

뉴스 동영상 요약을 위한 앵커 장면 자동 추출 알고리즘

정진국, 이태연, 낭종호, 김경수*, 하명환*, 정병희*

서강대학교 컴퓨터학과, *KBS 기술 연구소

{jguk, leetaey}@mlneptune.sogang.ac.kr, jhnang@ccs.sogang.ac.kr

Automatic Detection of Anchorperson Shots for News Video Abstraction

Jin-Guk Jeong Tae-Yeon Lee Jong-Ho Nang Gyung-Su Kim* Myung-hwan Ha* Byung-hui Jeong*

Dept. of Computer Science, Sogang University, *KBS Technical Research Institute

요 약

최근 많이 사용되는 대용량의 뉴스 비디오의 편리한 검색 및 관리 방법이 필요하게 되면서, 뉴스 비디오 데이터를 자동으로 분석하여 저급 수준의 정보로부터 고급 수준의 내용 정보를 자동으로 추출하는 기술이 필요하게 되었다. 특히 뉴스를 요약하는 데 있어서는 이런 기술이 더 유용하게 쓰일 수 있다. 앵커, 그래픽, 인터뷰, 기자보도, 회견/연설 장면 등이 뉴스 비디오의 고급 수준 내용 정보가 될 수 있는데 그 중에서도 앵커 장면은 뉴스의 기사를 나누는 고급 수준의 정보로서 중요한 의미를 갖게 된다. 본 논문에서는 이러한 앵커 장면을 자동으로 추출하는 방법을 제안한다. 앵커 장면의 공통된 특징을 이용하여 검출하게 되는데 첫째 특징은 한 뉴스 프로그램을 진행하는 앵커는 동일하다는 점이고 두 번째 특징은 동일한 스튜디오안이라는 점이다. 본 논문에서는 앵커를 판별하는 방법으로 얼굴의 검출방법과 옷 색깔의 히스토그램 비교방법을 이용한다. 본 논문의 알고리즘을 여러 개의 KBS 9시 뉴스 비디오 데이터에 적용하여 실험한 결과 Recall 과 Precision 모두 96% 이상 나오는 것을 알 수 있었다.

1. 서론

최근 인터넷을 중심으로 디지털 비디오의 활용이 급속히 확산되고 대중화됨에 따라 VOD(Video On Demand)나 디지털 비디오 라이브러리와 같이 비디오 데이터를 이용하는 시스템의 개발이 활발히 이루어지고 있다. 이런 이유로 인해 디지털 비디오를 분석하는 기술에 대한 연구 또한 활발히 이루어지고 있는데 디지털 비디오를 분석하는 기술에서 가장 기본이 되는 것은 자동 인덱싱 기술 [10]이다. 자동 인덱싱 기술이라는 것은 디지털 비디오의 기본이 되는 단위인 신과 샷을 자동으로 검출하는 기술이다. 하지만 실제로 시스템에서 디지털 비디오가 이용될 때에는 자동 인덱싱 기술과 더불어 사용자가 원하는 고급 수준 내용 정보를 자동으로 추출하는 기술 또한 필요하게 된다.

각 장르에 따라 중요한 고급 수준 내용 정보는 다르게 될 것이다. 예를 들어 추구와 같은 장르에서는 골 장면이 중요한 의미를 가지게 될 것이고 액션 영화와 같은 장르에서는 액션 장면이 중요한 의미를 가질 것이다. 뉴스와 같은 장르에서는 앵커 장면, 그래픽 장면 등 여러 가지 고급 수준 내용 정보가 있을 수 있지만 그 중에서도 앵커 장면이 가장 중요하다. 앵커 장면은 각 기사

비디오 데이터의 가장 처음에 위치해 전체 기사 내용을 요약하고 기사와 기사 사이를 구분하는 역할을 한다. 본 논문에서는 이러한 앵커 장면을 자동으로 추출하는 알고리즘을 제시한다.

앵커 장면을 찾아내는 방법은 현재 많은 논문 [1,5,6]들에서 다양하게 제시되어 왔다. 영역 모델들의 비교를 통해 앵커 장면을 찾는 방법 [5], 템플릿을 이용하여 찾는 방법 [6] 등 여러 방법이 제시되었는데 본 논문에서는 뉴스 장르의 특징을 이용하여 앵커 장면을 자동으로 추출하는 방법을 제시하였다. 본 논문에서 이용한 특징은 한 뉴스내에서 앵커의 복장의 색깔은 일정하다는 것과 앵커 장면은 스튜디오 내에서 이루어진다는 것이다. 그리고 마지막으로 앵커 샷의 길이가 다른 샷에 비해 길다는 것이다. 이러한 특징을 이용한 본 논문의 알고리즘을 KBS 9시 뉴스 비디오 데이터에 적용한 결과 Recall 과 Precision 모두 96% 이상 나오는 것을 알 수 있었다.

2. 앵커 샷 추출 알고리즘

2.1 뉴스의 구조

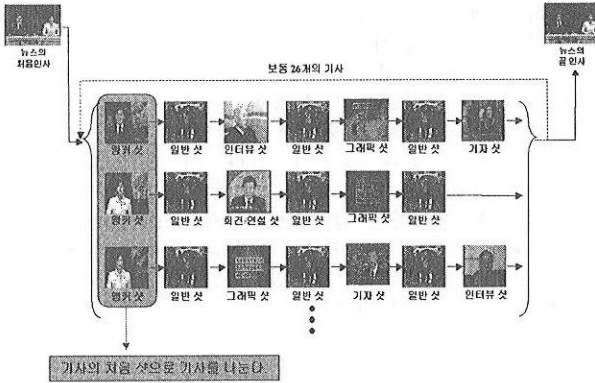


그림 1 뉴스 비디오 구조

<그림 1>은 뉴스 비디오의 구조를 나타낸 것이다. 그림을 보면 알 수 있듯이 뉴스는 기사들이 순차적으로 연결이 되는 단순한 구조를 가지고 있다. 즉, 기사들의 구성은 기사마다 인터뷰 샷이 있는 경우도 있고, 회견/연설 샷이 있는 경우도 있고, 기자 보도 샷이 중간에 나오는 경우도 있는 등 다른 모습을 보이지만 앵커 샷으로 시작하는 공통적인 특징을 보인다. 앵커 샷은 기사 앞에 존재하여 기사 전체의 내용을 요약하기 때문에 기사를 나누거나 뉴스 동영상을 요약하는 데 있어 중요한 정보가 된다.

2.2 뉴스 동영상에서의 샷의 분류

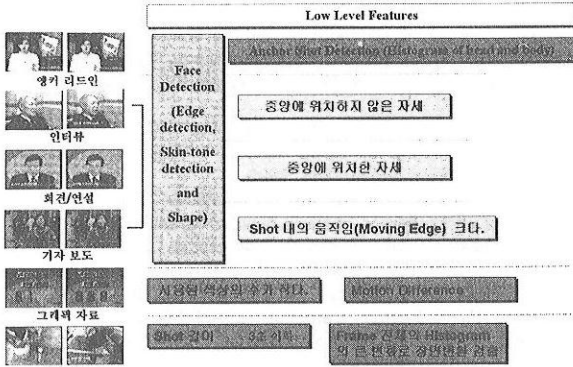


그림 2 뉴스에서의 샷 종류에 따른 특징

본 논문에서는 뉴스 동영상의 샷의 종류를 여섯 가지 - 앵커 샷, 인터뷰 샷, 회견/연설 샷, 기자 보도 샷, 그래픽 샷, 일반 자료 샷 - 로 나누었다. <그림 2>는 샷 종류에 따른 시각적인 특성을 나타낸 것이다. 앵커 샷은 일정한 스튜디오 안에서만 진행이 되기 때문에 배경의 움직임이 없다는 특징이 있고, 앵커의 히스토그램과 옷의 색깔이 한 뉴스 안에서

일정하다는 특징이 있다. 인터뷰 샷, 회견/연설 샷, 기자 보도 샷은 움직임이 없는 사람이 나온다는 공통점이 있지만 인터뷰 샷은 사람이 앞을 보지 않고 비스듬한 자세를 취한다는 특징이 있고 회견/연설 샷은 사람이 앞을 본다라는 특징이 있다. 기자 보도 샷은 앞에서 얘기하는 사람의 움직임은 없지만 배경의 움직임이 크다는 특징이 있다. 마지막으로 그래픽 샷은 자연 이미지가 아니라 인공적으로 만들어진 샷이기 때문에 사용된 색상의 수가 작고 또한 사람의 손에 의해서 조작되는 샷이기 때문에 움직임이 다른 샷들에 비해 규칙적이라는 특징이 있다.

2.3 앵커 샷 추출 알고리즘

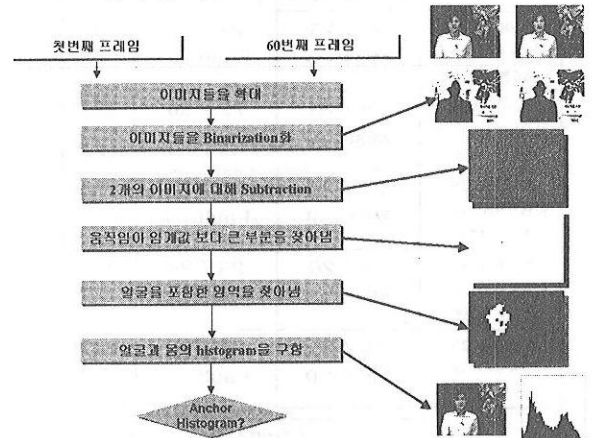


그림 3 앵커 샷 추출 알고리즘

<그림 3>은 본 논문에서 제안된 앵커 샷 추출 알고리즘을 나타낸 것이다. 오른쪽의 그림들은 각 단계의 결과들을 나타내고 있고 맨 마지막의 그림에서는 앵커 얼굴의 히스토그램을 나타내고 있다. 기본적으로 앞 절에서 언급했던 앵커 샷의 특징을 이용한다. 즉, 오브젝트를 추출하여 그 오브젝트가 앵커인지 아닌지를 앵커의 특징을 이용하여 알아보는 것이다. 앵커의 특징은 첫 화면에 앵커가 나온다는 가정을 하고 앵커가 인식하는 첫 장면에서 추출해 낸다. 앵커 샷인지를 알아보기 위해 샷의 첫번째 프레임과 60번째 프레임을 이용한다. 우선 정확하게 움직임을 찾아내기 위해 이 두 이미지를 확대한다. 그 후 이미지들을 이진 이미지로 만들고 2개의 이미지를 비교하여 움직임이 많은 부분을 찾아낸다. 이 때 움직임이 많다고 찾는 부분이 앵커의 입이 되고 이것을 기준으로 앵커의 옷과 얼굴 부분을 찾아낼 수 있는 것이다. 옷과 얼굴 부분을 구분하는 것은 HSV Color 영역에서 피부색의 범위를 이용한다. 옷과 얼굴의 영역을 찾아낸 후에는 첫 화면에서 찾아낸 앵커의

특징을 이용하여 비교한 후에 결정한다. 비교는 칼라 히스토그램의 표준편차와 최대치를 기록하는 빈의 위치를 이용한다. 표준편차는 히스토그램의 모양을 구별하기 위해서 이용하는 것이고 최대치 빈은 히스토그램 강약의 정도를 고려하기 위해서 이용하는 것이다.

3. 실험 결과 및 분석

표 1 실험 결과

TYPE	Anchor Shot		총 샷 개수
	Recall (A / (A + C))	Precision (A / (A + B))	
2000. 1.12. KBS 9시 뉴스 (46분 54초)	23 / 23 Miss : 0	23 / 25 Fault : 2	198
2000. 3. 9. KBS 9시 뉴스 (58분 16초)	25 / 25 Miss : 0	25 / 30 Fault : 5	221
2000. 4.25. KBS 9시 뉴스 (57분 37초)	24 / 24 Miss : 0	24 / 25 Fault : 1	273
2000. 5.30. KBS 9시 뉴스 (57분 37초)	26 / 26 Miss : 0	26 / 26 Fault : 0	250
2000. 6.27. KBS 9시 뉴스 (55분 55초)	25 / 25 Miss : 0	25 / 26 Fault : 1	243

	Anchor Shot 인 경우	Anchor Shot 이 아닌 경우
맞은 Shot	A	B
못 맞은 Shot	C	D

이 장에서는 2장에서 제안한 앵커 샷 추출 방법을 여러 실험 데이터를 이용해 실험한 결과를 보이고 분석함으로 해서 그 유용성을 평가하도록 하겠다. 실험에서 사용한 비디오 데이터는 MPEG-1 형식의 데이터이다. <표 1>은 5개의 KBS 9시 뉴스 샘플을 실험한 결과를 보여준다. 표를 보면 알 수 있듯이 Recall과 Precision 모두 96% 이상 보임을 알 수 있다. 3월 9일자 뉴스를 보면 Precision이 다른 실험 데이터에 비해 떨어지는 것을 알 수 있는데 이는 기사 중에서 앵커와 비슷한 옷이 나오는 장면이 있는 것과 앵커의 옷이 처음에 특징을 추출하는 데 있어 잘 나타나지 않는 것 때문이다.

4. 결론 및 앞으로의 연구 방향

최근 많은 멀티미디어의 사용으로 인해 멀티미디어 데이터를 분석하기 위한 여러 가지 알고리즘이 제안되어 왔다. 특히 최근에는 저급 수준 내용 정보를 이용하여 고급 수준 내용 정보를 추출하는 연구 또한

활발히 진행되고 있는데 본 논문에서는 여러 가지 영역 중에서 뉴스를 그 대상으로 하여 앵커 샷을 자동으로 추출하는 알고리즘을 제안하였다. 본 논문의 알고리즘은 뉴스 고유의 특징을 이용하여 우선 샷을 여섯 가지로 나누고 그 중에서 앵커 샷의 특징 - 샷의 길이, 앵커의 옷과 얼굴의 칼라 히스토그램 - 을 이용하여 추출하는 알고리즘을 제안하였다. 실험 결과 본 논문에서 제안하는 알고리즘은 recall 과 precision 모두 96% 이상 나와 뉴스 동영상 요약에 유용하게 적용할 수 있음을 알 수 있었다. 뉴스 기사내에서 앵커와 비슷한 옷이 나오는 경우나 처음 앵커의 특징을 추출하는 데 있어 특징이 잘 나타나지 않는 경우에 잘못 추출하는 경우가 있었는데 앞으로의 연구에서는 비디오 정보만이 아닌 오디오, 캡션등의 정보를 이용하여 보다 견고한 알고리즘을 제안하도록 할 것이다.

5. 참고 문헌

- [1] H. Luo and Q. Huang, "Automatic Model-Based Anchorperson Detection," *Proceedings of SPIE on Storage and Retrieval for Media Databases, Vol. 4315*, pp.536-544, 2001.
- [2] J. Nang, S. Hong and Y. Ihm, "An Efficient Video Segmentation Scheme for MPEG Video Stream using Macroblock Information," *Proceedings of ACM Multimedia '99*, 1999.
- [3] C. Y. Low, Q. Tian and H. Zhang, "An Automatic News Video Parsing, Indexing and Browsing System," *Proceedings of the fourth ACM International conference on Multimedia*, pp.425-426, 1996.
- [4] A. Hanjalic, Reginald L. Lagendijk, J. Biemond, "Semi-Automatic News Analysis, Indexing and Classification System based on Topics Preselection," *Proceedings of SPIE on Storage and Retrieval for Image and Video Databases VII, San Jose (Ca, USA)*, 1999.
- [5] H. J. Zhang, S. Y. Tan, S. W. Smoliar and Y. Gong, "Automatic Parsing and Indexing of News Video," *Proceedings of ACM Multimedia Systems*, pp.256-266, 1995.
- [6] A. Hanjalic, R. L. Lagensijk, J. Biemond, "Template-based detection of anchorperson shots in news programs," *Proceedings of IEEE International Conference on Image Processing '98, Chicago (IL, USA), Vol. 3*, pp.148-152, 1998.
- [7] H. Wang, S. F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, Issue: 4*, pp.615-628, 1997.
- [8] Changick Kim and Jenq-Neng Hwang, "Fast and Robust Moving Object Segmentation in Video Sequence," *Proceedings of IEEE International Conference on Image Processing '99, Kobe, Japan*, 1999.
- [9] Y. Avrithis, N. Tsapatsoulis, S. Kollias, "Broadcast News Parsing using Visual Cues: A Robust Face Detection Approach," *Proceedings of IEEE International Conference on Multimedia and Expo 2000, Vol. 3*, pp.1469-1472, 2000.
- [10] H. J. Zhang, C. Y. Low and S. W. Smoliar, "Video Parsing and Browsing using Compressed Data," *Proceedings of Multimedia Tools and Applications, Vol. 1*, pp.89-111, 1995.