

# 간단한 특징에 기반한 실시간 얼굴 검출

\*임옥현, 이우주, 이경일, 이배호  
전남대학교 컴퓨터정보통신공학과

## The Real-Time Face Detection based on Simple Feature

\*Ok-Hyun Lim, Wu-Ju Lee, Kyung-Il Lee, Bae-Ho Lee  
Dept. of Computer Engineering, Chonnam National University

### 요 약

본 논문에서는 간단한 사각형 특징과 계층적 분류기를 이용하여 실시간으로 얼굴을 검출하는 방법을 제안하고자 한다. 우리는 다섯 가지 형태의 기본적인 특징 모델을 바탕으로 20\*20 크기의 훈련 영상에 적용하여 많은 초기 특징 집합을 구성하였다. AdaBoost(Adaptive Boosting) 알고리즘을 이용한 학습을 통하여 초기 특징 집합 중에서 얼굴 검출하는데 강인한 집합들만을 선택하였다. 제안된 알고리즘을 이용한 실제 실험에서 90% 이상의 높은 검출율을 확인하였고 초당 10프레임의 실시간 검출에도 성공하였다.

### 1. 서론

패턴 인식의 주요 연구 분야인 얼굴 검출은 대상의 얼굴 크기, 형태, 위치, 배경 및 조명 등과 같은 복잡한 문제점을 포함하고 있어 현재 다양한 방법이 소개 되었음에도 불구하고 좀 더 효과적인 방법을 찾기 위한 연구가 계속되고 있다. 특히 실시간 얼굴 검출은 영상감시 시스템, 원격회의 시스템, 얼굴 인식 시스템 등의 다양한 분야에서 응용되고 있어 최근 많은 연구가 진행되고 있다.

얼굴 검출을 위한 방법은 윤곽선 정보, 칼라 정보, 움직임 정보를 이용한 방법과 템플릿 정합, 신경망, SVM(Support Vector Machine)과 같은 통계적 방법들이 있다. 이와 같은 기존의 다양한 방법은 얼굴 검출율을 많이 향상시켰다. 그러나 얼굴을 검출하는데 연산시간이 많이 걸린다는 단점 때문에 실시간 얼굴 검출에서는 많은 문제점이 나타났다.

최근 이러한 문제를 해결하기 위한 연구가 시작되었고 간단한 특징 모델을 기반으로 한 방법들이 소개 되었다[1][2]. 특징 기반의 얼굴 검출 방법은 학습단계에서는 신경망과 같은 통계학적 방법보다 적은 시간을 요구하고 실제 검출에 있어서는 주위 환경과 조명

변화에 강인한 장점을 가지고 있다.

본 논문은 간단한 특징 모델 기반 분류기를 이용하여 실시간으로 얼굴을 검출하는 방법을 제안한다. 우리는 다섯 종류의 간단한 특징 모델을 사용하여 초기 특징 집합을 구성하였다. 이렇게 얻어진 초기 특징 집합과 얼굴과 배경 훈련 영상을 계층적 분류기를 통하여 얼굴 패턴에 대한 중요한 구조적 특징을 추출하였다. 계층적 분류기는 각각의 단계에서 AdaBoost 알고리즘을 이용하였다.

본 논문에서는 2장에서 다섯 종류의 특징 모델을 설명하고, 3장에서 SAT(Summed Area Table)를 이용한 특징값 계산 방법 설명한다. 4장에서 AdaBoost 알고리즘을 이용한 분류기 생성을 설명하고 5장에서는 계층적 분류기에 설명하며, 6장과 7장에서는 실험 결과를 분석하고 결론을 맺는다.

### 2. 특징 모델

본 논문에서는 얼굴 검출에 간단한 특징을 이용하였다. 픽셀 기반 검출에서는 훈련 데이터의 양이 유한하면 학습을 하는데 많은 어려움이 있었으나 특징 기

반 검출은 적은 훈련 데이터의 양으로도 코드화를 통하여 얼굴 검출에 쉽게 접근할 수 있다. 또한 특징 기반 검출은 픽셀 기반의 검출 보다 빠르게 처리할 수 있는 장점도 가지고 있다. 그러나 특징 기반 검출은 복잡한 특징을 사용하였을 경우 오히려 코드화에 복잡성이 증가하여 픽셀 기반 보다 처리 속도가 느려지는 경우도 있다. 본 논문에서는 이와 같은 점을 고려하여 다섯 종류의 간단한 특징 모델만을 사용하였다.

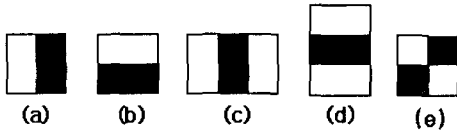


그림 1. 다섯 종류의 간단한 특징 모델

본 논문에서는 훈련 영상의 표준 크기를 20\*20로 하였다. 제안된 다섯 종류의 특징 모델은 훈련 영상에 다양한 크기로 적용되어 초기의 특징 집합을 구성하게 된다. 초기 특징 집합의 수는 44300개이다.

### 3. SAT를 이용한 특징값 계산

특징 모델을 이용하여 내부의 특징값을 구하기 위해서는 내부의 모든 픽셀의 연산이 필요하다. 이러한 과정은 여러 가지 특징들과 원 영상 이미지의 연산에서 반복적으로 일어난다.

특징값 계산이 대부분 반복적으로 일어나기 때문에 본 논문에서는 학습하기 전에 훈련 영상의 집합을 인테그랄 이미지(Integral Image)로 변환하였다. 원 훈련 영상을 인테그랄 이미지로 변환하는 이유는 방대한 특징 집합과 계산에서 연산량을 줄이기 위해서다. 인테그랄 이미지는 영역의 합 테이블(SAT: Summed Area Table)과 매우 유사한 방법이다[3].

SAT는 영상 내에 존재하는 특정한 좌표 영역(x, y)의 값은 입력 영상의 원점(0, 0)에서부터 좌표 영역(x, y)를 포함하는 사각형 영역 내에 존재하는 모든 픽셀 값을 더한 것이며, 식(1)과 같이 정의된다.

$$S(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \quad (1)$$

특징값의 계산은 영역의 합 테이블에 의해 생성된 인테그랄 이미지를 이용한다.

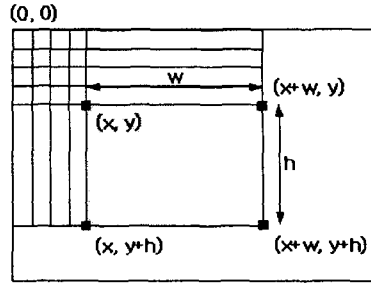


그림 2. 특징값 계산

그림2에서 사각형 (x, y, w, h)의 전체 면적 픽셀값(RecS)을 계산하면 식(2)와 같이 정의된다.

$$RecS(x, y, w, h) = S(x+w, y+h) - S(x+w, y) - S(x, y+h) + S(x, y) \quad (2)$$

선택되어진 전체 영역의 특징값(RecF)은 식(2)이 사용하여 사각형 픽셀값(RecS)과 검정색 사각형 픽셀값(RecSb)에 대한 차로 구하며, 식(3)과 같이 정의된다.

$$RecF(x, y, w, h) = RecS - RecSb \quad (3)$$

### 4. AdaBoost 알고리즘

AdaBoost(Adaptive Boosting)알고리즘의 기본 개념은 약한 분류기(Weak classifier)를 선형적으로 결합하여 최종적으로 높은 검출 성능을 가진 강한 분류기(Strong classifier)를 생성하는 것이다[4]. 약한 분류기는 얼굴과 배경 영상을 가장 잘 나누는 하나의 특징을 선택하는 역할을 한다. 약한 분류기는 식(4)와 같이 정의된다.

$$h_j(x) = \begin{cases} 1 & \text{if } f_j(x) < \theta_j \\ -1 & \text{if } f_j(x) \geq \theta_j \end{cases} \quad (4)$$

식(4)에서  $f_j(x)$ 는 영상과 특징을 이용하여 특징값을 구하는 특징 함수이며,  $\theta_j$ 는 얼굴과 배경을 구별하는 임계값(threshold)이다.

강한 분류기는 약한 분류기의 선형적 결합 형태로 여러 개의 특징을 결합해서 실질적으로 얼굴의 패턴을 구별하는 역할을 한다.

표1. AdaBoost 알고리즘

1. Given  $N$  examples  $(x_1, y_1), \dots, (x_N, y_N)$   
with  $x \in R^k, y_i \in \{-1, 1\}$
2. Initialize weights  $w_1(i) = \frac{1}{N}, i = 1, \dots, N$
3. Repeat for  $t=1, \dots, T$ 
  - (a) Train weak learner using weight  $w_t$
  - (b) Get weak hypothesis  $h_t = X \rightarrow \{-1, +1\}$   
with error  $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$
  - (c) choose the classifier using error  $(\epsilon_t, h_t)$
  - (d) weight update  

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right)$$

$$w_{t+1}(i) = w_t(i) \cdot \begin{cases} e^{-\alpha} & \text{if } h_t(x_i) = y_i \\ e^{\alpha} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

$$= w_t(i) \cdot \exp(-\alpha_t y_i h_t(x_i))$$
4. Output the final hypothesis (strong classifier)  

$$H(x) = \text{sign} \left( \sum_{i=1}^T \alpha_i h_i(x) \right)$$

분류기는 얼굴과 배경 패턴으로 구성된 훈련 영상과 다섯 종류의 특징에 의해 생성된 초기 특징 집합을 이용하여 학습된다.

약한 분류기( $h_t$ )는 표1의 3번과 같이 가중치( $w_t$ )를 이용하여 학습을 한다. 입력 받은 훈련 영상을 잘못 분류하면  $w_t(i)$ 를 증가시키고, 옳게 분류하면  $w_t(i)$ 를 감소시킨다. 이때 가장 낮은 에러율을 가지는 사각형 특징을 선택한다. 하나의 특징이 선택되면 다시 훈련 영상을 입력 받아들이고 이와 같은 과정을 계속 반복( $T$ 번)하여 가중치를 변환하고 에러율이 가장 낮은 특징을 선택한다. 결과 이것은 약한 분류기를 단계적으로 생성하게 되고 이것들은 선형적으로 결합하여 강한 분류기를 생성한다. 강한 분류기는 표1의 4번에 최종 출력과 같이 정의된다.

### 5. 계층별 분류기 학습

본 논문에서는 그림3과 같은 계층적 구조에 기반을 둔 학습 방법을 제안한다. 기존의 AdaBoost알고리즘을 이용해서 하나의 강한 분류기를 생성하기 위해서

는 많은 수의 훈련 영상이 필요했다. 매우 큰 훈련 영상 집합은 학습 시간을 증가시키게 된다. 또한, 훈련 영상 집합 내의 학습데이터가 변경되면 결과에 변화가 오기 때문에 다시 처음부터 학습을 해야 하는 단점이 있다.

본 논문에서는 간 단계별로 비 얼굴 패턴 일부를 제거하고 다음 단계에서 새롭게 훈련 영상을 구성하였다. 특히 각 단계의 학습과정에서 오검출된 영상들은 얼굴과 비슷한 패턴을 갖는 영상이기 때문에 오검출 영상 집합을 만들어 다음 단계에서 반복적으로 적용한다.

계층적 구조는 일반적으로 각 스테이지에서 많은 수의 훈련 영상을 제거하므로 다음 스테이지를 학습할 때 연산량을 대폭 감소시키는 장점이 있다.

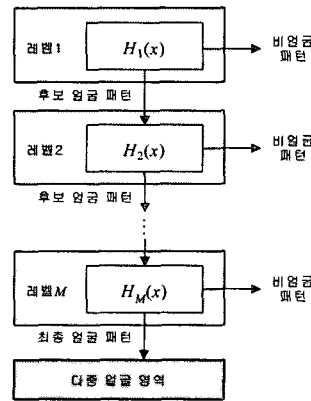


그림 3. 계층적 분류기 학습

### 6. 실험 및 결과 분석

실험에서 사용한 훈련 영상 집합은 3000개의 얼굴 영상과 25000개의 배경 영상으로 구성되며, 각각 영상은 20\*20 크기의 그레이 영상이다. 그리고 다섯 종류의 간단한 특징 모델을 이용하여 44300개의 초기의 특징 집합을 생성하였다. 각 특징들은 20\*20 영상 내에서 다양한 위치에 존재한다. 훈련 영상 집합과 특징 집합은 본 논문에서 제안한 AdaBoost알고리즘의 입력 값으로 사용된다.

본 실험에서 학습의 목표를 검출율 90%로 정하고 검출율이 90%이상일 때 까지 단계별로 학습을 하였다. 학습 결과로 그림4와 같이 24계층에서 검출율이 90%를 보였다. 이때 초기의 특징 집합 44300개 중 얼굴에 강인한 특징 집합 2134개를 얻을 수 있었다.

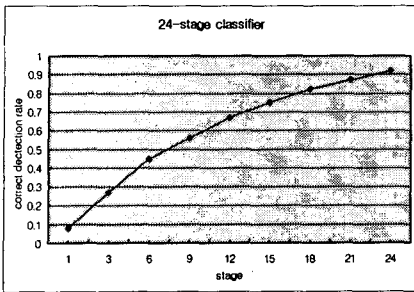


그림 4. 단계별 검출율

본 실험에서는 이렇게 얻어진 특징 집합과 강한 분류기를 이용하여 실제 환경에 적용하였다. 실제 환경에서 조명과 복잡한 환경에 관계없이 91%의 높은 검출율을 보였다. 또한 초당 10프레임의 실시간 검출도 성공하였다.



그림 5. 실시간 얼굴 검출

## 7. 결론

본 논문은 간단한 특징 모델 기반 분류기를 이용하여 실시간으로 얼굴을 검출하는 방법을 제안하였다. 제안된 알고리즘은 계산의 효율성과 검출 성능을 동시에 만족시키며, 얼굴 영역에서 발생하는 다양성을 수용할 수 있었다.

제안된 알고리즘은 SAT를 이용하여 특징값을 효율적으로 계산하였다. 분류기 학습은 AdaBoost 알고리즘을 이용하여, 높은 정확도를 가진 분류기를 계층적으로 구성한다. 또한 중요한 얼굴 패턴은 다음 레벨에 반복적으로 적용함으로써 우수한 검출 성능을 가진다.

이와 같은 방법으로 생성된 특징 기반 분류기는 얼굴 패턴을 추출하는데 결정적인 역할을 하는 특징을 이용하여 얼굴을 검출한다. 실시간 얼굴 검출은 특징 기반 분류기를 통해, 빠르고 효율적으로 얼굴 영역을 찾아냈다. 제안된 알고리즘은 실험에서 우수한 계산

효율성과 검출 성능을 보였고, 실시간 처리에서 초당 10프레임의 속도를 나타내었다.

## [참고문헌]

- [1] Paul Viola, M. Jones, "Robust real-time object detection", *International Conference on Computer Vision*, 2001.
- [2] S.Z. Li, L. Zhu, Z.Q. Zhang, and H.J. Zhang, "Statistical Learning of Multi-View Face Detection", In *Proc. 7th European Conference on Computer Vision*, Copenhagen, Denmark. May 2002.
- [3] F. Crow, "Summed-area tables for texture mapping", In *Proceedings of SIGGRAPH*, Vol.18(3), pp. 207-212, 1984.
- [4] Y. Freund, R.E. Schapire, "A Short Introduction to Boosting", *Journal of Japanese Society for Artificial Intelligence*, 14(5), pp. 771-780, 1999.
- [5] M. Oren, C. Papageorgiou, E. Osuna, P. Sinha, T. Poggio. "Pedestrian detection using wavelet template", In *Proceedings IEEE Conf on Computer Vision and Pattern Recognition*, pp193-199, 1977.