

논문 2006-43SP-3-3

앵커 객체 추출을 이용한 앵커 프레임 검출

(Anchor Frame Detection Using Anchor Object Extraction)

박 기 태*, 황 두 선**, 문 영 식***

(Ki Tae Park, Doo Sun Hwang, and Young Shik Moon)

요 약

본 논문에서는 뉴스 비디오에서 앵커 프레임 검출을 위한 알고리즘을 제안한다. 제안된 알고리즘은 다음과 같이 4단계로 구성된다. 첫 번째 단계에서, 뉴스 비디오를 비디오 샷들로 분할하기 위해 누적 히스토그램(cumulative histogram) 기법을 이용하여 샷 경계(shot boundary)를 검출한다. 두 번째 단계에서는 각 비디오 샷 경계에서 얼굴 영역들을 찾기 위해서 피부 컬러(skin color) 정보를 이용하고, 세 번째 단계에서는, 앵커 객체를 추출하기 위해서 사람의 상체 부분의 컬러 정보를 이용하여 앵커 후보 프레임을 검출하며, 마지막 단계에서, 후보 프레임들에 대해서 앵커 프레임과 비앵커 프레임을 분류하기 위해서 그래프 이론을 이용한 클러스터 분석 알고리즘을 적용한다. 실험 결과를 통해서 제안한 알고리즘이 효과적으로 앵커 프레임을 검출하는 것을 보여준다.

Abstract

In this paper, an algorithm for anchor frame detection in news video is proposed, which consists of four steps. In the first step, the cumulative histogram method is used to detect shot boundaries in order to segment a news video into video shots. In the second step, skin color information is used to detect face regions in each shot boundary. In the third step, color information of upper body regions is used to extract anchor object, which produces candidate anchor frames. Then, from the candidate anchor frames, a graph-theoretic cluster analysis algorithm is utilized to classify the news video into anchor-person frames and non-anchor frames. Experiment results have shown the effectiveness of the proposed algorithm.

Keywords: anchor frame detection, object extraction, graph-theoretic cluster algorithm

I. 서 론

멀티미디어의 기술이 상당히 발전함에 따라 대용량의 정보들이 멀티미디어화 및 디지털화 되고 있다. 또한 이러한 멀티미디어 정보를 효과적으로 검색할 수 있는 방법들이 요구되고 있다. 특히, 비디오에 대해서는 사용자가 효율적인 정보 검색을 할 수 있도록 하기 위해서 많은 멀티미디어 검색 시스템들은 색인 방법을 사

용한다. 이러한 방법은 video on demand (VOD), 홈쇼핑, 그리고 디지털 도서관등의 분야에서 활용할 수 있으며, 내용기반 비디오 검색 기술이 필요하게 되는데, 일반적으로 컷(cut) 검출을 기반으로 하는 샷을 검출함으로써 수행되어진다.

프레임들의 시퀀스로 구성된 뉴스 비디오는 고유의 특징을 갖고 있다. 뉴스 비디오를 분석해 보면 앵커가 요약된 기사를 전달하는 앵커 구간과 그 기사를 상세히 설명해 주는 보도구간이 하나의 뉴스 기사로 구성되는 것을 확인 할 수 있다. 본 논문에서는 하나의 뉴스 비디오를 의미 있는 여러 개의 뉴스 기사 단위로 분할한다.

뉴스 비디오를 의미 있는 단위로 분할하기 위해서, 많은 방법들이 제안되었다. Hanjalic은 뉴스 동영상에서 앵커 장면을 검출하기 위하여 템플릿을 결정하고 그 템플릿에 대해서 유사도를 검사하는 방법을 제안하였다^[1].

* 학생회원, *** 평생회원, 한양대학교 컴퓨터공학과
(Department of Computer Science and Engineering, Hanyang University)

** 정회원, 삼성종합기술원
(Samsung Advanced Institute of Technology)

※ 본 연구는 한국과학재단 특정기초연구
(R01-2006-000-10876-0)지원으로 수행되었음.

접수일자: 2005년10월18일, 수정완료일: 2006년3월31일

이 방법은 템플릿을 결정하기 위해서 너무 많은 시간이 소요되며, 첫 번째 앵커 장면이 어느 정도의 시간이 지난 후에 나온다는 정보가 필요하다. 또한, 템플릿이 잘못 결정된다면 전혀 예상치 못한 결과를 초래할 수 있다. Wang은 앵커 장면을 배경 영역과 얼굴 영역으로 나누어서 두 단계를 거쳐서 앵커 장면을 검출하였다^[2]. 먼저, 배경 영역과 얼굴 영역에 대해서 모델을 결정한다. 그런 다음, 결정된 모델에 만족하는 장면을 앵커 장면으로 검출한다. 이 방법은 고정된 배경 색상을 가진 뉴스에서는 좋은 성능을 보이지만, 배경이 다른 뉴스이거나 배경이 바뀌는 뉴스에서는 만족스럽지 못한 결과를 얻는 단점이 있다. Albiol은 앵커를 검출하기 위해 뉴스 동영상내에서 사람이 나타나는 동시에 음성이 나오는 샷을 추출한 후 얼굴 인식 알고리즘과 음성 인식 알고리즘을 융합한 방법을 제안하였다^[3]. 이 방법은 얼굴 및 음성 인식 방법을 사용하고 있으므로, 특정한 앵커 모델이 반드시 필요하다. 따라서, 특정한 앵커에 대해서는 우수한 결과를 보이지만, 앵커가 주기적으로 교체된다거나 할 경우 매번 새로운 앵커 모델을 만들어야 하는 문제점이 있다. 또한, Lan은 뉴스 동영상에서 카메라 움직임, 프레임간의 차이의 평균 및 분산, 평균 프레임 휘도, 그리고 에지 변화율의 평균 및 분산과 같은 저 수준의 특징들로 이루어진 특징 벡터를 SVM(support vector machine)을 이용하여 후보 앵커 프레임 샷들을 추출 한다^[4]. 그런 다음, 시각적, 청각적 정보와 사람의 출현 정보를 이용하여 뉴스 동영상에서 주기적으로 나타나는 샷들을 앵커 프레임 샷, 회담, 기상예보, 그리고 광고와 같이 네 부분으로 클러스터링한 후 실제 앵커 프레임 샷들을 검출한다. 이 방법은 앵커 프레임 검출 뿐만이 아니라 뉴스 동영상을 네 부분으로 분할하여 뉴스의 구조를 잘 표현한다. 그러나, 이 방법은 우리나라 뉴스 프로그램의 형식과는 많은 차이점이 있어서 적용하기가 곤란하다. 또한, 뉴스 동영상을 네 부분으로 분할할 때 샷의 연속 구간(interval)이 짧은 경우는 앵커 샷이지만 검출하지 못하는 단점이 있다. 그리고, Gao는 장면 전환점을 이용하여 앵커 장면을 검출하는 방법을 제안하였다^[5]. 먼저 장면 전환점을 검출한 다음 그 전환점들의 장면들에 대해서 최소 신장 트리(MST : Minimum Spanning Tree)를 사용한 클러스터링 기법 수행하여 앵커 장면을 검출한다. 이 방법은 각각의 장면 전환 프레임을 서로 비교하여 유사도를 검사하는데, 이들 프레임들은 배경의 변화와 뉴스 캡션으로 인한 오류를 줄이기 위해서 각각 16개의 블록으로

분할한 후 같은 위치의 블록들 간의 유사도를 측정하여 가장 차이가 많이 나는 8개의 블록을 제외한 다음 나머지 8개의 블록들에 대해서만 유사도를 측정한다. 이 유사도는 특정 임계치의 조건에 따라 MST를 구성하고, 클러스터링된 결과에서 두 개 이상의 프레임이 존재하는 집합들에 대해서만 앵커 장면으로 검출한다. 그러나 이 방법은 조건에 의해서 앵커는 아니지만 앵커와 유사한 장면이 여러 번 나올 경우 오검출하는 문제점이 있다. 또한, 앵커가 중간에서 나타나는 장면처럼 단 한 번 나오는 장면은 검출 할 수 없다. 그러므로 본 논문에서는 앵커의 얼굴 정보와 배경 및 앵커 객체(상반신)의 색상 정보를 추출함으로써 앵커의 위치와 나타나는 횟수에 무관한 앵커 프레임 검출을 위한 방법을 제안한다.

본 논문의 구성은 II장에서 제안하는 알고리즘에 대한 전반적인 기술을 하고, III장에서는 실험에 대한 결과를 분석하고, 마지막 IV장에서는 결론 및 향후 과제에 대해서 기술한다.

II. 제안하는 방법

뉴스 비디오에서 앵커 프레임 추출을 통한 뉴스 시퀀스를 뉴스 기사 단위로 분할할 수 있는 방법을 제안한다. 그림 1은 제안된 알고리즘의 전반적인 흐름도를 보여주고 있다.

본 논문에서 제안된 알고리즘은 4단계로 구성되어 있다. 첫 번째 단계에서 누적 히스토그램 비교 방법을 사용하여 샷 경계를 검출한다. 이 방법은 뉴스 시퀀스를 샷들로 분할하기 위하여 사용된다. 두 번째 단계에서는 비디오 샷 내에서 피부색 컬러 정보를 이용하여 앵커 얼굴 정보를 추출하고, 세 번째 단계에서는 앵커 객체의 정보 추출하기 위해서 앵커의 상반신의 컬러 정보를 추출한 후 앵커 후보 장면들을 추출하고, 마지막 단계에서는 배경 컬러 정보들에 대해서 그래프 이론 기반의 클러스터 분석 방법인 MST 방법을 이용하여 앵커 장면과 비앵커 장면을 분류한다.

1. 장면 전환 검출

뉴스 비디오 분석을 통해서 앵커 장면은 장면 전환점에서 나타나고, 앵커 구간에서는 장면의 전환이 나타나지 않은 것을 발견할 수 있다. 그러므로 앵커 프레임을 검출하기 위해서, 장면 전환 검출 기법을 고려한다. 현재까지 장면 전환 검출 기법은 픽셀 밝기 비교법, 히스

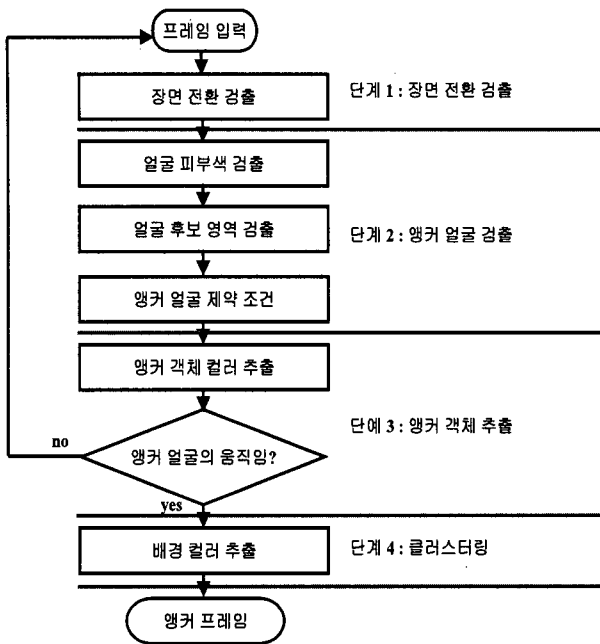


그림 1. 제안한 알고리즘의 흐름도
Fig. 1. Flowchart of the proposed algorithm.

토그램 비교법, 그리고 에지 비교법등이 연구되어 왔다.

본 논문에서는 이들 방법들 중에서 계산량과 정확성에서 비교적 좋은 성능을 보이는 누적 히스토그램 비교법에 의한 장면 전환 검출 기법을 사용한다^[6,7]. 장면 전환에는 단순 컷(abrupt shot change)과 점진 컷(gradual shot change)으로 분류할 수 있다. 본 논문에서 사용하고자 하는 누적 히스토그램 기법은 단순 컷에서는 매우 좋은 성능을 보이지만 점진 컷을 검출하기에는 어려운 점이 있다. 따라서 점진적으로 변화는 장면 전환에서는 컷을 미 검출 할 수 있다. 일반적으로 뉴스 비디오에서 앵커 장면에서는 점진 컷이 거의 나타나지 않지만, 좀더 정확하고 장면 전환 검출 시간을 줄이기 위해서 B 프레임과 P 프레임을 제외한 I 프레임만을 고려한다.

2. 얼굴 영역 검출

얼굴 영역 검출은 얼굴을 이용한 영상처리를 위한 기본적인 알고리즘으로 많은 연구가 되었다. 그러나 이들 중 대부분의 방법은 임의의 영상에서 얼굴을 찾아내기 위한 방법이기 때문에 정확한 얼굴을 찾기 위해 많은 계산량을 필요로 한다. 하지만, 뉴스 비디오에서 최대한 정확하고 가능한 한 적은 계산량으로 앵커 얼굴 영역을 검출할 수 있어야 한다. 뉴스 비디오 분석을 통해서 우리는 다음과 같은 얼굴 제약조건을 정의하여 얼굴 영역 검출 시 사용한다.

- 앵커 얼굴 영역의 크기는 일정하다.
- 앵커 얼굴 영역의 가로와 세로 크기의 비는 항상 일정하다.
- 앵커의 얼굴은 화면의 구석에 나타나지 않는다.
- 앵커 얼굴 영역의 움직임은 거의 없다.

최근, 비디오에서 얼굴 색상을 이용한 얼굴 검출 알고리즘들이 많이 연구되고 있으며, 좋은 결과를 보이고 있다^[8-10]. 본 논문에서도 기본적으로 얼굴 색상을 이용하여 얼굴 영역 검출 방법을 사용한다. 그림 2는 얼굴 영역 검출 알고리즘에 대한 전반적인 흐름도를 보여준다.

얼굴 색상은 YCbCr 컬러 모델의 Cb, Cr 부분에서 일정한 범위에 존재하는 특징을 가진다. 이를 이용하여 Cb와 Cr 화소 정보의 특정 범위만을 얻을 수 있는데, 본 논문에서는 많은 실험을 통하여 Cb는 [95,130], Cr은 [140,150]을 사용한다. 이들 두 조건을 만족하는 영역만을 추출하면 얼굴 생상 정보만을 포함하는 이진 영상을

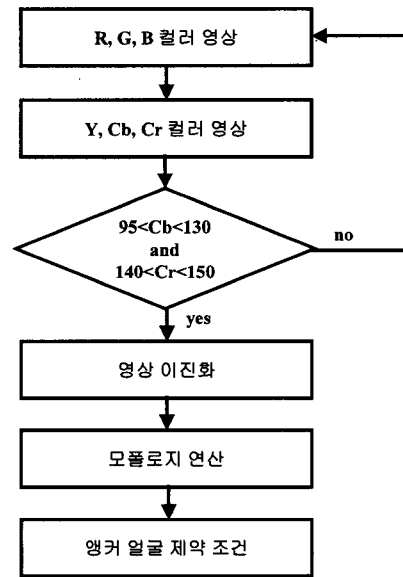


그림 2. 얼굴 영역 검출에 대한 흐름도
Fig. 2. Flowchart of face region detection.

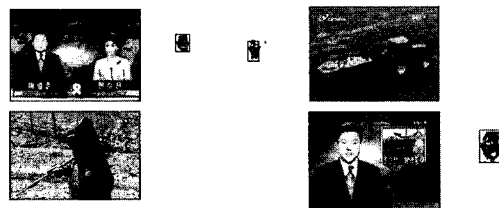


그림 3. 얼굴 영역 추출
Fig. 3. Face region extraction.

로 만들 수 있고, 이 영상에 대해 모폴로지 연산을 수행하여 잡음을 제거한다. 마지막으로 얼굴 제약 조건을 적용하여 실제 얼굴 영역을 검출한다. 얼굴 영역 검출에 사용되어진 임계치는 많은 실험에 의해서 결정되었다. 그림 3은 얼굴 영역을 검출한 결과를 보여준다.

3. 앵커 객체 추출

앵커 객체를 추출하기 위해서 앵커의 상반신 영역의 컬러 정보를 이용한다. 그림 4는 앵커 객체를 추출하기 위한 사용되는 템플릿 영상을 보여준다. 템플릿 영상은 앵커 얼굴의 제약 조건을 만족하는 첫 번째 장면으로 선택한다. 그림 5는 앵커 객체를 추출하기 위한 알고리즘을 보여준다.

일반적으로 뉴스 비디오에서 앵커들은 단조로운 색상의 옷을 입고 있는 것을 확인할 수 있다. 이것은 옷의 색상이 앵커 객체를 추출할 수 있는 색상 정보로 사용되어 질 수 있음을 의미한다. 우리는 그림 4에서 보는 것과 같이 앵커 객체의 색상을 추출하기 위해 이미 추출된 얼굴 영역의 크기에 비해 너비를 300%, 높이를 150% 확장한 앵커의 상반신을 객체로 정의한다.

객체의 색상 정보를 이용하기 위해서 앵커 객체에 대해서 컬러 히스토그램이 구성된다. 그러나 이렇게 구성



그림 4. 템플릿 영상
Fig. 4. Template image.

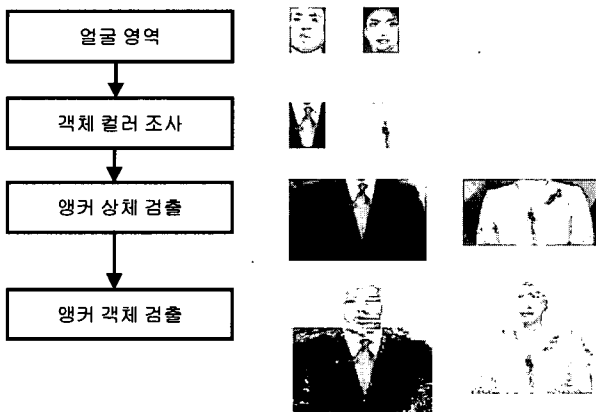


그림 5. 앵커 객체 추출 알고리즘
Fig. 5. Algorithm for extracting anchor object.

된 컬러 히스토그램은 객체의 크기에 따라 표현하는 정보가 달라지므로 객체의 크기에 강건한 방법으로 컬러 히스토그램을 객체의 크기로 나누어서 정규화 한다^[11].

$$H_{NRO(n)} = \frac{H_{RO(n)}}{S} \quad n = 1, 2, \dots, 128 \quad (1)$$

식 1에서, H_{NRO} 는 정규화된 히스토그램이고, H_{RO} 는 앵커 객체의 자체 히스토그램이다. S 는 앵커 객체의 크기를 의미하며, n 은 히스토그램 빈들의 개수이다.

식 1을 이용하여 얼굴 영역 추출 알고리즘을 이용해서 추출한 각 프레임에 대해서 앵커 객체를 추출한다. 그런 다음, 템플릿 영상에서 구성한 컬러 히스토그램과 얼굴 영역을 포함한 각 프레임에서 구성한 컬러 히스토그램 사이의 유사도를 측정하여 앵커 후보 프레임을 추출한다. 유사도 검사를 위해서 식 2에서처럼 히스토그램 인터섹션 기법을 사용한다.

$$HS(H_{NRO}, H_{NCO}) = \sum_{i=1}^n \min(H_{NRO(i)}, H_{NCO(i)}) \quad (2)$$

식 2에서, H_{NCO} 는 앵커 후보 프레임에서 추출된 객체 정규화된 컬러의 히스토그램이고, $HS(H_{NRO}, H_{NCO})$ 는 템플릿 영상과 앵커 후보 프레임간의 유사도를 나타낸다. n 은 컬러 히스토그램의 빈 개수를 의미한다. 이 유사도가 임계치보다 큰 값을 가질 때 앵커 후보 프레임으로 선언한다. 왜냐하면, 앵커 프레임들은 앵커 객체의 컬러가 변하지 않기 때문이다. 그러나 앵커 후보 프레임들 중에는 앵커 프레임이 아니면서 앵커 객체의 컬러와 유사한 프레임들을 많이 발견할 수 있다. 이러한 프레임들을 앵커 프레임으로 추출하지 않기 위해서 우리는 ‘앵커의 얼굴은 2초 이상 거의 변함이 없다’ 라는, 즉 움직임의 변화가 없는 정보를 이용한다. 이것은 식 3을 이용하여 계산한다. 하지만, 앵커 장면에서의 화면 전체의 프레임 차이 값을 계산하지 않고 앵커 얼굴 영역에 대해서만 프레임의 차이 값을 계산한다.

$$FRFD = \frac{\sum_{x=1}^M \sum_{y=1}^N FR(x, y, t-1) - FR(x, y, t)}{M * N * 255} \quad (3)$$

식 3에서, 얼굴 영역 프레임 차이값(face region frame difference : $FRFD$)은 얼굴 영역에 대한 프레임 차이 값을 의미하고, M, N 은 각각 얼굴 영역의 가로와



그림 6. 앵커 객체 추출 및 움직임 정보를 이용한 앵커 프레임 검출 결과

Fig. 6. Anchor frame detection using anchor object extraction and anchor's motion.

세로의 길이를 나타낸다. 그리고 $FR(x, y, t)$ 는 t 프레임에서의 얼굴 영역내 (x, y) 위치의 밝기 값을 의미한다. 만약 $FRFD$ 값이 임계치를 초과한다면, 해당 프레임을 비앵커 프레임으로 선언한다. 그림 6은 $FRFD$ 값을 이용한 결과를 보여준다.

그러나 그림 6에서 보는 것과 같이 비 앵커 프레임이 앵커 프레임으로 오검출 되는 문제를 해결하기 위해서 다음과 같은 클러스터링 기법을 이용한다.

4. 최소 신장 트리를 이용한 클러스터링

이전까지의 결과에서 오검출된 프레임을 제거하기 위해서 최소 신장 트리를 이용한 클러스터링 알고리즘을 수행한다^[5]. MST를 이용한 클러스터링을 하기 위해서 앵커 프레임들을 노드(node)로 정의하고, 배경 컬러 유사도 값을 노드들을 연결하는 에지(edge)의 가중치로 정의한다. 본 논문에서는 배경 컬러의 유사도를 측정하기 위해서 제한된 두 영역의 배경 컬러를 사용한다. 이 두 부분은 앵커 프레임에 대한 분석을 통해서 변화가 거의 없는 영역임을 확인할 수 있다. 한 영역은 얼굴 영역 주변의 배경 영역이고, 다른 한 영역은 프레임 영상의 중앙 상단 부분의 배경 영역이다. 그림 7과 그림 8은 에지의 가중치 계산에 사용될 두 영역을 보여준다.

식 4는 두 배경 컬러 영역을 사용하여 유사도를 측정함으로써 가중치를 계산한다.

$$SW(pre_B, cur_B) = \frac{1}{2}(W_{face} + W_{top})$$

$$W_{face} = \sum_{i=1}^n \min(H(pre_B_{face})_{(i)}, H(cur_B_{face})_{(i)})$$

$$W_{top} = \sum_{i=1}^n \min(H(pre_B_{top})_{(i)}, H(cur_B_{top})_{(i)})$$

(4)

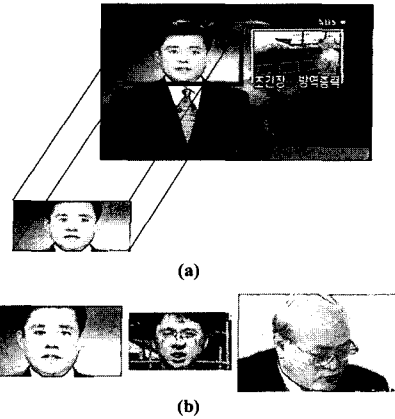


그림 7. 얼굴 주변의 배경 컬러 영역 (a) 확장된 얼굴 영역 (b) 앵커 얼굴 후보

Fig. 7. Background color region around a face. (a) Expanded face region (b) Candidate anchor face



그림 8. 중앙 상단 부분의 배경 컬러 영역
Fig. 8. Background color region of upper-center area.

식 4에서, SW 는 두 노드들 사이의 유사도 가중치를 나타낸다. pre_B 는 현재의 이전에 해당하는 프레임들을 나타내고, cur_B 는 현재 프레임을 의미한다. $H(pre_B_{face})$ 는 얼굴 주변 영역의 정규화된 컬러 히스토그램을 나타내고, $H(pre_B_{top})$ 는 중앙 상단 영역의 정규화된 컬러 히스토그램을 의미한다. n 은 컬러 히스토그램의 빈 개수를 의미한다.

프레임을 나타내는 노드와 노드들을 연결하는 에지의 가중치를 배경 컬러 영역들을 이용하여 구한 후에 최소 신장 트리(MST)를 이용하여 클러스터링을 수행함으로써 실제 앵커 프레임을 검출한다. 그림 9는 본 논문에서 사용한 MST를 이용한 클러스터링 알고리즘을 나타내고 있다.

그림 9에서 단계 1은 입력 노드(프레임) 1을 루트(root) 노드로 트리를 생성한다. 단계 2에서는 입력 노드 2와 루트 노드 1에 대해 두 노드 사이의 유사도 비교를 하여 실험에 의한 임계치 T 와 비교해서 크면 노드 2를 루트(root) 노드 1의 자식 노드로 연결하고 임계치 T 보다 작으면 입력 노드 2를 루트(root) 노드로 하는 다른 트리를

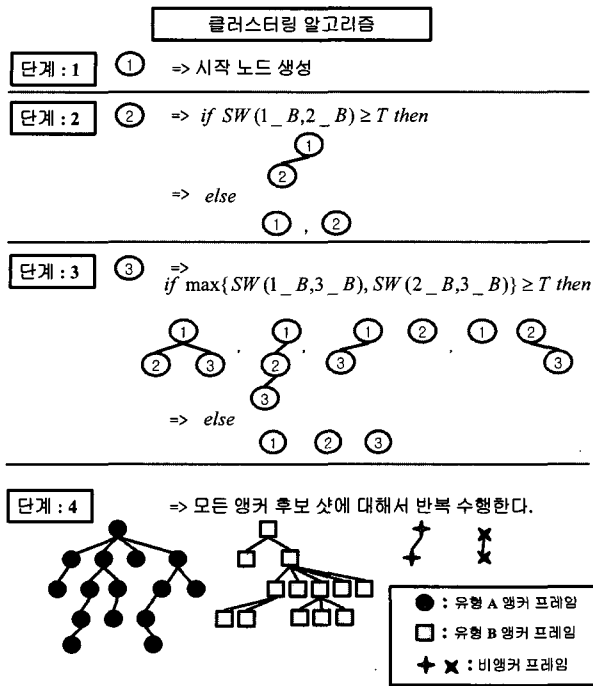


그림 9. 최소 신장 트리를 이용한 클러스터링 방법
Fig. 9. Clustering algorithm using MST.

생성한다. 이와 같이 입력으로 들어오는 모든 노드에 대해서 그림 9의 알고리즘을 반복적으로 수행함으로써 유사한 배경 컬러를 갖는 장면들의 그룹으로 클러스터링 할 수 있다.

III. 실험 결과

본 논문에서 사용된 뉴스 비디오는 점진적인 샷을 포함하거나 다양한 배경을 가진 5개의 비디오 클립이다. 그림 10은 앵커 프레임을 검출한 결과를 보여준다. 기존의 Gao 방법에서는 앵커가 가운데 나오는 장면이 뉴스 비디오에서 한번 나오는 경우는 미검출하는 문제점이 있지만 본 논문에서 제안한 방법에서는 앵커 객체 정보와 배경 컬러 정보를 사용함으로써 앵커가 나타나는 위치와 횟수에 무관하게 검출하는 것을 확인할 수 있다.

배경 컬러를 이용하는 방법은 오검출이 다른 방법에 비해 적지만, 서로 다른 두 개 이상의 배경을 사용하는 뉴스 비디오에서는 미검출을 할 수 있다. 그러므로 템플릿을 이용한 방법과 배경 컬러를 이용한 방법에서 효과적인 앵커 프레임 검출이 되지 않는다. 하지만, 제안한 방법은 앵커의 객체 컬러 정보를 이용하여 후보 장면을 검출한 후 배경 컬러 정보를 사용하여 클러스터링을 수행함으로써 만족할만한 성능을 보였다. 그림 11은



그림 10. 앵커 프레임의 검출 결과
Fig. 10. Result of anchor frame extraction.



그림 11. 다른 배경에서의 앵커 프레임 검출 결과
Fig. 11. Result of anchor frame detection with different backgrounds.

배경이 변화하는 뉴스 비디오에서의 앵커 프레임 검출 결과이다.

정확한 실험 결과를 평가하기 위해서는 기존 방법들과 직접적인 비교가 되어야 하지만, 서로 다른 특정한 뉴스 프로그램의 형식에 의거한 방법들이 사용되고 있고, 또한 각각의 방법들에서 사용되고 있는 파라미터 값들이 실험결과에 결정적인 영향을 미치고 있어서 직접적인 실험비교가 어렵다. 그래서, 본 논문에서는 제안하는 방법의 성능을 평가하기 위해서, 정확율(precision)과 재현율(recall)을 사용한다. 정확율은 전체 검출된 프레임에 대해서 실제 앵커 프레임의 비율로 정의한다. 그리고 재현율은 뉴스 비디오에서의 전체 앵커 프레임에 대해서 실제 앵커 프레임의 비율로 정의한다. 정확율과 재현율에 대한 수식은 식 5와 식 6과 같다.

$$precision = \frac{n_c}{n_c + n_f} * 100 \tag{5}$$

$$recall = \frac{n_c}{n_c + n_m} * 100 \tag{6}$$

n_c 는 정확한 앵커 프레임을 검출한 수이고, n_f 는 오검출한 수이며, n_m 은 미검출한 수이다. $n_c + n_f$ 는 모든 검출

표 1. 앵커 장면 검출 성능 결과

Table 1. Performance results of anchor frame detection.

뉴스 비디오 클립	실제 앵커 프레임	추출된 앵커 프레임	오검출	미검출
뉴스 비디오 1	22	21	0	1
뉴스 비디오 2	17	16	0	1
뉴스 비디오 3	10	10	0	0
뉴스 비디오 4	33	33	1	0
뉴스 비디오 5	33	28	0	5

표 2. 정확율과 재현율

Table 2. Precision and recall.

	정확율	재현율
평균	99.4 %	94.9 %



그림 12. 오검출이 발생한 앵커 장면 검출 결과

Fig. 12. Result of anchor frame detection with fault detection.

된 프레임의 수이고, n_c+n_m 은 뉴스 비디오에 있는 전체 앵커 프레임의 수이다.

제안하는 방법의 성능 결과는 표 1과 표 2에 요약하였다.

표 1에서 보는 바와 같이 비디오 1과 비디오 2에서 앵커 프레임을 미검출 하였다. 이것은 앵커가 나타나는 장면에서 점진적인 장면 전환으로 인하여 장면 전환 검출 알고리즘의 오류로 발생한 것이다. 비디오 4에서는 앵커 프레임과 유사한 특징을 갖고 있는 프레임으로 인해 오검출하였다. 즉, 객체 컬러와 배경 컬러 그리고 움직임 정보의 모든 조건이 앵커 프레임과 유사하였다.

그림 12는 앵커 프레임과 유사한 특징을 갖고 있는 프레임을 오검출한 결과를 나타낸다. 비디오 5에서는 상대적으로 많은 미검출 결과를 보이고 있는데, 왜냐하면 뉴스 비디오가 월드컵 축구 관련 뉴스에서 야외 배경을 뉴스의 배경으로 사용함으로써 모든 배경의 앵커 프레임을 검출할 수 없었다.

표 2에서는 제안한 방법이 정확율과 재현율에서 우수한 성능 결과를 보여주고 있다.

IV. 결론 및 향후 과제

본 논문에서는 앵커 객체의 컬러 정보와 배경 컬러 정보에 대한 클러스터링 기법을 사용하여 앵커 프레임을 검출하는 방법을 제안하였다. 실험 결과로 제안한 방법이 기존의 방법들에 비해 앵커가 나타나는 위치와 횡수에 무관한 검출과 다양한 배경에서도 만족할만한 결과를 얻을 수 있었다. 또한 앵커 프레임을 검출함으로써 뉴스 비디오를 앵커 구간과 보도 구간을 하나의 사건 단위로 구조화 할 수 있다.

향후 연구과제로 본 실험에서는 실행 시간을 고려하지 않고 정확성에 중점을 두었기 때문에 실행 시간을 줄이기 위해서 압축 영역에서 접근하는 방법을 고려해야한다. 또한 다양하게 변화는 배경에서도 앵커 프레임 검출의 성능을 향상 시킬 수 있는 방법에 대한 연구가 이루어져야 할 것이다.

참고 문헌

- [1] A Hanjalic, R. L. Lagendijk, and J. Biemond, "Template-based detection of anchorperson shots in news programs," IEEE Image Processing, ICIP Vol. 3, pp. 148-152, 1998.
- [2] W. Wang and W. Gao, "A fast anchor shot detection algorithm on compressed video," Advanced in Multimedia Information Processing - PCM, pp. 873-878, October 2001.
- [3] A. Albiol, L. Torres, and E. J. Delp, "The indexing of persons in news sequences using audio-visual data," in Proc. of IEEE ICASSP 2003, vol. 3, pp. 137-140, April 2003.
- [4] D. J. Lan, Y. F. Ma, and H. J. Zhang, "Multi-level anchorperson detection using multimodal association," in Proc. of IEEE ICIP 2004, vol. 3, pp. 890-893, August 2004.
- [5] X. Gao and X. Tang, "Unsupervised video shot segmentation and model free anchorperson detection for news video story parsing," IEEE Trans. Circuits and Systems for Video Tech., pp. 765-776, September 2002.
- [6] 황두선, 이종철, 조위덕, 문영식, "누적히스토그램과 에지 정보를 이용한 장면 전환 검출," 대한전자공학회 하계종합학술대회 논문집 1호, pp. 211-214, 2002.
- [7] K. T. Park, D. S. Hwang and Y. S. Moon, "Anchor frame detection in news video using anchor object extraction," IEICE Tans. Fundametal, vol. E88-A, pp. 1525-1528, June

2005.

- [8] H. Ishii, M. Fukumi, and Akamatsu, "Face detection based on skin color information in visual scenes by neural networks," in Proc. of IEEE SMC '99 Conf. on Systems, Man, and Cybernetics, 1999.
- [9] J. Brand, J. S. Mason, M. Roach, and M. Pawlewski, "Enhancing face detection in colour images using a skin probability map," in Proc. of 2001 International Symposium Conf. on Intelligent Multimedia, Video and Speech Processing, pp. 344-347, 2001.
- [10] K. K. Sung and T. Poggio, "Example based learning for view based human face detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 39-51, January 1998.
- [11] J. B. Oh and Y. S. Moon, "Content based image retrieval based on scale space theory," IEICE Tans. Fundametal, June 1999.

 저 자 소 개



박 기 태(학생회원)

2000년 한양대학교 전자계산학과
학사 졸업.

2002년 한양대학교 컴퓨터공학과
석사 졸업.

2002년~현재 한양대학교 컴퓨터
공학과 박사 과정.

<주관심분야 : 비디오 색인 및 검색, 내용기반 영
상검색, 영상처리>



문 영 식(평생회원)-교신저자

1980년 서울대학교 전자공학과
학사 졸업.

1982년 한국과학기술원 전기 및
전자공학과 석사 졸업.

1990년 University of California
at Irvine / Dept. of
Electrical and Computer
Engineering 박사 졸업.

1982년~1985년 한국통신연구소 연구원

1990년~1992년 생산기술연구원 선임 연구원

1992년~현재 한양대학교 컴퓨터공학과 교수

<주관심분야 : 영상처리, 패턴인식, 비디오 색인
및 검색, 내용기반 영상검색>



황 두 선(정회원)

2001년 한양대학교 전자계산학과
학사 졸업.

2003년 한양대학교 컴퓨터공학과
석사 졸업.

2003년~현재 삼성 종합 기술원
연구원

<주관심분야 : 비디오 색인 및 검색, 영상처리>