

# 실시간 얼굴 표정 인식을 위한 새로운 사각 특징 형태 선택기법

## New Rectangle Feature Type Selection for Real-time Facial Expression Recognition

정성욱\*, 김도형, 안광호, 정명진

(Sung Uk Jung, Do Hyoung Kim, Kwang Ho An, and Myung Jin Chung)

**Abstract** : In this paper, we propose a method of selecting new types of rectangle features that are suitable for facial expression recognition. The basic concept in this paper is similar to Viola's approach, which is used for face detection. Instead of previous Haar-like features we choose rectangle features for facial expression recognition among all possible rectangle types in a  $3 \times 3$  matrix form using the AdaBoost algorithm. The facial expression recognition system constituted with the proposed rectangle features is also compared to that with previous rectangle features with regard to its capacity. The simulation and experimental results show that the proposed approach has better performance in facial expression recognition.

**Keywords** : rectangle feature, feature selection, facial expression recognition, AdaBoost, pattern classification

### I. 서론

인간과 로봇과의 직관적인 의사소통 방법에는 여러 방법이 존재한다. 음성, 얼굴 표정, 몸짓 등이 그 대표적인 예이다. 본 논문은 직관적인 의사소통 방법의 하나로써 그 중요성이 점차 부각되고 있는 얼굴 표정 인식에 대한 것이다. 얼굴 표정 중에 놀람과 같은 표정은 수백분의 1초 이내로 잠깐 나타났다가 사라지고, 슬픔과 같은 표정은 몇 초 정도 지속 된다[1]. 따라서 얼굴 표정 인식에 있어서 실시간이란 개념은 오래 지속되는 얼굴 표정만이 아니라 빠르게 나타났다가 사라지는 얼굴 표정까지 인식 할 수 있는 것이라고 할 수 있다. 현재까지 얼굴 표정을 인식하기 위하여 많은 접근 방법이 시도되었다. 얼굴 표정 인식 방법에 대한 분류는 어떤 점을 크게 고려하느냐에 따라서 많은 분류 방법이 존재 한다[2,3]. 얼굴 표정의 특징 추출을 기준으로 기존 연구를 분류하면 크게 두가지 방법인, 모델 기반(model-based) 방법과 영상 기반(image-based) 방법으로 나눌 수 있다.

모델 기반 방법의 대표적인 것으로 Essa와 Pentland는 삼차원 그물 모양(메시, mesh) 모델을 이차원 얼굴 영상에 맞추어 얼굴 표정을 나타낼 때 근육 움직임의 최대치로 얼굴 표정을 구분하였고[4], Lanitis 등은 얼굴을 AAM(Active Appearance Models) 방법을 이용하여 얼굴 조명 변화(facial illumination)와 얼굴 변형(deformation change)과 같은 서로 다른 영역에서 정보를 얻어 얼굴 표정을 구분하였다[5]. Huang과 Huang은 경사 기반(gradient based) 방법을 이용하여 PDM(Point Distribution Model)방법으로 얼굴의 모양을 구분하였다[6]. 또한 Zhang 등은 부분적으로 유동적인 얼굴 특징점으로부터 얼굴 표정을 구분하였다[7].

영상 기반 방법으로는 전체 얼굴을 하나의 특징으로서 사용하여 가버 웨이블릿(gabor wavelet)으로 분류한 Lisetti와 Rumelhart[8], 그리고 Fellenz 등의 방법이 있다[9]. 또한 지역적인 방법으로 Padgett와 Cottrell은 입과 눈 주위의 영역으로부터 얼굴 표정을 추출하여 PCA(Principal Component Analysis) 방법을 이용하였고[10], Bartlett는 영상 분할의 영상 밝기 프로파일을 이용해서 얼굴을 구분하였다[11]. Lien 등은 조밀도의 흐름(dense flow)을 PCA방식으로 구분하고, 블록 기반 밀도(block based density) 방법을 이용하여 얼굴 표정을 구분하였다[12].

위와 같은 방법들은 모델 기반인 경우 가려짐(occlusion) 현상에 강인하지만 얼굴 특징들을 정합하는 시간과 정밀한 얼굴 표정을 구분해 내기 위해서 입력영상을 높은 해상도로 받아들이기 때문에 실시간으로 얼굴 표정을 인식하는 데는 부적절하다는 단점이 있다.

최근의 방법으로 Viola와 Jones는 사각 특징을 이용한 영역별 영상 밝기 정보를 특징점으로 이용하여 AdaBoost 알고리즘으로 학습시켜 얼굴을 검출 하는데 있어서 빠른 시스템을 구현하였다[13]. Wang 등은 이와 같은 방법을 얼굴 표정 인식에 적용시켜 7가지의 얼굴 표정을 구분하는 시스템을 실시간으로 구현하였다[14]. 하지만, AdaBoost를 이용하여 구현한 시스템들은 사각 특징점을 사용하는데 있어서 기존의 얼굴 검출할 때 쓰는 사각 특징점을 사용하거나 다소 변형하여 사용하기 때문에[14,15] 이에 대한 분석과 새로운 사각 특징점이 필요하다. Viola의 얼굴 검출 알고리즘에서 사용한 기본 사각 특징들은 얼굴 표정의 어느 한 특징만을 분류의 기준으로 삼기 때문에 더 많은 특징을 고려할 수 있는 새로운 사각 특징점이 고려의 대상이 된다.

본 논문의 목적은 얼굴 표정을 구분할 수 있는 새로운 모양의 사각 특징(rectangle feature) 형태를 제안하는 것이다. 다시 말하면, 7가지 기본 얼굴 표정(평온, 기쁨, 화남, 슬픔, 놀람, 역겨움, 공포)에 대해서 각각의 표정을 잘 구분하는 사각 특징 형태를 제안하고, 이것을 이용하여 AdaBoost 알

\* 책임저자(Corresponding Author)

논문접수 : 2005. 9. 27., 채택확정 : 2005. 12. 8.

정성욱 : 한국전자통신연구원(brcastle@etri.re.kr)

김도형, 안광호, 정명진 : 한국과학기술원 전자전산학과

(bborog@cheonji.kaist.ac.kr/akh@cheonji.kaist.ac.kr/mjchung@ee.kaist.ac.kr)

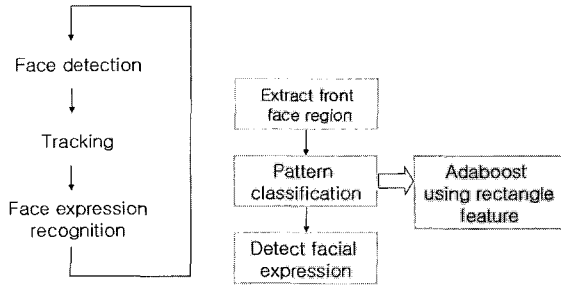


그림 1. 얼굴 표정 인식 시스템 전체구조.  
Fig. 1. Overall structure of facial expression recognition system.

고리즘으로 학습을 시켜 강분류기(strong classifier)의 인식률을 높이고, 실시간으로 얼굴을 추적하여 얼굴 표정을 인식하는 것이다.

본 논문의 구성은 2장에서 구현된 시스템의 전체구조, 3장에서 제안된 특징 형태를 선택하는 방법론을 Viola의 방법에 의거하여 언급하였다. 그리고 4장에서는 3장의 방법론을 이용하여 선택된 특징 형태를 기존의 사각 특징 형태와 각 감정에 대해서 구체적으로 대조, 비교하여 본 논문에서 제안된 사각 특징 형태가 얼굴 표정을 분류하는데 더 효과적임을 살펴보고, 나머지 5장에서는 제안된 방법의 모의 실험과 실험 결과를 이용하여 그 타당성을 검증하였다. 끝으로 6장에서 본 연구의 결론을 내린다.

II. 전체 구조

얼굴 표정 인식 시스템의 전체 구조는 그림 1과 같다. 이 때 필요한 전처리 과정이 얼굴 검출인데, 이는 실시간성이 보장된 Viola의 검출 방법을 도입하여 사용하였다. 초기 얼굴이 검출되면, 그 부근에서 얼굴 검색 창을 설정하고, 그 영역에 대해서 얼굴 추적을 실시한다. 얼굴이 검출될 때마다 정면얼굴인지 확인하고, 얼굴 표정을 검출할 정도의 충분한 크기일 경우 (본 논문에서는 50×50 화소) 패턴 분류 과정을 거쳐 얼굴의 표정을 인식하게 된다. 패턴 분류 과정은 사각 특징점을 이용하여 얼굴표정을 검출하게 되는데, 얼굴 검출 때 쓰였던 사각 특징점(7개의 모양)과는 달리 미리 학습된 각 표정에 대한 사각 특징점들(42개의 모양)을 이용하여 얼굴 표정을 검출하게 된다.

III. 특징점 추출

각 표정에 가장 높은 인식률을 가지는 사각 특징점을 찾기 위해서 3×3 까지의 모든 사각 특징점의 경우(그림 2)에 대해서 변경된 JAFFE 영상 1065장을 이용하여 학습시켰다 [16]. 즉, 하나의 사각 특징점을 모든 얼굴 표정에 대해서 학습시켜 오류 비율을 측정한 후, 총 316개의 모든 모양에 대해서 각 표정에 대해 가장 낮은 오류 비율을 갖는 사각 특징점을 각 표정 당 각각 5개씩 추출하였다.

1. AdaBoost 학습 알고리즘

AdaBoost는 많은 수의 잠재적인 특징점에서 소수의 효과적인 약분류기를 찾는 간결한 알고리즘이다[13]. 구현된 시스템에선 얼굴 표정을 잘 구분 짓는 사각 특징점의 모양과,

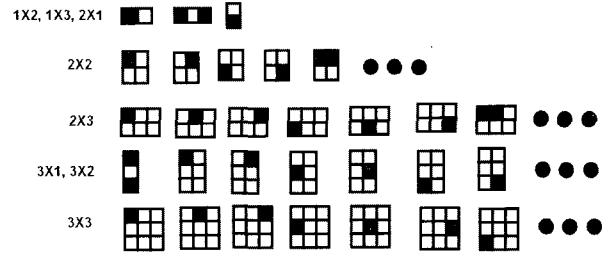


그림 2. 학습에 사용되는 3×3 구조까지의 모든 가능한 사각 특징점 모양.

Fig. 2. All possible rectangle feature types up to the 3×3 structure size used for training.

구해진 사각 특징점의 모양을 이용해서 적은 수의 효과적인 특징점을 찾는 과정, 그리고 분류기를 학습하는데 이 알고리즘을 이용하였다. 강분류기를 학습시키는 부스팅 알고리즘은 아래와 같다.

- 표본 영상  $(x_1, y_1), \dots, (x_n, y_n)$  가 있고, 여기서  $x_i (i=1, 2, \dots, n)$ 는 학습하려는 영상이고, 음(negative) 영상을 0, 양(positive) 영상을 1이라고 표현하면  $y_i=0$  또는  $1 (i=1, 2, \dots, n)$  로 나타낼 수 있다.
- 초기 가중치(weight)는 다음과 같이 결정한다.

$$w_{1,i} = \begin{cases} \frac{1}{2m} & \text{단, } y_i = 0 \\ \frac{1}{2l} & \text{단, } y_i = 1 \end{cases} \quad (1)$$

여기서  $m$  과  $l$  은 음 영상의 표본과 양 영상의 표본 개수이다.

- For  $t=1, \dots, T$ ,
- 1) 가중치를 아래와 같은 방법으로 정규화 한다.

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}} \quad (2)$$

- 2) 각각의 특징점  $j$ 에 대하여, 약분류기  $h_k$  를 학습시킨다.

$$h_k(x) = \begin{cases} 1, & \text{단, } p_k f_k(x) < p_k \theta_k \\ 0, & \text{나머지 경우} \end{cases} \quad (3)$$

약분류기  $h_k(x)$ 는 특징점 값인  $f_j$ , 임계값인  $\theta_k$ , 그리고 극성을 나타내는  $p_k$ 로 이루어진다. 오류값은 다음과 같이 결정된다.

$$\epsilon_k = \sum_i w_i |h_k(x_i) - y_i| \quad (4)$$

- 3) 가장 적은 오류 ( $\epsilon_t$ )를 갖는 분류기( $h_t$ )를 선택한다.

$$\epsilon_t = \min \epsilon_k, \quad k=1, 2, \dots, n \quad (5)$$

- 4) 가중치를 갱신한다.

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_i}, \quad \beta_t = \frac{\epsilon_t}{1-\epsilon_t} \quad (6)$$

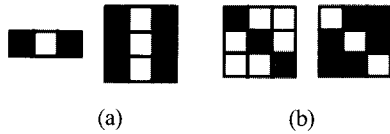


그림 3. 중복되는 사각 특징점 모양의 예 (a) 크기의 변화에 따라 겹치는 경우, (b) 영상 반전에 의해서 겹치는 경우.

Fig. 3. Examples of overlapped rectangle feature types (a) the types are independent of size variation, (b) the types consider unsigned feature value.

여기서 만약 표본  $x_i$ 가 바르게 분류됐다면  $e_i=0$ , 반대의 경우엔  $e_i=1$  이다.

- 최종 학습된 강분류기는

$$H(x) = \begin{cases} 1, & \text{단, } \sum_{t=1}^T \alpha_t h_t(x) \geq 0.5 \times \sum_{t=1}^T \alpha_t \\ 0, & \text{나머지 경우} \end{cases} \quad (7)$$

여기서,  $\alpha_t = \log \frac{1}{\beta_t}$

이런 과정을 거쳐서  $T$  개의 약분류기는 간단한 특징점을 이용해서 찾는다. 최종의 강분류기는  $T$  개의 약분류기의 가중 선형 조합(weighted linear combination)으로 이루어진다.

2. 사각 특징점 모양 추출 과정

- 3×3 행렬 형태의 모든 사각 특징점 모양을 구한다.

(그림 2)

• 위에서 구한 사각 특징점 모양 중 서로 겹치는 모양은 배제한다.

1) 크기변화의 경우

그림 3(a)의 두개의 사각 특징점은 서로 다른 모양을 가지고 있지만, 사각 특징점은  $u, v$  방향으로 크기를 변화하면서 약분류기를 찾기 때문에 서로 같은 결과가 나오게 된다.

2) 영상 반전에 의한 경우

그림 3(b)는 서로 다른 극성을 가지고 있으나 같은 사각 특징점값을 나타낸다. 왜냐하면 AdaBoost에 의한 분류방법은 두개의 집단을 구분하는 것이기 때문에 두개의 사각 특징 모양은 서로 극성만 다를 뿐 같은 분류기준을 가진다.

• 각 표정을 가장 잘 구별하는 35개의 사각 특징점 모양과 7개의 기본 사각 특징점 모양을 이용해서 얼굴 영상 데이터베이스를 가지고 다시 학습시켜 얼굴 표정에 대한 강분류기를 구한다.

- For  $p=1, \dots, \#$  of 얼굴 표정 집단

표본영상  $(x_1, y_1), \dots, (x_n, y_n)$ 가 있고, 여기서  $x_i$ 는 학습하려는 영상이고, 구분하고자하는 얼굴 표정 영상을 1, 다른 얼굴 표정 영상을 0, 이라고 표현하면  $y_i=1, 0$  ( $i=1, 2, \dots, n$ )으로 나타낼 수 있다.

- For  $q=1, \dots, Q,$

초기 가중치(weight)는 다음과 같이 결정한다.

$w_i = \frac{1}{2m}, \frac{1}{2l}$  단,  $y_i=1, 0$ . 여기서  $m$ 과  $l$ 은 목적 얼굴 표정과 다른 얼굴 표정의 표본수이다.

- For  $t=1, \dots, T,$

1) 가중치를 정규화 한다.

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}} \quad (8)$$

2)  $q$ 번째 사각특징점 모양에서 각각의 특징점  $j$ 에 대하여, 약분류기  $h_j$ 를 학습시킨다.

$$\epsilon_j = \sum_i w_i |h_j(x_i) - y_i| \quad (9)$$

3) 가장 적은 오류( $\epsilon_{q,t}$ )를 갖는 분류기( $h_{q,t}$ )를 선택한다.

4) 가중치를 갱신한다.

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}, \quad \beta_t = \frac{\epsilon_{q,t}}{1-\epsilon_{q,t}} \quad (10)$$

여기서 만약 표본  $x_i$ 가 바르게 분류됐다면  $e_i=0$ , 반대의 경우엔  $e_i=1$  이다.

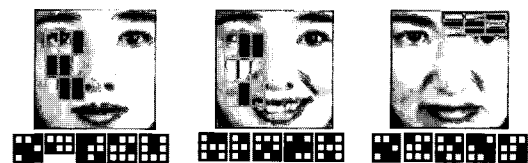
5) 아래와 식으로  $\alpha_{p,q,t}$ 를 계산한다.

$$\alpha_{p,q,t} = \log \frac{1-\epsilon_{q,t}}{\epsilon_{q,t}} \quad (11)$$

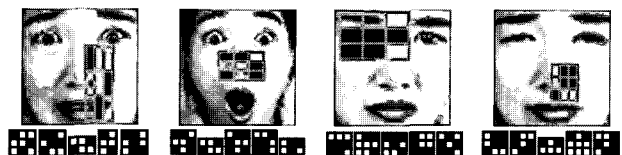
• 각각의  $p$ 에 대해서 가장 높은 값으로  $\alpha_{p,q,t}$ 를 정렬한다. 그리고  $\alpha_{p,q,t}$  값이 큰 사각 특징 모양을 선택한다.

3. 선택된 사각 특징점

그림 4는 앞의 학습의 결과로 추출한 평운을 가장 잘 구분하는 사각 특징점을 나타낸 것이다. 가장 오류 비율이 낮은 것부터 왼쪽으로 나열 하였다. 그림 4의 사진은 학습에 사용된 영상에서 그 표정을 가장 잘 구분할 수 있는 위치와 크기에 놓은 것이다. 그림 4(a)에서 알 수 있듯이 평운을 가장 잘 구분하는 사각 특징점은 입 바로 위와 볼 부분



(a) Neutral (b) Happiness (c) Anger



(d) Sadness (e) Surprise (f) Disgust (g) Fear

그림 4. 각 표정에 대하여 선택된 5가지의 사각 특징점 모양.  
Fig. 4. Five selected rectangle feature types for each facial expression.

그리고 눈의 흰자위 부분을 고려하고 있음을 알 수 있다. 이것은 일반적으로 사람은 웃을 때, 슬플 때나 화날 때 입꼬리를 변화시키거나 눈 주위가 변하게 되므로, 이것을 고려한 것이라고 할 수 있다. 그림 4(b)은 기쁨을 가장 구분하는 5가지의 사각 특징점들을 나열하였다. 그림에서는 가장 오류 비율이 낮은 사각 특징점을 도시하였다. 그림에서 보듯이 웃는 얼굴의 눈과 입 꼬리를 모두 고려하고 있는 것을 알 수 있다. 따라서 7가지 사각 특징점으로 하나의 부분만 고려하는 것보다는 충분히 효과적으로 기쁜 얼굴을 구분해 낼 수 있다고 생각된다.

그림 4(c)는 화난 얼굴을 가장 잘 구별할 수 있는 사각 특징점들을 나열하였다. 화난 얼굴의 가장 큰 특징중의 하나는 눈썹 모양의 변화라고 할 수 있다. 화난 얼굴은 눈썹이 V자를 그리고, 눈썹사이에 골짜기 모양의 주름이 생기는 것이 특징이다. 그림 4(c)에서 보듯이 화남을 구분하는 사각 특징점은 눈썹과 눈썹사이를 동시에 고려하고 있는 것을 알 수 있다. 그림 4(d)는 슬픔을 구별하는 사각 특징점들을 나타낸 것이다. Ekman은 슬픈 표정을 지을때 입꼬리는 내려가거나 떨리고, 눈썹 안쪽은 서로 당겨지고 위로 올라가며 눈썹 밑의 피부는 삼각형이 된다고 설명하였다. 하지만 그림 4(d)는 이러한 특징들을 잡아내는 사각 특징점이 아니다. 이것은 슬픔표정을 나타내는 얼굴표정의 특징들이 너무 강도가 약하기 때문에 이 자체가 다른 표정과 구분할 수 있는 단서가 될 수 없기 때문이다. 따라서, 학습을 통해서 가장 높은 인식률을 가진 사각 특징점을 찾은 것이 그림 4(d)이다.

그림 4(e)는 놀람을 구별하는 사각 특징점을 나타내었다. 놀란 표정을 나타내는 특징은 여러 가지가 있다. 그림에서 볼 수 있듯이 이 사각 특징점은 코 주변에 위치하고 있음을 알 수 있다. 사람이 놀람 표정을 나타낼 때 강도의 차이는 있겠지만 일반적으로 눈을 크게 뜨고, 입을 벌리게 된다. 그것으로 인해 눈 밑과 입 주변의 피부가 당겨지게 되어서 그 부분의 밝기가 높아지게 된다. 그림 4(e)에서는 피부가 당겨진 부분을 정확하게 고려하고 있음을 알 수 있다. 역겨운 표정은 가장 구분하기 힘든 표정중의 하나이다. 사람마다 역겨움을 나타내는 표정이 일관되지 않고 또한 그 강도도 모두 다르기 때문이다. 특히, 이 표정은 화난 표정과 구분하기가 힘들다. 그림 4(f)의 사각 특징점은 학습을 통해서 역겨움을 가장 잘 구분할 수 있는 것을 찾은 것이다. 여기서 이 사각 특징점은 눈 주위부분과 코의 주름을 고려하고 있음을 알 수 있다. 그림 4(g)는 공포를 구분하는 사각 특징점을 나열하였다. 공포는 얼굴 표정 중에서 가장 구분하기 힘든 얼굴 표정이다. 따라서 여러 영상을 통한 학습을 통해서 그림 4(g)와 같은 5가지의 사각 특징점들을 찾았다.

앞에서 열거한 각 얼굴 표정에 대한 사각 특징점들은 각 얼굴 표정의 특징들을 충분히 반영하고 있음을 알 수 있다. 기존의 7가지의 사각 특징점이 각 표정에 대해서 한 부분만을 고려하고 있다면 새로이 제시한 35개의 사각 특징점은 얼굴 표정의 두 부분 이상을 고려하고 있다. 실험적으로 학습을 통해서 기존의 7개의 사각 특징점보다 낮은 오류

비율을 가지고 있고, Ekman이 정리한 자료와 비교해 봐도 충분히 합리적임을 알 수 있다.

IV. 특징점 분석

앞 절에서는 구한 35가지의 사각 특징점이 Ekman의 자료를 이용해서 얼굴표정을 분석하는데 합리적이라는 것을 보였다. 이번 절에서는 구한 사각 특징점을 Viola가 제시한 4가지의 사각 특징점과 대각 사각 특징점을 합친 7가지 모양의 사각 특징점들과 비교 분석을 통하여 성능을 확인하였다. 먼저 기존의 7개의 사각 특징점들을 모든 영상에 대해서 학습을 통해서 각각의 얼굴 표정을 가장 잘 구별해 낼 수 있는 사각 특징점들을 찾았다. 구한 사각 특징점들을 42개 모양의 사각 특징점들과 같은 영상에 대해서 사각 특징점값을 비교함으로써 정량적인 분석을 하였다.

1. 약분류기 비교

그림 5(a)는 기본 사각 특징점 중에서 기쁨 얼굴을 가장 잘 구별하는 사각 특징점을 보인 것이고, 그림 5(b)는 42가지의 사각 특징점 중 기쁨 얼굴을 가장 잘 구별하는 사각 특징점을 찾아서 각각의 표정에 적용한 것이다.

각 표정의 특징점값(사각형내의 흰색 부분의 밝기와 검은 색 부분의 밝기의 차)은 그림 6에 도시하였다. 그림 6(a)는 그림 5(a)의 결과를 보여주고, 그림 6(b)는 그림 5(b)의 결과를 나타낸다. 또한 그림 6에서 나타난 점선은 기쁨을 가장 구별하는 사각 특징점의 임계값을 나타낸다. 이 임계값은 학습을 통해서 구해졌다.

그림에서 볼 수 있듯이 기본 사각 특징점은 7가지 표정 중 5가지 표정만을 기쁨이 아닌 것으로 구별한다(원안의 점 부분). 제안된 사각 특징점은 기쁨을 제외한 모든 감정들을 기쁨이 아닌 것으로 구분하는 것을 볼 수 있다. 하지

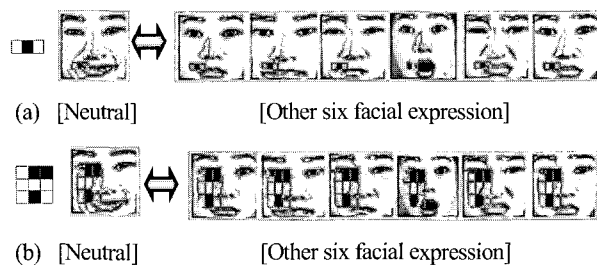


그림 5. 기쁨 사각 특징점 모양을 각각의 표정에 대입한 그림 (a) 7가지 사각 특징점 모양의 경우에 기쁨 표정과 다른 6가지의 표정을 가장 적은 오류를 가지고 구별하는 사각 특징점 모양 (b) 42가지 사각 특징점 모양의 경우에 기쁨 표정과 다른 6가지의 표정을 가장 적은 오류를 가지고 구별하는 사각 특징점 모양.

Fig. 5. Applying happy rectangle feature types to each facial expression; (a) the selected features that distinguish happy facial expression with the least error from other expressions in case of 7 (Viola's) feature types, (b) the selected feature that distinguish happy facial expression with the least error from other expressions in case of the proposed 42 feature types.

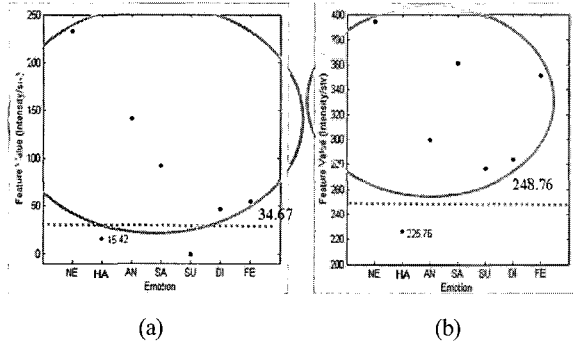


그림 6. 각 표정에 기쁨 사각 특징점 모양을 대입해서 특징점값을 비교한 결과 (a) 7가지모양의 사각 특징점의 경우, (b) 제안한 42가지의 사각 특징점의 경우.

Fig. 6. Comparison of feature values in applying happy rectangle features to each facial expression (a) in case of 7 (Viola's) feature types, (b) in case of the proposed 42 feature types.

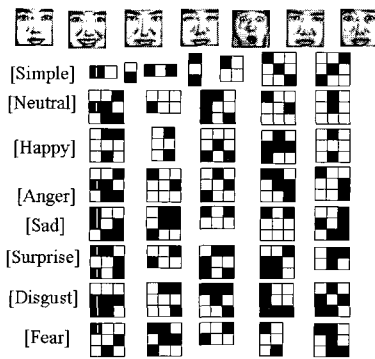


그림 7. 42가지 모양의 사각 특징점 모양 (첫행의 7개의 사각 특징점 모양은 Viola의 모양과 대각 사각 특징점 모양).

Fig. 7. 42 types of rectangle features (the first row is 7 rectangle features, that is, for of Viola's and three of diagonal types).

만, 이 분석은 하나의 약분류기에 대한 분석이므로, 전체 강분류기의 결과와는 다를 수도 있다. 그럼에도 불구하고, 각각의 약분류기가 표정을 구분하는 인식률이 높아진다면 전체의 강분류기의 인식률도 높아진다는 것은 자명한 일이다.

2. 학습데이터에 대한 효과적인 특징점 개수

앞서 3장에서 구한 35개의 사각 특징점과 기본 사각 특징점을 합하여 총 42개 모양의 사각 특징점을 이용하여 모든 표정에 대해서 다시 학습을 시켰다. 학습결과로 각 표정에 대해서 80개의 약분류기를 찾았다. 그림 7은 학습시키는 사각 특징점들을 나열한 것이다.

80개의 약분류기를 이용하여 각 표정의 얼굴에 대해서 몇 개의 약분류기만으로 전체 학습영상을 그 표정으로 검출하는지 실험해 보았다. 그림 8에서 보면 42개 모양의 사각 특징점을 이용한 경우가 적은수의 약분류기를 사용했을 경우 더 많은 학습영상들은 올바른 얼굴표정으로 검출하고 있음을 알 수 있다. 또한 전체 학습영상을 올바른 얼굴표정으로 검출하는 것도 기본 사각 특징점의 경우와 비교해 보면 상대적으로 적은수의 약분류기를 이용하여 검출해 낸다.

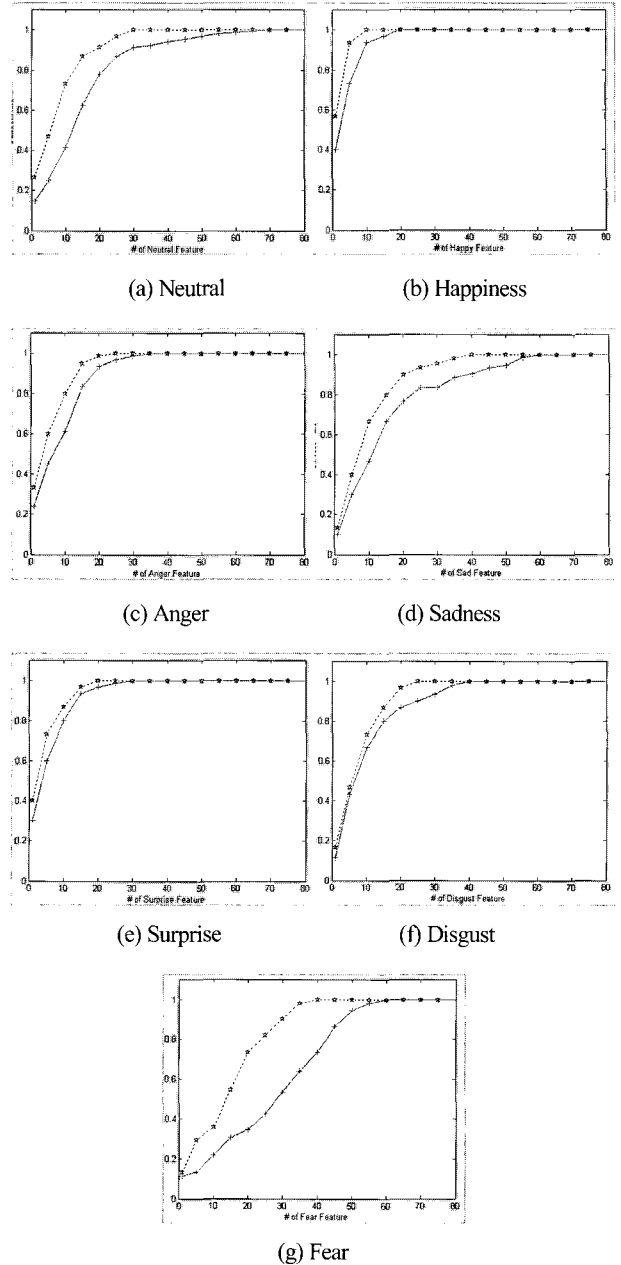


그림 8. 일곱 가지 표정의 학습영상에 대한 검출률; '+'는 7가지 특징점 모양에 대한 결과, '\*'는 제안한 42가지 특징점 모양에 대한 결과.

Fig. 8. The detection rate of 7 facial expression in training image set; '+' line is the result in case of 7 (Viola's) feature types, '\*' line is the result in case of the proposed 42 feature types.

그림 8에서 기쁨, 놀람, 화남과 같은 얼굴 표정은 전체를 검출하는데, 42개 모양인 경우 20개의 약분류기가 필요하고, 잘 구분이 안되는 평온, 슬픔, 역겨움, 공포 같은 감정들은 약간의 차이는 있지만 30개 정도의 약분류기로 모두 검출해 내고 있음을 알 수 있다. 반면에 기본 사각 특징점들을 사용한 경우는 모든 감정을 검출해 내는데 60개 정도의 약분류기가 필요하다.

V. 모의 실험 및 실험 결과

1. 인식률 비교

앞장에서 살펴본 바와 같이, 학습된 영상을 모두 올바른 얼굴 표정으로 검출하는데 있어서 기본 사각 특징점을 이용하는 경우에는 60개의 약분류기를 사용할 경우 모든 표정을 검출하고, 42개의 사각 특징점을 사용하는 경우에는 30개의 약분류기를 사용하여 모든 표정을 올바르게 검출할 수 있다. 따라서, 인식률을 비교하는데 있어서 사용한 약분류기 수는 42개의 사각 특징점을 사용하는 경우 30개의 약분류기를 사용하였고, 일곱개의 기본 사각 특징점을 이용하는 경우에는 60개의 약분류기를 사용하였다. 인식률 비교에 사용한 데이터베이스는 AR 얼굴 영상 데이터베이스, PICS (The Psychological Image Collection at Stirling) 데이터베이스, 그리고 Ekman 얼굴 영상 데이터베이스를 사용하였다. 총 정면얼굴의 407개의 영상을 검출하는데 사용하였다(그림 9).

인식률 비교에 사용된 데이터베이스는 실제로 학습에 사용된 데이터베이스와는 달리 임의로 추출된 데이터베이스이다. 얼굴의 약간의 각도, 회전, 그리고 안경을 착용한 사진, 수염 등 학습에 사용된 데이터베이스와는 여러 가지 면에서 틀리다. 또한, JAFFE 데이터베이스는 동양인(일본인)들로 이루어진 반면에 인식률 비교용 데이터베이스는 모든 인종으로 구성되었다. 실제로 학습에 사용된 영상은 영상 회전과 모든 인종을 고려하지 않고, 영상수도 많은 수가 아니기 때문에 인식률이 높게 나오지는 않았다. 하지만 기본 사각 특징점과의 비교를 하는 것이기 때문에 결과로서 서로 비교할 수 있다. 그림 10은 각 감정에 대한 인식률을 보인 것이다.

그림 10은 인식률 비교용 데이터베이스를 이용하여 일곱개의 기본 사각 특징점과 42개의 사각 특징점과의 인식률을 비교하였다. 정확도(correct rate)측은 전체 영상을 모두 올바른 얼굴 표정으로 검출했을 때를 1이라 한다면 그 상대적인 비율을 나타낸 것이다. 감정(emotion)측은 각각의 얼굴 표정의 약자로 나타내었다. 즉, NE는 neutral, HA는 happy, AN은 anger, SA는 sad, SU는 surprise, DI는 disgust, FE는 fear를 나타낸다. 그림에서 알 수 있듯이 얼굴의 특징이 잘 드러나는 기쁜 표정과 놀란 표정은 인식률이 상대적으로 다른 얼굴 표정과 비교하여 높은 것을 알 수 있다. 전체적으로 보면 기본 사각 특징점을 이용해서 추출한 약분류기를 이용한 것이 상대적으로 인식률이 낮음을 확인할 수 있다. 감정에 대해서 적게는 5%, 크게는 10%이상 인식률이 차이가 난다. 위의 결과로서 기본 사각 특징점을 사용하는 것보다 42개의 모양의 사각 특징점을 사용하는 것이 얼굴 표정을 인식하는데 더욱 효과적임을 알 수 있다.

2. 실행속도 결과

실제로 본 논문에서 구현한 얼굴 인식 시스템의 수행 속도는 표 1과 같다. 구현된 시스템은 펜티엄4 2.8GHz의 CPU를 사용하고 카메라로부터 320×240 영상을 받아들인다. 이 입력영상에서 사람이 있고, 정면을 바라보았을 경우 초기얼굴 검출에 걸리는 시간이 250~300ms 정도 소요된다. 초기에 얼굴을 검출하는 과정은 전 영역에 대해서 후보지역을

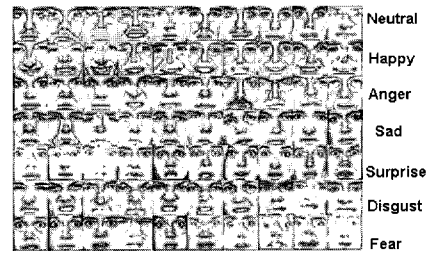


그림 9. AR, PICS, Ekman 데이터 베이스에 얻은 407개의 정면 얼굴 영상 표본.

Fig. 9. Samples of test database total 407 frontal facial images selected form the AR face database, PICS database and Ekman's face database.

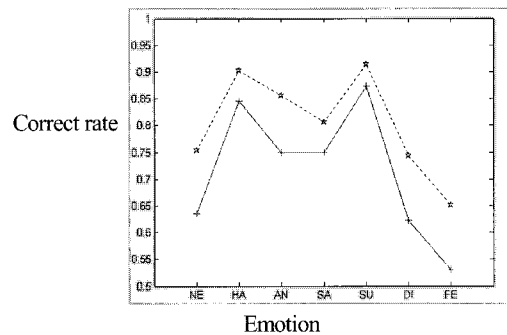


그림 10. 얼굴 표정 검출률 비교 결과; '+'는 7개의 사각 특징점의 모양을 이용해서 검출했을때의 결과, '\*'는 제안한 42개의 사각 특징점을 이용해서 검출했을 때의 결과.

Fig. 10. Comparison of the rate of facial expression recognition; '+' line shows the correct rate of each facial expression in case of 7 (Viola's) feature types, '\*' line shows the correct rate of each facial expression in case of the proposed 42 feature types.

표 1. 구현된 얼굴 표정 인식 시스템의 실행 속도.

Table 1. Processing time of implemented facial expression recognition system.

	Average Processing per image (320×240)
Initial face detection	250~300 ms
Face tracking & facial expression recognition	40~50 ms

추출하고 그 후보영역에 패턴분류를 해야 되기 때문에 얼굴 추적에 걸리는 시간보다 많이 걸리게 된다. 일단 얼굴이 검출되면 추적모드로 바뀌어서 추적 창 내에서만 패턴 분류를 하기 때문에 수행시간을 현저히 줄일 수 있다. 따라서 구현된 시스템은 초기 얼굴이 검출 되었을 경우 얼굴을 추적하고 얼굴 표정을 검출하는데 1초당 20~25장의 영상을 처리한다. 표 2은 7개의 사각 특징점으로 학습시킨 60개의 사각 특징점값과 42개의 사각 특징점을 이용하여 학습시킨 30개의 사각 특징점값을 이용하여 50×50 얼굴 영상 200장을 처리하는데 걸리는 시간을 나타낸 것이다. 위 표에서는

표 2. 7개의 사각특징점의 경우와 42개의 사각 특징점인 경우의 실행 속도.

Table 2. Processing time comparison between 7 rectangle features and 42 rectangle features.

	Average processing time per 200 images (50 × 50)
7 simple rectangle feature case (60개 features)	250~266 ms
42 rectangle feature case (30개 features)	234~250 ms

영상을 불러드리는데 걸리는 시간과 얼굴 영상을 처리하는데 걸리는 시간을 모두 측정한 것이다. 표 2에서 알 수 있듯이 두 가지 경우의 처리속도는 거의 차이가 없음을 알 수 있다.

3. 얼굴 표정 인식 실험

그림 11의 각각의 사진은 임의의 한 사람의 여러 표정을 한 개의 영상으로 만든 후 특정한 얼굴 표정을 검출하도록 실험한 사진이다. 그림 11(a)에서는 평온과 역겨움 표정을 그림 11(b)에서는 놀람과 평온, 좌측 하단은 기쁨과 놀라움, 그림 11(c)에서는 기쁨과 놀람, 그림 11(d)는 화남과 슬픔을 각각 검출하도록 설정한 후 실험한 결과이다. 그림에 보여지는 것처럼 정확하게 각 표정들을 검출하고 있음을 알 수 있다.

빠른 시간 내에 얼굴 표정을 찾는 것은 얼굴 표정 인식 시스템에 있어서 필수적인 요소라 할 수 있다. 얼굴 표정은 기본적으로 몇 초 내에 사라지므로 그 시간 내에 얼굴 표정을 검출하는 것이 필요하다. 그림 12에서는 7개의 감정(평온, 기쁨, 화남, 슬픔, 놀람, 역겨움, 공포)에 대해서 얼굴 표정을 검출하는 과정을 실험해 보았다. 먼저 평온으로부터 한 가지 감정씩 돌아가면서 실험을 해보았다.

VI. 결론

본 논문에서는 새로운 사각 특징점을 이용하여 얼굴 표정 인식을 하는 방법을 제시하였다. AdaBoost 알고리즘을 이용하여 얼굴 표정 인식을 하는 기존의 연구에서 쓰이던 기본 사각 특징점을 확장한 42개의 사각 특징점을 제안하였다. 구한 42개의 사각 특징점을 이용하여 얼굴 인식 시스템을 구성하고, 기존의 기본 사각 특징점과의 성능비교를 통해 42개의 사각 특징점을 사용하여 얼굴 표정 인식을 하는 것이 더 나은 성능을 나타낸다는 것을 보였다.

기존의 사각 특징점은 얼굴 표정의 어느 한 가지 특징만을 나타내는 약분류기를 생성한다. 하지만 42개의 사각 특징점은 얼굴 표정의 두 가지 이상의 특징을 고려함으로써 약분류기의 인식률을 높여 전체 강분류기의 인식을 상승에 도움을 주었다. 사람의 표정을 7가지(평온, 기쁨, 화남, 슬픔, 놀람, 역겨움, 공포)으로 구분해서 각각의 표정에 대해 가장 잘 구분하는 5개씩의 사각 특징점을 AdaBoost 학습 알고리즘을 이용해서 찾고, 이 구한 35개의 사각 특징점과 기본 사각 특징점 7개를 합한 42개의 사각 특징점을 이용하여 새로운 약분류기를 찾아서 얼굴 표정 인식에 이용하였다.

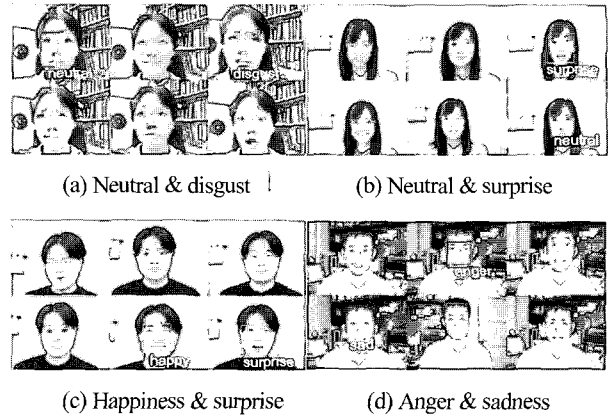


그림 11. 다양한 얼굴표정 영상에서 얼굴표정 검출 실험. Fig. 11. Facial expression recognition experiment from various facial expressions.

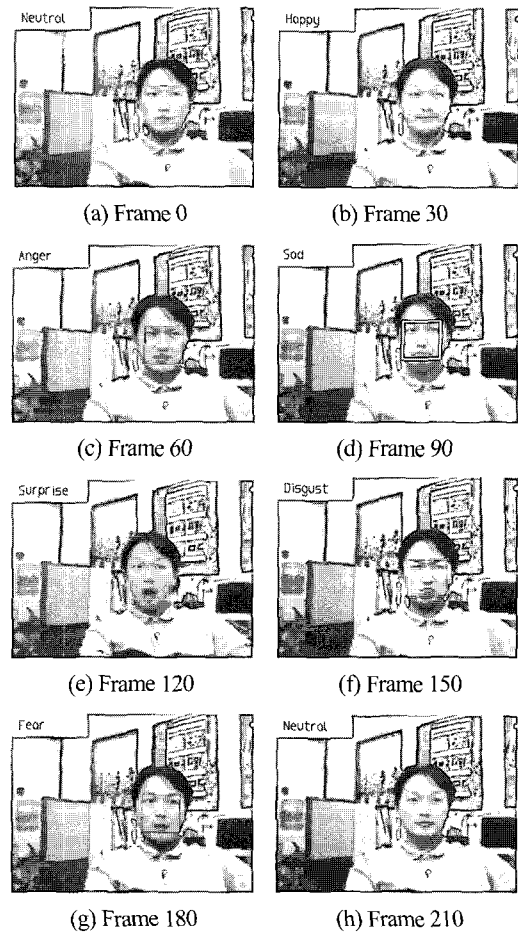


그림 12. 7가지 얼굴 표정에 대한 인식 실험. Fig. 12. Experiment of recognizing the basic 7 facial expressions.

참고문헌

[1] P. Ekman and W. V. Friesen "Unmasking the face," Malor Books press, 2003.  
 [2] M. Pantic, L. J. M. Rothkrantz, "Automatic analysis of facial expression: the state of art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp.

1424-1445, 2000.

[3] B. Fasel, J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[4] I. A. Essa, A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, 1997.

[5] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743-756, 1997.

[6] C. L. Huang and Y. M. Huang. "Facial expression recognition using model-based feature extraction and action parameters classification," *Journal of Visual Communication and Image Representation*, vol. 8, no. 3, pp. 278-290, 1997.

[7] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," *The Second IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 454-459, 1998.

[8] C. L. Lisetti and D. E. Rumelhart, "Facial expression recognition using a neural network," *The 11th International Flairs Conference*. AAAI Press, 1998.

[9] W. A. Fellenz, J. G. Taylor, N. Tsapatsoulis, and S. Kollias, "Comparing template-based, feature-based and supervised classification of facial expressions from static images," *Circuits, Systems, Communications and Computers*, pp. 5331-5336, 1999.

[10] C. Padgett and G. W. Cottrell, "Representing face image for emotion classification," In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, vol. 9, pp. 894-900, Cambridge, MA, 1997. MIT Press.

[11] M. S. Bartlett, "Face image analysis by unsupervised learning and redundancy reduction," *PhD thesis, University of California*, San Diego, 1998.

[12] J. J. Lien, T. Kanade, J. F. Cohn, C-C. Li, "Automated facial expression recognition based FACS action units," *The Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 390-395, 1998.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 511-518, 2001.

[14] Y. Wang, H. Ai, B. Wu, C. Huang, "Real time facial expression recognition with adaboost," *17th IEEE International Conference on Pattern Recognition*, vol. 3, pp. 926-929, 2004.

[15] G. C. Littlewort, M. S. Bartlett, J. Chenu, I. Fasel, T. Kanda, H. Ishiguro, & J. R. Movellan, "Towards social robots: automatic evaluation of human-robot interaction by face detection and expression classification," In S. Thrun & L. Saul & B. Schoelkopf, (Eds.) *Advances in Neural Information Processing Systems*, vol 16. pp. 1563-1570, MIT Press. 2004

[16] The Japanese Female Facial Expression(JAFFE) Database, <http://www.irc.atr.jp/~mlyons/jaffe.html>



**정성욱**

2003년 고려대학교 공과대학 전기전자 전파공학부 학사. 2005년 한국과학기술원 전자전산학과 석사. 2005년~현재 한국전자통신연구원 연구원. 관심분야 Face Expression Recognition, Computer Vision, Face Detection/Recognition, Human

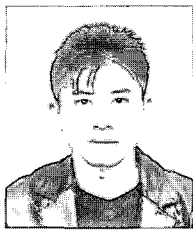
Robot Interaction.



**김도형**

1998년 전북대학교 공과대학 제어계측 공학과 학사. 2000년 한국과학기술원 전자전산학과 석사. 2000년~현재 한국과학기술원 전자전산학과 박사과정. 관심분야 Human Computer Interaction, Robot Emotion, Computer Vision, Active

Vision, Visual Servoing.



**안광호**

2002년 고려대학교 공과대학 전자공학과 학사. 2004년 한국과학기술원 전자전산학과 석사. 2004년~현재 한국과학기술원 전자전산학과 박사과정. 관심분야 Pattern Recognition, Face detection, Computer Vision, Human

Robot Interaction.



**정명진**

1973년 서울대학교 공과대학 전기공학과 학사. 1977년 미시간대학교 전기공학과 석사. 1983년 미시간대학교 제어공학과 박사. 1976년 국방과학연구소 연구원. 1981년 1월~1983년 8월 미시간대학교 CRIM 연구 조교. 1983년 10

월~현재 한국과학기술원 전자전산학과 교수. 관심분야 Service Robot. Human Robot Interaction.