

규칙기반 뉴스 비디오 앵커 TIT 검출방법: KBS와 MBC 9시 뉴스를 중심으로

유현우^{1*}, 이명의²

Rule-Based Anchor Shot Detection Method in News Video: KBS and MBC 9 Hour News Cases

Hun-Woo Yoo^{1*} and Myung-Eui Lee²

요약 본 논문은 뉴스 비디오를 색인, 검색하기 위한 관리 시스템의 기본 기술인 앵커 샷을 검출하는 방법을 제안한다. 이를 위해 현재 가장 많은 사람이 시청하는 'KBS 9시 뉴스'와 'MBC 9시 뉴스'의 앵커 샷의 특징 요소를 분석하여 4단계의 규칙기반 검출방법을 제안한다. 먼저 전처리로 비디오의 샷 경계를 검출하고 첫 번째 프레임에 키 프레임으로 선택한 후에 다음의 4가지 조건을 모두 만족하면 해당 샷을 앵커 샷으로 판단한다. 1) 키 프레임에 앵커의 얼굴이 존재하는가의 여부, 2) 에지의 분포가 구조적으로 적합한지의 여부, 3) 배경의 색상 정보를 추출하여 기존의 앵커 모델의 색상과 유사한지의 여부, 마지막으로 4) 샷 내의 움직임 비율이 일정 임계치 이하인지의 여부를 판단한다. 제안된 방법의 성능을 보이기 위해, 총 108분 분량의 서로 다른 날에 저장된 KBS와 MBC의 9시 뉴스 비디오에 대해 실험한 결과 평균적으로 0.97의 정확도와 1.0의 회수율, 0.98의 F-값을 얻을 수 있었다.

Abstract In this paper, an anchor shot detection method, which is a basic technology for managing news videos for index and retrieval purposes is proposed. To do that, two most popular news program such as 'KBS 9 Hour News' and 'MBC 9 Hour News' are analyzed and 4-step rule based detection method is proposed. First, in the preprocessing, video shot boundaries are detected and the 1st frame of each shot is extracted as a key frame. Then, the detected shot is declared as an anchor shot, if all the following 4 conditions are satisfied. 1) There is an anchor face in the key frame of a shot. 2) Spatial distribution of edges in the key frame is adequate. 3) Background color information of the key frame is similar to the color information of an anchor model. 4) Motion rate in the shot is low. In order to show the validity of the proposed method, three 'KBS 9 Hour News' and three 'MBC 9 Hour News', which have total running time of 108 in minute and are broadcasted at different days, are used for experiments. Average detection rates showed 0.97 in precision, 1.0 in recall, and 0.98 in F-measure.

Key Words : Anchor Shot Detection, Face Detection, Edge Detection, Anchor Model, Motion Information

1. 서론

최근에 컴퓨터와 멀티미디어 기술의 눈부신 발전으로 방대한 양의 비디오 소스들이 생성되고 있으며, 이들 소스들의 효과적인 색인과 검색을 위한 기술이 필요하게 되었다. 특히 뉴스 검색과 같은 서비스는 최근에 많은 연

구자들의 관심을 받고 있는데, 이는 수요도 많을 뿐더러 현재와 과거의 뉴스를 시간에 구애 받지 않고 빠른 시간 내에 쉽고 편리하게 청취할 수 있다는 장점이 있기 때문이다. 이러한 목적을 달성하기 위해서는 기존 비디오를 이야기 중심으로 분할할 수 있어야 한다. 구체적으로 비디오를 카메라 시작과 끝마침 같은 카메라 동작 변화에 따른 샷으로 구분하고, 샷 들의 내용에 따라 앵커 샷과 비-앵커 샷으로 구분해야 한다. 왜냐하면 뉴스 비디오는 각 사건마다 앵커가 등장하고, 사건을 설명하는 화면과

¹연세대학교 컴퓨터과학과

²한국기술교육대학교 정보기술공학부

*교신저자: 이명의(melee@kut.ac.kr)

기자의 보도가 있는 후 다시 다음 사건을 설명하는 앵커가 등장하는 방식으로 구성되어 있기 때문이다. 따라서 앵커 샷을 검출하는 것은 이야기 중심으로 뉴스 비디오를 색인하고 검색하기 위한 가장 핵심적인 기술이다. 본 논문에서는 이러한 앵커 샷을 검출하는 새로운 알고리즘을 제안한다.

앵커 샷을 검출하는 기존 방법으로는 템플릿을 이용하는 방법이 있다 [1, 2, 3]. Günsel 과 Ferman [1]은 정규 템플릿 매칭 방법을 이용하여 앵커 샷을 검출하였는데 뉴스 스튜디오의 조명 상태와 앵커 피부색에 따라 매우 민감한 차이가 발생한다는 단점이 있다. Ma [2]등은 일반화된 휴 변환 (Hough Transform)에 의해 앵커 영상에서 에지 곡선들을 추출하여 템플릿으로 이용하였고, Xu [3]등은 앵커 화면의 배경이 일정하다는 가정 하에서 템플릿을 이용하였다. 그러나 위의 방법들은 모두 계산량이 많다는 단점이 있다.

다른 한편으로 단순히 화면의 특성만을 가지고 앵커 샷을 찾아내는 방법이 아닌 샷 내의 움직임 정보를 이용하는 방법들이 연구되고 있다. 그 중에서 Luo와 Huang [4]은 앵커 화면은 얼굴이 움직임이 없다는 점을 이용하여 앵커라고 추정되는 곳에 움직임 벡터를 얻어내어 움직임이 거의 없으면 앵커 샷으로 설정하는 방법을 제안하였다. 그러나 이러한 방법은 앵커 이외에도 움직임이 없는 영상이 존재 할 수 있어 잘 못 인식할 가능성을 내포하고 있다.

또한 비슷한 샷들을 군집화 하여 일정한 규칙에 의해 앵커 샷으로 이루어진 군집을 추출하는 방법도 이용되고 있다 [5, 6, 7, 8, 9, 10]. Luan [5] 등은 HSV 색상 모델에서 유사한 색상들로 샷 들을 군집화하고 군집내의 샷들의 평균 지속시간, 샷의 개수, 샷들 간의 평균거리를 이용하여 앵커 샷을 판단하였다. Gao [6, 7] 등은 위와 유사한 특징을 그래프 이론에 근거한 군집 (GTC) 분석 기법을 이용하여 앵커 샷을 검출하였고, Santo [8] 등도 역시 그래프 이론에 근거하여 군집의 시간적 거리를 나타내는 'lifetime'이라는 개념을 사용하였는데, Gao와 Tang의 방법보다 보다 향상된 성능을 나타낸다고 보고하였다. Chaisorn [9] 등은 보다 광범위한 접근을 시도하였는데, 일단 샷 경계가 검출되면 결정 트리를 이용하여 앵커 샷 외에 뉴스도입부의 하이라이트, 기자의 인터뷰 장면, 경제뉴스, 상업광고, 일기예보, 스포츠 등의 미리 정한 13개

의 군집군 중 하나로 할당 하고 HMM (Hidden Markov Model) 을 이용하여 앵커 샷을 검출하였다. Zhai [10] 등도 앵커 샷과 일기예보, 스포츠 등을 분할하는 방법을 사용하였는데, 이들은 샷 연결 그래프 (SCG)를 이용하여 비디오를 표현하고 그래프 내에 있는 사이클을 찾아내는 방법을 이용하였다. 그러나 위에서 열거한 방법들은 기본적으로 유사한 특징을 가지는 군집군으로 분류하는 과정에서 유사한 특징을 가지는 비-앵커 샷이 존재하면 이를 앵커 샷으로 판단할 수 있는 단점이 있다. 그 외에 앵커 샷 검출 방법으로 Qi [11] 등이 제안한 방법과, Bertini와 Bimbo [12] 등이 제안한 방법이 있다.

2. 앵커 샷 검출

현재 국내에서 가장 많은 사람들이 시청하는 뉴스 프로그램으로는 저녁 9시 황금 시간대의 'KBS 9시 뉴스'와 'MBC 9시 뉴스'가 있다. 따라서 본 논문에서는 이 2개의 뉴스 프로그램의 앵커 샷을 검출하는데 초점을 맞추고 있다.

KBS와 MBC 9시 뉴스 비디오의 앵커 샷은 몇 가지 공통적인 특징을 가지고 있다. 그림 1과 그림 2에서 보듯이 앵커 화면은 매우 구조적으로 구성되어 있다. 1) 화면 좌상부 영역에 일정한 크기로 앵커 얼굴이 위치한다는 것이고, 2) 화면 우상부 영역에는 현재 뉴스를 축약적으로 설명하는 그래픽이 존재하며, 3) 그래픽 바로 아래에는 뉴스와 관련된 텍스트가 표시된다는 것이다. 또한, 4) 일반적으로 뉴스가 진행되는 스튜디오는 변하지 않으므로 앵커의 배경은 특별한 경우 외에는 항상 일정하다는 것이다. 마지막으로 5) 앵커 샷은 앵커들이 고정된 상태로 뉴스를 진행하기 때문에 샷 내의 움직임 정보가 매우 적다는 것이다. 따라서 본 논문에서는 위의 공통점들을 앵커 샷과 비-앵커 샷을 구별하는 특징으로 삼아 검출하고자 한다.

먼저 전처리 단계로서 샷 경계 검출을 수행하고, 샷의 처음 프레임을 키 프레임으로 선택한다. 샷 경계 검출은 샷의 명도 히스토그램과 에지 히스토그램의 상관 계수를 이용하여 상관 관계가 일정 임계치를 넘으면 새로운 샷으로 구별한다[13]. 위의 전처리 단계를 끝내면, 선택된 키 프레임 (즉, 앵커 화면)에 대해 앵커가 존재하는지를 얼굴 검출로 확인하고, 얼굴이 존재치 않으면 비-앵커 샷으로 구별한다. 얼굴이 존재하는 경우 2번에 걸친 에지



그림 1. 앵커 화면 (KBS 9시 저녁 뉴스와 MBC 9시 저녁 뉴스)

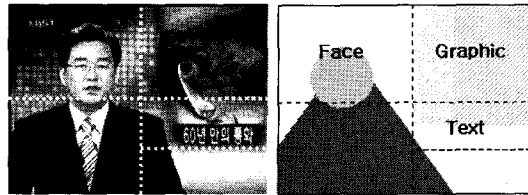


그림 2. 앵커 화면과 구조적 특징

개수의 비교 (화면의 좌면의 영역과 우면의 영역의 에지 개수의 크기 비교와 우-하부 영역의 상부 (텍스트 위치한 곳)에 위치하는 에지 개수와 우-하부 영역의 하부에 존재하는 에지 개수의 크기비교)를 통해 비-앵커 샷을 걸러낸다. 다음에는 앵커 얼굴이 존재하는 좌상부 영역의 색상을 모델로 저장하고 이 모델과 현재 샷의 배경이 유사하지 않을 경우 비-앵커 샷으로 걸러진다. 만약 위의 3가지 조건을 모두 만족한 경우에는 마지막으로 샷 내의 움직임 정보를 추출하여 일정 임계치 이상이면 앵커 샷이 아닌 것으로 간주하고, 이하이면 최종적으로 앵커-샷으로 간주한다. 위의 과정은 그림 3에 요약되어 있다.

분석하여 쉽고 간단하게 앵커 샷을 검출하였다는 점과, 기존의 템플릿 매칭 방법과는 달리 앵커의 색상 모델로서 앵커 화면에서 비교적 변화가 없는 좌상부 영역만을 사용하였다는 점이다.

조건1: 얼굴 검출

앵커 화면은 그림 1과같이 필연적으로 앵커가 등장하므로 얼굴의 존재여부를 검출하여 앵커 샷 인지를 확인할 수 있다. 또한 KBS와 MBC 9시 뉴스는 그림2에서와 같이 화면의 4분면에서 좌상부 영역에만 얼굴이 존재한다. 따라서 본 연구에서는 얼굴탐색영역을 좌상부 영역으로만 국한함으로써 탐색시간을 줄이도록 하였다. 이 방법은 일반적으로 중앙에 얼굴이 위치하는 인터뷰 등의 샷 과도 구별이 가능하므로 잘못된 얼굴 검출 (false detection)을 피할 수 있다. 얼굴 검출 알고리즘으로는 잘 알려진 Rowley [14] 등의 방법을 이용하였다.

조건2: 에지 분포

그림2와 같이 앵커 화면은 좌상부 영역에 앵커의 얼굴이 존재하고, 우상부 영역에 현재의 사건과 관련된 그래픽이, 우-하부 영역의 상부에는 연관된 텍스트를 전방위에 표시함으로써 윤곽이 뚜렷한 구조적 특징이 있다. 기타 영역은 배경으로 후방에 위치하고 있으므로 윤곽이 뚜렷하지 않다. 이를 근거로 앵커 샷의 키 프레임에 대해 에지 검출을 수행하면 프레임의 좌면 영역의 에지 개수와 우면 영역의 에지 개수가 명확한 차이를 보인다. 일반

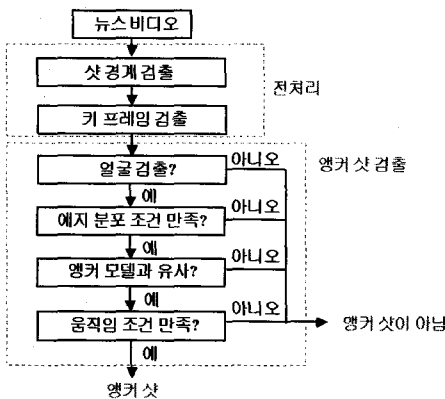


그림 3. 앵커 샷 검출 순서

제안한 방법이 기존의 방법과 다른 점은 'KBS'와 'MBC' 9시 저녁 뉴스의 앵커 화면에 대해 에지 분포의 구조를

적으로 그림 4와 같이 좌면 영역은 주로 앵커의 윤곽만이 보이고 우면 영역은 복잡한 그래픽과 텍스트에 해당하는 윤곽이 보이므로 우면 영역 에지의 개수가 보다 많음을 알 수 있다. 에지 검출로는 잡음의 영향을 줄이기 위해 캐니 에지 연산자 [15]를 이용하였다. 그림 5는 KBS와 MBC 9시 저녁 뉴스의 앵커 샷의 키 프레임에 대한 좌우면 영역의 에지 개수를 비율로 계산하여 비교한 것이다. 그림에서 보듯이 명확하게 우면 영역 에지의 개수가 많음을 알 수 있다. 따라서 본 논문에서는 허용 오차를 고려하여 아래의 식 (1)을 만족하면 앵커 샷에 대한 1차적인 에지 분포 조건을 통과하는 것으로 간주한다. 실험에서는 허용 오차를 0.05로 하였다.

$$ER_{LEFT} : (ER_{RIGHT} - \alpha) \quad (1)$$

여기서, ER_{LEFT} 와 ER_{RIGHT} 는 각각 전체 에지에 대한 좌면 영역과 우면 영역의 에지 개수의 비율을 나타내고, α 는 허용 오차를 나타낸다.

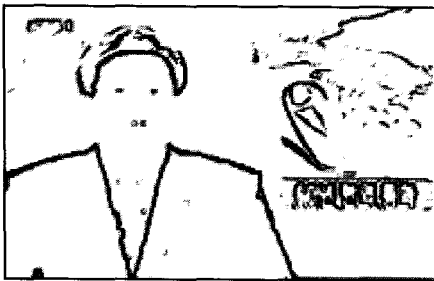
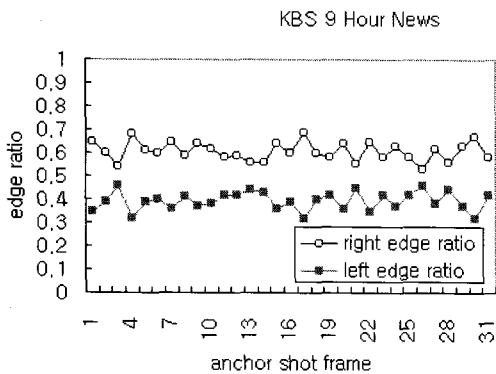
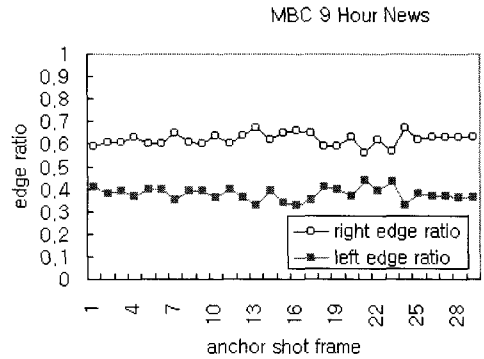


그림 4. 앵커 화면의 캐니 에지 검출 결과



(a)



(b)

그림 5. 앵커 화면에서 좌면 영역과 우면 영역의 에지 개수 비교 (a) KBS 9시 저녁 뉴스 (b) MBC 9시 저녁 뉴스

또한 주목해야 할 점은 우하부 영역에 텍스트가 존재한다는 것이다. 따라서 해당 위치에 텍스트의 존재 여부가 앵커 샷 검출의 중요한 열쇠가 될 수 있다. 그러나 이 방법은 기존의 텍스트 검출 알고리즘의 안정성 문제와 텍스트 검출에 필요한 계산량이 증가한다는 단점이 있다. 그리고 그림 6과 같이 텍스트가 우하부 영역이 아닌 우상부 영역에 위치하는 경우도 있을 뿐만 아니라, 텍스트 자체가 뒤틀려 표시되는 경우도 있기 때문에 검출에 실패할 가능성이 아주 많다. 따라서 본 연구에서는 차선으로 우하부 영역을 상하로 둘로 분할 한 후 일반적으로 텍스트가 존재하는 상부의 에지 개수와 텍스트가 존재하지 않는 하부의 에지 개수를 비교하여 상부의 에지 개수가 하부에 비해 얼마나 많은지를 앵커 샷 검출의 조건으로 이용하였다. 그림 7은 우하부 영역의 상부와 하부의 에지 개수를 비율로 비교한 것이다. 그림 7(a)의 'KBS 뉴스'에서는 모든 프레임에서 상부의 에지 개수와 하부의 에지 개수가 명확한 차이를 보인다. 반면에 상대적으로 그림 7(b)의 'MBC 뉴스'에서는 일부 프레임(1번, 6번, 13번)에서 차이가 매우 미미함을 볼 수 있다. 본 논문에서는 허용 오차를 고려하여 아래의 식 (2)를 만족하면 앵커 샷에 대한 2차적인 에지 분포 조건을 통과하는 것으로 간주한다. 1차 조건과 마찬가지로 실험에서는 허용 오차를 0.05로 하였다.

$$ER_{BRBottom} : (ER_{BRTop} - \beta) \quad (2)$$

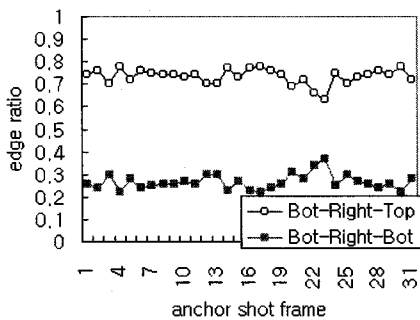
여기서, $ER_{BRBottom}$ 와 ER_{BRTop} 은 우하부 영역의 에

지에 대한 우-하부 영역의 하부와 우-하부 영역의 상부의 에지 개수의 비율을 나타내고, β 는 허용 오차를 나타낸다.



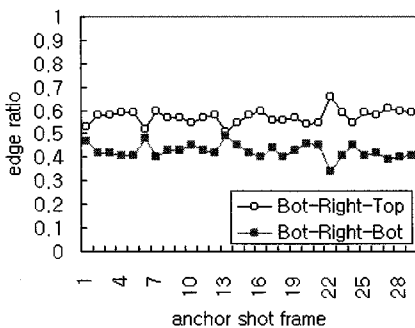
그림 6. 텍스트가 우-상부 영역의 그래픽 내에 뉘어진 상태로 위치한 앵커 화면

KBS 9 Hour News



(a)

MBC 9 Hour News



(b)

그림 7. 앵커 화면에서 우-하부 영역의 상부와 하부의 에지 개수를 비교한 것(a) KBS 9시 저녁 뉴스 (b) MBC 9시 저녁 뉴스

조건3: 앵커 모델과 유사

경우에 따라서는 얼굴 검출과 에지 분포 조건을 만족한다고 해도 앵커 샷에 해당되지 않는 경우가 있다. 한

예로, 그림 8에 나타난 프레임은 앵커 샷이 아님에도 불구하고 위의 2조건을 모두 만족한다. 따라서 보다 정확한 검출을 위해 초기에 미리 앵커 샷의 색상 히스토그램을 모델로 저장하고, 얼굴 검출과 에지 분포 조건을 모두 통과한 샷의 경우 이 모델과 비교하여 유사하지 않으면 앵커 샷이 아닌 것으로 처리한다.

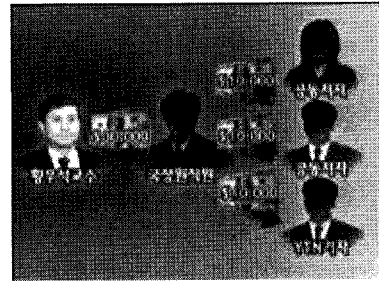


그림 8. 앵커 화면이 아니지만 얼굴 검출과 에지 분포 조건을 만족하는 것

이 때 저장할 모델의 색상 히스토그램은 화면 전체의 정보를 이용하는 것보다 비교적 변화가 없는 영역의 정보를 이용하는 것이 좋다. Zhai [10] 등은 그림 9와 같이 얼굴과 착용하고있는 옷의 일부를 몸체 영역 (Body Region)으로 정의하여 이 영역의 색상을 모델로 저장하여 비교하였다. 그러나 이 방법은 앵커가 매일 유사한 색상의 옷을 입고 진행한다고 볼 수 없으므로 효과적이지 못하다.



그림 9. 몸체 영역 (Body Region)

그림 2를 보면 좌-상부의 영역은 앵커의 얼굴과 배경 정보만 포함되기 때문에 비교적 색상의 변화가 없는 반면, 우-상부의 영역은 뉴스 사건마다 다른 그래픽이 표시

된다. 좌·하부 영역의 경우도 앵커가 매일 동일한 옷을 착용하고 뉴스 진행을 하지 않기 때문에 부적절하며, 우·하부 영역의 경우도 사건마다 상이한 텍스트가 표시되기 때문에 이용할 수 없다. 따라서 본 논문에서는 얼굴과 배경을 포함하는 좌·상부 영역만의 색상 히스토그램 정보를 모델로 저장한다. 색상 히스토그램으로는 조명 변화에 민감하지 않도록 RGB보다는 HSI색상 정보를 이용하였다. 이를 위해 19차원의 모델 색상 히스토그램, $h_{model}[i], i = 0, 1, 2, \dots, 18$ 을 각 빈(Bin)마다 0값으로 초기화 한 후, 모델의 좌·상부 영역에 존재하는 화소의 RGB 값을 HSI값으로 변환하고, 변환한 H (색상)과 S (채도)를 그림 10과 같이 H는 30도 간격으로, S는 3개로 양자화한 19개의 영역 중에 어디에 해당하는지를 찾은 후, 연관된 19차원의 모델 색상 히스토그램의 빈 값을 1씩 증가시켰다. 위의 방법을 좌·상부 영역의 모든 화소에 적용하여 각 빈에 해당 색상에 해당하는 영역내의 화소의 개수가 저장되도록 하였다. 마지막으로 빈별 개수를 영역의 크기로 나누어 각 빈이 영역에서 차지하는 비율 값 $[0, 1]$ 을 가지도록 정규화 시켰다 ($H_{model}[i]$).

$$H_{model}[i] = \frac{h_{model}[i]}{W_{TL} \times H_{TL}}, \quad \text{for } i = 0, 1, \dots, 18 \quad (3)$$

여기서, $H_{model}[i]$ 는 모델 색상 히스토그램의 i 번째 빈의 값이고, W_{TL} 와 H_{TL} 은 각각 프레임의 좌·상부 영역의 가로크기와 세로 크기이다.

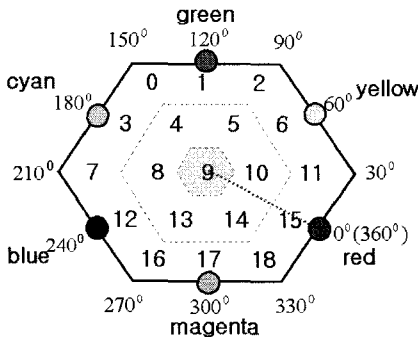


그림 10. HSI 색상 영역 (육각형내의 변호는 모델 색상 히스토그램의 빈에 해당)

개별 방송국의 뉴스에서는 남자와 여자 앵커가 존재하

므로 KBS 뉴스에서 남/여 앵커 샷의 좌·상부 영역 2개와 MBC 뉴스에서 남/여 앵커 샷의 좌·상부 영역 2개, 즉 총 4개의 색상 히스토그램 모델을 저장한다. 따라서 얼굴 검출과 에지 분포 조건을 모두 통과한 샷에 대해서 좌·상부 영역의 색상 히스토그램을 추출하여 정규화하고 ($H_{compare}[i]$) 이를 모델의 색상 히스토그램과 비교하여 식 (4)와 같이 4개의 모델 중에 하나라도 각 빈간 거리 차이의 합이 일정 임계치 t_{model} 이하가 되면 앵커 모델과 유사하다고 간주한다.

$$\sum_{i=0}^{18} |H_{model}[i] - H_{compare}[i]| \leq t_{model} \quad (4)$$

여기서, $H_{model}[i]$ 은 모델의 색상 히스토그램의 i 번째 빈을 나타내고, $H_{compare}[i]$ 는 얼굴 검출과 에지 분포 조건을 통과한 비교 대상 샷의 좌·상부 영역의 색상 히스토그램의 i 번째 빈을 나타낸다.

물론 방송사에 따라 일정 시점이 끝나면 뉴스 스튜디오의 배경을 바꾸거나, 특별 방송의 경우 스튜디오 외부에서 진행할 경우가 있다. 따라서 기존과 다른 모델 히스토그램이 발생할 수도 있는데, 그러나 이런 경우에도 현실점에서 모델을 다시 정의하여 사용하면 된다.

조건4: 작은 움직임 값

대부분의 앵커 샷 들은 정적인 상태에서 뉴스를 진행하기 때문에 움직임이 매우 적다. 따라서 얼굴 검출, 에지 분포, 유사한 앵커 모델의 3가지 조건을 모두 만족한다고 하더라도 샷 내의 움직임이 일정한 임계치 t_{motion} 을 넘을 경우 앵커 샷이 아닌 것으로 선언하였다. 프레임 크기가 $N \times M$ 이고, 샷이 n 개의 프레임으로 구성되어 있을 경우, 샷 내의 15프레임 간격으로 떨어져 있는, $\Delta t = 15$, 프레임간의 동일 위치에서 화소 단위의 밝기 값의 절대차가 일정한 값을 넘으면 해당 화소는 움직임이 있다고 간주하였다. 먼저 프레임 내의 모든 화소 중에 움직임이 있는 화소의 비율을 계산하여 프레임간의 움직임 정보를 계산하였다. 다음에 샷내의 모든 프레임에 대해서 동일하게 움직인 정보를 계산하고 합산한 후 평균하여 평균 움직임 값을 얻었다. 최종적으로 이 값이 일정 임계치

t_{motion} 보다 낮으면 앵커 샷으로 판단하였다.

$$\frac{\sum_{i=1}^{n-\Delta t} D_{i,i+\Delta t}}{n - \Delta t} \leq t_{motion} \quad (5)$$

여기서, $D_{i,i+\Delta t}$ 는 프레임 i 와 프레임 $i + \Delta t$ 간의 움직임이 있는 화소의 비율이다.

앵커 샷의 움직임 값은 평균적으로 KBS는 0.28, MBC는 0.27로 나타났으며, 앵커 샷이 아닌 경우는 평균적으로 KBS는 0.63, MBC는 0.69의 움직임 값을 가짐으로써 명확한 차이가 있었다. 그림 11의 윗 부분은 전체 970개의 샷을 가진 KBS 9시 뉴스 비디오의 움직임 값과, 1065개의 샷을 가진 MBC 9시 뉴스 비디오의 움직임 값을 보여준다. 아래 부분은 이 중에서 일부 구간(KBS의 100-220 번째 샷, MBC의 160-280번째 샷)만을 확대해 보여준 것이다. 앵커 샷만을 O로 표시하였는데 모두 낮은 움직임 값을 가지고 있음을 알 수 있다. 그 외에 X로 표시된 것도 앵커 샷이 아니지만 낮은 움직임 값을 가지고 있는데, 대부분이 인터뷰 샷이나 기자의 보고 장면 등이다. 비록 이런 샷들은 움직임 조건은 만족하지만 이전 단계인 에지 분포나, 앵커 모델과 유사한지 등의 조건을 통과하지 못하므로 앵커 샷으로 처리되지 않는다.

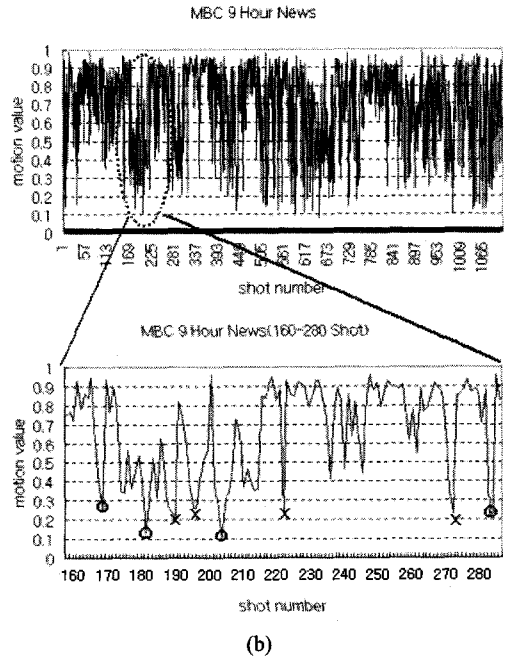
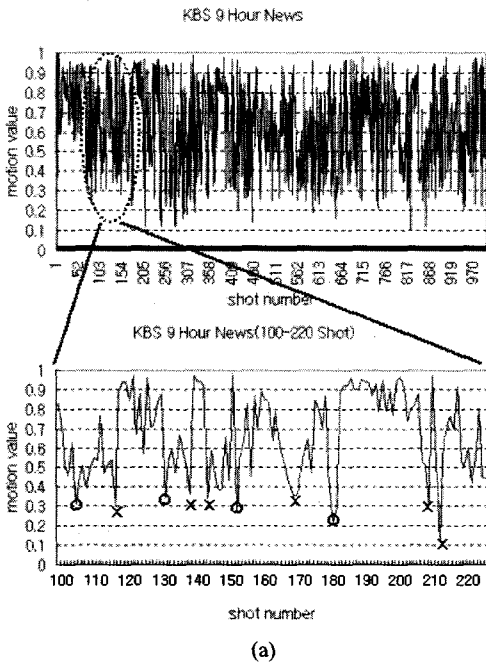


그림 11. 뉴스 비디오의 움직임 값: (a) KBS 9시 저녁 뉴스 (b) MBC 9시 저녁 뉴스

3. 실험

제안된 방법의 우수성을 평가하기 위해 매일 방송되는 뉴스 프로그램에서 가장 높은 시청률을 보이는 KBS와 MBC 9시 저녁 뉴스를 대상으로 실험을 실시하였다. 각각 KBS뉴스3개, MBC 뉴스 3개씩을 이용하였는데 제안된 방법의 신뢰성을 높이기 위해 모두 다른 날짜에 방송된 프로그램을 이용하였다. 표1에서 보는 바와 같이 KBS는 총 54분 분량(KBS 1은 46분, KBS 2는 4분, KBS 3는 4분)에 1162개의 샷(KBS 1은 999개, KBS 2는 45개, KBS 3는 118개)으로 이루어져 있고, 이 중에 총 37개의 앵커 샷(KBS 1은 31개, KBS 2는 2개, KBS 3는 4개)이 포함되어 있다. 반면에 MBC는 총 54분 분량(MBC 1은 46분, MBC 2는 4분, MBC 3는 4분)에 1248개의 샷(MBC 1은 1114개, MBC 2는 45개, MBC 3는 89개)으로 이루어져 있고, 이 중에 총 35개(MBC 1은 29개, MBC 2는 2개, MBC 3는 4개)가 앵커 샷이다.

앵커 샷을 검출하기 위해서는 먼저 비디오의 샷 경계 검출을 실시하여야 되는데, 본 논문에서는 이전 연구에서 사용한 색상과 명도의 상관계수를 이용하였다 [13]. 본



논문에서 설정하여야 할 변수로는 앵커 모델과 유사한지의 확인을 위한 임계치 t_{model} 과 움직임 정보를 판단하는 임계치 t_{motion} 이 있다. 각 방송별로 KBS는 Video 1으로 MBC는 Video 4를 가지고 실험했을 때 가장 적절하다고 판단한 0.72와0.43로 설정하였다.

3.1 평가 척도

평가 방법은 전통적으로 정보 검색 시스템에 사용되는 정확도 (Precision)과 회수율 (Recall), 그리고 F-값을 이용하였다. 정확도 (P)는 제안된 방법에 의해 앵커 샷으로 검출된 샷들 중에서 실제로 앵커 샷인 개수가 얼마인지를 비율로 나타낸 것이고, 회수율 (R)은 비디오 내에 포함된 앵커 샷 중에서 제안된 방법에 의해 얼마나 많은 앵커 샷이 검출되었는지의 비율을 나타낸다.

$$P = N_C / (N_C - N_F) \quad (6)$$

$$R = N_C / (N_C - N_M) \quad (7)$$

여기서, N_C 는 정확히 검출된 앵커 샷의 개수 (Correct), N_M 은 앵커 샷인데 검출하지 못한 개수 (Missed), 그리고, N_F 는 앵커 샷이 아닌데 앵커 샷으로 잘못 검출한 경우의 개수 (False)를 나타낸다.

때로는 P는 높게 나왔는데 R이 낮은 경우와, 반대로 R은 높게 나왔는데 P가 낮은 경우가 있다. 이 때는 어느 경우가 보다 좋은 성능을 나타내는지 판단하기가 어렵다. 더욱이 P와 R 간에는 서로 상충되는 면 (trade-off)이 있기도 하다. 따라서 이를 극복하기 위한 척도로 아래와 같이 정의된F-값을 종종 사용한다.

$$F_{SCORE} = (2 \times P \times R) / (P + R) \quad (10)$$

여기서 P와 R은 정확도와 회수율이다.

3.2 결과 및 분석

실험결과를 표1에 정리하였다. 여기서 Running Time은 분으로 표시된 방송시간을, N_{TS} 는 전체 샷 개수, N_{AS} 는 앵커 샷 개수, N_C 는 정확히 검출된 앵커 샷 개

수, N_F 는 잘못 검출된 앵커 샷 개수, N_M 은 검출하지 못한 앵커 샷 개수, P는 정확도, R은 회수율, F는 F-값을 나타낸다. 평균 정확도는 0.97이었고, 평균 회수율은 1.0, F-값은 0.98이었다. 모두 좋은 결과를 나타내었는데 제안된 방법의 우수성을 나타낸다. 본 논문에서 사용되는 임계치 t_{model} 과 t_{motion} 을 Video 1과 Video 4에 맞추도록 적절히 설정하였는데 예상대로 두 비디오에서 모두 검출률이 1.00으로 완벽한 결과를 나타내었다. 하지만 Video 3에서는 잘 못 검출된 샷이 1개 있는데 기존의 4가지 조건을 모두 만족하는 그림 12와 같은 화면이 앵커 샷으로 잘 못 판단되었다.



그림 12. 앵커 샷으로 잘 못 판단된 샷

KBS와 MBC외에 많은 사람들이 시청하는 뉴스 프로그램으로 'SBS 8시 저녁 뉴스'가 있다. 제안된 방법은 KBS와 MBC의 9시 저녁 뉴스의 앵커 샷을 검출하는데 초점을 맞춘 방법이지만 앵커 화면의 구조가 약간 다른 SBS 뉴스에는 적용했을 때 어느 정도의 성능을 나타낼지 알아 보았다. 3개의 비디오를 사용했는데 총 51분 분량 (SBS 1은 43분, SBS 2는 4분, SBS 3는 4분)에1060개의 샷 (SBS 1은 934개, SBS 2는 76개, SBS 3는 50개)으로 이루어져 있고, 이 중에 총 30개의 앵커 샷 (SBS 1은 25개, SBS 2는 3개, SBS 3는 2개)이 포함되어 있다. 실험 결과를 표2에 정리하였다. 평균 정확도는 0.98, 평균 회수율은 0.89, F-값은 0.93이었다. 이결과는 표 1에 비해서는 성능이 다소 저하되었지만 가능성이 충분히 있음을 보여 준다. Video 7에서 8개의 앵커 샷을 놓쳤는데 이 중 5개는 2단계의 에지 분포 조건을 만족하지 못했기 때문이다. 이는 SBS의 앵커 화면의 에지 분포가 KBS, MBC와는 다소간 다른 형태를 가지고 있음을 나타낸다. 놓친 나머지 3개는 4단계 움직임 값을 통과 하지 못한 것 1개와, 샷의

표 1. 앵커 샷

Data	Running Time (min)	N_{TS}	N_{AS}	N_C	N_F	N_M	P	R	F
Video 1 (KBS 1)	46	999	31	31	0	0	1.0	1.0	1.0
Video 2 (KBS 2)	4	45	2	2	0	0	1.0	1.0	1.0
Video 3 (KBS 3)	4	118	4	4	1	0	0.8	1.0	0.88
Video 4 (MBC 1)	46	1114	29	29	0	0	1.0	1.0	1.0
Video 5 (MBC 2)	4	45	2	2	0	0	1.0	1.0	1.0
Video 6 (MBC 3)	4	89	4	4	0	0	1.0	1.0	1.0
평균							0.97	1.0	0.98

표 2. 앵커 샷

Data	Running Time (min)	N_{TS}	N_{AS}	N_C	N_F	N_M	P	R	F
Video 7 (SBS 1)	43	934	25	17	1	8	0.94	0.68	0.79
Video 8 (SBS 2)	4	76	3	3	0	0	1.0	1.0	1.0
Video 9 (SBS 3)	4	50	2	2	0	0	1.0	1.0	1.0
평균							0.98	0.89	0.93

경계를 검출하지 못하여 앵커 샷인지의 여부를 판단하지 못한 것 1개, 그리고 뉴스 끝 무렵에 앵커가 중앙에 위치하여 얼굴 검출이 제대로 되지 않은 것 1개 때문이다. 또한 앵커 샷은 아니지만 기존의 4가지 조건을 모두 만족하여 앵커 샷으로 잘못 판단된 것이 1개 있었다.

4. 결론

본 논문에서는 4단계 규칙에 의한 앵커 샷 검출 방법을 제안하였다. 전처리 단계로 샷 경계 검출을 실시하고, 최초 프레임에 해당 샷의 키 프레임으로 선택하였다. 검출된 샷이 앵커 샷에 해당하는지는, 1) 키 프레임의 좌상부 영역에 얼굴이 존재하고, 2) 좌면 영역과 우면 영역의 에지 분포와 우하부 영역의 상부와 하부의 에지 분포가 일정 기준에 적합하며, 3) 기존의 앵커 샷의 색상 모델과 유사하여야 하고, 4) 샷 내의 움직임 값이 일정 임계치 이하여야 한다는 4가지 조건을 모두 만족하면 앵커 샷으로 판단하였다. 총 108분의 'KBS 9시 저녁 뉴스'와 'MBC 9시 저녁 뉴스' 프로그램에 대해 실험한 결과 평균적으로 0.97의 정확도, 1.0의 회수율, 0.98의 F-값이라는 좋은 성능을 얻을 수 있었고, 'SBS 저녁 8시 뉴스' 프로그램에 대해서도 가능성이 있는 결과를 보여주었다. 그러나 제안된 방법도 몇 가지 보완해야 할 점이 있었다. 특

히 실험을 하면서 음성 정보를 이용한 검출 조건도 추가된다면 보다 성능을 향상시킬 수 있다고 생각되었다. 또한, 본 논문은 기본적으로 KBS와 MBC 9시 저녁 뉴스만을 대상으로 하였다는 제약이 있다. 따라서 기타 CNN, ABC 뉴스 등의 앵커 샷 검출은 얼굴이 중앙에 위치할 수도 있다는 점과, 에지 분포를 기준으로 하는 조건들이 충족되지 못 하는 경우가 있을 수도 있기 때문에 제안된 방법을 적용하기에는 한계가 있다.

참고문헌

- [1] B. Günsel and A.M. Ferman, "Video Indexing Through Integration of Syntactic and Semantic Features," in *Proc. 3rd Workshop on Application of Computer Vision*, pp. 90-95, 1996.
- [2] Y.-F. Ma, X.-S. Bai, G.-Y. Xu, Y.-C. Shi, "Research on Anchor Person Detection Method in News Video," *Journal of Software*, vol. 12, no. 3, pp. 377-382, 2001.
- [3] J. Xu, H.-B. Hu, D.-R. Zhou, "Research on Identification of Anchorperson Shots in News Video," *Computer Engineering*, vol. 28, no. 3, pp. 165-166, 2002.
- [4] H. Luo and Q. Huang, "Automatic Model-based Anchorperson Detection," in *Proc. SPIE on Storage*

and Retrieval for Media Databases, vol. 4315, pp. 536-544, 2001.

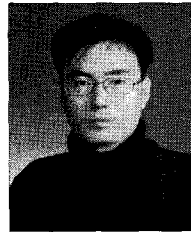
- [5] X.-D. Luan, Y.-X. Xie, L.-D. Wu, J. Wen, and S.-Y. Lao, "AnchorClu: An Anchorperson Shot Detection Method Based on Clustering," in *Proc. International Conference on Parallel and Distributed Computing, Applications and Technologies*, pp. 840-844, 2005.
- [6] X. Gao and X. Tang, "Unsupervised Video-Shot Segmentation and Model-Free Anchorperson Detection for News Video Story Parsing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 765-776, 2002.
- [7] X. Gao, J. Li, and B. Yang, "A Graph-Theoretical Clustering based Anchor Shot Detection for News Video Indexing," in *Proc. International Conference on Computational Intelligence and Multimedia Applications*, pp. 108-113, 2003.
- [8] M.D. Santo, G. Percannella, C. Sansone, and M. Vento, "An Improved Algorithm for Anchor Shot Detection," in *Proc. International Conference on Image Analysis and Processing*, pp. 679-686, 2005.
- [9] L. Chaisorn, T.-S. Chua, and C.-H. Lee, "The Segmentation of News Video Into Story Units," in *Proc. International Conference on Multimedia and Expo*, pp. 73-76, 2002.
- [10] Y. Zhai, A. Yilmaz, and M. Shah, "Story Segmentation in News Videos Using Visual and Text Cues," in *Proc. International Conference on Image and Video Retrieval*, pp. 92-102, 2005.
- [11] W. Qi, L. Gu, H. Jiang, X.R. Chen, and H.J. Zhang, "Integrating Visual, Audio and Text Analysis for News Video," in *Proc. International Conference on Image Processing*, pp. 520-523, 2000.
- [12] M. Bertini and A.D. Bimbo, "Content Based Annotation and Retrieval of News Videos," in *Proc International Conference on Multimedia and Expo*, pp. 483-486, 2000.
- [13] H.-W. Yoo and D.-S. Jang, "Automated Video Segmentation Using Computer Vision Technique," *International Journal of Information Technology and Decision Making*, vol. 3, no. 1, pp. 129-143, 2004.
- [14] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions*

on PAMI, vol. 20, no. 1, pp. 23-38, 1998.

- [15] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on PAMI*, vol. 8, no. 6, pp. 679-698, 1986.

유 헌 우(Hun-Woo Yoo)

[정회원]

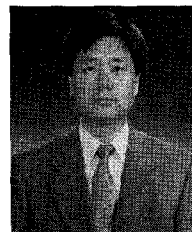


- 1992년 2월 : 인하대학교 전기공학과 (공학사)
- 1994년 8월 : 인하대학교 전기공학과 (공학석사)
- 2001년 8월 : 고려대학교 산업공학과(공학박사)
- 1994년 7월 ~ 1997년 10월 : LG전자 생산기술센터 기술개발 연구소 (주임연구원)
- 2000년 3월 ~ 2003년 2월 : 코스모정보통신 연구개발 본부 (수석연구원)
- 2003년 3월 ~ 2006년 8월 : 연세대학교 인지과학연구소 연구교수
- 2006년 10월 ~ 현재 : 연세대학교 컴퓨터과학과 연구교수

<관심분야>
컴퓨터비전, 영상처리, 멀티미디어

이 명 의(Myung-Eui Lee)

[정회원]



- 1985년 2월 : 인하대학교 전기공학과 (공학사)
- 1987년 2월 : 인하대학원 기기및 제어 (공학석사)
- 1991년 8월 : 인하대학원 기기및 제어 (공학박사)
- 1991년 8월 ~ 1995년 8월 : 현대전자 산업전자연구소 (선임연구원)
- 2004년 1월 ~ 2005년 1월 : U.C.Berkeley (객원교수)
- 1995년 9월 ~ 현재 : 한국기술교육대학교 정보기술공학부 교수

<관심분야>
Real-time Control System, Satellite Attitude & Orbit Control System, Embedded system, Computer Simulation