# 색상에 기반한 실시간 얼굴 검출 및 추적 시스템

임옥현°, 이우주, 이배호
전남대학교 컴퓨터정보통신공학과
ohlim@moiza.chonnam.ac.kr

# The Real-Time Face Detection and Tracking System based on Skin-Color

Okhyun Lim°, Wuju Lee, Baeho Lee
Dept. of Computer & Information Engineering, Chonnam National University

**요 약**

본 논문에서 색상을 기반으로 한 알고리즘으로 얼굴을 검출하고 검출된 얼굴을 움직이는 Pan-Tilt 카메라상에서 추적하는 방법을 제안하고자 한다. 얼굴 검출 알고리즘은 얼굴색의 특징인 피부색상을 이용하여 후보영역을 검출하고 후보 영역에서 얼굴형태의 특징인 타원 형태를 이용하여 최종적으로 얼굴을 검출하였다. 얼굴 추적은 영상에서 검출된 얼굴의 크기 및 위치 정보와 Pan-Tilt 카메라의 위치정보를 이용하여 항상 얼굴이 카메라의 중심에 위치하도록 하였다. 우리는 실제 실험에서 초당 10프레임 이상의 실시간 얼굴 검출 및 추적에 성공하였다.

## 1. 서 론

패턴 인식의 주요 분야인 얼굴검출은 대상 얼굴 크기, 형태, 위치, 배경 및 조명등과 같은 복잡한 문제점을 포함하고 있으며, 이로 인해 다양한 방법이 소개되었음에도 불구하고 좀 더 효과적인 방법을 찾기 위해 연구가 활발히 진행되고 있다. 기존의 대부분의 검출 연구들은 대상의 입력 영상에 제약을 준 상태에서 연구가 이루어져서 대상이 제약조건을 벗어날 경우 검출율이 크게 떨어지게 되었다. 실시간으로 영상을 검출하기 위해서는 대상에 특별한 제약을 가하지 않고 복잡한 배경에서 일반적인 입력 영상을 가지고도 얼굴 영역을 쉽게 분리할 수 있어야만 한다.

본 논문에서는 아무런 제약을 가하지 않은 복잡한 배경을 포함하는 일반적인 영상에서 피부색과 타원형인 얼굴형태를 이용하여 얼굴 영역 검출할 수 있는 알고리즘을 제안한다. 사람의 피부색은 YCbCr 색상 모델에서 다른 사물과 구별되는 불변의 특징이기 때문에 얼굴 영역을 검출하는데 좋은 특징이 될 수 있다. 또한 타원형의 얼굴형태는 이렇게 변환된 영상에서 얼굴을 최종적으로 검출하는데 훌륭한 기준이 될 수 있다. 얼굴 검출은 크게 피부색상 추출과 추출된 피부색상을 이용한 얼굴 후보 영역 추출 단계, 영역에서 얼굴을 결정하는 최종 얼굴 영역 추출 단계로 이루어진다. 이렇게 검출된 얼굴은 Pan-Tilt 카메라를 이용하여 광범위 시야에서 얼굴을 추적한다. 얼굴 추적은 얼굴의 크기 및 위치 정보와 Pan-Tilt 카메라 자체 위치 정보를 이용하여 실시간 추적이 가능하도록 했다.

본 논문은 2장에서는 색상 기반의 얼굴 검출 방법을 설명하고, 3장에서는 Pan-Tilt 카메라를 이용한 광범위 시야에서 얼굴 추적 방법을 설명하고, 4장에서는 제안한 방법을 이용하여 실험결과를 분석하고, 마지막으로 5장에서는 결론을 제시한다.

## 2. 얼굴 검출 방법

### 2.1 YCbCr색상 모델

카메라로 입력된 영상에서 얼굴을 검출하기 위해서는 입력 영상으로부터 각 픽셀의 RGB값을 식(1)에 대입하여 YCbCr색 공간으로 변환하여야 한다. YCbCr 색상모델에서 일반적인 피부색상은 전체 색상 영역 중 매우 좁은 영역에 분포하므로 다른 색상 모델에 비해 좀 더 효율적으로 얼굴 영상을 다룰 수 있다.

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

### 2.2 피부색상 추출

얼굴의 주요 특징 중 피부색은 다른 사물과 구별되는 특정한 색상 영역을 가지고 있다. 물론, 비슷한 색상을 포함하고 있는 사물들도 많이 존재하지만 대부분의 경우 얼굴 색상과는 많은 차이를 나타내고 있다. 따라서 얼굴 색상을 이용할 경우 복잡한 배경에서 얼굴 영역을 추출하기 위한 매우 훌륭한 필터가 될 수 있다. 본 논문에서는 표본 얼굴 색상 구축을 위해 90개의 얼굴 영상을 사용하였고, 표1은 각각의 얼굴 색상의 CbCr 평면에서의 평균값을 나타내고 있다.

표 1. Sample Skin-tone

| Skin | Number | Cb | Cr |
|---|---|---|---|
| Light | 30 | -25.8 | 33.2 |
| Medium | 30 | -23.5 | 25.7 |
| Dark | 30 | -23.2 | 30.4 |
| Skin-tone | 90 | -24.2 | 29.8 |

표1에서 각 피부색상의 평균 Cb, Cr값이 매우 유사함을 알 수 있으며, 표본 피부색상의 중점은 (-24, 30)으로 결정하였다.

입력영상에서 피부색상 추출은 표본 피부색상을 바탕으로 2D-Gaussian함수를 사용하여 피부색상을 모델링하였고 이를 입력 영상에 적용하여 임의의 임계값을 기준으로 하여 얼굴 색상 여부를 판별하여 피부색상 추출을 수행하였다.

$$G(x,y)=\frac{e^{-\frac{1}{2(1-\rho_{xy}^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2-\frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho_{xy}^2)}} \quad (2)$$

얼굴 색상 영역은 CbCr 평면에서 제 4사분면 위에 존재하므로 이 부분만을 2D-Gaussian 함수로 모델링하였다. 식(2)에서 G(x, y)는 $(\mu_x, \mu_y)$일 때 최대값을 갖으며, 최대값은 식(3)과 같다.

$$G(\mu_x,\mu_y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho_{xy}^2)}} \quad (3)$$

그러나, 식(3)에서와 같이 G(x, y)의 최대값이 너무 작으므로, 이를 1.0이 되도록 일반화하고, 표본 피부색상을 적용하여 이를 입력 영상에서 피부색을 결정하는 임계값으로 사용한다. 임계값(Th)은 식(4)에 의해 계산된다.

$$Th=p(x,y)\left(2\pi\sigma_x\sigma_y\sqrt{(1-\rho_{xy}^2)}\right)G(x,y) \quad (4)$$

p(x, y)는 CbCr 평면(x, y) 위치에 대한 표본 피부색상의 발생확률이다. 이 임계값을 이용하여 입력영상에 대하여 얼굴 색상여부를 판별하였다.

### 2.3 영역분할 및 후보 영역 추출

피부색상 추출은 픽셀 단위로 이루어지므로 영역을 이루지 못한다. 얼굴형태를 비교하기 위해서는 넓게 퍼진 픽셀들을 통합하여 영역으로 묶고 각각을 다시 덩어리로 분리해야한다.

본 논문에서는 후보 영역을 추출하기 위해 추출된 얼굴 색상 영역을 Region-growing 알고리즘을 적용하였다. Region-growing 알고리즘은 영역을 분리하여 같은 영역내의 모든 구성요소들은 동일한 값을 갖게 하고, 영역의 크기에 대한 정보를 얻을 수 있게 한다. 우리는 얼굴의 크기 20*20픽셀로 제한하고 이보다 적은 영역은 비얼굴 영역으로 간주하고 제거하였다. 이 과정을 거치면 영상은 일정한 크기 이상의 얼굴 색상 영역만을 포함하게 된다.

본 논문에서는 얼굴 후보 영역을 추출하기 위해 얼굴의 타원형태를 이용하였다. 우리는 임의의 영역에 대해 그 영역의 기울기와 크기를 갖는 타원형 파라미터를 구하는 best-fit ellipse 알고리즘을 적용하였다. 주어진 영역의 전체적인 기울기 $\theta$, 타원에서 장축 a, 단축 b는 다음 식(5)와 같이 구할 수 있고 이를 바탕으로 최종 얼굴 후보 영역을 추출한다.

$$a=\left(\frac{4}{\pi}\right)^{\frac{1}{4}}\left(\frac{I_{\max}^3}{I_{\min}}\right)^{\frac{1}{8}},\ b=\left(\frac{4}{\pi}\right)^{\frac{1}{4}}\left(\frac{I_{\min}^3}{I_{\max}}\right)^{\frac{1}{8}}$$
$$\theta=\frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \quad (5)$$

본 논문에서는 인간의 해부학 특성상 a:b는 1:2까지만 허용하고 $\theta$값은 $-45° < \theta < 45°$로 제한하였다.

### 2.4 최종 얼굴 영역 추출

최종 얼굴 추출은 후보 영역들 중 실제 얼굴 영역을 포함하고 있는 영역을 검출하는 것이 목표이다. 본 논문에서는 각 영역에 대해 얼굴의 존재 여부를 검사하기 위해 신경망을 사용하며, 신경망 학습을 위해 모멘텀 역전파 알고리즘을 사용하였다.

신경망의 입력으로 각 후보 영역을 사용하였고, Average shrinking알고리즘을 사용하여 20*20 픽셀크기로 정규화한 후 패턴을 생성하였다.

신경망 학습을 위해서는 다량의 얼굴 패턴 및 비 얼굴 패턴이 요구되며, 본 논문에서는 CMU 얼굴 데이터베이스와 인터넷을 이용하여 총 310개의 얼굴 영상 패턴을 추출하였다. 비얼굴 패턴을 추출하기 위해 풍경사진과 같은 비얼굴 영상에서 임의의 사각영역을 캡쳐하여 이를 패턴화하였다. 신경망은 총 20*20 크기에서 마스크 60개를 제외한 340개의 입력 뉴런, 100개의 은닉뉴런, 1개의 출력 뉴런으로 구성되었으며, 활성화 함수로 bipolar sigmoid함수를 사용하였다. 또한 학습률 $\alpha$=0.02, 모멘텀 상수 $\beta$=0.04, 최대 허용 에러 e=0.01을 학습파라미터로 사용하였으며, weight 값은 -0.5 ~ 0.5 사이의 임의의 값으로 초기화하여 사용하였다. 신경망의 출력은 -0.1 ~ 1.0 사이의 범위를 가지며, 본 논문에서 0.0을 얼굴 영역 결정 임계값으로 사용하였다.

## 3. 얼굴 추적 방법

움직이는 카메라를 이용하여 얼굴을 추적에는 프레임간의 차영상, 얼굴의 움직이는 방향, 흐름 정보를 이용한다. 그러나 카메라와 추적할 얼굴이 같이 움직인다면 알고리즘이 복잡해지고 이로 인하여 실시간 얼굴 추적이 어렵게 된다. 본 논문에서는 프레임간의 얼굴의 크기 및 위치정보와 Pan-Tilt 카메라의 위치정보를 이용하여 얼굴이 화면의 중심에 오는 방법을 통한 얼굴 추적을 제안한다.

실시간 얼굴 추적은 검출된 영역의 중심을 전체 영상의 중심으로 옮기는 방법을 기본으로 사용하였다. 검출된 얼굴의 중심을 찾기 위해 타원으로 검출된 얼굴 영역을 포함하는 사각형을 형성한다.

구성된 얼굴 영역 사각형의 크기(w, h)와 시작점(x, y)를 이용하여 얼굴의 중심을 계산한다. 얼굴 추적은 얼굴의 중심(x', y')을 영상의 중심(X, Y)에 위치하도록 하면서 이루어진다. 이렇게 하기 위해서는 Pan-Tilt 카메라가

얼굴의 중심을 윈도우의 중심으로 이동시켜 주어야한다.

영상의 중심과 얼굴의 중심의 위치 정보를 이용하여 Pan-Tilt를 구동하기 위한 파라미터를 결정한다. 이때 얼굴의 움직임을 부드럽게 하기 위하여 얼굴의 중심과 영상의 중심의 거리에 따라 가우시안 분포 가중치를 적용한다.

얼굴 검출 실패로 인한 얼굴 추적이 순간적으로 실패하였을 경우 이것을 보완하기 위하여 10프레임간에 얼굴의 위치정보와 크기정보를 저장하여 보존하였다.

### 4. 실험 결과 및 분석

본 논문에서 제안된 알고리즘을 적용하여 정지 영상에서의 얼굴 검출과 Pan-Tilt 카메라를 이용한 실시간에서의 얼굴 검출 및 추적을 실험 분석하였다.

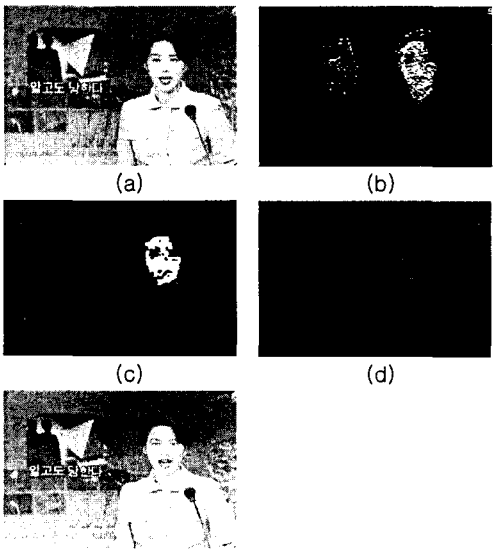


그림 1. 얼굴영역 검출 (a)원본 영상 (b)피부색상 추출 (c)영역분할 (d)후보영역 (d) 최종 얼굴영역 검출

정지 영상에서의 얼굴 검출은 인터넷에 게제된 사진, 디지털 카메라 사진 등 150개 영상을 이용하였다. 150개의 영상에는 219개의 영상이 존재하였고 표2는 실험결과이고 그림1은 실험영상 중 하나의 실험 과정을 나타낸 것이다.

표 2. 정지 영상에서 실험 결과
(총 얼굴개수: 219, 총 후보영역개수: 261)

| | |
|---|---|
| 얼굴 후보 영역 추출 | 202/219(92.2%) |
| 추출 실패 | 17/219(7.8%) |
| 얼굴 검출 | 191/202(94.5%) |
| 신경망 거부 | 11/202(5.5%) |
| 비얼굴 영역 검출 | 10/42(23.8%) |

Pan-Tilt 카메라를 이용한 실험 결과는 정지 영상을 이용한 검출율보다 낮은 대략 84%를 나타내었다. 검출율이 낮은 이유는 정지 영상보다 화질이 낮고 주위환경이 정지영상보다 복잡하였기 때문이다. 얼굴 추적은 10프레임으로 이루어져 실시간 속도에 가깝게 유지할 수 있었다.

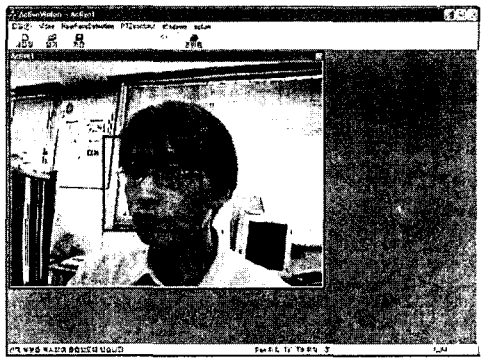

그림2. Pan-Tilt카메라를 이용한 실험 결과

### 5. 결론

본 논문에서는 얼굴색의 특징인 피부색상과 타원의 얼굴형태를 이용하여 얼굴을 검출할 수 있는 알고리즘과 얼굴의 위치와 Pan-Tilt의 위치 정보를 이용하여 얼굴을 추적할 수 있는 알고리즘을 제안하고 실험하였다.

실험결과 제안한 알고리즘은 정지영상에 대하여 90%에 가까운 검출율을 보였고 Pan-Tilt 카메라를 이용한 실시간 검출 및 추적에서도 84%의 검출율과 10프레임의 추적 속도를 보였다. 검출 실패의 원인을 살펴보면 피부색 추출에 대한 부분이 가장 높았다. 따라서 검출율을 더욱 높이기 위해서는 피부색 추출에 대한 연구가 계속 이루어져야 한다.

### 참고문헌

[1] 유태웅, 오일석, "색채 분포 정보에 기반한 얼굴 영역 추출", 정보과학회논문지, Vol 24, pp. 180-192, 1997

[2] C. Garacia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis", IEEE Trans. on Multimedia, Vol. 1 , No. 3, pp. 264-277, 1999

[3] R. Burunelli and T. Poggio, "HyperBF networks for gender classification", in Proc., DARPA Image Understanding Workshop, p. 311-314, 1992

[4] Randy crane, A Simplified Approach to Image Processing, Prentice-Hall, 1997