

# 비디오 등장인물 검색을 위한 얼굴검출

준회원 백승호\*, 김준환\*, 정회원 유지상\*

## Face Detection for Cast Searching in Video

Seung-ho Paik\*, Jun-hwan Kim\* Associate Members, Ji-sang Yoo\* Regular Members

### 요약

드라마와 같은 비디오에서 사람의 얼굴은 일반적으로 자주 등장하며 비디오 내용을 분석하기 위한 유용한 정보를 제공한다. 얼굴검출은 얼굴인식 및 얼굴영상의 DB 관리와 같은 응용분야에 중요한 역할을 한다. 본 논문에서는 비디오 등장인물 검색을 위한 얼굴검출 기법을 제안하였다. 전체 과정은 크게 세단계로 구성되며 첫 번째 장면 전환 검출단계, 두 번째 얼굴영역 검출단계, 마지막으로 얼굴의 특징점인 눈과 입 검출단계로 구성되며, 색상에 기반한 얼굴영역 검출단계에서 발생된 얼굴 특징점을 눈과 입의 검출에 적용하였다. 실험결과 다양한 환경에서 성공적으로 얼굴을 검출하며, 기존의 색상기반 얼굴검출 방법에 비해 측면영상에서 24%의 성능향상을 보였다.

**Key Words** : face detection, scene change, facial feature, video

### ABSTRACT

Human faces are commonly found in a video such as a drama and provide useful information for video content analysis. Therefore, face detection plays an important role in applications such as face recognition, and face image database management. In this paper, we propose a face detection algorithm based on pre-processing of scene change detection for indexing and cast searching in video. The proposed algorithm consists of three stages: scene change detection stage, face region detection stage, and eyes and mouth detection stage. Experimental results show that the proposed algorithm can detect faces successfully over a wide range of facial variations in scale, rotation, pose, and position, and the performance is improved by 24% with profile images comparing with conventional methods using color components.

### I. 서론

최근 초고속 통신망 및 대용량 저장매체의 발달로 언제 어디서나 자신이 원하는 정보에 접근할 수 있게 되었고 이를 통해 사용자에게 다양한 멀티미디어 정보가 제공되고 있다. 이러한 환경 속에서 일반 사용자들은 기존의 일반 TV, 혹은 케이블 TV에서와 같이 수동적인 시청방식에서 탈피하여 능동적으로 자신이 원하는 프로그램만을 시청할 수 있는 환경을 요구하고 있다. 이러한 요구를 만족하려면

먼저, 데이터베이스 안에 존재하는 멀티미디어 정보를 보다 효율적이고 정확하게 인덱싱(indexing)하는 기술이 개발되어야 한다.

기존의 인덱싱 기술은 멀티미디어 데이터를 사용자가 직접 분석하여, 적절한 내용을 텍스트 형태로 입력하는 방식이었다. 그러나 이 방식은 객관성이 결여되어 있으며, 또한 관리자에게 많은 시간과 노력을 요구한다. 이에 따라 멀티미디어 정보검색을 효율적으로 검색하기 위해, 비디오로부터 색인에 사용될 적절한 장면들을 분할하는 장면전환점 검출기

\* 광운대학교 전자공학과 디지털 미디어 연구실 ([bsh1602, kjunan81]@image.gwu.ac.kr, jsyoo@daisy.kw.ac.kr),

논문번호 : KICS2005-05-217, 접수일자 : 2005년 5월 27일

※본 연구는 2005 교내학술연구 지원과 정보통신부의 출연금으로 수행한 IT-SoC 핵심설계인력 양성사업의 결과로 수행되었음.

법과, 각 장면으로부터 의미 있는 객체추출에 대한 기법들이 연구되고 있다. 드라마, 영화와 같은 동영상은 그 특성상 인물이 중심이 되며, 인물에서 중요한 특징인 얼굴을 검출하고 인식하는 기술은 일반 사용자를 대상으로 한 PVR(personal video recorder), 모바일 폰과 멀티미디어 공급자를 위한 비디오 관리 시스템에 이용될 수 있다.

기존에 연구되었던 얼굴 검출기법으로는 주성분 분석(PCA), 신경망(neural networks), Support Vector Machines(SVM)과 같은 훈련을 통한 검출방법과 색상기반, Template Matching, 움직임 검출(motion extraction)과 같은 비훈련 검출방법이 있다<sup>[1-6]</sup>. 훈련을 통한 얼굴검출은 얼굴과 얼굴이 아닌 영상들의 훈련을 통해 검출하는 방법이므로 많은 영상표본이 필요하며, 이미 얼굴이 아닌 영상들이 정의되어야 하는 단점이 있다. Template Matching은 영상의 이동, 확대, 축소, 회전, 조명의 변화에 민감하다는 단점이 있으며, 움직임 검출은 장면 안에 움직이는 물체가 다수 존재할 경우 이를 구분해야 하며 카메라 자체의 움직임으로 인한 모션을 감지해야 하는 어려움이 따른다.

본 논문은 동영상에서 등장인물의 얼굴을 검출하기 위해 색상 정보 기반의 방법을 사용한다. 색상정보를 이용한 얼굴검출 방법은 얼굴의 특징을 빠르고 쉽게 얻을 수 있고, 다른 방법에 비해 계산량을 줄일 수 있어 실시간으로 구현될 수 있는 장점을 가지고 있다. 그러나 색상정보를 이용한 방법은 조명과 같은 외부 환경에 민감하여 영향을 받으며 얼굴과 유사한 색상의 분포를 모두 검출해 낸다는 단점을 가지고 있다. 따라서 본 연구에서는 실시간의 전제 조건을 만족하면서 강인성을 좀 더 보장할 수 있는 기법에 연구의 목표를 두었다.

본 논문의 2장에서는 장면전환 검출을 설명한다. 3장에서는 얼굴검출을 설명한다. 4장에서는 시스템 구현 및 실험 결과를 분석한다. 마지막 5장에서 본 논문의 문제점과 앞으로의 연구방향을 언급하며 결론을 맺는다.

## II. 장면전환 검출(Scene Change Detection)

장면전환 검출 기법의 전체 순서도는 그림 1과 같다. 동영상 데이터를 보다 효율적으로 관리하기 위해 검색이나 특징을 추출하는 기본적인 단위로 영상을 분할하여 구조화할 필요가 있다. 동영상의

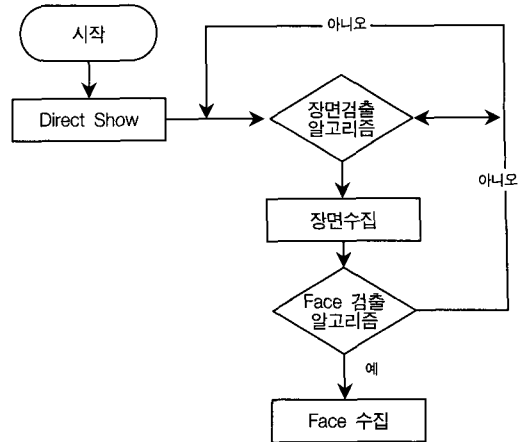


그림 1. 장면전환 검출 기법의 전체 순서도

구조는 크게 4개의 계층으로 구성되는데 최하위는 개개의 정지영상을 가리키는 프레임(frame)이고 다음은 샷(shot)으로, 한 카메라의 연속적인 촬영에 의한 프레임의 모임을 말한다. 그리고 내용상 관련이 있는 인접한 샷의 집합을 장면(scene)이라 하고, 최상위 계층으로 연속적인 장면들의 집합인 시퀀스(sequence)로 구성이 된다.

장면전환은 컷(cut)과 같은 급진적인 장면변화와 디졸브(dissolve), 페이드(fade) 등의 점진적인 장면변화로 구분된다. 가장 대표적인 급진적인 장면전환 추출에는 히스토그램간의 차이를 이용한 방법과 화소차를 이용한 방법이 있다. 화소단위 비교방법은 동일한 샷 내에서는 화소값의 변화가 적다는 성질을 이용하는 것으로, 연속하는 한쌍의 프레임에서 대응하는 화소값을 비교하여 얼마나 많은 변화가 발생하였는가를 측정한다. 화소단위 비교방법은 화소단위로 하기 때문에 움직임이 많은 영상이면 오검출이 많아지는 단점이 있다.

히스토그램 비교방법은 동일한 샷으로 분류되는 프레임들의 색상분포는 유사하다는 성질을 이용하여 컷을 검출한다. 이 방법은 각 프레임의 컬러 히스토그램 차이를 계산한 후, 식 (1)과 같이 그 차이 값이 임계값을 넘으면 컷으로 판단한다.

$$\sum_{j=1}^N |H_i(j) - H_{i+1}(j)| > T \quad (1)$$

여기서  $H_i(j)$ 는  $i$  번째 프레임에서  $j$  레벨의 히스토그램 값을 나타내고,  $N$ 은 히스토그램의 레벨 수를 나타낸다. 인접한 프레임들의 차이를 정해놓은 임계값  $T$ 와 비교함으로써 컷을 검출한다.

히스토그램 비교 방법은 구현이 쉽고 연산이 적

어 효율적이며 움직임이 있는 장면에서도 전체적인 명암도는 크게 변하지 않으므로 물체의 움직임이나 카메라 이동에 덜 민감하다. 그러나 플래시 효과처럼 밝기가 순간적으로 변화하는 경우에는 대처하기 곤란하고, 다른 샷이지만 색상분포가 유사할 때에는 오검출이 발생하는 단점이 있다<sup>[6]</sup>. 드라마와 같은 동영상은 프레임간의 유사성이 높으므로 원활한 얼굴검출을 위해 움직임과 카메라 이동에 덜 민감한 히스토그램간의 차이를 이용하여 장면전환 검출을 수행한다.

입력된 동영상에서 일정시간 마다 프레임을 획득하고 다양한 크기의 동영상을 일정 크기로 변환을 해준다. 그림 2와 같이 크기가 변화된 영상의 RGB 성분에 대한 각각의 히스토그램을 구하고 이전 프레임과 현재 프레임의 차이 값을 계산하여 임계값보다 클 경우 장면전환이라 한다<sup>[7]</sup>. 임계값을 넘어서 장면전환으로 판단된 프레임은 영상의 재생 순서대로 저장되며, 임계값을 넘지 못한 프레임은 저장하는 순서를 거치지 않는다.

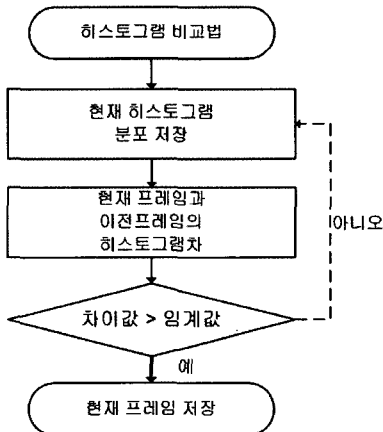


그림 2. 장면전환 검출 순서도

### III. 얼굴 검출

얼굴 검출과정은 크게 얼굴의 색상값을 이용한 얼굴 후보영역 검출과 얼굴 특징점의 색상특징을 이용한 얼굴 특징점 검출로 이루어진다.

#### 3.1 얼굴 후보영역 검출(Face Localization Detection)

동영상에서 획득된 입력영상은 일정 크기의 RGB 컬러영상이다. 먼저 획득된 영상에서 색상 정보만을 이용하여, 조명에 대한 영향을 줄이기 위하여 주어

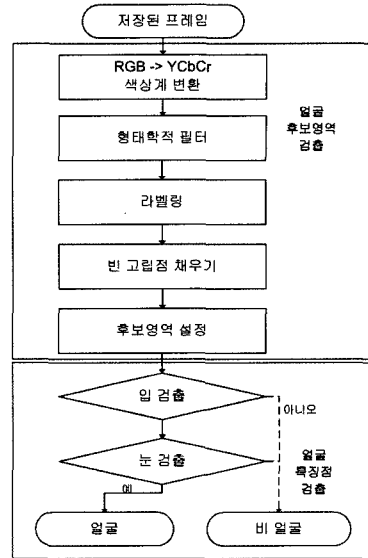


그림 3. 얼굴검출 순서도

진 영상을 RGB 색상계로부터 YCbCr 색상계로 변환한다<sup>[9]</sup>. 얼굴검출의 전체적인 과정은 그림 3과 같다.

입력 영상의 RGB 색상계에서 YCbCr 색상계로의 변환식은 식 (2)과 같으며 YCbCr 정보 중에서 Cb, Cr 성분에 대해 색상 영역분할(color segmentation) 과정을 거쳐 살색 값을 갖는 영역만을 추출해 내는데 이때 Cb, Cr 각각의 선택범위는 식 (3)와 같다.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

$$f(x, y) = \begin{cases} 255 & \text{if } (100 < Cb < 127) \cap (129 < Cr < 158) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

그림 4의 주어진 입력 영상 (a)에 대해 색상 영역분할 과정을 거친 결과가 그림 4(b) 영상이다. 그림 4(b) 영상을 보면 세밀한 잡음 형태가 존재하는 것을 볼 수 있는데 이런 잡음의 형태를 제거하여 얼굴 영역만을 분리해 내기 위하여 형태학적 필터(morphological filter) 중에서 식 (4)에 정의된 열림(opening) 연산을 수행한다<sup>[10]</sup>.

$$\text{열림} : r_n(f) = \delta_n(\epsilon_n(f)) \quad (4)$$

여기서  $\epsilon_n$ 과  $\delta_n$ 은 각각 녹임(erosion)과 불림(dilation) 연산자 이며 식 (5)와 (6)과 같이 각각 정의된다.

녹임 :  $\varepsilon_n(f)(x) = \text{Min}\{f(x+k), k \in M_n\}$  (5)

블림 :  $\delta_n(f)(x) = \text{Max}\{f(x-k), k \in M_n\}$  (6)

여기서  $f(x)$ 는 입력영상이고  $x$ 는 2차원 좌표점을 의미하며  $M_n$ 은 크기가  $n$ 인 이치형태소 (flat structuring element)를 나타낸다.

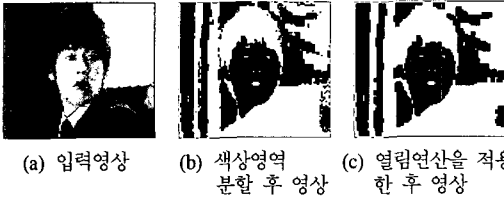


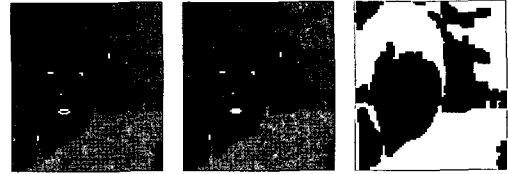
그림 4. 살색 영역의 분리

열림 연산을 수행 후 잡음이 많이 제거된 영상에는 얼굴영역 외에도 몇 개의 다른 영역이 포함될 수 있다. 예를 들어, 영상에 손이 있을 경우 이 손이 하나의 독립된 영역으로 구별될 것이다. 또 배경에는 원목가구와 황토색 벽지, 커튼과 같이 피부와 비슷한 색의 큰 물체가 있을 때 하나의 영역을 이룰 것이다. 이런 경우 각 영역을 구별해 내기 위한 작업이 필요하다. 이를 위해 후보영역에 라벨링 (labeling)을 수행한다<sup>[11]</sup>. 라벨링 과정에서는 형태학적 필터를 통해서 제거하지 못한 잡음을 제거하기 위해 라벨링된 면적이 일정크기보다 작을 경우는 얼굴 후보영역으로 검출하지 않도록 한다.

잡음은 아니지만 일정 크기보다 더 작을 경우 눈과 입에 대한 특징들이 잘 나타나지 않으므로 라벨링된 면적은 전체 영상크기의 최소 1/20 이상이 되도록 한다. 라벨링된 후보영역은 실제 얼굴영역일 경우 눈과 입의 Cb, Cr 값은 얼굴의 Cb, Cr 값과 상당히 다르므로 이진화 과정에서 그림 4(b)와 (c) 같이 눈과 입의 부분이 남게 된다. 본 논문에서는 이러한 부분을 고립점이라 부르며 눈과 입에 대한 고립점을 검출하여 얼굴 특징점을 검출하는데 검토 요소로 활용한다. 고립점 검출은 다음과 같다. 그림 5 (a)와 같이 라벨링된 영상에서 배경부분만 라벨링을 수행한다. 그리고 그림 5 (b)와 같이 배경부분이 라벨링된 영상에서 그림 5 (c)와 같이 고립점을 찾아낸다. 고립점이 있을 경우, 그림 5 (d)와 같이 고립점만 Conditional Dilation<sup>[12]</sup>을 통해 그 내부를 채우게 되고 최종 얼굴 후보 영역을 위해서 그림 5 (e)와 같이 고립점을 얼굴과 같은 값으로 처리하여 후보영역에서 얼굴 내부의 공간을 모두 채우게 된다.



(a) 라벨링된영상 (b)배경부분 라벨링



(c) 고립점 검출 (d) conditional dilation (e) 고립점을 채운 영상

그림 5. 고립점 검출 및 고립점 채우기

얼굴영역 설정은 라벨링된 영역이 모두 포함될 수 있도록 설정한다. 그림 6은 얼굴 후보 영역으로 결정된 영상이다.



그림 6. 얼굴 후보 영상

### 3.2 얼굴 특징점 검출(Face Feature Detection)

#### 3.2.1 얼굴 특징점 부각

다양한 얼굴 특징 가운데 눈과 입은 얼굴인식을 위해 사용되는 대표적인 특징점이다. 이러한 눈과 입 검출을 위해서 형판(template), 에지 검출, 히스토그램 분석 등 다양한 방법이 제안되어 왔다<sup>[13]</sup>. 그러나 이러한 방법들은 크기, 회전, 그림자, 조명 등에 강인하지 못한 문제를 가지고 있다. 형판은 미리 정해 놓은 얼굴특징에 관한 형판을 가지고 상관 관계를 조사하여 검출하는 방법이 있다. 그러나 많은 크기와 회전에 따른 형판을 만들었다 하더라도 비교하여 찾는 시간이 많이 소요되며, 데이터를 저장해야 할 공간도 많이 필요한 단점이 있다. 에지검출을 이용할 경우 에지가 조명에 많은 영향을 받고 잡음에 의해 손상되며, 특히 그림자가 얼굴에 있을 경우 강한 잡음 에지로 얼굴 특징을 검출하기 어려운 단점이 있다. 히스토그램 분석은 일반적으로 눈이 얼굴 영역에서 어두운 분포를 가지고 있다는 것을 이용한 것으로 이 또한 조명에 강인하지 못하다.

이러한 문제를 해결하고자 본 논문에서는 얼굴의 색차 성분을 이용하여 얼굴 특징점을 검출하는 방법을 적용하였다. 눈 영역은 색차성분 중 그림 7의 (a)와 같이 Cb성분에서 상대적으로 높은 화소 값을 갖고, Cr성분과 휘도성분인 Y성분에서 낮은 화소 값을 갖는 특성이 있고 입은 얼굴에서 상대적으로 빨간색 성분을 많이 포함하고 있으므로 그림 7의 (b)와 같이 상대적으로 Cr성분에서 높은 화소 값을 갖고 Cb성분과 Y성분에서 낮은 화소 값을 갖는다. 따라서 이러한 특징이 최대한 반영되도록 식 (7), (8)을 이용 그림 7의 (c), (d), (e), (f)와 같이 눈과 입을 부각시켰다<sup>41</sup>. 그러나 주변 배경 화소값이 눈, 입보다 더 높은 값을 가질 수 있으므로 그림 5의 (e) 고립점이 채워지고 이진화 된 값들만 부각 시키도록 하였다.

$$Eye = \frac{1}{3} \{ (Cb^2) + (\widetilde{Cr})^2 + (Cb/Cr) \} \quad (7)$$

$$Mouth = Cr^2 \cdot (Cr^2 - \eta \cdot Cr/Cb)^2, \quad \eta = 0.95 \cdot \frac{\frac{1}{n} \sum_{(x,y) \in FG} Cr(x,y)^2}{\frac{1}{n} \sum_{(x,y) \in FG} Cr(x,y)/Cb(x,y)} \quad (8)$$

여기서 Cb, Cr 값은 모두 [0, 255] 로 정규화 된 값이고  $\widetilde{Cr}$ 은  $255 - Cr$ 이며 n은 얼굴후보영역 마스크 FG의 화소개수이다. η는 Cr<sup>2</sup>의 평균값과 Cr/Cb 평균값의 비율이다.

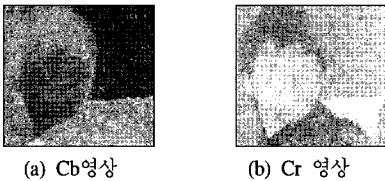


그림 7. 입력영상

그림 8은 위의 그림 7을 이용하여 실험한 결과이다.

이러한 얼굴 후보영역의 특징점 부각과정에서 실제 얼굴의 경우 그림 8와 같이 눈과 입이 부각되는 특징을 가지게 되고 실제 얼굴이 아닐 경우에도 후보영역에서 눈, 입과 비슷한 성분을 가지는 화소가 255로 정규화 되서 특징점으로 검출이 될 수 있다. 따라서 특징점 검출과정에서 고립점을 이용 검출을 더욱 강화할 수 있다.

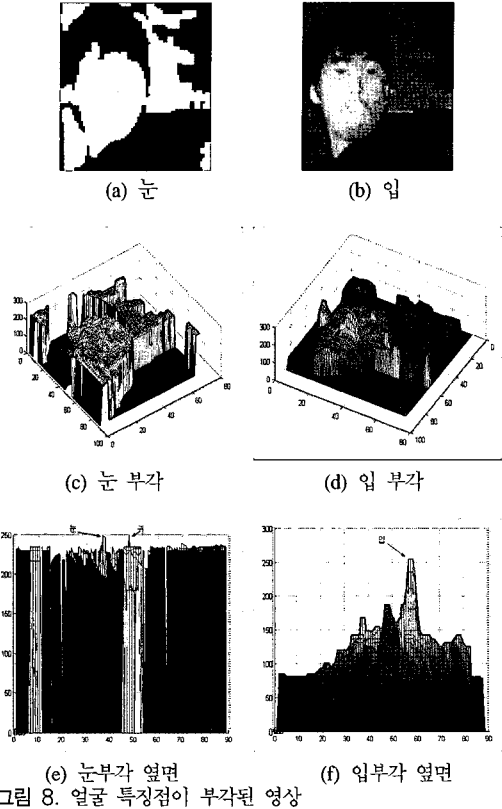


그림 8. 얼굴 특징점이 부각된 영상

### 3.2.2 입 검출

효율적인 입 검출을 위해 높은 화소 값을 많이 갖도록 불림연산을 한다. 입의 검출척도로 사용하기 위해 입의 특징이 부각된 영상에서 최고 화소값을 찾는 것이 아니라 일정집단의 누적값으로 찾기위해 additional value (AV)를 제안한다. AV는 0~255 사이의 흑백 영상과 7\*5 크기에 해당되는 빈 형판과의 누적 값이다. 형판의 크기는 얼굴영상이 다양함으로 실험적으로 최적의 크기로 계산된다. 누적값을 적용함으로써 검출결과의 신뢰도를 높인다.

$$MV(m, n) = \sum_{x=0}^{X_i-1} \sum_{y=0}^{Y_i-1} I[x+(m-3)][y+(n-2)] \quad 3 \leq m < X_i - 3, 2 \leq n < Y_i - 2 \quad (9)$$

식 (9)에서 I[x][y]는 입력영상의 얼굴영역을 X<sub>i</sub>와 Y<sub>i</sub>는 빈 형판의 수직 및 수평 크기를 나타내며 AV 값이 가장 큰 값을 갖을 때 입의 위치로 예측한다. 예측된 입의 중심점에서 3\*3 검색을 수행 후 라벨링 단계에서 발생된 고립점이 있으면 올바른 입의 위치로 인식하고 고립점이 없으면 해당 위치를 모두 '0'으로 만들어 다시 입을 부각 시키도록

한다. 그 중심점을 찾아 3\*3 검색을 하도록 한다. 이와 같은 과정을 반복해서 입을 검출하고 일정 반복 수행 후 일치하는 위치가 없으면 후보영역에서 입은 없는 것으로 인식한다.

3.3.3 눈 검출

눈 검출 역시 입과 마찬가지로 방식을 사용하지만 5\*3 크기의 bin형판을 사용한다.

$$MV(m,n) = \sum_{x=0}^{X_i-1} \sum_{y=0}^{Y_i-1} I[x+(m-2)][y+(n-1)]$$

$$2 \leq m < X_i - 2, 1 \leq n < Y_i - 1 \quad (10)$$

식 (10)에서  $I[x,y]$ 는 입력영상의 얼굴영역을  $X_i$ 와  $Y_i$ 는 bin 형판의 수직 및 수평 크기를 나타내며 AV 값이 가장 큰 값을 갖을 때 눈의 위치로 예측한다. 가장 큰 AV 값이 나타나는 지점을 눈의 위치로 예측하고 중심점에서 3\*3 검색을 통해 라벨링 단계에서 발생한 고립점을 찾는다. 단 눈의 경우, 우선 입을 검출하므로 눈을 부각 후 입의 위 부분만 검색한다. 눈의 경우는 눈이 두개 이므로 먼저 한 개의 눈을 찾고 검출된 입의 위치와 비교하여 다른 눈을 찾는다. 다른 눈을 찾을 경우 마찬가지로 입과 먼저 찾은 눈의 위치를 이용하여 검색 범위를 제한한다.

IV. 실험결과 및 분석

본 논문에서는 표 1에서와 같이 여러 드라마 동영상 사용하여 실험을 수행하였다. 장면전환점 경우는 일반적으로 4가지로 구분을 한다. 장면전환점이 컷(cut)과 같이 온전한 얼굴을 가진 영상과 페이드(fade in, fade out), 디졸브(dissolve) 같이 영상이 점진적으로 변하는 경우에는 일정시간이 지난 뒤에 얼굴이 나타난다. 이전화면은 점점 작게, 새로 나타날 화면은 점점 많이 나오는 와이프(wipe)의 경우 영상에서 얼굴이 일부 잘리는 특징이 있다. 따라서 제안된 알고리즘의 성능검증을 위해서 온전한 얼굴을 가진 장면전환 검출영상과 이를 이용한 얼굴검출에서 얼굴의 측면을 고려하여 입과 눈이 최소 2개 이상 빠르게 검출 될 경우를 정상 얼굴로 판단한다. 제안된 알고리즘의 성능검증을 위해 Visionics사의 FaceIt 와 비교 평가 하였다<sup>[4]</sup>. 검출률 측정을 위한 실험 비디오는 가을동화 6회분으로 사용하였

고 그 밖에 아일랜드 1회분, 쾌걸춘향 13회분, 봄날 13회분, 등 여러 드라마 동영상 사용하였다.

표 1. 실험에 사용한 동영상 데이터

	아일랜드	쾌걸춘향	봄날	가을동화
크기(pixel)	640*480	640*480	640*480	640*480
프레임수 (# frames)	114532	113049	103505	72069
재생시간 (mm:ss)	63:41	62:52	57:33	50:05

실험환경은 펜티엄4 1.8GHz프로세서, 512MByte 메모리 상에서 Visual C++ 6.0과 Directshow 라이브러리를 이용하여 구현하였다. 실험에서 장면전환점을 판별하는 임계값은 고정임계값을 사용하였다. 제안하는 기법을 통해 비교된 값들이 임계값을 넘었을 경우에 장면전환점으로 판정을 하였다. 임계값은 실험을 통해 얻어진 최적의 값을 사용하였다. 본 논문에서 제안한 알고리즘의 효율성을 평가하고 성능 측정을 위해, 검출률을 식 (11)과 같이 정의 하였다.

$$\text{검출률 (\%)} = \frac{F_N}{C_N} \times 100 \quad (11)$$

여기서  $F_N$ 는 얼굴검출 영상에서 실제 얼굴이 있는 영상 수를 가리키고,  $C_N$ 은 장면전환 영상에서 실제 얼굴이 있는 영상의 수를 나타낸다. 표 2에서  $F_T$ 는 얼굴검출 영상의 총 수를 가리키고  $C_T$ 는 장면전환 영상의 총 수를 나타낸다. 가을동화 6회분 및 여러 동영상에서 얼굴검출 성능은 표 2와 같다.

표 2. 실험데이터의 결과

임계값	아일랜드	쾌걸춘향	봄날	가을동화
$C_T$	364	241	206	327
$F_T$	113	86	65	131
$C_N$	112	76	59	87
$F_N$	95	66	49	75
검출률 (%)	84.8	86.8	83.0	86.2

검출된 영상들은 그림 9, 그림 10과 같은 정면영상 뿐만 아니라 그림 11과 같은 옆면, 심지어 그림 12와 같이 측면얼굴까지 검출하는 것을 볼 수 있다. 영상은 배경의 화소 값이 얼굴과 비슷해서 상당히 큰 얼굴후보 영역을 설정 뒤에도 불구하고 얼굴의 특징점을 검출하는 것을 볼 수 있다. 이는 후보 영

역 설정 시 고립점들을 모두 채워주고, 실제 후보영역 모두를 사용하는 것이 아니라 이진화 된 후보영역만 사용하므로 배경 화소 값으로부터 얼굴의 특징점들이 잘 부각 될 수 있다. 마지막으로 고립점을 이용한 얼굴 특징점의 검증단계를 통해 배경 화소 값이 일부 얼굴영역으로 설정 되더라도 배경부분이 특징점으로 검출 되지 않으며, 얼굴의 기하학적 특징을 이용해 눈과 입 사이의 거리를 측정하여 그 거리가 상대적으로 비정상적일 경우 검출에서 제외 하였다. 검출된 입의 위치를 이용 눈의 위치 영역을 제한하여 효율적인 눈의 검출 가능성을 높였다.

가을동화 6회 분에서 얼굴 검출 성능은 표 3과 같다. 가을동화에서 제안한 알고리즘의 검출률은 86.2% 로 FaceIt의 53.8% 보다 약 30% 높았다.

실험에서 1프레임에 대한 얼굴검출 시간은 표 4와 같다. 아일랜드에서 실험했을 경우 처리시간은 평균 0.609초 로 다른 알고리즘 5.78초<sup>[4]</sup>에 비해 처리시간이 짧아서 실시간으로 구현이 된다. 이와 같이 처리시간이 단축된 것은 다른 알고리즘에서는 각각의 얼굴 후보영역들을 합쳐서 하나의 큰 얼굴 후보를 생성하는 과정이 있으나 제안한 방식에서는 얼굴 후보영역 설정시 라벨링된 영역들을 모두 포함할 수 있게 하였으므로 합치는 과정에서 발생하는 시간을 단축할 수 있었다.

다양한 얼굴방향에 대한 검출 성능 비교는 표 5와 같다. 아일랜드에서 실험했을 경우 정면 정면에 가까운 영상과 반 옆면 영상에서는 다른 알고리즘<sup>[4]</sup>에 비해 비슷한 성능을 보이나 측면에 대해서는 높은 성능을 보인다. 고립점 사용의 차이가 측면 영상에서 큰 영향을 준 것으로 판단된다. 측면영상은 눈과 입이 1개씩 나오거나 후보영역 설정시 눈과 입 모두 배제된 후보영역이 등장할 수 있다. 제안한 기법에서는 얼굴 후보영역을 설정할 때 부수적으로 발생하는 고립점을 이용하고 얼굴 영역을 최대한으로 설정함으로써 측면에서 얼굴 특징점이 포함되도록 하였다.

표 3. 가을동화 6회에서 얼굴 검출 성능 비교

	제안된 기법	FaceIt[14]
검출률 (%)	86.2%	53.8%

표 4. 얼굴검출 시간

	제안된 기법	다른 알고리즘[4]
영상 개수	113	227
영상 크기	200*200	150*220
시간(초):평균	0.609±0.27	5.78±4.98

표 5. 얼굴 방향에 대한 검출 성능비교

얼굴 방향	영상의 개수		검출률(%)	
	제안된 기법	다른 알고리즘[4]	제안된 기법	다른 알고리즘[4]
정면 (Frontal)	46	66	91.30	89.40
정면에 가까운 (Near-Frontal)	36	54	88.88	90.74
반 옆면 (Half-Profile)	23	75	78.26	74.67
측면 (Profile)	7	11	42.85	18.18
총 계	112	206	84.8	80.58



(a) 원본

(b) 검출된 얼굴



(c) 원본

(d) 검출된 얼굴

그림 9. 정면 영상



(a) 원본

(b) 검출된 얼굴



(c) 원본

(d) 검출된 얼굴

그림 10. 정면에 가까운 영상

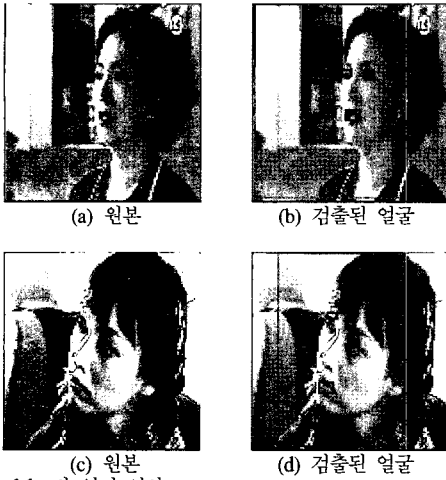


그림 11. 반 옆면 영상

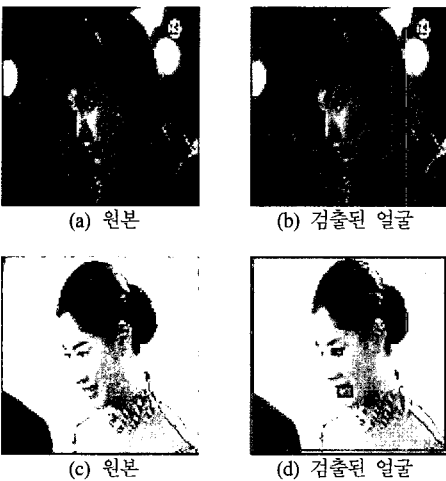


그림 12. 측면 영상

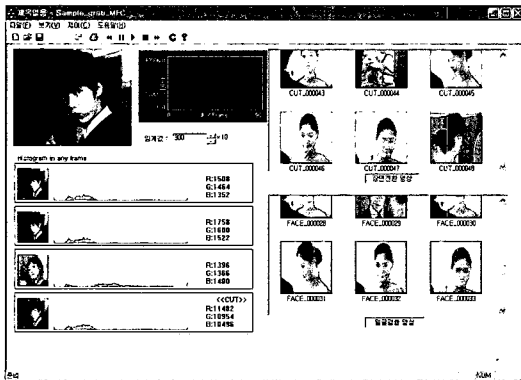


그림 13. 등장인물 검출 프로그램

### V. 결론 및 향후과제

본 논문에서는 등장인물이 중요한 특징이 되는

드라마 동영상을 입력으로 사용하여 프레임간의 히스토그램 비교를 통해 장면을 검출하고, 검출된 장면에서 얼굴을 검출하기 위한 기법을 제안하였다. 먼저 얼굴 후보영역을 검출하고 후보 영역 내에서 얼굴의 특징점인 눈과 입을 검출하기 위해 눈과 입에서 얼굴의 색차 성분의 특징을 이용하였으며 마지막으로 후보영역에서 라벨링 시 발생한 고립점을 검증요소로 활용하였다. 제안된 알고리즘은 학습을 필요로 하는 방식보다 계산량과 얼굴의 표본이 필요하지 않으므로 메모리가 많지 않은 이점이 있지만 색상기반 방식은 다양한 외부 조명에 영향을 받는다. 이와같이 색상 기반에 의한 한계를 모두 극복하는 것은 매우 어려운 일이지만 향후 여러 가지 외부환경에 더 강인하며, 실시간으로 구현 할 수 있는 새로운 기법들이 연구 개발되어야 할 것이다.

### 참고 문헌

- [1] B. Menser and F. Muller, "Face detection in color images using principal components analysis", *IEEE Conf. Image Processing and Its Applications*, vol 2. pp. 620-624, 1999.
- [2] H. A. Rowley, S. Baluja and T. Kanade, "Neural network based face detection", *IEEE Trans Pattern Analysis and Machine Intelligence*, vol 20, pp. 23-38, 1998
- [3] E. Osuna, R. Freund, and F. Girosi. "Training Support Vector Machines: an Application to Face Detection", *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp. 130-136, June 1997
- [4] Rein-Lien Hsu; Abdel-Mottaleb; M. Jain, A.K., "Face Detection in Color Images," *IEEE trans. Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 696-706, May 2002
- [5] A. Yuille, D. Cohen and P. Hallinan, "Feature extraction from faces using deformable template", *Proc. IEEE Computer Society conf. on Computer Vision and Pattern Recognition*, pp. 104-109, 1989
- [6] Ronghua Liang, Chun Chen, Jiajun Bu, "Real-time facial features tracker with motion estimation and feedback", *IEEE Systems, Man and Cybernetics Conf.*, vol. 4, pp.3744-3749, Oct, 2003,



- [7] 문영호, 노정진, 유지상, “대역별 웨이블릿 계수 특성을 이용한 장면전환점 검출기법”, *한국통신학회논문지* 제29권 제10호 pp.1414-1424, 2004
- [8] D. Swanberg, C.F. Shu, R. Jain, “Knowledge Guided Parsing In Video Database”, *In Proc. Of SPIE Symposium On Electronic Imaging : Science And Technology*, pp.12-24, San Jose. CA., 1993.
- [9] Chun-Hung Lin, Ja-Ling Wu, “Automatic Facial Feature Extraction by Generic Algorithms”, *IEEE Trans. Image Processing*, Vol,8. No.6, pp.834-845, June, 1999.
- [10] J. Serra, P. Soille(eds.), *Mathematical Morphology and Its Applications to Image Processing*, pp.69-76, Kluwer Academic Publishers. Printed in the Netherlands, 1994.
- [11] Y. Ishiyama, C. Funaoka, F. Kubo, H. Takahashi and F. Tomita, “Labeling Board Based on Boundary Tracking”, *Pattern Recognition, 11th IAPR International Conference on Proceedings*. Vol. IV. pp. 34-38, 1992
- [12] Rafael C. Conzalez and Richard E. Woods “Digital Image Processing”, *Prentice Hall*, p535-536, 2002
- [13] 김영구, 노진우, 고한석 “복잡한 배경의 칼라영상에서 Face and Facial Feature 검출”, *대한전자공학회 하계종합학술대회 논문집* 제25권 제1호, pp.69-72, 2002
- [14] 김기남, 김형준, 김희율, “방송 비디오 등장인물 자동 분석 시스템”, *방송공학회논문지* 제9권 제2호, pp.164-173, 2004

백 승 호 (Seung-ho Paik)

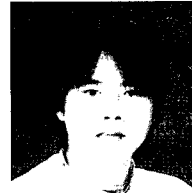
준회원



2005년 2월 광운대학교 전자공학과 졸업  
2005년 3월~현재 광운대학교 전자공학과 석사과정  
<관심분야> 영상인식, 동영상 코덱

김 준 환 (Jun-Hwan Kim)

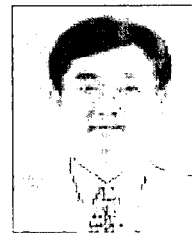
준회원



2005년 2월 광운대학교 전자공학과 졸업  
2005년 3월~현재 광운대학교 전자공학과 석사과정  
<관심분야> 영상처리, 동영상 코덱

유 지 상 (Ji-sang Yoo)

정회원



1985년 2월 서울대학교 전자공학과 졸업(공학사)  
1987년 2월 서울대학교 대학원 전자공학과 졸업(공학석사)  
1993년 5월 Purdue 대학교 전기공학과 졸업(Ph.D.)

1993년 9월~1994년 8월 현대전자산업(주) 산전연구소 선임연구원  
1994년 9월~1997년 8월 한림대학교 전자공학과 조교수  
1997년 9월~2001년 8월 광운대학교 전자공학과 조교수  
2001년 9월~현재 광운대학교 전자공학과 부교수  
<관심분야> 웨이블릿 기반 영상처리, 영상압축, 영상인식, 비선형 신호처리