

논문 2012-49SP-3-3

사각 특징을 추가한 Viola-Jones 물체 검출 알고리즘

(Viola-Jones Object Detection Algorithm Using Rectangular Feature)

서 지원*, 이 지 은*, 곽 노 준**

(Ji-won Seo, Ji-eun Lee, and Nojun Kwak)

요 약

실시간 물체 검출에 매우 효과적이라고 알려져 있는 Viola-Jones 알고리즘에서는 약분류기를 구성하기 위해 Haar 모양의 특징들을 사용한다. 이러한 Haar 모양 특징은 각각 양의 영역과 음의 영역에 해당하는 두 개 이상의 사각형의 조합으로 구성되며 양의 영역에 해당하는 화소값들의 합과 음의 영역에 해당하는 화소값들의 합의 차에 의하여 특징값을 계산한다. 본 논문에서 새롭게 제안하는 사각 특징은 두 개 이상의 사각 영역으로 구성되는 Haar 모양 특징과는 달리 단일한 사각 영역으로 구성되어 영역 내의 화소값들을 총합과 분산을 특징으로 사용한다. 이러한 사각 특징들을 기존의 Haar 모양 특징과 함께 사용하면 물체의 특징을 인접하는 밝은 영역과 어두운 영역의 조합으로만 선택했던 기존의 방법으로 인해 그동안 배제되어 온 새로운 특징을 선택할 수 있으며 그 결과 계산상의 손실 없이 물체 검출의 성능을 높일 수 있다.

Abstract

Viola-Jones algorithm, a very effective real-time object detection method, uses Haar-like features to constitute weak classifiers. A Haar-like feature is made up of at least two rectangles each of which corresponds to either positive or negative areas and the feature value is computed by subtracting the sum of pixel values in the negative area from that of pixel values in the positive area. Compared to the conventional Haar-like feature which is made up of more than one rectangle, in this paper, we present a couple of new rectangular features whose feature values are computed either by the sum or by the variance of pixel values in a rectangle. By the use of these rectangular features in combination with the conventional Haar-like features, we can select additional features which have been excluded in the conventional Viola-Jones algorithm where every features are the combination of contiguous bright and dark areas of an object. In doing so, we can enhance the performance of object detection without any computational overhead.

Keywords : Viola-Jones, 물체 검출, Haar 모양 특징, 사각 특징

I. 서 론

물체 검출이란 패턴 인식의 한 연구 분야로 기계가 사람의 도움 없이 영상에서 특정 물체를 검출하게 하는 기술이다. 이러한 물체 검출 시스템의 성능을 판단하는 가장 중요한 항목은 검출하고자 하는 대상이 되는 영역을 얼마나 정확하게 찾아내는가에 대한 것이다. 다시

말하면 검출하고자 하는 대상이 되는 물체 외의 물체나 배경을 대상 물체라고 판단하는 오인식을 방지하면서 검출하고자 하는 대상 물체의 영역을 정확하게 판단하는 것이 물체 인식의 궁극적인 목표이다. 이러한 물체 검출을 위한 가장 대표적인 알고리즘으로 약 10년 전 Paul Viola와 Michael Jones가 제시한 Viola-Jones 알고리즘을 들 수 있다^[1-3].

Viola-Jones 알고리즘은 사용되는 Haar 모양 특징의 종류에 따라 형성되는 강분류기의 성능이 크게 달라지는 특성이 있어 Haar 모양 특징을 다양하게 변형시켜 더욱 정확한 강분류기를 만들고자 하는 노력이 계속되어 왔다^[4]. 또한 Haar 모양 특징을 변형시켜 기울어진

* 학생회원, ** 정회원, 아주대학교 전자공학과
(Department of Electrical Engineering, Ajou University)

※ 이 연구는 한국연구재단 지원 (KPF-2011-0005324)에 의해 이루어졌음.

접수일자: 2011년8월5일, 수정완료일: 2012년4월25일

얼굴을 검출하기 위한 연구^[5]와 알고리즘의 속도와 정확도를 향상시키기 위한 연구^[6]도 함께 진행되어 왔다. 본 논문에서는 새로운 형태의 Haar 모양 특징인 사각 특징을 제안한다. 사각 특징은 기존의 Haar 모양 특징이 두 개 이상의 밝고 어두운 사각 형태의 영역으로 조합된 것과는 달리 단일 밝기 영역의 사각 형태로 이루어져 있다. 즉, 기존의 Haar 모양 특징이 밝고 어두운 영역의 조합을 물체의 특징으로 찾아오면서 간과했던 일정한 밝기 값을 가지는 화소들의 영역을 새롭게 특징으로 제안한 이 사각 특징에 의해 물체의 또 다른 특징으로 활용할 수 있게 되었다.

논문의 각 장의 내용은 다음과 같다. 제 II장에서는 Viola-Jones 알고리즘의 기본 원리를 설명하고, 물체 검출 과정에 대해 학습 단계와 테스트 단계로 나누어 살펴본다. 제 III장에서는 제안하는 사각 특징을 추가한 알고리즘을 수식과 함께 설명하고 물체 검출 과정을 학습 단계와 테스트 단계로 나누어 살펴본다. 제 IV장에서는 기존의 Viola-Jones 알고리즘과 제안하는 사각 특징을 추가한 Viola-Jones 알고리즘의 검출률과 오인식률을 얼굴 검출 실험과 자동차 검출 실험을 통해 비교 분석한다.

II. 기존의 Haar 모양 특징을 사용한 Viola-Jones 물체 검출

1. Viola-Jones 알고리즘에 대한 기본 배경

Viola-Jones 알고리즘은 크게 적분 영상(integral image), Haar 모양 특징, Adaboost 알고리즘, cascade 구조의 4가지 요소로 나누어 이해할 수 있다. 적분 영상과 Haar 모양 특징, cascade 구조는 연산량을 크게 줄여 실시간으로 물체를 검출할 수 있게 하며, Adaboost 알고리즘은 매우 간단한 하나의 특징만을 이용하는 단일 약분류기(weak classifier) 여러 개를 중요도에 따라 가중치를 두어 조합한 강분류기(strong classifier)를 만들어냄으로써 Viola-Jones 알고리즘이 물체를 더욱 정확하게 검출할 수 있도록 한다.

가. 적분 영상

Viola-Jones 알고리즘의 가장 첫 번째 과정은 입력 영상으로 변환하는 것이다. 적분 영상의 임의의 화소값은 해당 픽셀의 왼쪽 위에 위치한 모든 화소값들과 해당 화소값의 합으로 계산된다. 즉, 그림 1에서 동그라미

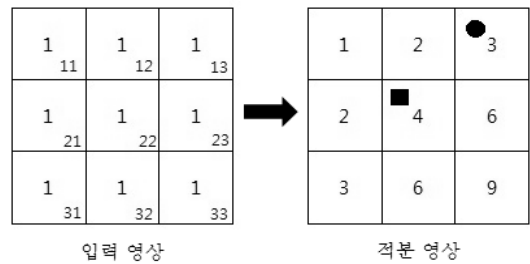


그림 1. 입력 영상을 적분 영상으로 변환
Fig. 1. Transformed integral images from input images.

로 표시된 적분 영상의 화소값은 입력 영상의 픽셀 11과 12, 13의 화소값의 합인 3이 되고, 네모로 표시된 적분 영상의 화소값은 입력 영상의 11과 12, 21, 22를 더한 4가 된다. 이러한 적분 영상을 이용하면 어떠한 사각형 영역 내의 화소값들의 합도 적분 영상 상의 네 점을 참조하여 빠르게 구할 수 있다.

나. Haar 모양 특징

Viola-Jones 알고리즘의 학습 단계에서 사용되는 Haar 모양 특징은 아래 그림 2와 같이 두 개 이상의 동일한 크기와 모양을 가진 사각형의 조합으로 이루어진다. Haar 모양 특징의 특징값은 일반적으로 흰색 사각형 영역 내의 화소값의 합과 검은색 사각형 영역 내의 화소값의 합의 차이로 표현된다. 이러한 Haar 모양 특징의 특징값은 적분 영상을 이용하여 간단히 계산된다. 예를 들면 그림 3에 나타난 특징값은 식 (1)에서 나타난 바와 같이 적분 영상에서 단지 6개의 점만을 참조함으로써 빠르고 간단하게 계산할 수 있다. 여기서 $i(p)$ 는 좌표 p 에서의 화소값을 의미하며 $ii(p)$ 는 좌표 p 에서의

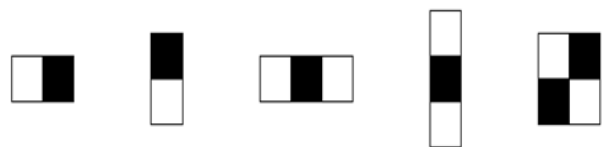


그림 2. Haar 모양 특징의 예
Fig. 2. The example of Haar-like feature.

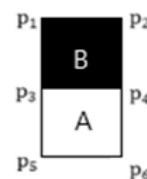


그림 3. 두 개의 사각형으로 구성된 Haar 모양 특징
Fig. 3. Haar-like feature composed by two of rectangular.

적분 영상의 최소값을 의미한다.

$$\begin{aligned} \text{Haar 특징 값} &= \sum_{p \in A} i(p) - \sum_{p \in B} i(p) \\ &= (ii(p6) + ii(p3) - ii(p4) - ii(p5)) \\ &\quad - (ii(p4) + ii(p1) - ii(p2) - ii(p3)) \quad (1) \\ &= ii(p6) + 2ii(p3) - ii(p5) \\ &\quad - 2ii(p4) - ii(p1) + ii(p2) \end{aligned}$$

다. Adaboost 알고리즘

Adaboost 알고리즘은 다양한 종류와 크기, 위치에 따라 만들어지는 수십만 개의 Haar 모양 특징들 중에서 분류 성능이 뛰어난 특징들만을 약분류기로 선별하고 이들을 조합함으로써 보다 성능이 향상된 강분류기를 만들기 위해 사용된다^[7].

Adaboost를 사용하여 임의의 샘플 x 를 두 개의 클래스 0과 1로 분류하는 문제에서 j 번째 약분류기 $h_j(x)$ 는 샘플 x 에 대한 특징 값 f_j , 임계 값 θ_j , 부등호의 방향을 표시하는 p_j 를 통해 식 (2)와 같은 수식으로 표현되며 분류 결과를 1(양의 샘플: positive sample) 또는 0(음의 샘플: negative sample)으로 나타낸다. 임계 값 θ_j 는 약분류기의 분류 성능이 최대가 되도록 하는 값으로 결정된다. 또한 양의 샘플에 해당하는 특징 값들이 음의 샘플에 해당하는 특징 값들보다 평균적으로 작은 값을 가진다면 부등호의 방향을 표시하는 p_j 는 1, 반대의 경우라면 -1로 결정된다.

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Adaboost 알고리즘의 과정은 다음과 같다.

Adaboost 초기화

i. 샘플 데이터 구성 : 검출하고자 하는 물체의 영상과 그 외의 영상인 $(x_1, y_1), \dots, (x_n, y_n)$ 을 구성한다. 여기서 x_i 는 각각의 샘플을, y_i 는 각 샘플이 속하는 클래스 정보에 따라 양의 샘플이면 1, 음의 샘플이면 0의 값을 갖도록 한다.

ii. 가중치 초기화 : i 번째 샘플에 대한 초기 가중치는 식 (3)과 같다. 여기서 m 과 l 은 각각 양의 샘플의 총 개수와 음의 샘플의 총 개수를 나타낸다.

$$w_{1,i} = \begin{cases} \frac{1}{2m} & \text{if } i = 1 \\ \frac{1}{2l} & \text{if } i = 0 \end{cases} \quad (3)$$

T개의 약분류기 선택 : iii ~ v를 T번 반복 수행한다.

iii. 가중치 정규화 : 각 단계 $t=1, \dots, T$ 에서 각각의 학습 샘플들의 가중치의 총 합이 1이 되도록 식 (4)와 같이 정규화한다.

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,i}} \quad (4)$$

iv. 약분류기 선택 : Haar 모양 특징의 종류와 크기, 위치에 따라 생성된 가능한 모든 수십만개의 특징들 중에서 임의의 어떤 j 번째 약분류기 h_j 를 학습시키고 가중치를 고려한 오분류률 ϵ_j 를 식 (5)와 같이 계산한다.

$$\epsilon_j = \sum_i w_i |h_j(x_i) - y_i| \quad (5)$$

v. 가중치 갱신 : 식 (6)에 따라 가중치를 갱신한다.

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

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t} \quad (7)$$

식 (6)에서 임의의 i 번째 학습 샘플인 x_i 가 약분류기를 통해 정확하게 분류되었다면 $e_i = 0$, 그렇지 않으면 $e_i = 1$ 이 된다. 식 (7)에서는 다음 단계의 가중치에 영향을 미치는 β_t 가 결정되는데 β_t 는 ϵ_t 에 비례하며 최대값은 오분류률이 최대가 되는 $\epsilon_t = 0.5$ 일 때 1이 된다. t 번째 약분류기 h_t 에 의해 분류된 임의의 샘플 x_i 가 정확하게 분류되었다면 $e_i = 0$ 이므로 $\beta_t^{1-0} = \beta_t$ 가 되어 다음 단계의 가중치는 감소한다. 반면에 분류결과가 틀렸다면 다음 단계의 가중치는 현 단계의 가중치와 동일하다. 각각의 학습 샘플에 대한 가중치가 재결정되면, 다시 iii으로 돌아가서 각각의 가중치를 정규화한다.

강분류기 조합

vi. iii ~ v를 T번 반복 수행하여 얻은 T개의 약분류기를 조합하여 만들어진 강분류기 $h(x)$ 는 식 (8)과 같다. 식 (9)는 각 단계에서 선택된 약분류기에 대한 가중치이며, β_t 값에 의해 결정된다. 즉, 약분류기의 가중치는 오분류률에 반비례하기 때문에 분류 성능이 좋은 약분류기일수록 큰 가중치를 갖게 하여 최종적인 강분류기의 성능을 더욱 향상시키는 역할을 한다.

$$h(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq 0.5 \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$\alpha_t = \log \frac{1}{\beta_t} \quad (9)$$

최종 강분류기를 식 (2)와 같은 약분류기들로부터 조합하기 위한 파라미터들은 다음과 같다.

- T : 강분류기를 구성하는 약분류기의 총 개수
- f_t : t 번째 약분류기가 되는 Haar 모양 특징
- p_t : t 번째 약분류기의 부등호 방향(1 또는 -1)
- θ_t : t 번째 약분류기의 임계 값
- α_t : t 번째 약분류기의 가중치

라. Cascade 구조

강분류기를 이용하여 물체를 검출할 때 매우 많은 수의 약분류기를 이용하게 된다. 이렇게 많은 수의 약분류기를 매번 사용한다면 물체 검출 결과를 실시간으로 출력할 수 없으므로 Viola-Jones가 cascade 구조를 제안하였다. Cascade 구조는 그림 4와 같이 비교적 적은 수의 약분류기들로 조합되는 강분류기 여러 개를 직렬로 연결하는 방법으로, 각 단계의 강분류기에서 물체로 판단되지 않은 샘플들은 걸러내고 물체로 판단된 샘플들은 다음 단계로 전달되어 모든 단계를 통과한 샘플들만이 물체로 검출된다.

이 과정에서 각 단계의 강분류기로 식 (8) 대신에 식 (10)을 사용하도록 한다. 식 (8)을 사용하는 경우, 예를 들어 각 단계의 강분류기의 성능이 모두 95%로 동일하다고 가정하면, $0.95^{10} = 0.60$ 이므로 최종 검출률은 60%밖에 되지 않는다. 그러므로 식 (10)에서 임계값 τ 가 0.5보다 작은 값을 가져야만 동일한 단계 수의 cascade 구조를 이용했을 때 식 (8)을 이용한 것보다 높은 검출률을 얻을 수 있다.

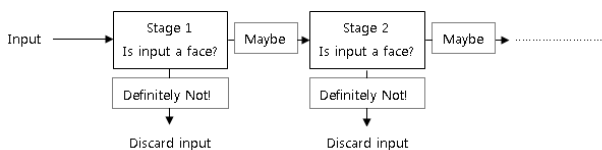


그림 4. Cascade 분류기
Fig. 4. Cascaded classifier.

$$h(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq \tau \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

2. Viola-Jones 알고리즘을 이용한 물체 검출

Viola-Jones 알고리즘을 이용한 물체의 검출 과정은 그림 5와 같이 학습 단계와 테스트 단계로 나누어진다.

가. 학습 단계

학습 단계에서는 입력 영상을 적분 영상으로 변환한 후, Haar 모양 특징의 종류와 크기 및 위치에 따라 수십만 개의 특징을 생성하여 학습 영상들로부터 각각 특징들의 중요도를 측정한다. 이 중 최상위 특징들만을 이용하여 Adaboost를 통해 가장 중요한 특징들만을 선택하고 물체 검출에 최적화된 강분류기를 구성한다.

학습 단계에서는 수십만 개의 특징들에 대해 각각의 중요도를 계산하여 중요한 특징들만을 선택하기 때문에 많은 시간이 걸린다.

나. 테스트 단계

테스트 단계에서 입력 영상을 적분 영상으로 변환한 후, cascade 구조에 입력으로 인가하여 특징값을 계산한다. 계산된 특징값들을 임계값을 기준으로 분류하여

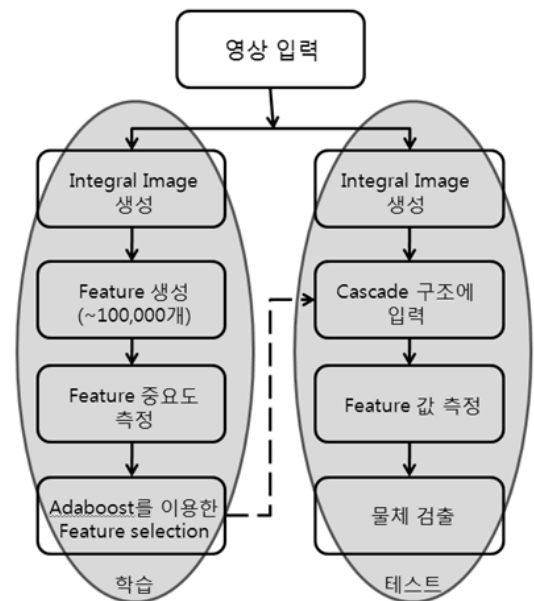


그림 5. Viola-Jones 알고리즘을 이용한 물체 검출 과정
Fig. 5. The process of object detection using Viola-Jones.

후보 영역이 검출하고자 하는 물체인진 아닌지를 판단한다.

테스트 단계에서는 학습 단계에서 미리 추출해 놓은 중요한 특징들만을 이용하여 중요도를 계산한 뒤, cascade 구조의 각 단계에서 오인식이라고 판단된 후보들은 탈락시키기 때문에 실시간으로 보다 정확하게 물체를 검출하는 것이 가능하다.

III. 제안하는 사각 특징을 추가한 물체 검출

제 III장에서는 제안하는 사각 특징을 추가한 알고리즘을 수식과 함께 설명하고 물체 검출 과정을 학습 단계와 테스트 단계로 나누어 살펴본다.

1. 제안하는 사각 특징

Viola-Jones 알고리즘에서 기존의 Haar 모양 특징을 이용할 경우 검출하고자 하는 물체의 영상에서 양의 영역과 음의 영역 화소값들의 차가 큰 부분에 놓여 있는 Haar 모양 특징들을 물체를 검출하기 좋은 특징으로 선택한다. 이러한 부분은 일반적으로 영상에서 변이가 큰 경계선(edge) 부분이다. 하지만, 일반적인 물체 검출에서 영상의 변이가 큰 경계선 부분뿐만 아니라 영상의 변이가 크지 않은 평평한 부분도 물체를 검출하는 데 중요한 특징으로 사용될 수 있다.

사각 특징을 이용할 경우 기존 Haar 모양 특징에서 이용하지 않은 물체의 색조 정보를 물체 검출에 이용할 수 있으며, 사각 특징의 분산값을 이용하여 일정 영역의 합의 값만을 이용할 때보다 좀 더 다양한 표현이 가능하므로 검출하고자 하는 물체를 좀 더 좋은 성능으로 검출할 수 있다.

예를 들어, 동양인의 얼굴을 검출하고자 할 때 기존 Haar 모양 특징을 이용한다면 대부분 눈 영역을 중요한 특징으로 간주하고 일반적으로 명도가 일정한 볼 부분은 중요한 특징으로 판단하지 않는다. 하지만 볼 부분은 얼굴 영상마다 거의 동일한 명도와 0에 가까운 아주 작은 분산값을 가지고 있으므로 이러한 특징을 이용하여 얼굴 검출 성능이 높아질 수 있다. 또한 검출하고자 하는 물체가 복잡한 형태를 지니고 배경이 거의 단색인 영상들인 경우 배경 영역의 분산이 0에 가깝기 때문에 물체에 해당하지 않는 영역을 보다 쉽게 제외시킬 수 있다.

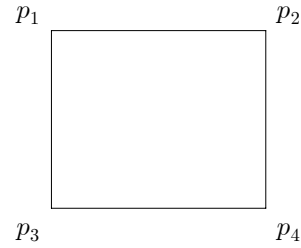


그림 6. 제안하는 사각 특징

Fig. 6. Suggested rectangular feature.

본 논문에서 제안하는 사각 특징은 기존의 흰색 사각형과 검은색 사각형으로 구성된 것이 아니라 그림 6과 같이 간단한 사각형 형태로 표현되며, 사각 영역 내의 화소값들의 총합과 분산값을 특징값으로 계산한다. 두 가지 사각 특징에 대한 식들은 아래 식 (11)과 식 (12)에 나타내었다. 식 (11)은 단순히 사각 영역 내의 화소값들의 총합을 특징값으로 계산한 것이며, 식 (12)는 사각 영역 내의 화소값들의 분산값을 특징값으로 계산한 것이다. 즉, 사각 영역 내의 화소값들의 분포를 통계학적인 수식을 이용하여 분석하고, 검출 대상인 영역과 검출 대상이 아닌 영역을 구별 가능하도록 특징을 구성한 것이다.

$$f_{sum} = \sum_{i=1}^n (x_i - m) \quad (11)$$

$$f_{var} = \frac{1}{n} \sum_{i=1}^n (x_i - m_A)^2 \quad (12)$$

여기에서 n 은 사각 영역에 포함되는 픽셀 개수를 나타내며 m 은 전체 영상의 평균 명도, m_A 는 사각 영역 픽셀들의 평균 명도를 나타낸다. 이러한 특징 f_{sum} 과 f_{var} 는 각각 적분 영상과 제곱의 적분 영상을 이용하여 쉽게 구할 수 있다. 예를 들어 그림 6과 같은 사각 특징을 사용할 때 f_{sum} 과 f_{var} 는 각각 다음의 식 (13)과 식 (14)와 같이 구할 수 있다.

$$f_{sum} = n(m_A - m) \quad (13)$$

$$f_{var} = \frac{ii(p_1) + ii(p_4) - ii(p_2) - ii(p_3)}{n} - m_A^2 \quad (14)$$

$$m_A = \frac{i(p_1) + i(p_4) - i(p_2) - i(p_3)}{n}$$

영상에서의 총 픽셀 수를 n_{total} 이라 하고 영상의 오른쪽 맨 아래 픽셀의 위치를 p_{end} 라 가정하면, $m = \frac{i(p_{end})}{n_{total}}$ 으로 나타낼 수 있다.

2. 사각 특징을 추가한 물체 검출

기존 Haar 모양 특징에 사각 특징을 추가한 Viola-Jones 알고리즘을 이용한 물체 검출 과정의 학습 단계와 테스트 단계는 기존의 방법과 같지만, 두 가지 종류의 사각 특징들이 추가되었기 때문에 생성되는 약분류기의 형태와 그들의 조합으로 만들어지는 강분류기의 성능이 기존과 달라진다.

가. 학습 단계

사각 특징을 추가한 알고리즘의 학습 단계도 기존 Viola-Jones 알고리즘과 동일한 방법으로 수행하나 그림 7에서 보는 바와 같이 기존 특징들에 두 가지 종류의 사각 특징의 크기와 위치에 따른 특징들이 추가된다. 먼저 입력되는 학습 영상을 적분 영상으로 변환한 후, 사각 특징이 추가된 Haar 모양 특징의 종류와 크기 및 위치에 따라 수십만 개의 특징을 생성하여 학습 영상들로부터 각각 특징들의 중요도를 측정한다. 이 중 최상위 특징들만을 이용하여 Adaboost를 통해 가장 중요한 특징들만을 선택하고 물체 검출에 최적화된 강분류기를 구성한다.

본 연구에서는 사각 특징의 효용성을 알아보기 위해 아래 그림 8과 같이 구성하여, 기존 Viola-Jones에 대한 실험을 수행할 때에는 ①~⑦ 가지 종류의 Haar 모양 특징을 이용하였으며 본 연구에서 제안하는 사각 특징을 추가한 실험을 수행할 때에는 ⑧ 사각 특징의 총합과 ⑨ 사각 특징의 분산을 포함하여 총 9가지 종류의 Haar 모양 특징을 사용하였다.

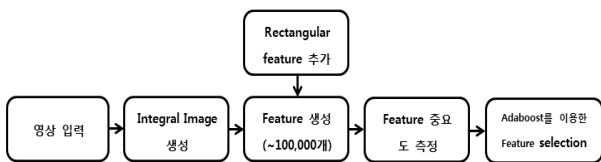


그림 7. 사각 특징을 추가한 Viola-Jones 알고리즘의 학습
Fig. 7. Training of Viola-Jones algorithm added rectangular feature.

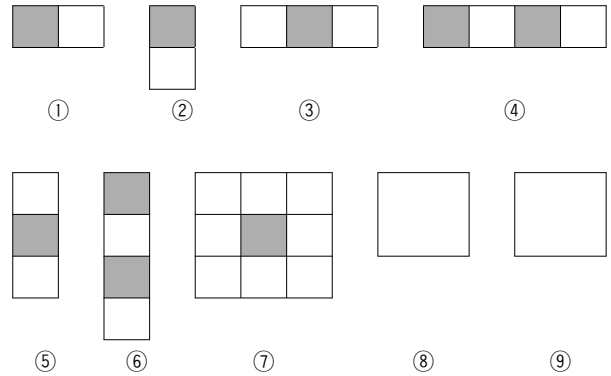


그림 8. 사각 특징을 추가한 Haar 모양 특징들
Fig. 8. Haar-like features added rectangular feature.

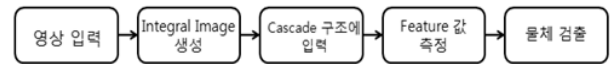


그림 9. 사각 특징을 추가한 Viola-Jones 알고리즘의 테스트
Fig. 9. Test of Viola-Jones algorithm added rectangular feature.

나. 테스트 단계

테스트 단계 또한 기존의 방법과 동일하나 학습 과정에서 사각 특징을 추가하여 강분류기를 조합했기 때문에 테스트 단계에서 후보 영역이 검출하고자 하는 물체인지 아닌지를 판단하기 위한 특징 값을 계산 할 때 식 (11)과 식 (12)를 이용하여 추가된 사각 특징에 의한 물체의 특징을 고려해야 한다. 그림 9는 사각 특징을 추가한 Viola-Jones 알고리즘의 테스트 과정을 나타낸 것이다.

IV. 실험

제 IV장에서는 기존의 Haar 모양 특징만을 이용한 Viola-Jones 알고리즘의 물체 검출 결과와 본 논문에서 제안한 사각 특징을 추가한 Viola-Jones 알고리즘 물체 검출 결과를 비교하여 성능을 분석한다.

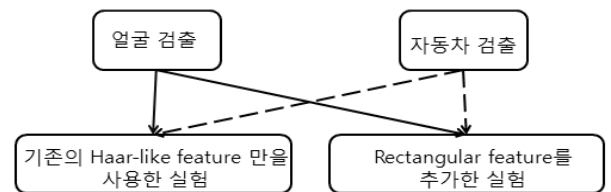


그림 10. 실험의 개요
Fig. 10. The outline of experiment.

그림 10에 나타난 바와 같이 얼굴 영상들과 자동차 영상들에 각각의 알고리즘을 적용하여 얼굴 검출 또는 자동차 검출을 수행하였다.

1. 얼굴 검출 실험

얼굴 검출 실험에 사용된 학습 데이터는 P. Carbonetto가 제공한 데이터로^[8], 총 12,788개로 구성되어 있다. 이는 4,916장의 얼굴 영상들(양의 샘플)과 7,822장의 얼굴이 아닌 영상들(음의 샘플)로 이루어져 있다. 모든 영상들은 24x24 크기의 그레이 영상이다. 그림 11은 학습 영상들의 일부를 나타낸 그림이다.

테스트에 사용된 데이터는^[9] 총 180장의 영상들로 이



그림 11. P. Carbonetto의 얼굴 영상들 중 일부 예
Fig. 11. Example of face images provided by P. Carbonetto.



그림 12. 테스트 영상들의 일부 예
Fig. 12. Example of test images.

루어져 있으며 각기 다른 크기의 그레이 영상이다. 이는 총 566개의 얼굴 영역을 포함하고 있으며, 얼굴 영역의 수는 검출률을 계산하기 위하여 실험자가 직접 산출하였다. 이 데이터에는 그림 12에서 보는 바와 같이 대부분의 얼굴 영역이 잘 정렬되어 있지 않거나 얼굴 영역이 없는 영상들도 포함되어 있기 때문에 정렬되지 않은 얼굴 영역이 좌우 90°범위 이내일 때 얼굴 영역으로 간주하였고, 얼굴 영역일지라도 그림 12의 위쪽 얼굴들처럼 상하 반전이 된 영상들은 학습 영상들과 큰 차이를 보이기 때문에 제외하였다.

가. 기존의 Haar 모양 특징만을 이용한 실험 결과

그림 8의 ①~⑦ 7가지 종류의 Haar 모양 특징만을 이용하여 4,916장의 얼굴 영상들과 7,822장의 얼굴이 아닌 영상들을 학습한 결과, 300개의 약분류기로 구성된 1개의 강분류기가 조합되었다.

테스트 단계에서 학습을 통해 얻은 단일 단계의 cascade 구조로 실험하면, 수십만 개의 후보 영역들 중 얼굴 영역이 아닌 영역들은 하나의 임계치에 의해서만 분류되기 때문에 오인식된 후보 영역들의 개수가 매우 많아진다. 그래서 300개의 약분류기를 임의로 조합하여 3단계의 cascade 구조를 구성하였고, 이에 대한 테스트 결과는 표 2에 보이듯이 356개의 얼굴 영역이 검출되었고 497개의 많은 오인식 영역도 발생하였다.

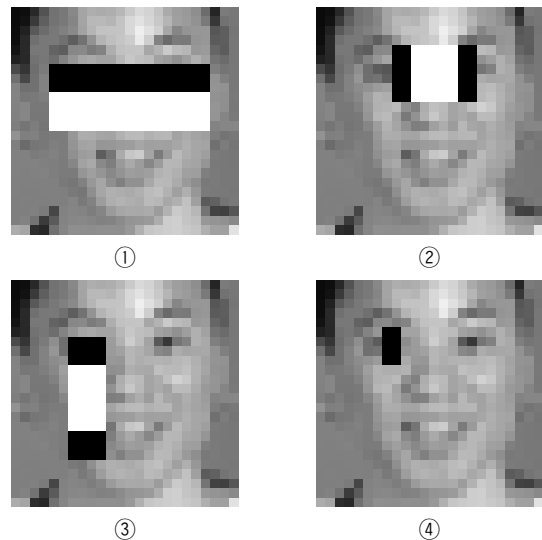


그림 13. 선별된 약분류기 중 일부 예
Fig. 13. Example of selected weak classifier.

나. 사각 특징을 추가한 실험 결과

그림 8의 ①~⑦ 7가지 종류의 Haar 모양 특징에 ⑧,

⑨의 사각 특징을 추가한 실험은 Haar 모양 특징만을 이용한 실험과 동일하게 수행하였다. 9가지 종류의 특징을 이용하여 학습을 통해 300개의 약분류기를 선별하였고 테스트 단계에서 임의로 3단계의 cascade를 구성하였다. 그림 13에서는 학습 단계에서 선별된 약분류기가 얼굴 영상들에서 어느 부분을 주요 특징으로 추출했는지를 알기 쉽게 하기 위한 예시를 보여준다. 그림 13의 ①, ②, ③번은 기존의 Haar 모양 특징을 약분류기로 선별한 예시로, 사람의 눈 영역이 뺨이나 코 영역과 함께 얼굴의 특징으로 추출된 것을 볼 수 있다. 반면 그림 13의 ④번은 본 논문에서 제안한 사각 특징을 약분류기를 선별한 예시로, 사람의 눈 영역이 단일 사각 특징으로 추출된 것을 볼 수 있다. 사각 특징을 추가한 실험에 대한 결과는 표 2에서 볼 수 있듯이, 357개의 얼굴 영역이 검출되었고 480개의 오인식 영역이 발생하였다.

다. 기존의 Haar 모양 특징만을 이용한 실험 결과와 사각 특징을 추가한 실험 결과와의 비교

표 1을 살펴보면 본 논문에서 제안한 사각 특징이 학습에서 선택된 총 300개의 약분류기 중에서 각각 9개와 92개로, 얼굴 영역을 분류하는 데에 좋은 특징으로 간주된다는 것을 확인할 수 있다.

기존 Haar 모양 특징만을 이용한 얼굴 검출 실험과 사각 특징을 추가한 얼굴 검출 실험의 테스트 결과는 표 2에서 확인할 수 있다. 표 2에서 보이는 바와 같이 사각 특징을 추가한 실험의 경우 기존의 Haar 모양 특징만을 이용한 실험에 비해 오인식 영역의 개수를 감소시키는 효과가 있으나, 그림 14와 같이 배경이 거의 단

표 1. 학습에서 선택된 각 Haar 모양 특징의 개수
Table 1. The number of each selected Haar-like feature in training.

기존 Haar 모양 특징		사각 특징 추가	
Haar 모양 특징	# of Haar 모양 특징	Haar 모양 특징	# of Haar 모양 특징
1	65	1	49
2	37	2	37
3	35	3	21
4	27	4	23
5	49	5	43
6	32	6	25
7	55	7	1
		8	9
		9	92
총	300	총	300

표 2. 얼굴 검출 테스트 결과 비교
Table 2. Comparison the results of face detection.

총 얼굴 개수	기존 Haar 모양 특징		사각 특징 추가	
	검출된 얼굴 개수	오인식 개수	검출된 얼굴 개수	오인식 개수
566	356	497	357	480

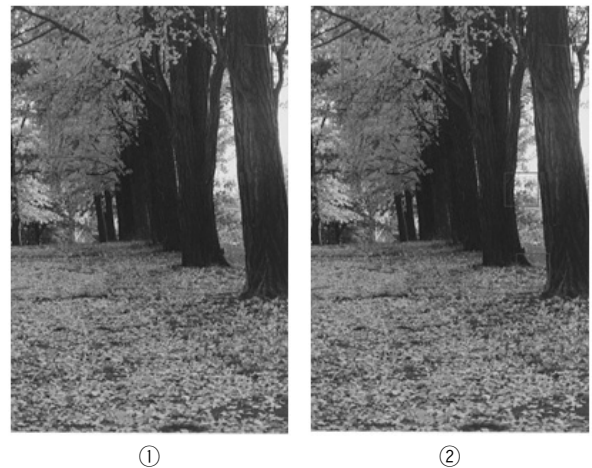


그림 14. 오인식이 증가한 예 (①: 기존 Haar 모양 특징 실험 영상 ②: 사각 특징을 추가한 실험 영상)
Fig. 14. The example that false positives are increased (①: test image using existing Haar-like features ②: test image added rectangular feature).

색으로 이루어져 있는 테스트 영상을 입력으로 인가한 경우, 학습에서 특징으로 뽑힌 분산값과 특징값이 비슷한 영역이 존재하기 때문에 기존 Haar 모양 특징만을 이용한 실험에 비해 오인식 영역이 증가하는 것을 볼 수 있다. 표 2에서 검출된 얼굴 영역의 개수와 오인식 영역의 개수에 대해 전체적으로 성능이 미미하게 향상되었는데, 이는 테스트 영상에서 좌우 90°범위 이내의 얼굴 영역만을 고려하더라도 상당수가 제대로 정렬되어 있지 않았기 때문에 검출 성능이 낮은 것이라 간주된다. 또한 validation 데이터를 사용하지 않고 cascade 구조를 세 단계만 진행하여 오인식 영역이 충분히 제외되지 않은 것으로 생각된다.

2. 자동차 검출 실험

자동차 검출 실험은 운전자의 차량을 기준으로 사각 지대 안에 존재하는 차량만을 검출하는 것을 목표로 시행하였다. 자동차의 학습 데이터는 그림 15와 같은 동영상의 각 프레임에서 자동차의 정면 영역과 자동차가 아닌 영역을 구분하여 20×20의 정사각형 형태로 잘라



그림 15. 동영상의 프레임
Fig. 15. Frame of video.



그림 16. 학습에 사용한 자동차 영상
Fig. 16. Vehicle image for training.



그림 17. 결과 분석을 위한 클래스 정보가 표시된 테스트 영상
Fig. 17. Test image marked class information for analyzing of the result.

서 획득하였다. 정사각형 형태로 잘라서 획득한 자동차 영상은 그림 16에서 볼 수 있다.

학습을 수행하기 위해 cascade 각 단계의 기준을 검

출률 99%이상, 오인식률 50%이하로 지정하였으며, 총 10단계의 cascade 구조를 수행하였다. 이 실험에서는 학습을 수행하는 중에 학습되지 않은 새로운 데이터에 대한 성능을 확인하기 위해 validation 데이터를 학습 데이터와 함께 이용하여 validation 데이터가 각 단계 강분류기의 검출률과 오인식률 기준을 만족할 때까지 약분류기들을 선택하도록 하였다.

테스트는 5개의 동영상을 이용하여 수행하였다. 각 동영상의 사이즈는 640×480으로 컬러 영상이며, 초당 30프레임으로 이루어져 있다. 영상에서 자동차가 존재하는 프레임들은 그림 17의 좌측 상단과 같이 표시하여 검출률을 객관적인 수치로 산출할 수 있도록 하였다.

가. 기존의 Haar 모양 특징만을 이용한 실험 결과
자동차 데이터를 기존의 Haar 모양 특징만을 이용하여 10단계로 학습한 결과, 총 41개의 Haar 모양 특징을 사각지대 안의 차량 영역과 차량이 아닌 영역을 구분하기에 좋은 약분류기로 선택되었다. 각 단계에서 뽑힌

표 3. 기존 Haar 모양 특징을 이용한 10단계 학습에 대한 결과
Table 3. The result of 10 cascade training using existing Haar-like features.

# of cascade	# of Haar 모양 특징	False positive ratio(%)	
		Training data	Validation data
1	4	0.1369	0.2447
2	4	0.2484	0.4569
3	4	0.1377	0.2630
4	4	0.2442	0.5708
5	4	0.4911	0.4772
6	4	0.3911	0.2043
7	4	0.0666	0.0698
8	4	0.0545	0.1074
9	4	0.0333	0.1113
10	4	0.3841	0.6316

표 4. 기존 Haar 모양 특징을 이용한 실험에 대한 동영상 테스트 결과
Table 4. The result of video test about the experiment of using existing Haar-like features.

테스트 영상	총 frame 수	차량 frame 수	검출 개수	차량 외 frame 수	오인식 개수
1	27209	9293	8131	17916	613
2	3682	2648	2477	1034	8
3	5704	3953	3443	1751	9
4	7134	3578	2991	3556	27
5	17705	4376	3791	13329	452

약분류기의 개수와 학습 데이터와 validation 데이터에 대한 오인식률을 표 3에 나타내었다.

표 4에 테스트 영상을 10단계의 cascade 구조에 인가하여 최종 검출된 자동차 영역의 개수와 오인식 영역의 개수를 프레임 수를 기준으로 나타내었다.

나. 사각 특징을 추가한 실험 결과

자동차 데이터를 사각 특징을 추가하여 10단계로 학습한 결과, 총 35개의 Haar 모양 특징을 사각지대 안의 차량 영역과 차량이 아닌 영역을 구분하기에 좋은 약분류기로 선택되었다. 각 단계에서 뽑힌 약분류기의 개수와 학습 데이터와 validation 데이터에 대한 오인식률을 표 5에 나타내었다.

표 6에 테스트 영상을 10단계의 cascade 구조에 인가하여 최종 검출된 자동차 영역의 개수와 오인식 영역의 개수를 프레임 수를 기준으로 나타내었다.

그림 18은 학습 단계에서 선별된 약분류기가 실제 자동차 영상의 어느 부분에 해당하는지를 보여주고 있다. 그림 18의 ①, ③번은 기존의 Haar 모양 특징을 약분류

표 5. 사각 특징을 추가한 10단계 학습에 대한 결과
Table 5. The result of 10 cascade training added rectangular feature.

# of cascade	# of Haar 모양 특징	False positive ratio(%)	
		Training data	Validation data
1	3	0.2396	0.2204
2	3	0.1461	0.3293
3	4	0.1611	0.2598
4	4	0.1842	0.2930
5	3	0.0977	0.2201
6	4	0.0065	0.4016
7	4	0.0184	0.2871
8	3	0.0725	0.4938
9	4	0.0406	0.0319
10	3	0.0012	0.0396

표 6. 사각 특징을 추가한 실험에 대한 동영상 테스트 결과

Table 6. The result of video test about the experiment of added rectangular feature.

테스트 영상	총 frame 수	차량 frame 수	검출 개수	차량 외 frame 수	오인식 개수
1	27209	9293	8731	17916	787
2	3682	2648	2588	1034	6
3	5704	3953	3713	1751	4
4	7134	3578	3186	3556	49
5	17705	4376	4252	13329	643

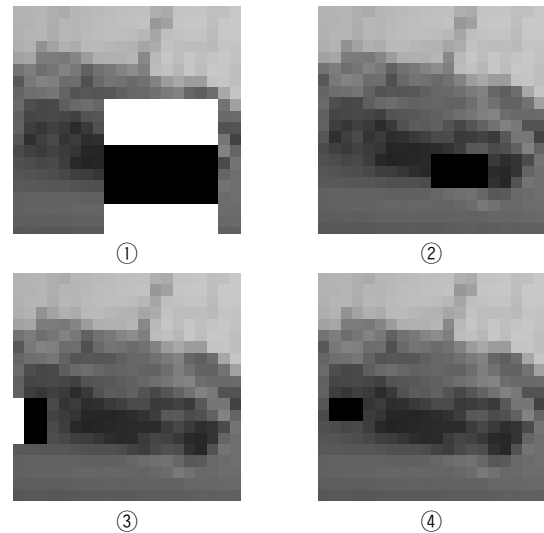


그림 18. 학습에 사용한 자동차 영상
Fig. 18. Vehicle image for training.

기로 선별한 예시로, 자동차의 번호판 영역과 바퀴 영역이 도로의 노면과 같은 배경 부분과 함께 자동차의 특징으로 추출된 것을 볼 수 있다.

반면 그림 18의 ②, ④번은 본 논문에서 제안한 사각 특징 약분류기로 선별한 예시로, 오로지 자동차의 번호판 영역이나 바퀴 영역이 단일 사각 특징으로 추출된 것을 볼 수 있다.

다. 기존의 Haar 모양 특징만을 이용한 실험 결과와 사각 특징을 추가한 실험 결과와의 비교

표 7을 살펴보면 본 논문에서 제안한 사각 특징을 추가한 실험은 기존의 Haar 모양 특징을 이용한 실험에서 총 41개의 약분류기를 사용하여야 원하는 검출률과 오인식률을 얻는 데 반해, 총 35개의 약분류기만을 사용하여도 같은 성능을 낼 수 있다. 또한 사각 특징을 추가하여 실험하면 학습에서 사각 특징이 각각 2개와 10개가 선택되어, 특히 분산 특징이 자동차 영역을 분류하는 데에 좋은 특징으로 간주된다는 것을 확인할 수 있다.

기존 Haar 모양 특징만을 이용한 자동차 검출 실험과 사각 특징을 추가한 자동차 검출 실험의 동영상 테스트 결과는 표 8에서 확인할 수 있다. 표 8에서 보이는 바와 같이 사각 특징을 추가한 실험이 기존의 Haar 모양 특징만을 이용한 실험보다 평균 검출률이 대략 6.7% 정도 증가하였고 평균 오인식률이 0.5% 증가하였다. 이를 통해 본 논문에서 제안한 사각 특징을 추가한 경우가 보다 좋은 검출 성능을 보이는 것을 확인할 수 있다.

표 7. 학습에서 선택된 각 Haar 모양 특징의 개수
Table 7. The number of each selected Haar-like feature in training.

기존 Haar 모양 특징		사각 특징 추가	
Haar 모양 특징	# of Haar 모양 특징	Haar 모양 특징	# of Haar 모양 특징
1	13	1	9
2	12	2	4
3	0	3	0
4	1	4	1
5	3	5	7
6	5	6	2
7	7	7	0
		8	2
		9	10
총	41	총	35

표 8. 자동차 검출 동영상 테스트 결과 비교
Table 8. Comparison the result of vehicle detection.

테스트 영상	기존 Haar 모양 특징		사각 특징 추가	
	검출률(%)	오인식률(%)	검출률(%)	오인률(%)
1	87.50	3.42	93.95	4.39
2	93.54	0.77	97.73	0.58
3	87.10	0.51	93.93	0.23
4	83.59	0.76	89.04	1.38
5	86.63	3.39	97.17	4.82
평균	87.672	1.77	94.364	2.28

V. 결 론

본 논문에서 제안한 사각 특징을 추가한 Viola-Jones 방법에서 두 가지 사각 특징은 사각 영역 내의 총합과 분산값을 나타낸다. 사각 특징을 추가한 실험을 통해 살펴보면 학습 단계에서 사각 특징을 중요한 특징으로 많이 뽑아내는 것을 볼 수 있다. 또한 제안한 실험 결과를 기존의 Haar 모양 특징만을 이용한 실험 결과와 비교해 본 결과, 얼굴 검출 실험에서 검출 영역의 개수는 미미한 성능 향상을 보였지만 오인식 영역의 개수가 더 적게 발생하였고, 자동차 검출 실험에서는 오인식률이 0.5%로 조금 증가하였으나 검출률이 6.7% 증가하여 좋은 성능을 도출한 것을 확인할 수 있다.

얼굴 검출 실험에서 검출률에 대해 미미한 성능 향상을 보인 이유는 테스트 영상에서 얼굴 영역의 상당수가 제대로 정렬되어 있지 않았기 때문이라 간주된다. 또한 validation 데이터를 사용하지 않고 cascade 구조를 세 단계만 진행하여 오인식 영역이 충분히 제외되지 않은 것으로 생각된다. 그러므로 얼굴 검출 실험에 대해

validation 데이터를 추가하고 여러 cascade 단계를 학습하는 연구가 필요하다.

참 고 문 헌

- [1] P. Vola and M. J. Jones, "Robust real-time face detection," International Journal of Computer Vision, 2004.
- [2] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, 2001.
- [3] P. Vola and M. J. Jones, "Robust real-time object detection," In CVPR, 2001.
- [4] Rainer Lienhart and Jochen Maydt, "An extended set of Haar-like features for rapid object detetion," IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep 2002.
- [5] 김석호, 김재민, 조성원, "Haar-like feature 변형을 이용한 기울어진 얼굴 검출," 대한전자공학회 2008년 하계종합학술대회, 987-988쪽, 2008년 6월
- [6] 로빈 칼리아, 이근동, 사미르, "A fast and accurate face detector utilizing the Viola-Jones rapid object detection technique," 대한전자공학회 2010년 하계종합학술대회, 415-417쪽, 2010년 6월.
- [7] LinLin Shen and Li Bai, "Adaboost Gabro feature selection for classification," In Proceedings of the image and vision computing, New Zealand, pp. 3, 2004.
- [8] P. Carbonetto, "Viola training data," 2001, <http://www.cs.ubc.ca/~pcarbo/viola-traindata.tar.gz>.
- [9] H. Rowley, S. Baluia, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Jan 1998, <http://www.cs.cmu.edu/~har/faces.html>

저 자 소 개



서 지 원(학생회원)
 2009년 아주대학교 전자공학부
 학사 졸업.
 2011년 8월 아주대학교
 전자공학부 석사 졸업.
 <주관심분야 : 컴퓨터 비전, 영상
 처리>



이 지 은(학생회원)
 2009년 아주대학교 전자공학부
 학사 졸업.
 2011년 아주대학교 전자공학부
 석사 졸업.
 2011년~현재 아주대학교
 전자공학과 박사과정.
 <주관심분야 : 패턴인식, 컴퓨터 비전, 영상 처
 리>



곽 노 준(정회원)
 1997년 서울대학교 전기공학부
 졸업.
 1999년 서울대학교 전기 컴퓨터
 공학부 석사 졸업.
 2003년 서울대학교 전기 컴퓨터
 공학부 박사 졸업.

2003년 3월~2006년 8월 삼성전자 정보통신
 총괄통신연구소 표준연구팀 책임연구원.
 2006년 9월~2007년 2월 서울대학교 전기컴퓨터
 공학부 BK조교수.
 2007년 3월~2011년 2월 아주대학교 전자공학부
 조교수.
 2011년 3월~현재 아주대학교 전자공학부
 부교수.
 <주관심분야 : 패턴인식, 컴퓨터 비전, 영상 처
 리>