이진영상은 u-v 평면상에 옆으로 길게 또는 둥근 패턴을 형성하며 분포하는 2차원 데이터로 간주할 수 있다(그림 6). 특히, 얼굴에 있어서 눈이나 입의 영역은 얼굴의 다른 부위와는 크게 구분되는 특징을 갖는다. 또 눈과 입은 사람에 따라 약간의 차이는 있으나 에지 데이터 분포의 주된 방향과 각 방향의 크기 및 비율 등은 일정한 범위 안에 존재한다고 볼 수 있다. 보다 세밀한 에지 데이터를 얻기 위해 눈 영역에 대한 학습 및 실험 샘플은 Canny 에지 추출기를 이용하였고, 입에 대한 것은 입의 특징을 가장 잘 표현할 수 있는 Sobel의 수평 에지 추출기를 이용하여 이진 영상을 얻었다.

신경망 학습을 위한 고유특징 입력 벡터의 생성과정은 다음과 같다. 학습 샘플의 이진 영상의 한 점 $\mathbf{p}_i$ 와 이러한 점들로 구성된 데이터 집합 $\mathbf{P}$ 로 부터 행렬 $\mathbf{M}(2\times2)$을 얻는다(그림 4 참조).

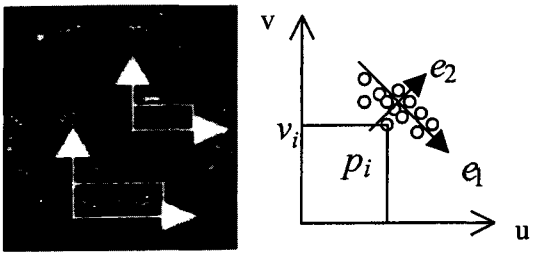


그림 4. 이진 영상의 데이터 분포와 고유벡터와의 관계
Fig. 4. The binary edge map and the corresponding eigenvectors for each region.

$$\mathbf{p} = (\mathbf{p}_0, \mathbf{p}_1, \cdots, \mathbf{p}_i, \cdots, \mathbf{p}_{N_P}) \tag{7}$$

$$\mathbf{M} = \mathbf{P} \cdot \mathbf{P}^T \tag{8}$$

여기서 N_p는 주어진 영상에서 '1'의 값을 갖는 화소들의 수이고, $\mathbf{p}_i = (u_i, v_i)^T$이다. $\mathbf{P}^T$는 $\mathbf{P}$의 Transpose이다. 행렬 $\mathbf{M}$에 대하여 (식 9)를 만족하는 λ_1, λ_2, e_1, e_2를 얻는다.

$$(\mathbf{M} - \lambda \mathbf{I}) \cdot \mathbf{e} = 0 \tag{9}$$

여기서 λ_1, λ_2는 고유 값($\lambda_1 > \lambda_2$)들이고, e_1, e_2는 각각 λ_1, λ_2와 관련된 고유 벡터들이다. e_1, e_2는 대상 영역 데이터들의 분포에 있어서 수직한 주 방향(Principal Direction)을 나타내고, λ_1, λ_2는 각 방향의 크기정도를 나타낸다[13]. 이렇게 얻어진 e_1, e_2, λ_1, λ_2으로 부터 확장된 정보와 이진 데이터의 u, v 방향 무게중심 값들로 구성된 고유특징을 얻어서 신경망 학습을 위한 입력 벡터 $\mathbf{x}_i$를 구성한다. 는 학습 샘플($\mathbf{x}_i$)들을 (식 11)과 같이 규준화 하여 얻은 $\mathbf{X}_{N,i}$로 구성된 입력 벡터들의 집합이다.

$$X = \{x_{N,i} \in R^9,\ i = 1, 2, \cdots, N_s\} \tag{10}$$

$$x_{N,i} = (\mathbf{x}_i - \bar{\mathbf{x}}_i) / \mathrm{var}(\mathbf{x}_i) \tag{11}$$

단,

$$x_i^1 = \lambda_2,\ x_i^2 = \lambda_1,\ x_i^3 = \lambda_2/\lambda_1,$$
$$x_i^4 = e_{2u},\ x_i^5 = e_{2v},\ x_i^6 = e_{1u},\ x_i^7 = e_{1v}$$
$$x_i^8 = \frac{1}{k}\sum_{i=1}^{k} u_i,\quad x_i^9 = \frac{1}{k}\sum_{i=1}^{k} v_i$$

여기서 N_s는 학습 샘플의 개수이며, k는 학습 샘플의 이진 영상에서 '1' 값을 갖는 화소 수이다. 즉, 신경망 학습을 위한 고유특징은 정규화 된 두개의 고유 값($x_{N,i}^1$, $x_{N,i}^2$), 두 고유 값의 비($x_{N,i}^3$), 두 고유 벡터의 u 방향성분 ($x_{N,i}^4$, $x_{N,i}^6$), v 방향성분 ($x_{N,i}^5$, $x_{N,i}^7$), u, v 방향의 무게중심 성분 ($x_{N,i}^8$, $x_{N,i}^9$) 등으로 구성된다. 고유 특징들은 분포된 데이터의 주축(Principle axis)상의 폭과 높이 정보 등을 포함하여 분포 데이터 분포 모습을 표현하고 있음을 알 수 있다. 제 IV장에서 입력 고유 특징 항목이 결과에 미치는 영향을 고찰한다.

2. Ensemble network의 출력 및 특징 영역 결정

제 II 장에서 찾은 *Left_Eye*, *Right_Eye*, *Mouth_window* 영역에 대하여 화소 단위로 이동하면서 학습 샘플 크기 영역 (눈: 20×10, 입: 40×20)에 대한 고유특징을 생성하여 해당 Ensemble network의 각 MLP에 입력한다. 각 MLP의 출력이 찾은 영역들을 (식 12), 그림 5와 같이 융합한다(4 MLPs를 사용한 경우).

$$\begin{pmatrix} u_c \\ v_c \end{pmatrix} = \frac{1}{N_E} \begin{pmatrix} \sum_{i=1}^{N_E} u_{l,c} \\ \sum_{i=1}^{N_E} v_{l,c} \end{pmatrix} \quad l = 1, \cdots, 4 \tag{12}$$

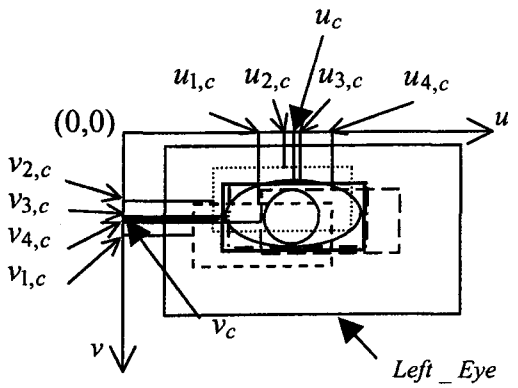
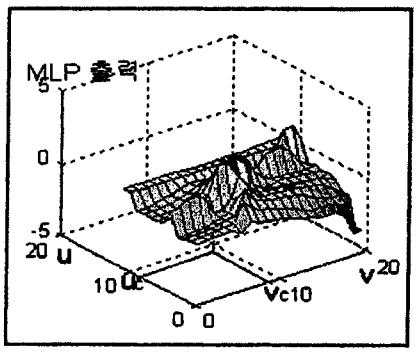


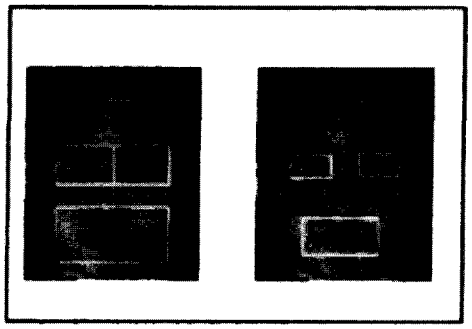
그림 5. Ensemble network의 출력
Fig. 5. The output of the ensemble network.

여기서 $u_{l,c}$, $v_{l,c}$는 주어진 window 영역 검색에서 Ensemble network내의 l MLP의 출력이 최대가 된 영역의 중심좌표이고, N_E는 해당영역에서 유효한 유사값을 얻은 MLP의 수이다. 그림 5에서와 같이 시험 영역(*Left_Eye*)에 대한 각 신경망의 출력

($yleft\_eye,l$ (x_k))들 중에서 가장 큰 값을 출력하는 영역 4곳의 중심 위치($u_{l,c}$, $v_{l,c}$)를 평균하여 최종적인 위치(u_c, v_c)를 얻는다. 동일한 방법으로 오른 쪽 눈과 입을 위한 Ensemble network에 대하여 위의 방법을 적용하여 각각의 중심위치를 얻는다. 그림 6은 입 영역 추출을 위한 Ensemble network내 MLP-1의 출력 예이다. 최고점을 보이는 곳의 위치가 ($u_{1,c}$, $v_{1,c}$)를 중심으로 하는 하는 영역의 시작점이다. 그림 6(b)는 대략적으로 찾은 영역(왼쪽그림)과 신경망이 최종적으로 추출한 입과 눈의 영역(왼쪽그림)을 보여준다.



(a)



(b)

그림 6. MLP의 출력 맵과 결과 예
(a) $y_{mouth}(\mathbf{x})$의 맵 (b) 특징 영역

Fig. 6. The MLP output profile map and the resulting regions of interest found.
(a) The profile map of $y_{mouth}(\mathbf{x})$. (b) The feature regions.

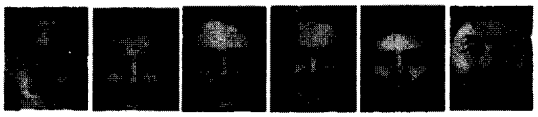
3. 성능 평가 방법

제안 시스템의 성능은 최종적으로 선택된 영역 내에 눈과 입이 정확하게 들어 오는 정도로 평가될 수 있을 것이다. 평가를 위한 기준 중심 좌표 데이터 ($u_{r,c}$, $v_{r,c}$)는 수동으로 시험영상에서 얻었다. 제안 시스템으로부터 얻은 각 부위의 중심 좌표(u_c, v_c)와 기준 좌표간의 거리는 유클리디안 거리 즉, $d=\sqrt{(u_{r,c}-u_{c)}^2+(v_{r,c}-v_c)^2}$로 평가하였다. 이후 d를 화소 거리라 칭하겠다. 참고로 이후 성능의 기준으로 삼을 화소 거리 값 8은 눈동자 크기 정도 된다.

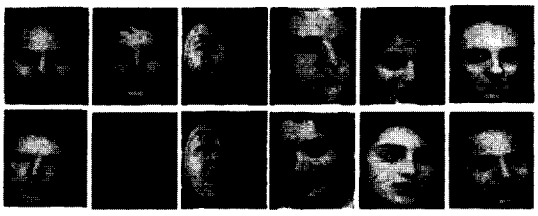
VI. 실험 결과 및 고찰

본 실험에 필요한 에지 영상 추출 및 MLP 학습 및 시험은 MATLAB 환경에서 수행되었다. 두 눈과 입에 사용된 3개의 Ensemble network은 각각 '9-50-30-1', '9-90-30-1', '9-27-9-1', '9-30-50-20-1' 등 4개의 MLP로 구성되어 있다. 신경망의 학습률은 모두 동일하게 0.01 값을 사용하였고, 학습 패턴 중에서 긍정적 샘플이 상대적으로 매우 적으므로 부정적 샘플에 균등하게 배치하여 학습하였다. 학습은 최대 4000 epoch 동안 이루어졌고, 학습은 1E-7~1E-8정도의 에러영역에서 수렴하고 있다.

실험에 사용한 영상들은 학습샘플을 추출하기 위해 사용한 영상들과 새로운 얼굴의 정면 및 자세의 변화가 있는 영상 등을 포함하여 모두 18명의 180개 영상을 사용하였으며, 남녀 성별 구별은 특별히 두지 않았다. 실험에 사용한 D/B는 40명의 400장의 영상으로 구성되어 있으나, 본 실험에서는 기본적으로 안경을 착용한 경우, 수염이 있는 경우, 대머리 인 경우, 눈이 너무 작은 경우 등을 제외한 18명을 대상으로 하였다. 시험 대상 영상은 정면 뿐만 아니라 오른쪽을 응시하는 측면, 왼쪽을 응시하는 측면, 아래를 내려다 보는 것 등으로 구성되어 있다(그림 7 (b) 참조).



(a)



(b)

그림 7. 학습과 실험에 사용한 이미지 예
(a) 학습 샘플 추출에 사용한 이미지
(b) 신경망 시험에 사용한 이미지

Fig. 7. The face images used for (b) learning (b) testing.

그림 7(a)에 있는 영상들로부터 추출한 눈과 입 부위를 긍정적 샘플로 하였다. 눈의 경우 왼쪽, 오른쪽 눈 각각 5개를 긍정적 샘플로 사용하였고, 입의 경우에는 벌린 입에 대한 학습을 위해 6개를 사용하였다 (그림 8(a)와 (c) 참조). 부정적 샘플은 그림 8(b)와 (d)에 있는 것과 같이 실험 전에 임의적으로 눈과 입의 주위에서 추출한 것과 1차 실험에 이용한 61장에서 크게 오류를 발생하는 샘플들을 모아서 구성하였다. 부정적 샘플은 반드시 실험대상의 영상이 아닌 임의의 사람 영상에서도 추출할 수 있는 것이므로 시험 샘플에서 추출하였다. 물론 최종 시험은 총 180장에 대하여 수행하였으므로 119장의 영상에서는 부정적 샘플을 추출하지 않은 것이다. 두 눈의 경우 부정적 샘플로 각각 70개를 사용하였다. 입의 경우에는 29개를 사용하였다.

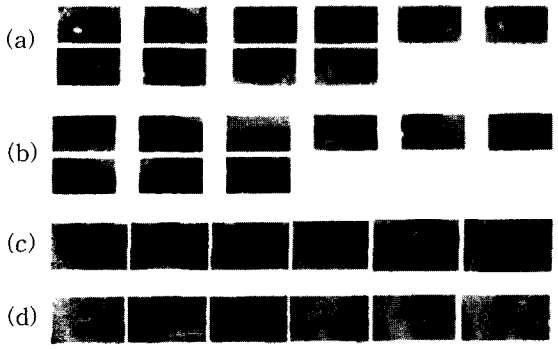


그림 8. 신경망 학습에 사용한 학습 샘플의 예
(a) 긍정적인 눈 학습 샘플 (b) 부정적인 눈 학습 샘플-일부 예 (c) 정상적인 입 학습 샘플 (d) 부정적인 입 학습 샘플-일부 예

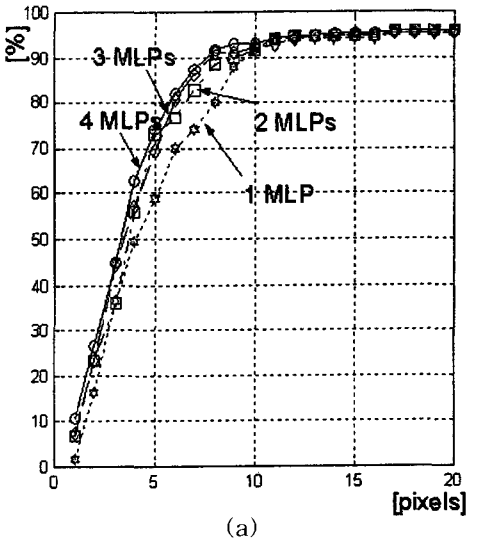
Fig. 8. Examples of the training set for MLPs.
(a) The positive eye samples. (b) Some negative eye samples. (c) The positive mouth samples. (d) Some negative mouth samples.

에지 영상을 위한 문턱 값은 Matlab에서 제공하는 값을 이용하였다. 수평 에지 영상을 위한 문턱 값은 약 0.0607, 수직 에지를 위한 문턱 값은 약 0.0783이다. 대략적인 영역의 탐색에 있어서 빠른 연산을 위해 실험 대상 영역의 값이 '1' 인 화소 수가 문턱 값 이상 갖지 않는 경우 시험에서 제외하도록 하였고, 긍정적인 샘플에 대한 유사 값이 0보다 큰 값을 찾지 못하는 경우에는 문턱 값을 낮추어서 다시 찾도록 하였다. 실험 결과를 보면, 두 눈과 입의 대략적인 영역이 잘못 찾아진 경우나 설정된 영역이 두 눈을 확실하게 포함하지 못하는 경우, 너무 어둡거나 눈이 작은 경우에는 눈을 찾지 못하는 경우가 발생한다. 입의 경우에는 얼굴과 배경사이의 에지에 의해 결과가 영향을 받고 있다.

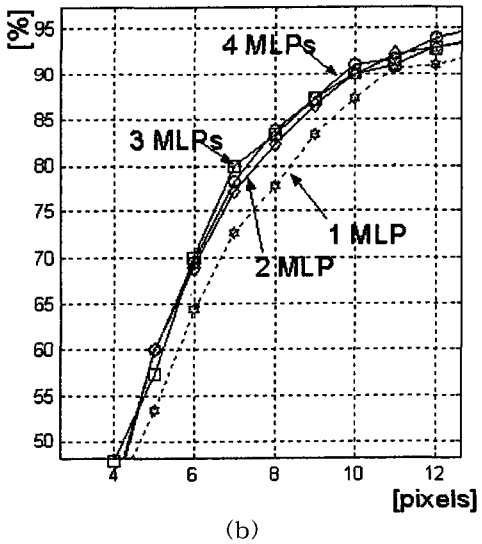
제안시스템의 전체적인 성능을 표현하기 위해 화소거리가 일정 범위에 속하는 영상의 누적 비율로 표현하였다. 즉, 일정 누적 비율 중에서 화소거리가 짧은 쪽의 누적 비율이 클수록 정확하게 추출한 것이 된다.

1. Ensemble network와 단일 MLP의 성능 비교

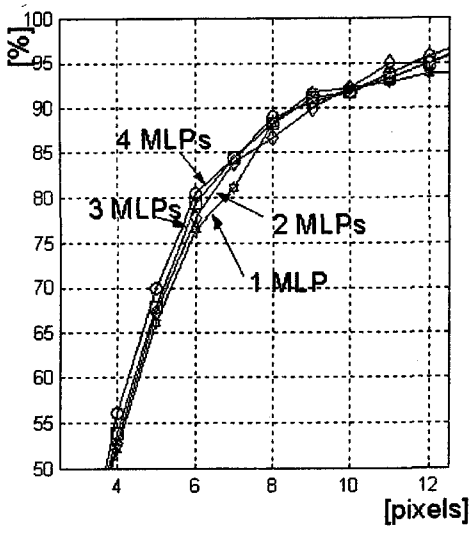
각 Ensemble network을 1~4개의 MLP로 구축하였을 경우의 성능을 비교하였다. 그림 9(a), (b), (c)는



(a)



(b)



(c)

그림 9. Ensemble network에서 사용하는 MLP 수에 따른 성능비교 (a) 왼쪽 눈 (b) 오른쪽 눈 (c) 입

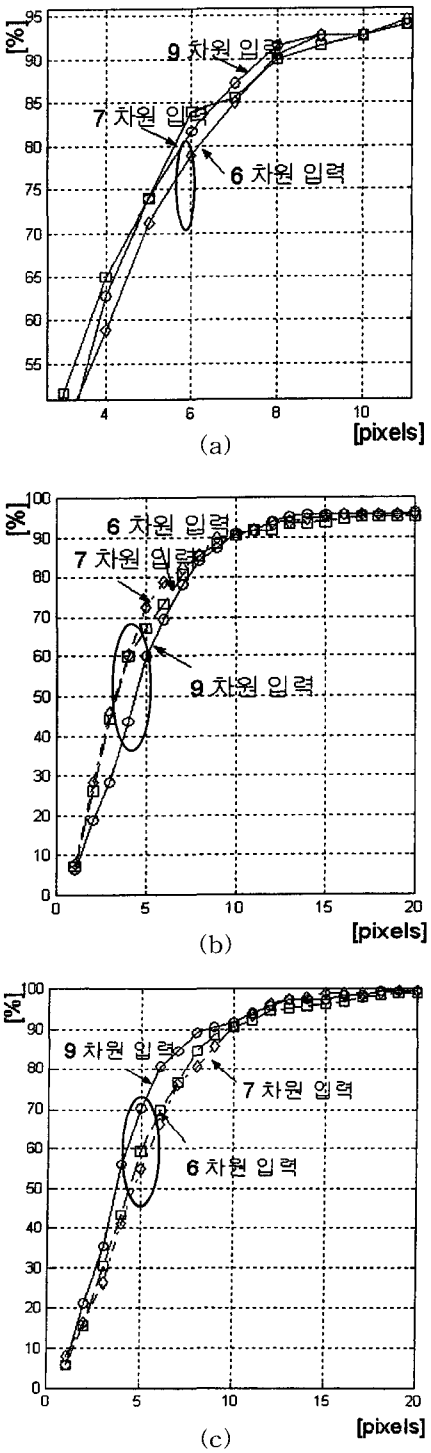
Fig. 9. Comparison of performances according to the number of MLPs used within the Ensemble network. (a) Left eye (b) Right eye (c) Mouth

각각 왼쪽 눈, 오른쪽 눈, 입에 대한 Ensemble network의 성능을 나타낸다. 전체적으로 1개의 network으로 분류하는 것보다는 Ensemble network로 구성하여 특징부위를 찾는 것이 정확성과 누적 비율에서 우수한 결과를 얻고 있으며, 각각의 MLP가 부족한 부분을 서로 보완하고 있음을 확인할 수 있다. 왼쪽 눈 Ensemble network의 경우에는 화소거리 8을 기준으로 보면 약 12%, 오른쪽 눈 Ensemble network의 경우에는 약 7%의 성능 향상을 나타내고 있다.

2. 고유특징벡터 구성에 따른 성능비교

신경망의 학습 패턴의 입력으로 사용한 고유특징에 따른 성능을 비교하였다. 선택한 고유특징은 앞장에서 전술한 바와 같이 고유 값과 고유 벡터, 고유 값의 비율, 무게 중심 좌표 등이었다. 실험에서는 고유특징을 모두 사용한 경우(9차원 입력), 무게 중심 좌표를 제외한 경우(7차원 입력) 와 고유 값의 비와 무게 중심 좌표를 제외한 경우(6차원 입력)에 대하여 실험하였다.

그림 10에 그 결과를 나타내었다. 실험 결과는 4 MLP를 사용한 Ensemble network에 대한 결과이다.



(a)

(b)

(c)

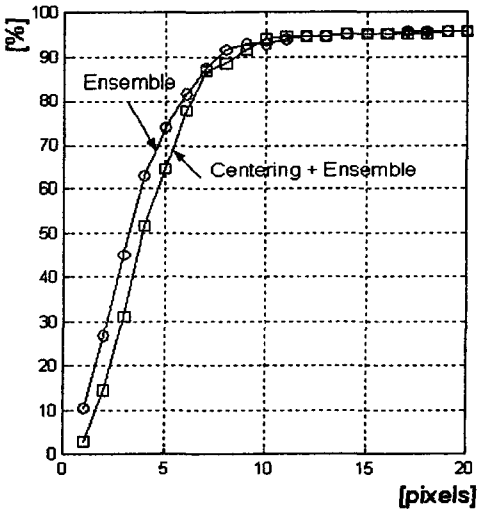
그림 10. 사용한 고유특징 수에 따른 성능비교 (a) 왼쪽 눈 (b) 오른쪽 눈 (c) 입

Fig. 10. Comparison of performances according to the number of eigenfeatures used as input. (a) Left eye (b) Right eye (c) Mouth

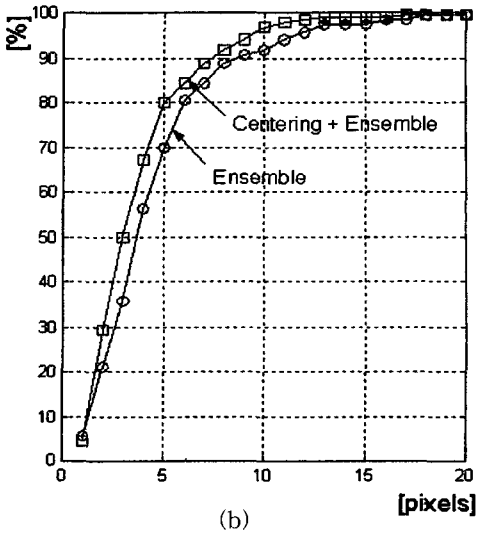
왼쪽 눈과 오른쪽 눈의 경우에는 시험영역에 대한 무게 중심 항목이 제외된 7차원 입력, 6차원 입력의 고유특징을 사용하는 것이 9차원 입력의 고유특징을 사용하는 것 보다 정확한 결과를 나타내고 있다. 즉, 화소거리가 짧은 쪽의 누적 비율이 클수록 정확성이 높은 것 이 된다(타원 표시 부분). 또한 7차원 입력의 경우가 6차원 입력의 경우보다 약간 우수한 것은 입력 벡터 중 고유 값의 비율이 나타내는 효과라 하겠다. 눈의 영역경우에 있어서는 자세나 눈의 깊이 등에 따라 그 주위에 그늘이 지는 경우 등이 에지 영상에 노이즈로 작용함에 따라 무게 중심의 좌표 정보가 크게 도움이 되지 않는 것으로 판단된다. 반면에 입의 경우에는 그늘 등에 대한 에지 노이즈가 크게 영향을 주지 않으므로 무게 중심 좌표가 성능 향상에 도움이 되고 있는 것으로 판단된다.

3. Centering 작업 유무에 따른 성능비교

Ensemble network내의 각 MLP가 선정한 영역의 무게 중심을 기준으로 영역을 다시 조정한 후 그 결과를 합성한 경우와 순수하게 MLP가 선정한 영역을 합성한 경우의 성능 비교이다. 그림 11 결과를 보면 입력고유 특징에 따른 성능비교와 유사한 결과를 얻고 있음을 확인할 수 있다. 즉, 두 눈의 Ensemble network의 경우에는 순수하게 MLP의 결과를 합성한 것이 우수하고, 입의 경우에는 무게중심을 선택하여 합성하는 것이 보다 우수한 결과를 보인다.



(a)



(b)

그림 11. Centering 작업에 따른 성능비교
(a) 왼쪽 눈 (b) 입

Fig. 11. Comparison of performances after the centering operation.
(a) Left eye (b) Mouth

위의 결과를 종합할 때, 두 눈의 Ensemble network는 7차원 입력를 이용하고, 입의 경우에는 9차원 입력 을 이용하여 Centering하여 합성한 결과를 이용하는 것이 가장 우수한 결과를 얻을 수 있음을 확인할 수 있다. 최종적으로 제안시스템의 성능은 왼쪽 눈 Ensemble network는 수동으로 설정한 눈의 중심과 8 화소거리 내에 속하는 누적비율이 90.0 %, 오른쪽 눈은 85.5 %, 입은 91.7 % 이다. 특징 부위 추출에 실패한 것 중에서 눈과 입 부위의 대략적인 위치 획득이 잘못된 경우가 왼쪽 눈의 경우 5 %, 오른 쪽 눈은 5.6 %, 입은 1.6 % 정도이다. 나머지는 수동으로 선택한 중심위치에서 8 화소 이상 초과한 거리를 선택하고 있다.

그림 12에 18명의 180장 영상에 대하여 얻은 결과 중에서 90장에 대한 결과를 보였다. 점선으로 표시한 두 장의 영상은 눈의 영역을 잘 찾지 못한 경우이다. 오른 쪽 하단의 영상은 대략적인 얼굴 영역의 검색에서 눈의 영역을 잘 못 추정한 경우이고, 왼쪽 상단은 그늘등에 의한 오류이다.

V. 결 론

흑백 얼굴 영상의 이진 영상에서 추출한 고유특징과

다수의 MLP 결과를 합성한 Ensemble network를 이용하여 눈과 입의 위치를 찾는 알고리즘을 제안하였다. 눈과 입의 에지 이진 영상의 데이터를 2차원 데이터의 집합으로 간주하여 얻은 고유 값과 고유 벡터, 무게 중심 값 등으로 눈과 입을 특징지을 수 있으며, 이러한 정보들을 확장하여 고유특징을 얻었다. 긍정적 샘플들과 부정적 샘플들의 고유특징으로 두 눈과 입을 위한 3개의 Ensemble network를 각각 학습하였다. 전 처리 과정으로 에지 영상의 수직, 수평 방향 특성을 이용하여 눈과 입 부위의 대략적인 검색영역을 축소하였다. 제안한 알고리듬의 특징을 요약하면 다음과 같다.

- 입력 영상에 대해 명암 보상이나 크기, 위치 등에 대한 정규화 작업을 하지 않는다.
- 눈 영역의 탐색 영역을 작게 설정하기 위해 기존의 연구들이 수평 에지 영상을 이용하여 대략적인 눈의 위치를 찾고 있으나, 수직 에지 영상의 수평 성분의 합을 이용하였다.
- 얼굴영역에서 눈과 입의 에지 이진 영상의 화소들을 2차원에 분포하는 데이터 집합으로 하여 구한 고유 값, 고유 벡터, 무게 중심의 정보를 확장하여 고유 특징을 추출하여 입력 차원을 줄였다.
- 두 눈과 입에 대해 각각1개씩 사용한 Ensemble network 학습은 눈의 경우 각각 5 개의 긍정적 샘플과 약 70 개의 부정적 샘플, 입의 경우 6 개의 긍정적 샘플과 약 29 개의 부정적 샘플로 학습하였다. 여기서 긍정적 샘플은 모두 정면 얼굴에서만 추출하였다.
- 각 MLP 출력은 선형출력함수로서 + 방향으로는 긍정적 샘플과 유사한 정도를 나타내고, - 방향으로는

그림 12. 시험에 사용된 영상과 특징 부위 추출 결과
Fig. 12. The test images along with their extracted fields.

부정적 샘플과 유사한 정도를 나타내도록 학습하였다. 가장 큰 유사성을 출력하는 시험 영역을 눈과 입이 있는 위치로 잠정 결정하며,

- 각 MLP의 출력으로부터 추출한 좌표들을 평균하여 최종적인 Ensemble network의 출력좌표를 얻어서 일반화 성능을 향상시켰다.
- 실험은 18명의 180개의 다양한 자세와 크기를 갖는 얼굴영상에 대해 수행하였으며, 추출의 정확성은 실험자가 수동으로 지정한 각 실험 샘플의 눈과 입의 중심점과의 Euclidean 화소 거리로 평가하여 객관성을 도모하였다.
- Ensemble network의 성능은 두 눈의 경우에 고유 값과 고유 벡터 값, 고유 값의 비율 등 7가지 정보로 사용하였을 때 수동으로 선정한 중심과의 거리 가 8 화소이내 인 것의 누적 비율이 각각 90.0 %, 85.5 %이었고, 입의 경우에는 고유 값과 고유벡터, 고유 값의 비율, 영역의 무게중심 좌표 등 9개를 사용하고 각 MLP의 찾은 영역에 대하여 무게중심으로 Centering한 후 다른 MLP의 결과와 합성하여 91.7 %의 누적 비율을 나타내었다.
- 제안한 알고리즘을 이용하는 경우, 5~6개의 정면 얼굴에서 추출한 긍정 학습 샘플로 Ensemble network을 학습하여도 학습에 사용하지 않은 정면 얼굴영상과 측면 자세의 영상에 있는 눈과 입의 위치를 찾을 수 있음을 보였다.

영상에서 눈과 입의 특징 부위를 찾는 것은 사용자 인증 및 인식 시스템, 감정 인식시스템, 컴퓨터 인터페이스등을 구현 하기 위한 기본적인 작업이며, 많은 응용 범위를 갖고 있다고 하겠다.

향후 연구는 제안한 알고리즘이 지니고 있는 단점, 얼굴영역 추출 실패에 따른 오류를 줄이기 위해 전처리 과정에서 얼굴 영상에서 살색 부위만 추출하여 정확한 검색영역을 제공하는 것과 눈 부위에서의 그늘에 의해 발생하는 오류를 줄이기 위한 방법이 필요할 것으로 보인다.

참고 문헌

[1] Yankang Wang, Hideo Kuroda, Makoto Fujimura and Akira Nakamura, "Automatic Extraction of Eyes and Mouth Fields from Monochrome Face Image using Fuzzy Technique," IEEE Conference on Universal personal Communications Record, ICUPC '95, pp. 778~782.

[2] T. Kanade, "Picture processing by computer complex and recognition of human faces. Technical report," Kyoto University, Dept. of Information Science, 1973.

[3] David J. Beymer, "Face Recognition under Varying Pose," A.I Memo No. 1461, 1993.

[4] 이상영, 함영국, 박래홍, "지식에 기초한 특징추출 x과 역전파 알고리듬에 의한 얼굴인식," 전자공학회논문지, 제 31권, B편, 제 7호, pp 119~128, 1994년 7월

[5] 이상영, 함영국, 박래홍, "뉴로-퍼지 알고리듬을 이용한 얼굴인식," 전자공학회논문지, 제 32권 B편, 제 1호, pp. 50~63, 1995년 1월

[6] R. Brunelli, T. Poggio, "Face Recognition: Features versus Templates," IEEE Transaction on PAMI, Vol. 15, No. 10, pp. 1042~1052, 1993.

[7] Paul Juell, Ron Marsh, "A hierarchical neural network for human face detection," Pattern Recognition, Vol. 29, No. 5, pp. 781~787, 1996.

[8] 윤호섭, 왕민, 민병우, "눈 영역 추출에 의한 얼굴 기울기 교정," 전자공학회논문지, 제 33권, B편, 제 12호, pp. 71~83, 1996년 12월

[9] 최동선, 이주신, "형태분석에 의한 특징 추출과 BP 알고리즘을 이용한 정면얼굴 인식," 전자공학회논문지, 제 33권, B편, 제 10호, pp. 63~68

[10] Martin Riedmiller, Heinrich Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," IEEE International Conference on Neural Networks, Vol. 1, pp. 586~591, 1993.

[11] David F. McCoy & Venkat Devarajan, "Aritficial Immume Systems and Aerial Image Segmentation." IEEE International Conference on Systems, Man, and Cybernetics 1997.

[12] Emile Fiesler and Russell Beale, "Handbook of Neural Computation," Oxford University Press, 1997.

[13] Sami Romdhani, "Face Recognition using Principal Components Analysis," MS thesis, http:// www.elec.gla.ac.uk/~romdhani/pca.htm.

저 자 소 개

柳 淵 植(正會員)

한양대학교 전기공학과 학사(1988년), 포항공과대학교 전자전기공학과 석사(1990년), LG전자(주) Digital Media Division, Digital Systems and Solution 개발실(1990~현재), 1998년~현재 포항공과대학교 전자·컴퓨터공학부 박사과정

吳 世 泳(正會員) 第 35卷 C編 第 9號 參照

현재 포항공과대학교 전자·컴퓨터 공학부 교수