내용기반 비디오 요약을 위한 효율적인 얼굴 객체 검출

준회원 김 종 성*, 정회원 이 순 탁*, 백 중 환**

An Efficient Face Region Detection for Content-based Video Summarization

Jong-Sung Kim*, Sun-Tak Lee*, Joong-Hwan Baek** *Regular Members*

요 약

본 논문에서는 효율적인 얼굴 영역 검출 기법을 제안하고 얼굴 객체 검출을 통해 인물 기반의 비디오 요약 시스템을 제공한다. 비디오 분할을 위해 비디오 시퀀스로부터 장면 전환점을 검출하고 분할된 장면들로부터 대표 프레임을 선정한다. 대표 프레임은 인접 프레임 간 변화량이 가장 적은 프레임으로 선정하였으며 추출된 대표 프레임에 대해서 얼굴 영역 검출 알고리즘을 적용하여 등장인물을 포함하는 프레임들을 요약 정보로 제공한다. 얼굴 영역 검출을 위해 피부색의 통계적 특성을 이용한 Bayes 분류기를 이용한다. 피부색 검출 결과 영상으로부터 수직 및 수평 투영 기법을 이용하여 영상 분할을 수행하고 후보군들을 생성한다. 생성된 후보군 중 오검출 영역을 최소화하기 위해서 이진 분류 나무(CART)를 이용하여 분류기를 생성한다. 특징 값으로는 SGLD(spatial gray level dependence) 매트릭스로부터 *Inertial, Inverse Difference, Correlation* 등의 질감 정보를 이용하여 최적의 이진 분류 나무를 생성한다. 실험 결과 제안된 얼굴 영역 검출 알고리즘은 복잡하고 다양한 배경에서도 우수한 성능을 보였으며, 얼굴 객체를 포함하는 프레임들을 비디오 요약 정보로 제공한다. 제안하는 시스템은 향후 화자 인식 기법을 이용하여 등장인물 기반의 비디오 분석 및 요약에 활용될 수 있을 것이다.

Key Words : video summarization, face detection, video indexing, CART, SGLD, spatial gray level dependence matrix.

ABSTRACT

In this paper, we propose an efficient face region detection technique for the content-based video summarization. To segment video, shot changes are detected from a video sequence and key frames are selected from the shots. We select one frame that has the least difference between neighboring frames in each shot. The proposed face detection algorithm detects face region from selected key frames. And then, we provide user with summarized frames included face region that has an important meaning in dramas or movies. Using Bayes classification rule and statistical characteristic of the skin pixels, face regions are detected in the frames. After skin detection, we adopt the projection method to segment an image(frame) into face region and non-face region. The segmented regions are candidates of the face object and they include many false detected regions. So, we design a classifier to minimize false region using CART. From SGLD matrices, we extract the textual feature

* 한국항공대학교 정보통신공학과 대학원 멀티미디어검색연구실(kmjgsg@mail.hangkong.ac.kr),
** 한국항공대학교 항공전자 및 정보통신공학부(jhbaek@mail.hangkong.ac.kr)
논문번호 : KICS2005-01-031, 접수일자 : 2005년 1월 17일
※ 본 논문은 산업자원부 한국산업기술평가원 지정 한국항공대학교 부설 인터넷정보검색 연구센터의 지원에 의함.

values such as Inertial, Inverse Difference, and Correlation. As a result of our experiment, proposed face detection algorithm shows a good performance for the key frames with a complex and variant background. And our system provides key frames included the face region for user as video summarized information.

Ⅰ. 서 론

동영상 압축 기술 및 통신 기술, 정보기기의 급격한 발달에 따라 디지털 비디오 정보의 활용이 급속하게 증가하고 있다. 따라서 비디오를 효과적으로 요약 및 검색하기 위한 방법들이 연구되고 있다[1],[2]. 최근에는 시각적인 내용 정보를 바탕으로 사용자가 쉽고 빠르게 원하는 정보를 브라우징 할 수 있는 비디오 요약에 관한 연구와 함께 표준화 활동이 활발하게 진행되고 있다[1],[3].

내용 기반의 비디오 요약 시스템을 위한 초기 단계로써 장면 전환점 검출에 관한 연구가 있었으며 우수한 성능을 갖는 검출 기법들이 있다[4]. 하지만 장면 전환점 검출 후 장면 마다 하나의 대표 프레임을 선정했을 때 정보의 양이 방대하여 다수의 군더더기를 내포하고 있다. 이러한 문제점을 해결하기 위해서 최근의 연구로는 비디오의 의미론적 분석을 위해 몇 개의 인접한 관련 장면(shot)을 그룹화하여 유사한 장면으로 묶는 기법들이 연구되었다[5]. 그룹화 된 장면(shot)을 하나의 씬(scene)이라고 하며 씬은 장면 단위의 비디오 분석보다 고차원의 비디오 구조를 제공하지만 비디오 내의 모든 인접한 장면들이 의미를 갖는 하나의 주제를 구성하기 위해 존재하는 것은 아니다.

이러한 문제점을 보완하기 위해 사용자의 특수한 요구에 대응할 수 있는 고차원의 비디오 요약을 위해 특정 객체를 포함하고 있는 장면을 검출하는 기법들이 요구되고 있다. 비디오의 내용을 고차원적이며 효율적으로 요약하기 위한 대표적인 방법으로 비디오 내에서 얼굴 영역 검출, 화자 인식 등의 연구가 진행되고 있다[6].

동영상에서 얼굴 영역 검출은 압축 영역과 비압축 영역에서의 검출 기법들이 있다[7]. 압축 영역에서의 얼굴 영역 검출은 특징 값의 제한된 종류로 인해 복잡한 배경을 갖는 상황에서 우수한 성능을 기대하기 어렵다. 따라서 본 논문에서는 등장인물 기반의 비디오 요약 시스템을 위해 비압축 영역에서의 비디오로부터 얼굴 영역 검출 알고리즘을 제안하며 분할된 장면들로부터 얼굴 영역을 포함하는 대표 프레임만을 추출한다.

본 논문의 구성은 Ⅱ장에서 제안하는 시스템 개요를 설명하고 Ⅲ장에서는 비디오의 기본 단위는 장면(shot) 단위로 분할하기 위해 CART를 이용한 장면 전환점 검출 방법과 대표 프레임의 선정 방법을 제안한다. Ⅳ장에서는 추출된 대표 프레임으로부터 얼굴 영역 검출을 위한 탐색 방법 및 특징 값으로 사용된 SGLD(spatial gray-level dependence)에 대해서 살펴본다. 마지막으로 Ⅴ장에서는 3개의 드라마 비디오에 대해서 제안한 장면 전환점 검출 방법과 얼굴 영역 검출 방법에 대한 실험 결과를 설명한다.

Ⅱ. 관련 연구 및 시스템 개요

비디오는 씬(scene)들로 구성되며 각각의 씬들은 다수의 장면(shot)으로 구성된다. 인접한 장면들이 하나의 씬을 구성하는데 이러한 씬을 검출하기 위해서 비디오를 장면으로 분할 후 클러스터링 기법을 이용하여 비디오를 구조화하는 연구가 진행되어 왔다[8]. 클러스터링 기법에서는 전체 비디오의 클러스터 수를 결정하기 곤란하다는 문제점이 있다. 또한, 씬 검출을 이용한 비디오 요약에서는 주석기반에서처럼 주관적인 경향이 있어 요약 결과에 대한 성능 평가 곤란하다.

따라서, 본 논문에서는 다수의 장면으로부터 추출된 대표 프레임 내의 중요한 객체인 인물을 추출하여 비디오 요약 정보를 제공한다. 드라마, 영화 등에서 인물을 포함하고 있는 장면이 콘텐츠의 중요한 내용을 내포하고 있다고 할 수 있다. 그림 1에서는 본 논문에서 제안하는 시스템을 보여주고 있으며 인물을 포함하고 있지 않는 장면들은 군더더기로(redundancy)로 고려되어 제거되거나 인물을 포함하는 장면과 분리된 후 요약 정보를 제공한다.

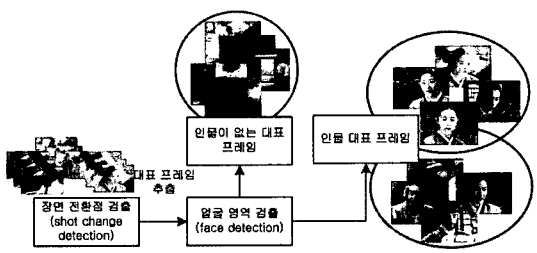


그림 1. 인물 객체 검출을 이용한 비디오 요약 시스템

Ⅲ. 내용기반 동영상 분할

3.1 장면 전환점 검출

비디오 분할을 위한 장면 전환점 검출은 압축 영역에서의 검출 방법과 비압축 영역에서의 검출 방법으로 나눌 수 있다. 압축 영역에서의 장면 전환점 검출은 I, B, P 픽쳐들로부터 DC 성분을 추출하여 특징 량을 사용한다. 압축 영역에서의 장면 전환점 검출은 압축 성분으로부터 특징 량을 추출해야하기 때문에 그 종류가 제한적이다. 그러나 비압축 영역에서의 장면 전환점 검출은 히스토그램 , 화소 기반의 방법으로부터 국부적 및 전역적인 특징 량 등 다양한 특징 값을 추출하여 이용할 수 있다. 또한 추출된 특징 값으로 장면 전환 여부를 결정하기 위해서 문턱 값(threshold value)을 이용하는 방법과 통계적 분류 기법을 이용하는 방법들이 있다.

본 논문에서는 비압축 영역에서의 특징값 추출과 추출된 특징 량을 통계적 분류 기법을 이용한 알고리즘을 제안한다.

3.1.1 장면 전환점 검출을 위한 특징 값

비디오 시퀀스에서 인접한 프레임 간의 컬러, 모양, 배경 등의 변화량을 검사하기 위해서 화소, 블록, 히스토그램 기법을 선택하였다.

히스토그램 기법(histogram difference)은 프레임 간의 컬러별 발생 빈도를 이용하여 비유사성을 검사하므로 구현 방법이 쉽고, 뛰어난 성능을 발휘한다. 본 논문에서는 히스토그램 기법을 이용하여 3가지의 특징 변수를 추출하였다. 그 3가지 특징 변수로 밝기(luminance) 정보의 국부 및 전역 히스토그램(LLH, LGH), 색상(hue) 정보의 전역 히스토그램(HGH)을 이용하였다.

화소 기반 기법(pixel-by-pixel difference)은 인접한 프레임과 대응되는 화소 차이를 계산하여 변화량을 측정하는 방법으로 밝기 성분에 대해서 화소 기반의 변화량(LPD)을 특징 변수로 이용하였다.

블록 기반 기법(block-based difference)은 인접한 프레임을 동일한 크기의 블록으로 나누어 인접한 블록사이의 유사도 함수를 이용하여 유사도(likelihood ratio)를 측정하는 방식이다. 블록 단위로 적용하기 때문에 카메라의 잡음, 미세한 움직임에 민감하지 않다. 블록 기반 기법을 이용한 특징 변수로써는 밝기 성분에 관한 유사도(LLR)를 이용하였다. 아래 수식 (1)은 m번째 프레임에 대한 유사도 LLR_m를 계산하는 유사도 함수이며 $\mu_{m,i}$와 $\sigma^2_{m,i}$는 m번째 프레임에서 i번째 블록에 대한 각각 평균과 분산을 의미한다.

$$LLR_m = \sum_{i=0}^{r} W_i$$

$$단,\ W_i = \begin{cases} 1 & LHR_j \geq th \\ 0 & LHR_j \geq th \end{cases} \qquad (1)$$

$$LHR_j = \frac{\left\{ \frac{\sigma^2_{m,j}+\sigma^2_{m-1,j}}{2} + \left(\frac{\mu_{m,j}-\mu_{m-1,j}}{2} \right)^2 \right\}^2}{\sigma^2_{m,j}+\sigma^2_{m-1,j}}$$

3.1.2 다중 특징 값을 이용한 분류

프레임 간 변화량에 관한 추출된 특징 값을 이용하여 이진 분류 나무(classification tree)로 장면 전환점을 분류하였다. 이진 분류 나무는 Bayes의 분류 기법들과는 다른 Nonmetric 분류 기법이므로 학습 시간과 분류 시간이 빠르다. 본 논문에서는 Breiman 등이 제안한 CART(Classification And Regression Tree)를 이용하여 장면 전환점 검출을 위한 분류기를 학습 시켰다[9]. 분류 나무는 뿌리 노드(root node), 가지(branch), 종단 노드(leaf node)로 구성되며 가지가 2인 분류 나무를 이진 분류 나무라 한다. 이진 분류 나무는 가지 생성(spit), 가지치기(pruning)과정을 통하여 형성된다. 가지 생성과정에서는 불순도(impurity)를 가장 크게 감소시키는 특징 값을 선택하여 가지를 생성하며 불순도 척도 함수에는 엔트로피(entropy), 지니(gini), 오분류(misclassification) 불순도 척도 함수가 있다. 본 논문에서는 두 클래스를 분류하는데 적합한 지니 불순도 함수를 이용하였으며 지니 함수는 수식 (2)와 같다. $i(N)$은 분류 나무의 노드 N에서의 불순도를 의미하며 노드 N에서 샘플들이 모두 동일한 클래스에 속해 있다면 불순도 $i(N)$는 0이 될 것이다.

$$i(N) = \sum_{i \neq j} P(w_i)P(w_j) = 1 - \sum_j P^2(w_j) \qquad (2)$$

분류 나무는 각 종단 노드의 불순도가 최소가 될 때까지 생성되며 가지치기 과정을 통하여 분류 나무 집합을 형성한다. 형성된 분류 나무 집합으로부터 표준 오차가 최소가 되는 나무가 최적화된 분류 나무라 할 수 있다.

본 논문에서는 추출된 5개의 특징 값을 이용하여 최적의 이진 분류 나무를 생성하여 급격한 장면 전환점 검출에 이용하였다. 그림 2는 최적화된 이진

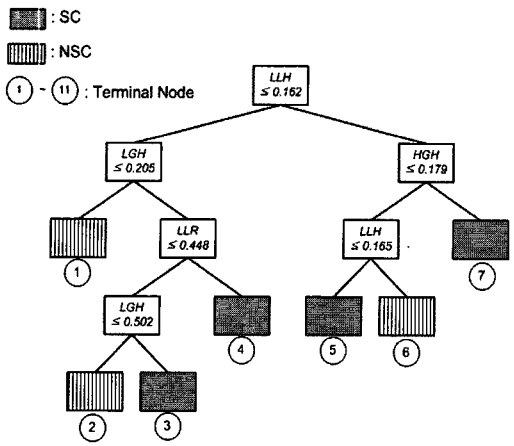


그림 2. 장면 전환점 검출을 위한 학습된 최적 이진 분류 나무

분류 나무이며 SC는 장면 전환이 발생한 프레임의 클래스를 말하며 NSC는 장면 전환이 아닌 프레임의 클래스를 의미한다.

3.1.3 대표 프레임의 선정

분할된 장면(shot)의 내용을 집약적으로 표현하기 위해서 대표 프레임을 선정한다. 샷의 내용 복잡도에 따라 대표 프레임은 하나 또는 그 이상의 프레임이 선정될 수 있다. Zhang 등은 하나의 장면 구간에서 첫 번째 프레임을 대표 프레임을 선정하는 방법과 카메라의 움직임을 고려하여 첫 번째 프레임과 마지막 프레임을 선정하는 방법을 제안하였다[10]. 또한, Wolf는 모션 정보를 구하여 국부적 최소가 되는 프레임들을 대표 프레임으로 선정하는 방법을 제안하였다[11]. 이 방법은 카메라의 움직임이 정적이거나 등장 인물의 움직임이 적은 구간일수록 중요도가 높은 프레임이라는 점에 착안한 방법이다. Gresle 등은 Wolf의 방법과 유사하게 샷 활동성에 기반한 대표 프레임을 선정하는 방법을 제안하였다[12]. Gresle 등이 제안한 샷 활동성이란 프레임간의 히스토그램 변화량을 구하여 국부적 최소가 되는 프레임을 대표프레임으로 선정하였다. 본 논문에서는 Gresle가 제안한 방법을 이용하였으며 변화량으로는 히스토그램만이 아닌 장면 전환점에 사용된 히스토그램 및 화소, 블록 기반 기법으로 추출된 프레임 간 변화량들의 평균치 $Var(m)$가 최소가 되는 프레임 m을 j번째 샷의 대표 프레임으로 선정하였다. 식 (3)은 대표 프레임 i를 선정하는 방법이며 LPD는 밝기 성분에 관한 화소의 변화량, LGH/LLH는 밝기 성분에 관한 히스토그램 전역/국부 변화량, HGH는 색상 성분에 관한 전역 히스토그램 변화량, LLR는 밝기 성분에 관한 블록 기반의 유사도를 의미한다.

$$key\ frame\ of\ j_{th}\ shot = \min_{m}\{Var(m)\},\quad m \in shot(j)$$
$$Var(m) = \frac{LPD(m)+LGH(m)+LLH(m)+HGH(m)+LLR(m)}{5} \qquad (3)$$

그림 3은 연속된 3개의 샷으로부터 선정된 대표 프레임을 보여주고 있다. 두 명의 등장인물이 대화하는 장면으로써 $k(j-1)$번째 프레임이 $k(j+1)$프레임과 유사함을 알 수 있다. 이것은 요약 정보의 군더더기(redundancy)로 고려 될 수 있으며 비디오 검색 속도를 느리게 하는 원인이 된다. 따라서 Yeung은 효율적이고 구조적인 비디오 요약을 위해서 장면 클러스터링 기법을 이용하여 유사한 장면을 하나의 스토리 단위로 그룹화 하였다[5]. 본 논문에서는 $k(j-1)$ 프레임과 $k(j+1)$ 프레임을 병합 가능성을 검사하기 위하여 장면 전환점 검출에서 사용된 최적화 이진 분류 나무 그림 2를 이용하였다.

그림 4는 6개의 연속된 대표 프레임들을 보여주고 있으며 $k(i+m)$와 $k(i+m+1)$, $k(j+m)$와 $k(j+m+1)$ 간의 유사도가 높은 것을 알 수 있다. 장면 전환점을 분류하는 이진 분류 나무에서 정의된 NSC는 유사도가 높은 클래스를 의미한다. 따라서 NSC로 분류되는 m내의 대표 프레임들을 하나의 대표 프레임으로 병합한다. 그림 5는 유사한 대표 프레임을 병합하는 알고리즘을 설명하며 ws는 비교할 샷 수를 의미한다. ws가 클수록 병합 과정에 요구되는 처리

그림 3. 드라마 비디오로부터 선정된 장면별 대표 프레임

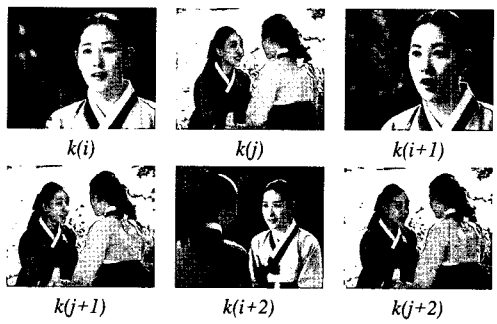


그림 4. 반복되는 유사 장면의 예

```
for(m=0; m<ws; m++)
  if |k(i+m)-k(i+m+1)| = Similar Keyframe
      add {k(i+m), k(i+m+1)} to Merged Keyframe[n]
  else
      n++
      make Merged Keyframe[n]
      add k(i+m+1) to Merged Keyframe[n]
```

그림 5. 유사 장면을 병합하기 위한 알고리즘

시간은 증가하게 되므로 실험적 결과로부터 가장 크게 병합 효과를 얻을 수 있는 크기로 설정하였다.

Ⅳ. 대표 프레임으로부터 얼굴 영역 검출

Ⅲ장에서 장편 영화, 드라마 등의 동영상으로부터 추출된 대표 프레임들은 수백 개 또는 수천 개에 이른다. 따라서 비디오의 의미론적 분석 기법과 특정 객체를 검출하여 사건(event) 중심으로 비디오를 요약하는 방법들이 제안되었다[13]. 영화, 드라마 등에서의 화자 인식은 사건 중심의 대표적인 방법이며 스포츠 비디오에서는 심판, 선수 등을 검출하여 비디오를 분석하는 연구가 있다. Defaux은 얼굴 검출 알고리즘을 이용하여 사람을 포함하고 있는 프레임을 대표 프레임으로 선정하는 방법을 제안하였다[14]. 본 논문에서는 대표 프레임들 중에서 특정 객체인 얼굴을 포함하고 있는 장면 검출 방법을 제안한다. 얼굴 영역 검출은 피부색의 통계적 분포 특성을 이용하여 피부색을 분류하고 SGLD(spatial gray-level dependence) 매트릭스를 이용하여 얼굴 영역을 분류하기위한 이진 분류 나무를 생성한다.

4.1 피부색 검출

그림 6은 피부색 검출을 위한 웹 이미지의 훈련 샘플에 대해서 추출된 피부색 영역을 보여주고 있으며 8×8 블록 내의 C_b, C_r의 평균값을 피부색 값으로 추출하였다. 그림 6의 (b)에서 흰 영역은 피부색의 학습 샘플로, 검은 영역은 피부색이 아닌 학습 샘플로 간주된다.

(a) (b)

그림 6. (a) 원영상 (b) 피부색 추출을 위한 마스크

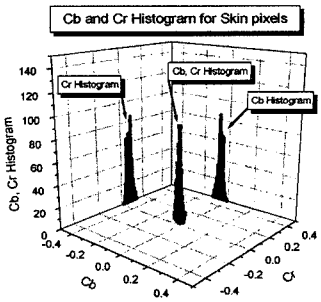


그림 7. 피부색 화소들에 대한 Cb, Cr의 히스토그램

그림 7은 훈련 샘플로부터 추출된 -0.5~0.5로 정규화된 C_b, C_r의 3차원 히스토그램을 보여주고 있다. 그림 6에서처럼 피부색의 학습 샘플들은 C_b, C_r 색 공간에서 공분산(σ)이 낮은 가우시안 형태의 분포 특성을 갖고 있다. 따라서 피부색 검출은 C_b성분과 C_r성분을 이용하여 식 (4)와 같이 2차원의 Bayes의 분류기를 이용하였다. C_b, C_r의 확률 밀도 함수는 2차원의 가우시안으로 가정하였으며 분류 단위는 8×8 블록에 대해서 피부색이 검출되어진다.

식 (4)에서 $p(x, y)$는 C_b, C_r에 대한 2차원 가우시안 확률 밀도함수이며 x는 해당 화소의 C_b값, y는 C_r 값을 나타낸다. D는 판별 함수를 의미하며 판별 함수가 0이상일 때를 피부색 클래스로 분류한다.

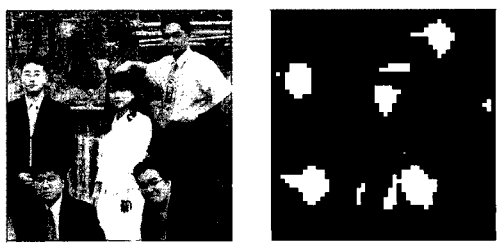
$$p(x,y)=\frac{e^{-\frac{1}{2(1-\rho^2_{xy})}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-\frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2_{xy}}} \quad (4)$$

$$D = p(x, y \mid Skin) - p(x, y \mid NonSkin)$$

if $D \geq 0$ *for sample* X

단, x : *Skin Class*, y : *NonSkin Class*

그림 8의 (a)는 피부색으로 분류되기 전의 영상을 보여주고 있으며 그림 7의 (b)는 8×8 블록들에 대해서 피부색 검출을 한 결과이다.



(a) (b)

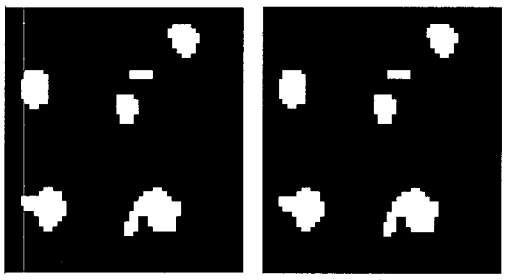
그림 8. 피부색 검출 결과 (a) 원 영상 (b) 8×8 블록에 대해서 피부색 검출 후의 이진화 영상

피부색을 검출 한 후 결과 영상은 이진화된 영상이며 피부 영역으로 분류된 영역을 1, 피부색이 아닌 영역을 0으로 한다. 이진화된 영상에는 팔과 같은 얼굴 영역이 아닌 다른 영역이 포함될 수 있으며 배경이 복잡하고 피부색과 유사한 색을 갖는 배경에 대해서는 오검출이 증가한다. 따라서 이러한 오검출 제거하기 위해서 Ying Dai는 SGLD 매트릭스를 이용하여 얼굴 영역 검출 알고리즘을 제안하였다[17]. 본 논문에서는 비디오 시퀀스에 적용하기 위한 보다 효율적이고 고속의 알고리즘으로 블록 단위의 탐색 과정과 피부색 검출 결과 영상에 대한 영상 분할 알고리즘을 이용하여 얼굴 영역 후보군들을 선정한다. 선정된 후보군들에 대해서 SGLD 매트릭스를 이용하여 얼굴 영역 오검출 영역을 최소화한다.

4.2 얼굴 영역의 후보군 선정

피부색 검출 결과인 이진화 영상인 그림 8의 (b)에서 잡음을 제거하기 위해서 미디언 필터를 이용하여 필터링 과정을 수행한다. 필터링 과정이 없을 경우 각 분할 영역이 증가되어 얼굴 영역의 후보군이 증가하게 된다. 따라서 1개의 블록으로 구성되는 수직 및 수평선들을 제거하고 필터링 속도를 고려하여 넓이가 3인 크로스 미디언 필터링 과정을 수행한다. 미디언 필터링 된 이진 블록 영상에 대해서 수평 및 수직 방향의 투영을 이용하여 영역 분할을 수행한다. 그림 9는 투영 결과로부터 영역 분할된 얼굴 영역 후보군을 보여주고 있다.

그림 9의 (a)는 전체 영상에 대해서 1차 투영한 결과이며 (b)는 1차 투영 결과로부터 각각의 분할 영역에 대해서 2차 투영으로부터 얻은 결과이다. 2차 투영결과로부터 얻은 분할 영역들은 얼굴 영역의 후보군들이며 탐색 과정을 통하여 팔, 손, 목 부분과 같은 얼굴 영역이 아닌 피부 영역들을 제거된다.



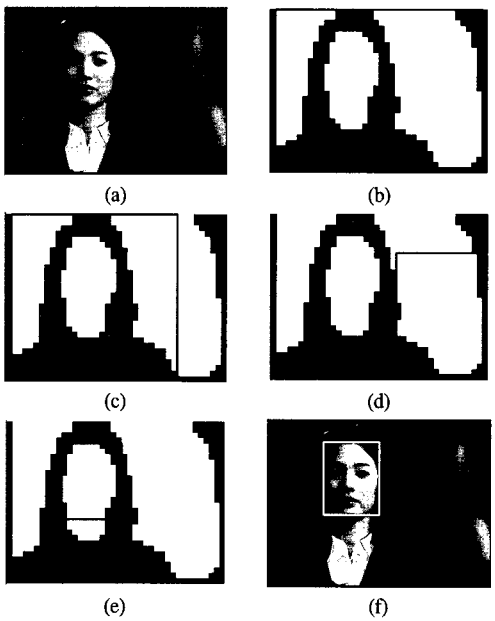
(a) (b)

그림 9. 투영을 이용한 영역 분할 (a) 전체 영상에 대해서 투영 결과 (b) 각 분할 영역에 대해서 2차 투영 결과

4.3 탐색 알고리즘

얼굴 영역 검출의 목적은 임의의 영상이 얼굴을 포함하고 있는지에 관한 여부를 결정하는 것이며 또한 각 얼굴 영역에 대한 위치와 크기를 파악하는 것이다[15]. 인간의 얼굴 영역은 타원 또는 호들의 결합 형태 등으로 추정된다[7]. 본 논문에서는 블록 단위의 탐색을 하기 위해서 사각형의 탐색창을 이용하였다. 탐색창은 전형적인 얼굴 형태의 비율인 1.4~1.6의 비율을 갖는 사각형을 탐색창으로 설정하였다. 탐색창의 크기는 2개의 블록 단위로 축소하면서 후보군에 대해서 얼굴 영역을 탐색한다.

그림 10의 경우는 (b)에서처럼 배경이 피부색과 유사할 경우 피부색 검출 결과인 이진 영상으로부터 얼굴 영역의 정확한 위치와 크기 파악이 부정확하다. 따라서 본 논문에서는 탐색창의 크기와 위치를 변화시키면서 얼굴 영역을 검출하는 방법을 제안한다. 또한 탐색창을 사용하는 방법은 검출 속도가 느린 단점이 있으나 블록 단위의 탐색창 위치 이동 및 크기 조절은 속도 개선에 매우 효율적이다[2]. 또한 후보군에 대해서 탐색창 내의 피부색 블록의 수가 일정한 비율 이상을 차지할 때(조건 1)와 탐색창 외의 테두리 부분에 대해서 피부색이 아닌 블록 수가 일정한 비율 이하일 때(조건 2)의 두 조



(a) (b) (c) (d) (e) (f)

그림 10. 탐색 과정과 얼굴 영역의 검출 결과 (a) 입력 영상 (b) 피부색 검출 후의 이진화 영상 및 분할 영역 (c) 초기 탐색 창 (d) 탐색 과정의 중간 단계 (e) 후보 영역에 대해서 얼굴 영역을 검출한 결과 (f) 원영상 및 얼굴 영역으로 검출한 탐색창

건을 만족하는 영역을 얼굴 영역이라고 가정한다. 그림 10의 (b), (c)에서는 조건 1을 만족하지 않으며 (d)에서는 조건 2를 만족하지 않는다. 하지만 탐색창의 위치와 크기가 (e)일 경우 조건 1을 만족하며 얼굴 영역이 아닌 머리카락 등의 배경으로 인해 조건 2를 만족하게 된다.

4.4 SGLD(spatial gray-level dependence) 매트릭스와 얼굴 영역 검출을 이진 분류 나무

SGLD 매트릭스는 질감 특징 분석에 사용되며 화소 *(i, j)* 위치에서의 *[0, L-1]*의 범위를 갖는 화소 값을 *I(i, j)*로 했을 경우 벡터 *(m, n)*에 대해 이웃하는 화소 값들의 발생 빈도 $P_{ab}(m, n)$를 식 (5)으로 정의된다[16]. $P_{ab}(m, n)$의 요소로 구성된 매트릭스를 SGLD 매트릭스라고 정의한다.

$$P_{ab}(m,n) = \#\{((i,j),(i+m,j+n)) \in (L_x \times L_y),\ I(i,j) = a,\ I(i+m,j+n) = b\} \quad (5)$$

$$N_{ab}(m,n) = P_{ab}(m,n)/(L_x \times L_y) \quad (6)$$

식 (5)에서 #은 집합 *a, b*에 대한 발생 빈도를 의미하며 L_x, L_y는 각각 영상의 폭, 높이를 의미한다. $P_{ab}(m, n)$의 정규화된 $N_{ab}(m, n)$는 식 (6)으로 근사화되며 정규화된 SGLD 매트릭스 $N_{ab}(m, n)$를 바탕으로 질감에 관한 특징 값들이 유도된다. 질감 특징들은 식 (7)~(11)과 같이 SGLD 매트릭스를 이용하여 측정되며 본 논문에서는 *Inertial, Inverse Difference, Correlation* 특징만을 이용한다. 수식 (5)에서 μ, σ는 각각 영상 전체에 대한 평균과 표준편차를 의미한다.

$$B_E(m,n) = Energy = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} N_{ab}(m,n)^2 \quad (7)$$

$$B_{ET}(m,n) = Entropy = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} N_{ab}(m,n) log N_{ab}(m,n) \quad (8)$$

$$B_I(m,n) = Inertia = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} (a-b)^2 N_{ab}(m,n) \quad (9)$$

$$B_D(m,n) = Inverse = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} \frac{1}{1+(a-b)^2} N_{ab}(m,n) \quad (10)$$

$$B_C(m,n) = Correlation = \frac{1}{\sigma^2}\sum_{a=0}^{L-1}\sum_{b=0}^{L-1} (a-\mu)(b-\mu) N_{ab}(m,n) \quad (11)$$

그림 10과 같은 수직 방향의 얼굴 영역에서의 질감 정보는 입, 코, 눈의 영향으로 수직 방향의 밝기 성분의 연속성이 떨어진다. 즉, 수직 방향으로 고주파 성분이 크게 나타나며 반면, 수평 방향의 고주파 성분은 적게 나타나는 질감 특징을 갖고 있다. 압축 영역에서의 얼굴 영역 검출 알고리즘에서 질감 특징 정보는 수직, 수평 방향으로 군집화된 DCT 계수를 이용하는 방법 등이 있다[7].

본 논문에서는 수직, 수평 방향의 질감 특징을 이용하기 위해서 SGLD 매트릭스를 이용하여 최적의 이진 분류나무를 형성하여 선정된 얼굴 영역 후보군에 대해서 오검출을 최소화하였다.

SGLD 매트릭스를 이용한 특징 값 중 $B_I(m, n)$ 배열은 근접한 두 화소 값 *a, b*의 변화량을 정도를 의미한다. 그림 10은 수평, 수직 방향의 선으로 구성된 영상 등에 대해서 *m=n=0, 1,* 2일 때 각각의 $B_I(m, n)$, $B_D(m, n)$, $B_C(m, n)$ 배열을 보여주고 있다.

그림 11에서 수직 방향의 선을 갖는 영상 (a)는 수평 방향으로의 인접한 화소간의 변화량이 수직

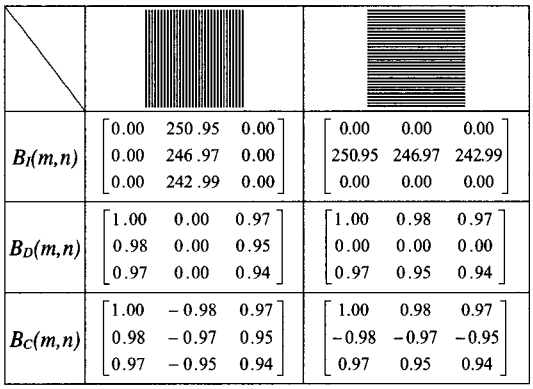


그림 11. SGLD 매트릭스로 측정된 inertial, inverse difference, correlation 등의 질감 특성

↓ n \ → m			
	(m=0,n=0)	(m=1,n=0)	(m=2,n=0)
	(m=0,n=1)	(m=1,n=1)	(m=2,n=1)
	(m=0,n=2)	(m=1,n=2)	(m=2,n=2)

(a)

1	2	5
3	4	7
6	8	9

(b)

그림 12. 배열의 인덱싱 및 1차원 변환 순서 (a) $B_I(m, n)$, $B_D(m, n)$, $B_C(m, n)$의 m, n에 대한 인덱싱 (b) 각 요소들의 1차원 나열 순서

방향의 변화량 보다 크다. 따라서 $B_I(m, n)$ 배열을 그림 12의 (a)와 같이 m, n에 대해서 인덱싱하면 $B_I(m=1, n) \geq B_I(m, n=1)$이 성립하며 수평 방향의 선을 갖는 영상 (b)에 대해서는 역으로 $B_I(m=1, n) \leq B_I(m, n=1)$이 성립한다. 따라서 m, n을 전치시켰을 때 대응되는 요소들을 연속적으로 비교하기 위해서 그림 12의 (b)와 같이 배열의 요소들을 1차원으로 나열한다. 즉, $B_I(7)$은 $B_I(m=2, n=1)$을 의미한다.

inverse difference 배열은 m, n 내에 있는 국부 영역에 대한 동질성을 나타낸다. 즉, m, n 내의 국부 영역이 동질의 화소로 구성될 경우 $B_D(m, n)$ 배열의 요소 값은 증가하며 이질의 화소로 구성될 경우 감소하게 된다. *inverse difference* 배열은 [0, 1]의 범위를 가지며 0일 경우 동질의 정도가 최소임을 의미한다. 또한, *correlation* 배열은 영상의 전체 영역에 대한 상관도를 의미한다. $B_C(m, n)$은 m, n에 있는 a, b가 전체 영상에 대하여 상관도가 높을수록 +1, 상관도가 낮을수록 0, 음에 관한 상관도가 높을수록 -1의 값을 갖는다. m, n내에 있는 영역이 전체 영상에 비해 상관도가 높을수록 $|B_C(m, n)|$은 1에 접근하며 상관도가 낮을수록 $|B_C(m, n)|$은 0에 접근하게 된다. 즉, 그림 13에서처럼 m, n에 있는 국부 영역이 전체 영상에 동질성(homogeneity)이 높을수록 $|B_C(m, n)|$의 값은 낮아지게 된다.

				
B_I	0.0 4.0 8.1 0.0 3.98 7.97 0.0 3.98 7.84	〃	〃	〃
B_D	1.0 0.97 0.94 0.98 0.95 0.92 0.97 0.94 0.91	〃	〃	〃
B_C	1.0 0.95 0.90 0.98 0.94 0.89 0.97 0.92 0.88	1.0 0.93 0.86 0.98 0.93 0.85 0.97 0.90 0.84	1.0 0.87 0.75 0.98 0.86 0.73 0.97 0.84 0.72	1.0 0.75 0.50 0.98 0.74 0.49 0.97 0.72 0.48

그림 13. 전체 영상에 대비 m, n 국부 영역의 $B_C(m, n)$ 배열의 동질성에 관한 특성

	원영상	전처리 결과 영상	원영상	전처리 결과 영상
				
$B_I(m,n)$	0.00 4.85 11.91 6.06 8.23 13.62 16.04 16.12 18.90		0.00 4.62 10.20 8.61 9.54 13.82 17.66 17.95 20.52	
$B_D(m,n)$	1.00 0.18 0.05 0.17 0.12 0.04 0.05 0.04 0.03		1.00 0.15 0.07 0.18 0.13 0.06 0.05 0.05 0.02	
$B_C(m,n)$	1.00 0.83 0.63 0.80 0.70 0.53 0.52 0.47 0.36		1.00 0.84 0.67 0.77 0.70 0.56 0.52 0.47 0.37	

(a) (b)

그림 14. 얼굴 영상에 대하여 전처리된 영상의 $B_I(m, n)$, $B_D(m, n)$, $B_C(m, n)$ 배열의 예

그림 14는 얼굴 영상에 대하여 미디언 필터링과 히스토그램 평활화를 이용하여 입력 영상의 다양한 조명 조건을 전처리하였다. 그림 14에서 전처리 결과 영상은 20×26으로 정규화된 영상이며 저해상도에 적당한 (m=2, n=2)을 설정하였다. 그림에서처럼 얼굴 영역에서 두 눈의 특성 때문에 수평 성분이 강하게 나타나므로 $B_I(m=K, n) \leq B_I(m, n=K)$이 성립한다. 따라서 $B_I(m, n)$ 배열을 일차원으로 나열하여 $(m+1)\times(n+1)-1$개의 $B_I(d) - B_I(b)$ ($단, d>b$)를 특징 값으로 한다. 또한 얼굴 영역의 눈, 코, 입 영역은 국부적으로 밀집되어 있는 특성이 있다. 따라서 그림에서처럼 $B_D(m, n)$, $B_C(m, n)$ 배열의 요소 값이 특정 범위에 분포하게 되며 m=n=0의 요소 값을 제외한 $(m+1)\times(n+1)-1$개의 배열 요소 값들을 특징 값으로 한다.

SGLD 매트릭스의 (m, n)의 값과 정규화 영상 크기를 설정하기 위해서 입력 영상 90×112에 대해서 정규화 영상을 20×26을 기준으로 가로/세로 1.5배씩 증가시키면서 (m, n)에 값에 따른 분류 성능을 평가하였다. 분류 방법은 이진분류 나무를 이용하였으며 교차 검증(cross validation) 방법을 이용하여 소환(recall) 비율과 정확(precision) 비율로 성능을 평가하였다. 또한, 분류를 위해 사용된 영상 샘플로는 AT&T의 얼굴 영상 데이터베이스의 400개 얼굴 영상과 Corel 영상들로부터 추출된 얼굴이 아닌 영상 300여개 영상을 이용하였다.

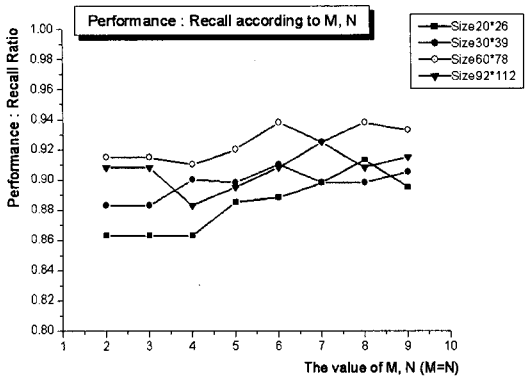


그림 15. 분류 성능 : 정규화 영상 크기 및 m, n에 따른 소환 비율(recall ratio)

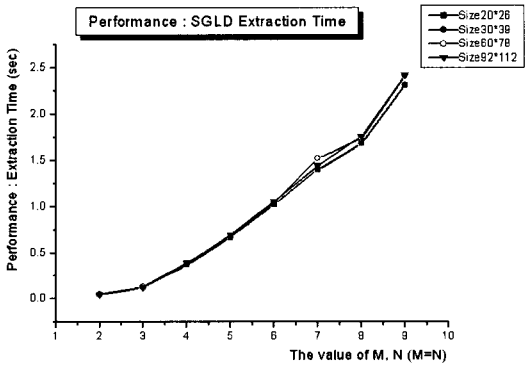


그림 16. 분류 성능 : 정규화 영상 크기 및 m, n에 따른 SGLD 매트릭스 추출 속도

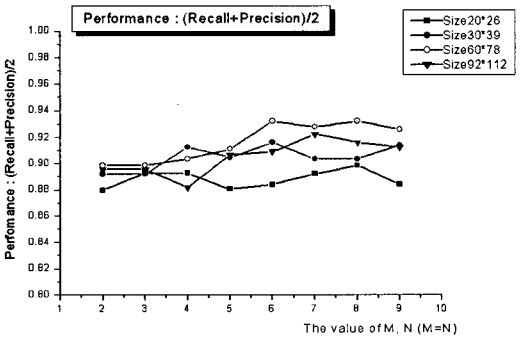


그림 17. 분류 성능 : 정규화 영상 크기 및 m, n에 따른 소환 비율 및 정확 비율의 평균

그림 15는 m=n의 값을 2~9까지 변화하면서 분류 나무의 성능을 평가하였다. 정규화 영상의 크기가 증가함에 따라 소환 비율이 향상되지만 원영상의 크기와 동일할 경우 성능이 감소하기 시작하는 것을 볼 수 있다. 또한 m=n의 값이 증가하면 성능 또한 향상되지만 그림 16에서처럼 SGLD 매트릭스 추출 처리 속도가 비례적으로 증가하는 것을 볼 수 있다. 따라서 m, n 값에 따른 성능과 처리속도는 보상관계(tradeoff)에 있으며 정규화 영상에 따른 처리 속도는 큰 변화가 없다. 따라서 높은 성능을 갖는 정규화 영상크기와 처리속도가 낮은 m, n의 값을 설정하여야 한다. 그림 17은 정확 비율과 소환 비율에 대한 평균을 보여 주고 있다. 얼굴 영역 검출에서는 얼굴 영역이 아닌 영역을 얼굴 영역으로 잘못 분류하는 정도를 나타내는 정확 비율 또한 성능 평가에 중요한 요소이다.

따라서 본 논문에서는 소환 비율과 정확 비율이 높은 60×78의 정규화 영상 크기를 선택 하였으며 m, n의 값은 6으로 선택하였다.

V. 실험 결과 및 검토

실험 동영상으로는 다양한 얼굴 영역을 포함하고 있는 드라마 비디오를 이용하였다. 표 1은 실험에 사용된 비디오에 대한 실제 장면 전환점 수 및 프레임 수 등의 정보를 나타내고 있다.

장면 전환점 검출을 위해 생성된 그림 1의 분류 나무를 이용하여 비디오 1~3을 장면 단위로 분할 후 각 장면으로부터 대표 프레임을 추출하였다. 표 2는 장면분할에 대한 결과를 보여주고 있다. 비디오 3에서는 본 논문에서 고려하지 않은 다수의 점진적 장면 전환점 때문에 소환 비율에 대한 성능 저하를 보였으며 객체와 카메라의 빠른 움직임으로 인해 정확 비율에 대한 성능 저하를 보였다.

표 3에서는 m의 값에 따른 대표 프레임 $K(i)$와 $K(i+ws)$ 사이에 있는 유사한 대표 프레임을 제거한 결과를 보여주고 있다. 표 3에서처럼 비디오 드라마

표 1. 실험 동영상

	프레임 율	총 프레임	장면 전환점
비디오 1	30frame/s	19,510	134
비디오 2	30frame/s	19,046	84
비디오 3	30frame/s	28,742	233

표 2. 장면 전환점 검출 결과 및 성능

	검출된 장면 전환점	장면 분할 성능	
		소환비율	정확비율
비디오 1	151	0.993	0.894
비디오 2	87	0.944	0.965
비디오 3	265	0.866	0.803

표 3. *ws* 크기에 따른 대표 프레임을 병합 결과

비디오 / *ws*	비디오 1	비디오 2	비디오 3
1	150	87	265
2	114	74	249
3	105	72	238
4	100	69	236
5	99	68	236
6	95	66	236
7	94	66	235
8	93	66	232

에서는 m=3~4일 때 병합이 가장이 많이 발생하며 이것은 빈번한 대화 장면들로 구성되어 있기 때문이다. 본 논문에서는 ws=4으로 설정하였으며 각각의 대표 프레임에서 얼굴 영역 검출을 실험하였다.

입력 영상을 60×78로 정규화한 후 m=n=6일 때의 $B_I(m, n)$, $B_D(m, n)$, $B_C(m, n)$ 배열로부터 추출된 각 $(m+1)\times(n+1)-1$개의 특징 값을 이용하여 이진 분류 나무를 그림 20에서처럼 생성하였다. 얼굴 영상(Face 클래스)의 학습 샘플로는 40명의 10개 포즈로 구성된 AT&T(Olivetti) 의 얼굴 데이터베이스를 이용하였다. 그림 18의 (b)는 AT&T의 얼굴 영상 (a)로부터 배경을 제거하여 얼굴 영역만을 고려하도록 학습 샘플을 구성하였다.

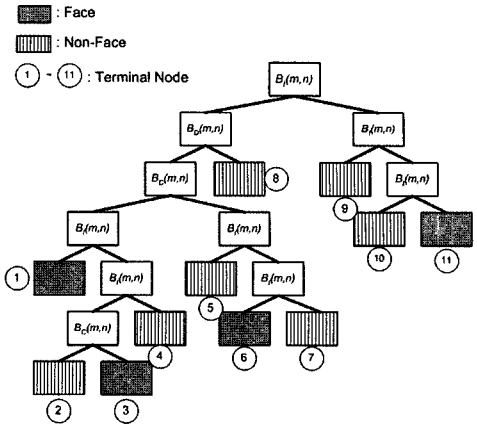
또한 얼굴 영역이 아닌 학습 샘플로는 Corel 영상들 중 스킨 영역으로 분류된 부분을 이용하였다. 그림 19의 (a)~(c)는 Corel 영상들 중 장미 그룹의 한 영상을 이용하여 학습 샘플을 생성한 예를 보여주고 있다.

(a) (b)

그림 18. Face 클래스의 학습 샘플의 예 (a) AT&T의 얼굴 영상 데이터베이스 (b) Face 클래스의 학습 샘플

(a) (b) (c)

그림 19. Non-face 클래스의 학습 샘플 (a) Corel영상 (b) 피부색 검출 결과 (c) Non-face 클래스의 학습 샘플



(a)

if $B_I(5, 0) - B_I(4, 4) \le -2.337$
  if $B_D(1, 0) \le 0.2885$
    if $B_C(2, 2) \le 0.731$
      if $B_I(4, 5) - B_I(5, 4) \le -0.318$
        then input image = "face class"
        terminal node = 1

(b)

그림 20. SGLD 매트릭스를 이용한 얼굴 영역 검출의 최적 이진 분류 나무 (a) 이진 분류 나무 (b) 종단 노드 ①에 대한 조건

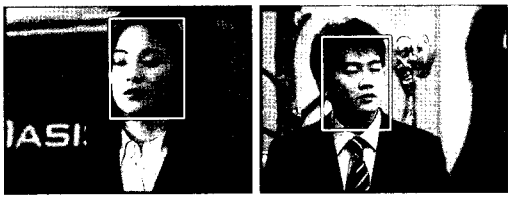
그림 19의 (b)는 Bayes의 피부색 분류기로부터 분류된 8×8 블록들을 보여주고 있으며 흰 영역은 피부색으로 분류된 영역을 의미한다. 따라서 그림 19의 (c)에서처럼 영상을 분리하여 얼굴 영상이 아닌 학습 샘플(Non-face 클래스)로 이용하였다. Corel 영상들 중 5개 그룹에 피부색 분류기를 적용하여 약 300개의 Non-face 클래스를 위한 학습 샘플을 생성하였다. 생성된 Face/Non-face 클래스의 학습 샘플을 이용하여 영상 크기를 60×78로 정규화 한 후 m=n=6에 대한 SGLD 매트릭스로부터의 특징 값을 추출하여 최적 분류 나무를 그림 20의 (a)에서와 같이 생성하였다. 이진 분류 나무에서 종단 노드 ①, ③, ⑥, ⑪로 분류되는 입력 영상은 얼굴 클래스로 분류되며 종단 노드의 조건에 대해서 노드 ①에 대한 예를 그림 (b)에 나타내었다.

표 4는 얼굴 영역 검출 알고리즘을 이용하여 비디오 1~3의 대표 프레임에 존재하는 얼굴 영역을 검출한 결과를 나타낸다. 비디오 1, 2에서는 6%이상의 얼굴 영역을 포함하고 있는 대표 프레임이 약 50%에 가까웠으며 비디오 3의 경우 얼굴 영역을 포함하고 있는 대표 프레임보다는 사물, 자연 풍경 등을 포함하고 있는 프레임이 더 많이 존재하였다. 얼굴 영역 검출은 얼굴 영역이 프레임 크기의 6%

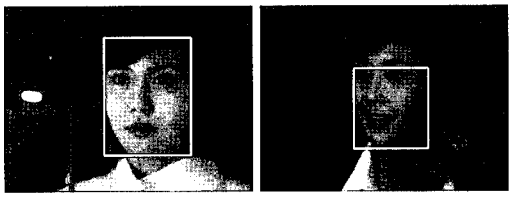
표 4. 얼굴 영역 검출 결과

	대표 프레임	얼굴 영역의 수	성 능	
			소환 비율	정확 비율
비디오 1	100	51	0.944	0.85
비디오 2	69	38	0.927	0.927
비디오 3	236	77	0.895	0.762

(a) 비디오 1에서 올바르게 검출된 얼굴 영역의 예



(b) 비디오 2에서 올바르게 검출된 얼굴 영역의 예



(c) 비디오 3에서 올바르게 검출된 얼굴 영역의 예

그림 21. 실험 비디오에서 올바르게 검출한 얼굴 영역의 예

그림 22. 실험 비디오에서 오검출한 얼굴 영역의 예

이상일 때를 고려하였으며 검출된 얼굴 영역이 20% 이상일 때를 클로즈 업 된 프레임으로 설정하여 데이터베이스에 인덱싱하였다.

그림 21의 (a)~(c)는 각각의 실험 비디오에서 올바르게 검출된 얼굴 영역을 보여주고 있으며 복잡한 배경과 피부색과 유사한 영역이 존재함에도 불구하고 얼굴 영역을 정확하게 검출하였다. 그림 22는 각 실험 비디오에서 잘 못 검출한 예로써 피부색 및 SGLD를 이용한 질감 정보가 얼굴 영역과 유사한 경우였다.

Ⅵ. 결 론

본 논문에서는 비디오를 장면 단위로 분할한 후 장면을 가장 잘 대표할 수 있는 대표 프레임으로 변화량이 가장 작은 프레임을 선정하였다. 장면 전환점 검출에서는 3개의 실험 비디오 대해서 평균 95%의 높은 소환비율을 보였다. 또한 향후 화자 인식, 등장인물 중심의 비디오 요약, 등장인물 인식 등의 비디오 분석 기법을 위한 얼굴 영역 방법을 제안하였다. 얼굴 영역 검출을 위해 우수한 성능을 갖는 질감 정보를 추출하기 위해 SGLD 매트릭스를 분석하여 최적의 특징 값을 추출하였다.

복잡한 배경을 갖는 비디오 드라마를 이용한 실험에서 제안한 알고리즘이 평균 92%의 소환비율로 우수한 성능을 보였다. 향후 오검출 얼굴 영역을 줄이기 위한 개선된 기법과 음성 정보를 병합하여 화자 인식, 등장인물 인식 등의 발전된 비디오 요약 및 분석에 관한 연구가 요구된다.

참 고 문 헌

[1] B. S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7, John Wiley&Sons, 2002.

[2] Y. Alp Aslandogan, Clement T. Yu, "Techniques and Systems for Image and Video Retrieval", IEEE Trans. on Knowledge and Data Engineering, vol. 11, no. 1, pp. 56-63, Jan./Feb. 1999.

[3] J. H. Lee, G. G. Lee, W. Y. Kim, "Automatic Video Summarizing Tool using MPEG-7 Descriptors for Personal Video Recorder", IEEE Trans. on Consumer Electronics, vol. 49, no. 3, pp. 742-749, Aug. 2003.

[4] S. B. Hong, W. Nah, J. H. Baek, "Abrupt Shot Change Detection Using Multiple Features and Classification Tree", IDEAL 4th International Conference on Intelligent Data Engineering and Automated Learning 2003, LNCS 2690, pp. 553-560, March 2003.

[5] M. Yeung, B. L. Yeo, "Segmentation of Video by Clustering and Graph Analysis", Computer Vision and Image Understanding, vol. 71, no. 1, pp. 94-109, July. 1998.
[6] Azirel Rosenfeld, David Doermann, Daniel DeMenthon, Video Mining, Kluewer Academic Publishers, 2003.
[7] H. Wang, S. Fu. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video", IEEE Trans. on Circuits and System for Video Technology, vol. 7, no. 4, pp. 615-928, Aug. 1997.
[8] C. W. Ngo, Y. F. Ma, H. J. Zhang, "Video Summarization and Scene Detection by Graph Modeling", IEEE Tans. on Circuits and Systme for Video Technology, vol. 15, no. 2, pp 296-305, Feb. 2005.
[9] L. Breiman, J. H. Friedman, R. A. Olshen, Charles J. Stone, Classification and Regression Tree, CRC Press, 1998.
[10] H. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing", Pattern Recognition, vol. 20, no. 4, pp. 643-658, 1997.
[11] W. Wolf, "Key frame selection by motion analysis", IEEE Int. Conference on Acoustic, Speech, and Signal Processing, vol. 2, pp. 1228-1231, May 1996.
[12] P. O. Gresle and T, S, Huan, "Gisting of video documents: A key frame selection algorithm using relative activity measure", 2nd Int. Conference on Visual information Systems, 1997.
[13] Ying Li, C. C. Jay, Video Content Analysis using Multimodal Information, Kluewer Academic Publishers, 2003.
[14] Dufaux F., "Key frame selection to represent a video", IEEE Proceedings of International Conference on Image Processing, vol. 2, pp. 275-278, 2000.
[15] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey," IEEE Trans. on PAMI, vol. 24, no. 1, pp. 34-58, Jan. 2002.
[16] Ying Dai, Y. Nakano, "Face-Texture Model Based On SGLD and Its Application in Face Detection in a Color Scene", Pattern Recognition, vol. 29, no. 6, pp. 1007-1017, 1996.

김 종 성 (Jong-Sung Kim) 준회원

2003년 2월 한국항공대학교 항공통신정보공학과(공학사)
2005년 2월 한국항공대학교 대학원 정보통신공학과(공학석사)
2005년 1월~현재 LG전자 단말 연구소 연구원
<관심분야> 비디오 요약 및 검색, 영상 압축, 영상 처리, 패턴 인식, 멀티미디어

이 순 탁 (Sun-Tak Lee) 정회원

1998년 9월 대학원 졸업(MS)
1998년 6월-2000년 8월 한국휴렛팩커드, 에질런트테크놀로지스 계측기 연구소
2000년 8월~현재 (주) 텔레칩스 미디어 연구소 선임연구원
<관심분야> 비디오 데이터베이스, 비디오 요약, 비디오 코딩, 영상 처리, 패턴 인식

백 중 환 (Joong-Hwan Baek) 정회원

1981년 2월 한국항공대학교 항공통신공학과(공학사)
1987년 7월 오클라호마주립 대학원 전기 및 컴퓨터공학과(공학석사)
1991년 7월 오클라호마주립 대학원 전기 및 컴퓨터공학과(공학박사)
1992년~현재 한국항공대학교 항공전자 및 정보통신공학부 교수
<관심분야> 영상처리, 패턴인식, 영상압축, 멀티미디어