

## 칼라 참조 맵과 움직임 정보를 이용한 얼굴영역 추출

이 병 석, 이 동 규, 이 두 수  
한양대학교 전자통신전공학과 신호처리 연구실  
전화 : 02-2290-0358 / 핸드폰 : 019-480-2943

### Facial region Extraction using Skin-color reference map and Motion Information

Byung-Seok Lee, Dong-Gyu Lee, Doo-Soo Lee  
Dept. of Electronics Engineering, Hanyang University  
E-mail : spee@ihanyang.ac.kr

#### Abstract

This paper presents a highly fast and accurate facial region extraction method by using the skin-color-reference map and motion information. First, we construct the robust skin-color-reference map and eliminate the background in image by this map. Additionally, we use the motion information for accurate and fast detection of facial region in image sequences. Then we further apply region growing in the remaining areas with the aid of proposed criteria. The simulation results show the improvement in execution time and accurate detection.

#### I. 서론

우리는 현재 멀티미디어의 시대에 살고 있다. 멀티미디어에 있어서 가장 중요한 정보는 영상과 음성이며, 이러한 영상과 음성정보는 그 특성상 엄청난 양의 데이터를 가지고 있다. 특히, 영상에서 얼굴과 같이 우리가 관심을 가지는 영역을 추출해내는 과정은 여러 응용분야에서 중요한 연구과제로 부상하고 있다.

얼굴인식의 전처리 단계로서 얼굴영역의 위치를 빠르게 검출하면 인식속도를 높일 수 있으며 H.263과 같은

저 전송율 비디오 코딩 분야에서는 중요한 데이터가 되는 얼굴 영역을 추출해서 이 영역을 다른 영역보다 충실히 부호화를 수행하면 화질의 열화를 보완할 수도 있다.[1][2][7]

본 논문에서는 컴퓨터와의 인터페이스나 얼굴인식, 화상전화, 화상회의 등의 시스템에서 전처리 단계로서 중요한 데이터가 되는 얼굴영역을 정확하고 고속으로 추출할 수 있는 알고리즘을 제안한다.

화자의 얼굴을 추출하기 위한 기존의 방법으로는 응용분야에 따라 신경 망(neural network), 허프(Hough) 변환, 웨이블릿(wavelet) 변환, HMM(Hidden Markov Model)등을 이용하는 방법이 활발히 연구되고 있다.[4][5][8]

하지만, 위의 방법들은 정확한 얼굴영역을 추출할 수 있는 반면에 실시간 시스템과 동영상에 적용하기에는 수행시간이 너무 오래 걸리는 단점을 가지고 있다. 따라서, 본 논문에서는 정확성과 함께 신속성을 함께 고려하고자 칼라와 움직임 정보를 이용한다. 인간의 살색 정보를 이용하여 환경에 강인한 살색 참조 맵을 구성하여 신속히 배경에서 살색을 포함하는 영역을 추출하고, 각 프레임간의 차 정보를 이용하여 좀 더 정확한 얼굴 영역을 추출하게 된다.

본 논문의 구성은 다음과 같다. II장에서는 제안된 얼굴 영역 추출 방법에 관하여 각 단계별로 고찰하고 III장에서는 모의실험에 대한 결과를 보이며 IV장에서는 앞으로의 개선점으로 결론을 맺는다.

## II. 얼굴 영역 추출 알고리즘

제안된 알고리즘의 전체 순서 도를 다음의 그림 I에 나타내었다.

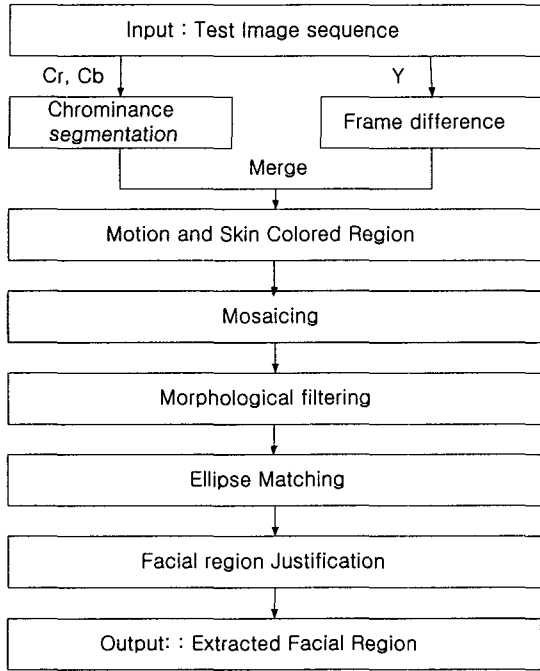


그림1. 전체 얼굴 영역 추출 알고리즘

입력영상이 들어오면 각 단계에서 제안된 알고리즘의 연산이 수행되며 최종적으로 얼굴 영역이 추출된 결과 영상을 얻게된다. 각 단계에서의 결과 영상은 순차적으로 다음 단의 입력으로 사용되게 된다.

### II.1 색상 정보를 이용

일반적으로, 칼라 공간에서의 사람의 살색 분포는 자연계의 다른 물체와 구분되는 뚜렷한 특징을 가지고 있다. 비록 사람과 인종마다 피부색은 다르지만 색차 평면상(chrominance plane)에서의 그 분포는 매우 비슷하면서도 협소한 영역을 차지하고 있고, 단지 밝기(intensity)의 차이만이 있어, 색차 성분을 이용하면 쉽게 살색 영역을 추출할 수가 있다.[1]

RGB 데이터의 영상이 입력되면 식1과 같이 YCrCb 포맷으로 변환한 후, 피부색에 해당하는  $R_{Cr}$ ,  $R_{Cb}$  값의 범위 내( $R_{Cr}=[133\ 173]$ ,  $R_{Cb}=[77\ 127]$ )의 픽셀만 추출하게 된다. 즉, 식 2와 같이 픽셀의 Cr과 Cb 값이

동시에 피부색에 해당하는 값에 위치하면 그 픽셀의 값을 1로 세팅하여 주고 그 이외에 해당하는 픽셀의 위치에서는 피부색이 아닌 영역으로 판정하여 0의 값으로 세팅을 하게된다.

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

$$O_i(x, y) = \begin{cases} 1, & \text{if } [Cr(x, y) \in R_{Cr}] \cap [Cb(x, y) \in R_{Cb}] \\ 0, & \text{otherwise} \end{cases} \quad \text{식(2)}$$

### II.2 움직임 정보를 이용

칼라 정보만을 이용하여 얼굴영역을 검출할 경우에는 살색 계통의 물체가 배경영역에 존재하여 함께 검출될 수가 있다. 따라서 보다 정확한 검출을 위하여 부가적인 정보를 이용할 필요가 있으며 본 논문에서는 카메라가 고정되어 있는 경우, 배경물체가 움직이지 않고 고정 되어있다는 사실에 착안하여 움직임 정보를 이용하였다.

다음의 식3과 같이 이전 프레임(k-1)의 영상과 현재 프레임(k)의 영상의 차 영상을 구하면 배경영역으로부터 사람의 움직임이 있는 대략적인 영역을 구할 수 있다.[7]

$$FD_{k, k-1}(x, y) = s(x, y, k) - s(x, y, k-1) \quad \text{식(3)}$$

즉, 휘도 신호만의 영상에서 현재의 프레임과 이전의 프레임의 동일한 위치에 있는 픽셀간의 차이를 구하면 움직임이 있는 픽셀에서만 값이 존재하게되어 복잡한 배경영역으로부터 간단하고 빠르게 움직임이 있는 영역만을 추출할 수가 있다. 이처럼 피부색 영역과 움직임 영역을 동시에 만족하는 영역을 얼굴의 1차 후보 영역으로 판정한다.

다음의 그림2에는 각 단계별 얼굴영역 추출 알고리즘의 결과 영상을 나타내었다.

(a)와 (b)는 각각 첫 번째 프레임과 두 번째 프레임을 나타내며, (c)는 두 프레임에서 휘도 성분만을 분리하여 계산된 차 영상이다. (d)는 움직임과 칼라정보를 이용하여 대략적인 얼굴 후보영역을 추출해서 얻어진 이진화된 결과 영상이며 (e)는 수행속도의 향상을 위하여 모자이크 영상을 만든 후 II.3절에서 설명하는 형태론적 필터링을 수행한 영상이다. (f)는 II.4절에서 설명할 템플릿 매칭방법을 이용하여 최종 얼굴 후보 영역을 추출한 결과 영상이다.

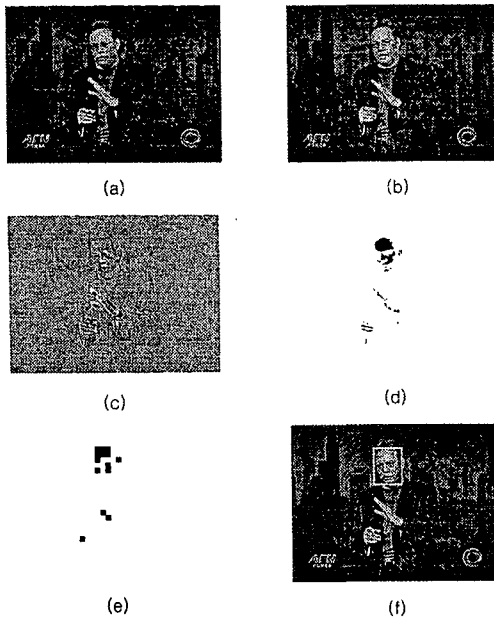


그림2. 각 단계별 얼굴영역 추출 알고리즘

### II.3 형태론적 필터링

검출된 1차 후보영역에 나타나는 잡음을 제거하기 위하여 형태론적(Morphological) 필터인 opening 연산과 closing 연산을 이진화된 마스크 영상 A에 적용한 후 미디언 필터링을 수행한다. Opening과 Closing은 각각 형태론적 침식연산과 팽창 연산을 이용하여 다음과 같이 정의된다.[6]

$$\text{Opening operator} : A \cdot B = (A \ominus B) \oplus B$$

$$\text{Closing operator} : A \cdot B = (A \oplus B) \ominus B$$

### II.4 타원형의 매칭(Ellipse Matching)

얼굴의 모양정보(shape constraints)는 칼라정보와 같이 인간마다 고유하면서도 일정한 특성을 가진다. 그 모양은 타원의 형태를 가지며, 본 논문에서는 그림4와 같이 타원형의 템플릿 대신 간단하면서도 수행속도가 빠른 일정한 가로대 세로 비[1, 1.4]를 가지는 직사각형의 템플릿을 사용하여 최종 얼굴후보영역을 판정하였다.[3][5]

템플릿의 크기는  $(M+2) \times (N+1)$ 의 픽셀 크기를 가지고 있으며, 실제 얼굴영역의 크기는  $M \times N$ 의 크기를 가진다.

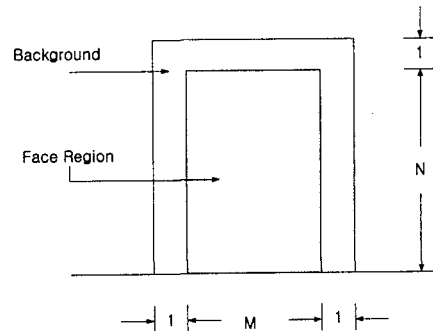


그림3. 사각형 템플릿

템플릿의 한 부분으로 배경영역을 고려한 이유는 얼굴영역에 인접한 픽셀의 칼라는 일반적으로 살색과 확연히 다른 칼라특성을 가지므로 바깥부분, 즉 배경영역의 직사각형에는 극소수의 픽셀만이 존재하여야만 한다.

템플릿이 매칭하기 위한 조건은 간단하다. 그림4와 같이 영상 위를 템플릿이 이동하면서 내부영역과 외부영역에 속하는 픽셀의 개수를 계산하여 그 수가 내부영역에는 많고, 외부영역에는 상대적으로 적게 존재하여야 한다. 그림4에서 (a)의 경우에는 얼굴영역에 많은 픽셀이 존재하고 배경영역에 적은 픽셀이 존재하는 매칭이 된 경우이며 (b)는 매칭이 안 된 경우이다.

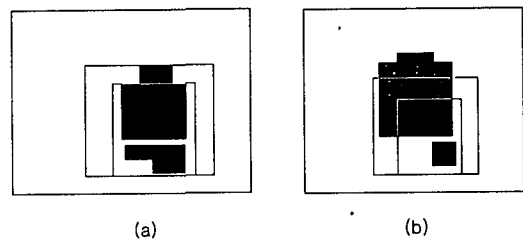


그림4. (a) a match (b) no match

## III. 실험결과

본 절에서는 제안한 알고리즘의 성능을 알아보기 위하여 얼굴이 포함된 다양한 종류의 칼라 영상을 가지고 시뮬레이션을 수행한 결과를 보인다.

실험 데이터로는 320x240 포맷의 움직임이 크며 얼굴과 손을 포함하는 전신 영상과 움직임이 적은 상반신 영상으로 구분하여 각기 다른 조명환경과 배경환경을 가지도록 하였다.

그림5에는 다양한 실험 데이터에 대하여 제안된 알

고리즘을 이용하여 얼굴영역을 추출한 결과 영상들을 보이고 있다.

(a)의 영상은 상반신의 영상으로 비교적 단순한 배경을 가지는 경우이며, (b)는 비교적 복잡한 배경을 가지는 경우이다. (c)와 (d)는 전신영상으로 (d)의 경우는 2개의 얼굴이 추출된 결과 영상을 보여주고 있다.

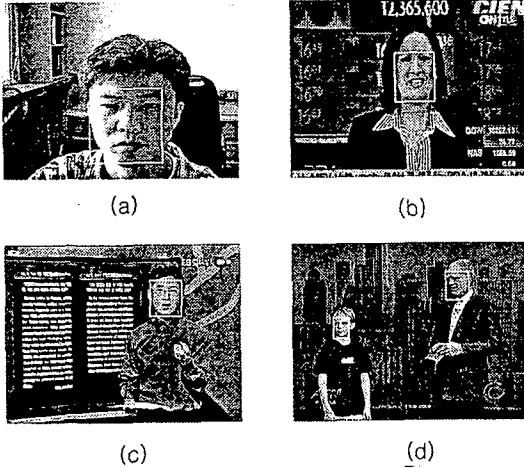


그림5. 최종 실험 결과 영상들

실험결과 에서 보듯이 조명의 영향에는 큰 차이를 보이지 않았지만 배경과 움직임의 정도에 따라 얼굴 영역 추출 결과에 차이를 보였다. 단순한 배경과 움직임이 거의 없는 경우에는 90%이상의 성공률을 보인 반면, 배경이 복잡해지고 움직임이 많아질수록 성공률은 약간 떨어짐을 알 수 있었다.

표1에는 실험에 사용된 영상과 얼굴 영역 추출 결과를 수치로 나타내었다.

표1. 실험 결과

영상의 종류		영상의 개수	성공률(%)	실패율(%)
전신 영상	단순한 배경	20	80	20
	복잡한 배경	20	75	25
상반신 영상	단순한 배경	20	95	5
	복잡한 배경	20	90	10

#### IV. 결론

본 논문에서는 영상의 칼라 성분과 움직임 정보를 이

용하여 동영상에서 관심영역인 얼굴영역을 빠르고 정확하게 검출할 수 있는 알고리즘을 제안하였다.

기존의 칼라정보만을 가지고 얼굴 영역을 검출하는 알고리즘들은 배경영역에 손과 같은 살색과 비슷한 물체의 움직임이 존재하는 경우 정확한 검출이 어려우므로 이를 보완하기 위하여 움직임 정보를 융합하여 검출을 수행하였다.

제안된 알고리즘이 타 알고리즘에 비하여 조명환경에 강인하며 빠르고 정확한 검출을 수행함을 실험 결과를 통하여 알 수 있었다. 하지만 움직임이 큰 영상이나, 살색성분을 가지면서 얼굴의 크기와 유사한 물체가 움직임을 가질 경우에는 정확한 검출이 어려웠다.

#### 참고 문헌

- [1] D Chi, King N. Ngan "Face Segmentation Using Skin-Color Reference map in Videophone Applications", IEEE Transactions on Circuits and Systems for Video Technology, vol.9 , June 1999
- [2] Bernd Menser, Michael Brunig "Face Detection and Tracking for Video Applications" IEEE Trans 2000
- [3] H. Wang, Shih-fu Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video" IEEE Trans on Circuits and Systems for Video Technology, August 1997
- [4] Henry A. Rowley, Shmeet Baluja, Takeo Kanade, "Neural Network Based Face Detection", Pattern Analysis and Machine Intelligence on IEEE Trans, Jan 1998
- [5] Jayashree Karlekar, U. B. Desai " Finding Faces in Color Images using Wavelet Transform" Image Analysis and Processing, 1999 Proceedings international conference on, 1999
- [6] Anil K. Jain, " Fundamentals of Digital Image Processing" Prentice Hall, 1989
- [7] A. M. Tekalp, " Digital Video Processing", Prentice Hall, 1995
- [8] A. N. Rajagoplan, K.Sunil Kumar, Jayashree Karlekar, " Finding Faces in Photographs" Computer Vision, 1998. Sixth International Conference on , 1998