

뉴스 동영상의 앵커 객체 추출을 이용한 앵커 장면 검출

박기태^o 황두선 이종설* 이석필* 문영식
 한양대학교 컴퓨터공학과, 전자부품연구원*

{parkkt^o, dswhang, ysmoon}@cse.hanyang.ac.kr, {leejs, lspbio}@keti.re.kr*

Anchor Frame detection in news video Using Anchor Object Extraction

Ki Tae Park^o Doo Sun Hwang Jong Sul Lee* Seok Pil Lee* Young Shik Moon

Dept. of Computer Science Engineering, Hanyang University, Korea Electronics Technology Institute*

요약

본 논문에서는 뉴스의 단위 구조화를 위해서 앵커 구간과 비앵커 구간으로 구분하는 효율적인 알고리즘을 제안한다. 이를 위하여 단순히 장면 전환 지점을 이용하기 보다는 화면에 존재하는 앵커의 얼굴을 찾는 후 앵커 장면의 객체 정보와 배경 컬러 정보를 이용한 클러스터링 기법을 이용하여 앵커 장면을 검출한다. 제안한 알고리즘을 다양한 뉴스 비디오에 적용한 결과, 그 성능이 우수함을 보였다.

을 보인다.

1. 서론

최근 멀티미디어 기술과 통신 기술의 발달은 컴퓨터의 이용 폭을 훨씬 증가시켰고, 이러한 환경에서의 많은 양의 동영상상을 잘 정리하여 서비스하기 위한 많은 방법이 연구되고 있다[1]. 동영상상의 색인은 영상의 내용에 따라 다른 접근을 하게 되는데, 특히 영화나 드라마, 또는 스포츠 영상과 달리 뉴스는 영상 자체가 상당히 구조적으로 구성되어 있다. 뉴스는 시간 순으로 앵커 구간과 비앵커 구간(보도자료 구간)으로 규칙적으로 구성되어 있다. 따라서, 이를 하나의 뉴스 단위로 설정하여 뉴스를 여러 조각의 작은 사건 단위 뉴스 영상으로 나누어 놓을 수 있다. 그러므로, 뉴스 비디오의 구조를 잘 구성하기 위해서는 앵커 장면의 정확한 검출이 이루어져야 한다. 본 논문에서는 앵커 장면을 검출하기 위하여 4단계 과정을 처리한다. 제안한 알고리즘을 다양한 뉴스 비디오에 적용한 결과 제안한 방법이 효율적임을 알 수 있다.

2. 제안한 방법

본 논문에서는 뉴스 동영상에서 효과적으로 앵커 구간을 추출하여 작은 사건 단위로 뉴스를 구조화하는 방법을 제안한다. 그림 1은 전반적인 알고리즘 구조도를 보여주고 있다. 제안한 알고리즘은 4단계로 구성이 되어있다. 첫 번째 단계는 누적 히스토그램을 이용하여 장면 전환이 되는 것들에 대해서 앵커 장면 검출 알고리즘을 수행한다. 두 번째 단계는 객체를 추출하기 위해서 먼저 얼굴 검출 방법을 사용하고, 세 번째 단계는 앵커 객체 추출을 통한 앵커 후보 장면을 검출한다. 그리고 마지막 네 번째 단계에서는 앵커 후보 장면을 중에서 클러스터링 알고리즘을 통하여 앵커 장면과 비앵커 장면으로 분류한다.

2.1. 장면 전환 검출

뉴스 동영상에 대한 분석을 통하여 앵커 장면은 장면이 전환하는 지점에서 나타나는 것을 알 수 있다. 그러므로, 먼저 장면 전환 검출을 위해서 누적히스토그램 비교법을 사용한다[2]. 이 방법은 앵커 구간과 비앵커 구간으로 나누어 뉴스를 작은 사건 단위의 뉴스로 구조화하기 위해서 처리 속도가 빠르고 비교적 좋은 성능

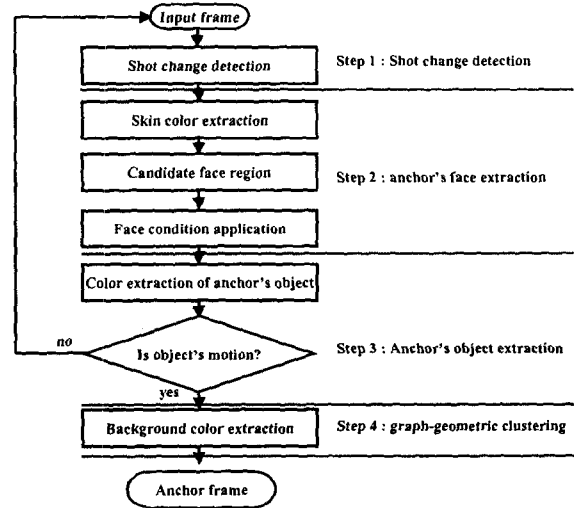


그림 1. 전반적인 알고리즘 순서도

2.2. 얼굴 영역 추출

얼굴 영역 추출 방법은 기존의 많은 방법들이 있으나, 동영상 의미 분석을 처리하기 위해서는 수행 속도가 빨라야 한다. 따라서, 일반적인 얼굴이 아닌 정형화된 뉴스의 앵커 얼굴이라는 점을 이용하여 앵커 얼굴 제약 조건을 고려해서 얼굴 영역을 추출한다. 얼굴의 피부 컬러는 YCbCr 컬러 모델에서 일정한 범위에서 나타나는 Cb, Cr 정보의 정보만을 추출한다[3]. 그러나, 얼굴보다 많은 잡음이 있기 때문에 모폴로지 연산을 수행하여 제거한다. 또한, 추출된 얼굴 후보 영역들중에서 다음의 제약조건을 만족하면 앵커 후보 얼굴 영역으로 간주한다..

- ① 앵커 얼굴은 일정한 크기를 가지고 있다.
- ② 앵커의 얼굴은 모서리 부분에 위치하지 않는다.
- ③ 얼굴 객체 면적에 대한 얼굴 사각형 크기의 비는 일정하다.

- ④ 앵커 얼굴의 가로, 세로의 비는 일정하다.
- ⑤ 앵커 얼굴의 변화는 거의 없다.

그림 2는 얼굴 영역을 추출한 결과를 보여준다.



그림 2. 얼굴 영역 추출 결과

2.3. 앵커 객체 추출

본 논문에서는 앵커 프레임의 검출을 위해서 앵커 객체의 컬러 정보를 이용한다. 그림 3은 앵커 객체를 추출하기 위해 사용되는 템플릿 영상을 보여주고 있다. 그리고, 그림 4는 앵커 객체 추출에 대한 알고리즘을 나타내고 있다.



그림 3. 템플릿 영상

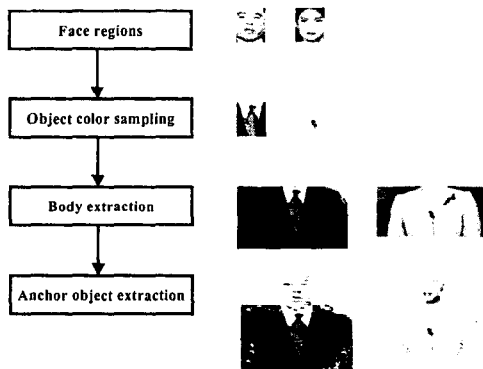


그림 4. 앵커 객체 추출에 관한 알고리즘

일반적으로 앵커는 단색의 옷을 입는다. 그러므로, 앵커 객체를 추출하기 위한 방법으로 객체의 컬러 정보를 사용하는 방법이 사용되고 있다. 템플릿 영상에서 객체의 영역은 얼굴 사각형을 좌, 우로 100%씩 크기를 확장하고, 위, 아래 방향으로 50%씩 크기를 확장한 상체 부분으로 정의한다. 앵커 객체에 대해서 컬러히스토그램을 구성하고 객체의 크기에 영향을 받지 않기 위해서 히스토그램을 양자화한다.

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

식 (1)에서, H_{NRO} 는 양자화된 히스토그램이고, H_{RO} 는 앵커 객체의 히스토그램이다. n 은 히스토그램의 빈(bin) 값을 의미한다. 뉴스 동영상에서 실제 앵커 장면을 찾기 위해서 앵커 후보 장면들에 대해서도 식 (1)과 같이 히스토그램(H_{NCO})을 구성한다. 그러나다

음 앵커 후보 장면이 앵커 장면인지 아닌지를 결정하기 위해서 히스토그램 인터섹션 기법을 사용한다.

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

식 (2)에서, $HS(H_{NRO}, H_{NCO})$ 는 템플릿 영상과 앵커 후보 장면들에 대한 유사도를 나타내고 있다. 이때, 앵커 장면들의 앵커 객체의 컬러는 아주 유사하기 때문에 일정한 임계값 이상이면 앵커 장면으로 검출한다. 그러나, 뉴스 동영상에서는 앵커와 비슷한 컬러를 가지는 객체를 많이 볼 수 있다. 그러므로, 오검출을 줄이기 위하여 '앵커 얼굴의 변화는 거의 없다' 라는 정보와 뉴스 동영상에서 앵커는 최소한 2초 이상 나온다는 사실을 이용하여 오검출된 앵커 장면을 제거 할 수 있다. 식 (3)은 프레임들간의 차이를 계산하는 식이다.

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

식 (3)에서, $FRFD$ 는 앵커 얼굴이 존재하는 위치 부분만을 잘라내어 그 영역의 프레임 차이값을 의미한다. M 과 N 은 프레임의 얼굴 사각형의 크기, $FR(x, y, t)$ 는 t 프레임에서의 얼굴 영역의 컬러 값이다. 이 때 프레임들의 차이값이 일정한 임계값보다 크면 앵커 장면에서 제거한다. 그리고, 그 결과를 그림 5에서 보이고 있다.



그림 5. 프레임 차이값을 이용한 결과

그러나, 그림 5에서 보는 바와 같이 앵커 장면이 아닌데 잘못 오검출된 것을 볼 수 있다. 이러한 오검출은 클러스터링 기법을 이용하여 해결한다.

2.4. 최소 신장 트리를 이용한 클러스터링

본 논문에서는, 오검출된 것을 제거하기 위해서 최소 신장 트리(MST : Minimum Spanning Tree)를 이용하여 클러스터링 작업을 수행한다. 많은 최소 신장 트리를 찾는 알고리즘들 중에서 Kruscal 알고리즘을 사용한다. 앵커 프레임을 하나의 노드로 정의하고, 노드들간의 배경 컬러에 대한 유사도를 에지(edge)의 가중치로 정의한다. 컬러 정보는 두 가지가 사용되는데, 첫 번째는 앵커 장면에서 얼굴 사각형 주위의 컬러 정보가 사용되고, 두 번째는 색상의 변화가 거의 없는 화면의 위쪽 컬러 정보가 사용된다. 그림 6과 그림 7은 얼굴 사각형 주위의 배경 컬러 정보 추출과 화면 위쪽 컬러 정보 추출하는 방법을 나타내고 있다. 이 두 배경 컬러들에 대한 유사도를 나타내는 가중치 구하는 방법은 식 (4)에 나타난다.

$$H(old_B_{ke(F,T)}, cur_B_{ke(F,T)}) = \sum_{i=1}^n \min(ld_B_k, cur_B_k)$$

식 (4)에서, H 는 두 노드들(프레임들)간의 유사도를 의미하는 가중치이고, n 은 히스토그램 빈(bin)을 나타낸다. B_F 와 B_T 는 각각 얼굴 사각형 주위의 배경 컬러 히스토그램과 화면 위쪽 배경 컬러

러 히스토그램을 의미한다.

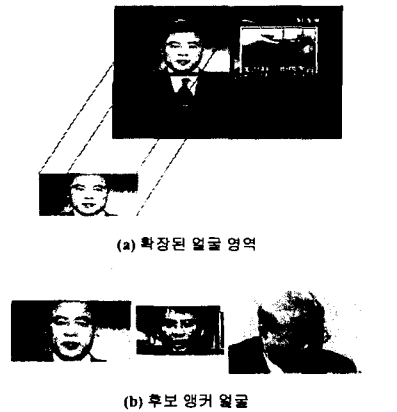


그림 6. 얼굴 사각형 주위의 배경 컬러 추출



그림 7. 화면 위쪽의 배경컬러 추출

그림 8은 최소 신장 트리를 이용한 클러스터링 알고리즘을 보여 주고 있다.

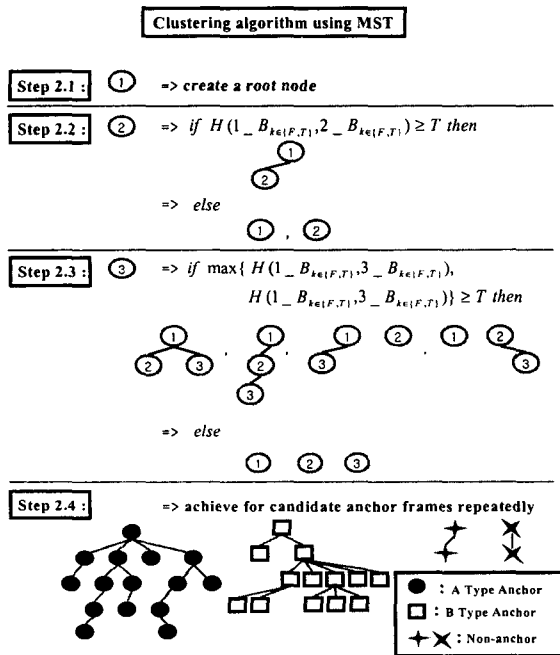


그림 8. 최소 신장 트리를 이용한 클러스터링

3. 실험 결과

앵커 장면 검출 실험을 위해 점진적인 장면 전환이 있거나 다양한 배경을 포함하고 있는 5가지 뉴스 비디오를 사용하였고, 실험 결과의 성능 평가는 precision과 recall을 구하여 측정하였다. 표 1은 앵커 장면 추출 결과를 나타내고 있고, 표 2는 앵커 장면 추출 결과에 대한 precision과 recall 결과를 나타내고 있다.

표 1. 뉴스 동영상별 앵커 추출 결과

News video clips	real existing anchor frame	extracted anchor frame	false	miss
Video 1	22	21	0	1
Video 2	17	16	0	1
Video 3	10	10	0	0
Video 4	33	33	1	0
Video 5	33	28	0	5

표 2. precision 및 recall 결과

	precision	recall
Average	99.4 %	94.9 %

video 5번 clip에서 미검출한 앵커 장면이 많은 이유는 고정된 스튜디오 배경이 아니라 다양한 변화가 발생하는 실제 거리 배경을 포함하고 있기 때문이다. 표 2는 전반적으로 우수한 결과를 보이고 있다.

4. 결론 및 향후과제

본 논문에서 제안하는 방법은 앵커의 객체를 추출하여 앵커 객체의 컬러 정보와 배경 컬러 정보를 이용한 클러스터링 기법을 사용하여 앵커를 검출한다. 실험 결과에서도 상당히 우수한 결과를 보이고 있다. 이렇게 검출한 앵커를 이용하여 하나의 뉴스 동영상을 작은 사건 단위로 구조화 할 수 있다. 그러나, 정확한 앵커 장면의 검출이 목표이므로 수행시간에 대한 문제를 고려하지 않았다. 따라서, 향후에는 MPEG 영상의 전체 복호화가 아닌 부분 복호화된 영상을 이용함으로써 전체 분석 시간의 향상을 고려해야 할 것이다.

참고문헌

- [1] Y. Tonomura, A. Akutsu, Y. Taniguchi, and G. Suzuki, "Structured Video Computing," *IEEE Multimedia*, pp. 34-43, 1994
- [2] 황두선, 이종설, 조위덕, 문영식, "누적히스토그램과 에지 정보를 이용한 장면 전환 검출," *대한전자공학회 하계종합학술대회 논문집 1호*, pp. 211-214, 2002년.
- [3] J. Huang, S. Ravi, M. Mitra, W. J. Zhu, and R. Zabih, "Image Indexing Using Color Correlograms," *In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 762-768, 1997년.
- [4] 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*, pp. 756-776, September 2002.