

신경망과 적응적 스킨 칼라 모델을 이용한 얼굴 영역 검출 기법

손 정 덕(孫正德)*, 고 한 석(高漢錫)**

* 고려대학교 영상정보처리학과, ** 고려대학교 전자공학과

Human Face Detection from Still Image using Neural Networks and Adaptive Skin Color Model

Jung Duk Son*, Han Seok Ko**

* Dept. of Visual Information Processing, Korea University

** School of Electrical Engineering, Korea University

E-mail : jdson@ispl.korea.ac.kr

Abstract

In this paper, we propose a human face detection algorithm using adaptive skin color model and neural networks. To attain robustness in the changes of illumination and variability of human skin color, we perform a color segmentation of input image by thresholding adaptively in modified hue-saturation color space (TSV). In order to distinguish faces from other segmented objects, we calculate invariant moments for each face candidate and use the multilayer perceptron neural network of backpropagation algorithm. The simulation results show superior performance for a variety of poses and relatively complex backgrounds, when compared to other existing algorithm.

I. 서론

최근에 하드웨어와 소프트웨어 기술의 눈부신 발전으로 고해상도 이미지 획득이 용이해지면서 영상으로부터의 정보추출과 처리기술이 많이 연구되고 있다. 특히, 자동적인 얼굴 인식 분야는 보안시스템이나 개인식별, 범인식별 시스템 등의 많은 적용 범위를 가지고 있어 활발히 연구되어지고 있다. 그런데 우수한 자

동 얼굴 인식 시스템을 개발하기 위해서는 정확한 얼굴 영역검출이 선행되어야한다. 그러나 얼굴 영역검출의 정확도는 조도의 영향, 얼굴의 기울어짐, 복잡한 배경과 카메라(영상획득매체)와의 거리에 따른 얼굴 크기의 변화 등의 요인에 커다란 영향을 받는다.

본 논문에서는 이러한 요인들에 강인한 알고리즘을 제안한다. 기존의 알고리즘에서는 RGB성분을 HSV 성분으로 변환하고 통계적 칼라 모델에 단일 기준의 Mahalanobis metric을 사용하여 조도의 영향에 민감하며 환경변화에 적응적이지 못하였다.

제안된 방법에서는 RGB 성분을 변환된 HSV 성분인 TSV 성분으로 변환하고 이미지 조절 (Image adjustment)을 통해 살색영역과 비살색 영역의 대조 (Contrast)를 높여주었으며 통계적 칼라 모델에 두 개 기준의 Mahalanobis metric을 사용하여 적응적으로 얼굴 후보 영역을 추출하였다. 또한 Hu[1] 에 의해 처음으로 소개된 invariant moments 기법 중에 Fourier Mellin moments를 사용하여 얼굴 기울어짐, 얼굴크기 변화에 강인하게 만들었고 최종적으로 이 모멘트를 신경망의 입력으로 사용하여 얼굴 영역을 추출한다.

II. 얼굴 후보 영역 검출

TSV공간에서 통계적인 칼라 모델을 생성하고 이 모델에 의해 이진영상(binary image)을 만들어 살색영역

을 검출한다. 검출된 영상에 대해 후처리작업을 거쳐 최종 후보영역을 얻는다.

1. 색공간의 결정

기존 방법들은 RGB 성분을 HSV 성분으로 변환하여 살색영역을 추출해주는 것이 대부분이었다. 그림 1 (a)

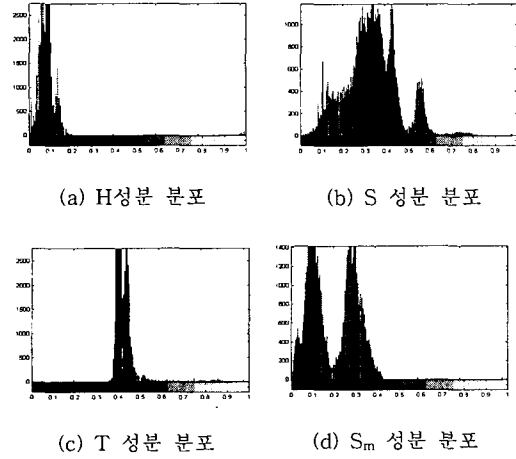


그림 1. 살색영역에서의 각 성분 분포(Histogram)

와 (b)에서 볼 수 있듯이 H, S 성분은 평균값이 가장자리에 치우쳐 있거나 넓게 퍼져있음을 알 수 있다. 하지만 살색에 해당하는 값들이 작은 범위 안에 집중되어 있을수록, 즉 퍼짐의 정도가 작아 분산치가 작을수록, 살색영역을 추출하는데는 유리하다. 또한 통계적인 칼라 모델인 Gaussian model은 각 성분의 살색영역 평균이 중심에 가까울수록 효과적이다. 이를 위해 HSV 성분을 변형시켜주어 그림 1.(c)과 (d) 와 같이 TSV 성분을 생성해주었다. TSV 성분에 대한 정의식은 다음과 같다.

$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4 & g' > 0 \\ \arctan(r'/g')/2\pi + 3/4 & g' < 0 \\ 0 & g' = 0 \end{cases}$$

$$S_m = [(r'^2 + g'^2)/5]^{1/2}$$

$$V = (r + g + b)/3 \quad (1)$$

여기서 T와 S_m는 각각 H와 S를 변형한 성분이며 RGB의 각 성분을 r, g, b 라고 할 때 r'=(r-1/3)이고 g'=(g-1/3)이다.

2. 적응적 칼라 모델 생성

RGB 성분에서 TSV 성분으로 변환한 후 T와 S 성분에 대한 Gaussian model를 생성하면 다음과 같다.

$$P(v, \mu, K) = \frac{1}{2\pi(\det K)^{1/2}} \exp[-\frac{1}{2} Q^2]$$

$$Q^2 = (v - \mu)^T K^{-1} (v - \mu) \quad (2)$$

여기서 $v = [T(i,j), S(i,j)]^T$ 이고 평균 $\mu = [m_T, m_S]^T$ 이며 K는 T와 S의 covariance matrix이다. 그림 3은 50명의 피부색 영상을 가지고 T와 S 성분을 식 (2)의 Gaussian model에 적용한 것이며 T 평균, $m_T = 0.4233$. 이고 S 평균, $m_S = 0.1908$ 이다.

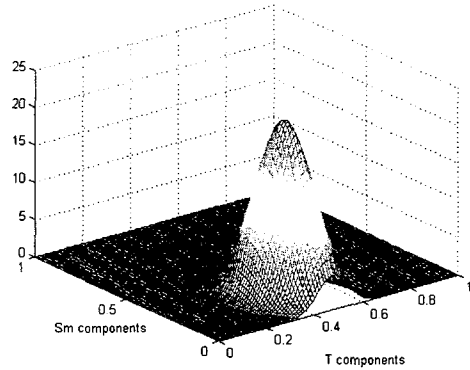


그림 2. T,S 성분에 대한 Gaussian model

기존의 방법들은 대부분 Mahalanobis metric이라고 불리는 식(2)의 Q²를 가지고 임계값을 정해주어 살색영역을 추출한다. 그러나 입력영상의 얼굴색은 보통 피부색보다 더 붉거나 어두울 수도 있고 밝을 수도 있다. 이러한 입력영상마다의 얼굴 색 차이를 보정해주는 작업이 필요하고 이를 위해 제안된 방법에서는 이미지조절(Image adjustment)과 Mahalanobis metric Q²에 두 개의 임계값을 사용하게된다. 우선, 이미지 조절 방법에서는 표 1에서 주어진 살색영역에서의 TSV 성분에 대한 임계값을 가지고 그림 3과 같이 비 살색영역에 대해 하한값이하의 값들은 0으로 매핑시키고 상한값이상의 값들은 1로 매핑시켜준다.

	T (색도)	S (채도)	V (명도)
상한값	0.77	0.48	1.0
하한값	0.34	0.0	0.2

표 1. 살색영역이 가지는 값의 범위

이렇게 함으로써 살색영역과 비 살색영역과의 대조(contrast)가 확실해져 살색영역 추출이 용이해진다. 그리고 실제로 살색영역 추출작업은 각 pixel 의 Q 값을 계산하고 이 값이 어떤 임계값 Q_{th}이상이면 그 pixel 값을 0으로, 임계값이하이면 1로 만들어주어

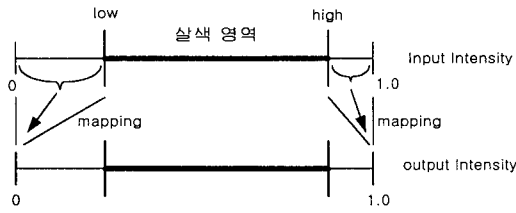


그림 3. 이미지 조절 (Image adjustment)

Binary image를 만들어 주게 된다. 그러면 임계값이 하의 pixel들이 살색영역으로 추출된다.

하지만 다양한 환경 변화에 대해 좋은 성능을 나타내기 위해서는 임계값을 적응적으로 변화시켜주는 것이 필요하며 이를 위해 식 (3) 와 같이 T성분의 평균값과 임계값 Q_{th} 를 변화시켜주었다. 표 1에서 볼 수 있듯이 다른 성분에 비해 T 성분이 분포 범위가 적어서 살색영역 추출에 가장 큰 영향을 미치므로 T 성분에 대해서만 평균값 변화를 시켜준다.

$$\begin{cases} m_{T_N} = m_{T_o} - \frac{m_{T_o} - m'_{T}}{a} & 0 < a < 1 \\ Q_{TH_N} = Q_{TH_o} + 0.07 & m_{T_o} - m'_{T} > 0 \end{cases}$$

$$\begin{cases} m_{T_N} = m_{T_o} + \frac{m'_{T} - m_{T_o}}{a} & 0 < a < 1 \\ Q_{TH_N} = Q_{TH_o} - 0.07 & m_{T_o} - m'_{T} < 0 \end{cases} \quad (3)$$

여기서 m_{T_N} 는 입력영상의 T 성분의 새로운 평균값이고 m_{T_o} 는 원래의 T성분 평균값이며 m'_{T} 는 입력영상의 T성분의 살색 범위값들의 평균이다. 그리고 Q_{TH_N} 는 Mahalanobis metric의 새로운 임계값이고 Q_{TH_o} 는 원래 임계값이며 a 는 임의의 상수이다.

이러한 방법으로 획득된 Binary image에 대해 median filtering를 하여 불필요한 에지(edge)를 제거해주고 clustering해주게된다. 이때 살색영역 중에 아주 작은 부분은 얼굴 후보 영역에서 제거하기 위해 전체 영상크기의 0.5%이하의 영역은 제거해준다. 제안된 방법에 의한 실험 예는 그림 4와 같다.

III. 신경망을 이용한 얼굴 영역 검출

얼굴 영역을 검출하기 위해서는 얼굴 후보 영역들에 대해 각 영역들이 얼굴인지 아닌지를 판단해주는 작업이 필요하다. 본 논문에서는 각 후보 영역의 invariant moments를 계산하여 모양에 대한 특징을 추출한다. 그리고 이 모멘트를 신경망의 입력으로 사용하여 최종



(a) 입력 영상



(b) 기존의 방법



(c) 제안된 방법

그림 4. 제안된 방법에 의한 살색영역 검출

결과를 도출하게 된다.

1. Fourier-Mellin moments(FMM's)

Invariant moments 방법은 Hu[1]에 의해 1961년에 처음으로 소개되었고 이후, 이미지 분석에 많이 사용되었다.[2] 이러한 모멘트에는 Zernike moments(ZM's), Legendre moments[3],[4] 등이 있으며 최근에 ZM's 보다 우수한 성능을 가진 FMM's이 개발되었다.[5] 본 논문에서는 식 (4)의 FMM's에 의해 얻어진 낮은 차수의 11개 모멘트를 후보영역의 특징벡터로 사용한다.[6]

$$M_{s,m} = \int_0^{2\pi} \int_0^{\infty} r^s f(r,\theta) \exp(-jm\theta) r dr d\theta \quad (4)$$

여기서 영상은 극좌표계로 나타내어지며 $f(r, \theta)$ 는 후보영역이고 $m=0, \pm 1, \pm 2, \dots$ 이며 s 는 일반적으로 복소수이다.

2. 신경망(Neural Networks)

본 논문의 신경망은 11개의 모멘트를 입력으로 사용하기 때문에 11개 노드(node)를 가진 입력 layer와 6개 뉴런을 가진 하나의 hidden layer와 출력 layer로 구성되어있다. 모든 뉴런의 값은 0에서 1까지의 값을 가진 log-sigmoid activation function에 의해 얻어진다. 따라서 후보영역중에 얼굴로 판단된 영역에 대해서는 출력값이 1이고 반대경우는 0이다. 신경망의 훈련시간을 단축하고 local minima에 빠질 확률을 줄이기 위해 adaptive learning rate와 momentum 기법을 사용해주

었으며 momentum 상수는 0.96, 초기 learning rate은 0.02로 설정해주었다.

VI. 실험 및 결과

얼굴 영역을 검출하는 과정은 그림 5와 같다. 여기서 검출된 얼굴영역에 대해 영역의 중심을 구하여 중심에서의 가장 먼 점까지의 거리를 계산하고 이 거리를 기준으로 영역을 사각형으로 표시해주었다.

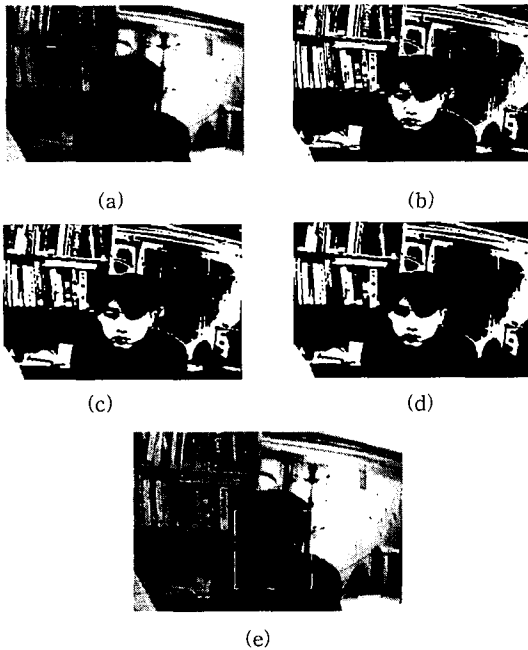


그림 5. (a) 입력영상 (b) 적응적 칼라 모델에 의한 이진영상 (c) median filtering된 영상 (d) 후보영역 검출 (clustering) (e) 얼굴영역 검출

V. 결론

본 논문에서는 TSV 공간에서 적응적 칼라 모델을 사용하여 조도의 영향을 최소화하였다. 또한 Invariant moments를 사용하여 얼굴모양의 변화나 복잡한 배경에 대해서도 비교적 좋은 결과를 얻었으며 신경망을 통한 자동적인 얼굴영역 검출이 가능해졌다.

향후 연구과제로는 그림 6에서 볼 수 있듯이 검출된 영역이 목등의 부위를 포함하여 커지게 되는데 이러한 검출 영역의 부정확도를 줄이고 계산량이 많아 느린 검출 속도를 개선하고자한다.

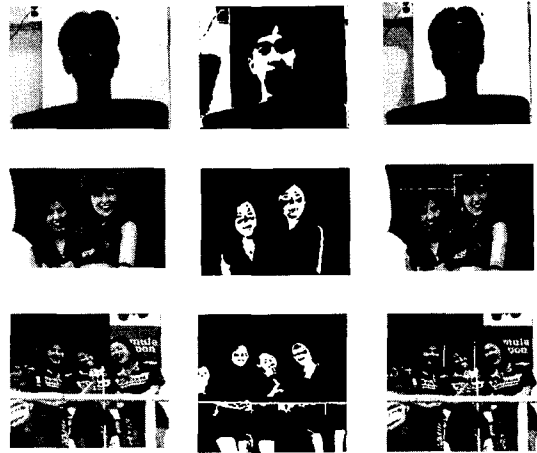


그림 6. 실험 결과 예

(본 연구는 한국과학재단(과제번호: 96-0602-01-01-3)의 지원으로 수행되었습니다.)

참고문헌

- [1] M. K. Hu, "Visual pattern recognition by moment invariants", *IEEE Trans. Inf. Theory*, Vol. IT-8, pp. 179-187, 1962.
- [2] A. Sluzek, "Using moment invariants to recognize and locate partially occluded 2D objects" *Pattern Recognition Letters* 7, pp. 253-257, April 1988.
- [3] C. H. Teh and R. T. Chin, "On Image Analysis by the Methods of Moments", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 10 No. 4, pp. 496-512, July 1988.
- [4] J. F. Boyce and W. J. Hossack, "Moment invariants for pattern recognition", *Pattern Recognition Letters* 1, pp. 451-456, July 1983.
- [5] Y. Sheng and L. Shen, "Orthogonal Fourier-Mellin moments for invariant pattern recognition", *J. Opt. Soc. Am. A*, Vol. 11, No. 6, June 1994.
- [6] J. C. Terrillon, M. David and S. Akamatsu, "Detection of Human Faces in Complex Scene Images by Use of a Skin Color Model and of Invariant Fourier-Mellin Moments", *Proc. of the 14th ICPR.*, Vol. 2, pp. 1350-1355, 1998.