TV 드라마 동영상에서의 장소를 기반으로 한 세그먼테이션

*류 재 석 ○, 낭 종 호
서강대학교 컴퓨터공학과
jaesugryu@gmail.com, jhnang@sogang.ac.kr

Video Segmentation Based on Place for TV Drama Video

Jaesug Ryu○, Jongho Nang
Dept. of Computer Science and Engineering. Sogang University

요 약

최근 급격한 인터넷의 발달과 동영상 컨텐츠 제공자가 많아짐에 따라 인터넷상의 숫자가 급격하게 많아지고 있다. 특히 스마트 기기의 발달로 인해 동영상을 제공받는 사용자들의 접근성도 나날이 늘어나고 있는 추세다. 이러한 이유로 동영상을 추천하기 위한 검색 시스템 기술은 중요시 되고 있다. 기존의 동영상 검색방법은 텍스트 기반의 검색 방법으로서 동영상을 사용자 관점에서 보았을 때 정확하게 표현을 하는데 많은 어려움이 있다. 본 논문에서는 검색 방법 중 장소를 기반으로 하는 내용기반의 검색 시스템을 위해 시리즈 동영상의 스트림을 장소단위로 세그먼테이션 하는 방법을 제안한다. 제안 방법은 로컬 특징점을 이용한 장소적 중복성을 계산하고 군집화를 검출하여 동영상 세그먼트를 표현하는 방법을 제안하였다.

1. 서 론

최근 급격한 인터넷의 발달과 동영상 컨텐츠 제공자가 많아짐에 따라 인터넷상의 동영상 숫자가 급격하게 많아지고 있다. 특히 스마트 기기(스마트 TV, 스마트 폰, 타블릿 PC)의 발달로 인해 동영상을 제공받는 사용자들의 접근성도 나날이 늘어나고 있는 추세다. 이러한 상황은 동영상 업로드 서비스 사이트들의 동영상 공급량과 수요량을 보면 간단히 알 수 있다. 따라서 나날이 늘어나는 접근성에 따라 동영상을 추천하기 위한 검색 시스템 기술은 중요시 되고 있다. 하지만 현재 보편적으로 사용자들에게 제공되고 있는 검색 시스템은 텍스트 기반의 검색 방법으로서 동영상 데이터에 사람이 직접 의미 정보를 기술하는 방법 즉, 태깅을 하는 방법이기 때문에 사용자의 관점이 불일치할 경우가 생기기도 하고, 또한 동영상 데이터는 복잡한 속성을 갖고 있기 때문에 정확한 표현을 할 수가 없는 단점이 있다. 따라서 동영상 데이터의 특징을 분석하여 추출하고 이를 기반으로 사용자들에게 검색 시스템을 제공한다면 더욱 객관적이고 검색시스템을 만들기 위한 비용도 줄어들게 될 것이다.

동영상의 데이터 특징을 추출하여 이를 기반으로 한 검색 시스템을 내용 기반의 검색이라고 한다. 기존의 내용 기반 동영상 검색 방법에는 사용자가 동영상의 대표 프레임을 질의대상으로 제공함으로써 해당 동영상을 찾는 방법이 대부분이다. 하지만 사용자 관점에서는 더 효과적인 검색을 위해 등장인물 또는 장소와 같은 의미적인 대상을 기준으로 검색하기를 원한다. 특히 장소 같은 경우 동영상마다의 특징을 갖고 있는 경우가 많다. 예를 들어 축구장 같은 경우 동영상의 대부분이 축구장이라는 장소를 배경으로 이루어지는 것과 같다. 한편, 동영상 중 시리즈 동영상 같은 경우 가상장소인 세트장을 배경으로 하는 경우가 많다. 이와 같은 경우 동영상의 다른 에피소드에서도 같은 배경을 사용하게 되어 그림 1과 같이 장소적 중복성이 생기게 된다. 따라서 장소를 기반으로 한 내용기반의 검색은 사용자에게 효과적인 검색 방법이 될 것이다.

본 논문에서는 위에서 언급한 장소를 기반으로 한 내용기반의 검색 개발을 위해 시리즈 동영상의 스트림을 장소단위로 세그먼테이션 하는 방법을 제안한다. 제안 방법은 동영상의 주요 프레임을 추출하고 전경과 배경을 구분하여 전경이 장소 단위로 세그먼테이션에 주는 영향을 최소화하였다. 그리고 SIFT[1] 특징점을 이용하여 장소적 중복성을 찾는 방법을 제안하여 검출성능 추출 및 분석을 하였다.

2. 관련 연구

2.1. 전경과 배경 분리

영상에서 전경과 배경영역을 분리하는 기술은 오래전부터 연구되어온 분야로 다양한 방법들이 존재한다. 이러한 방법은 크게 배경 차분 기반[2][3], 움직임 정보기반[4]과 모델 기반[5] 방법들로 구분할 수 있다. 배경 차분 방법은 전경을 제외한 배경 데이터를 미리 학습하여 배경 모델을 만들어 놓고 전경이 등장하게 되면 배경모델과 입력영상의 차이로 전경영역과 배경영역을 쉽고 빠르게 분리하는 방법이다. 하지만 이 방법의 경우 배경의 위치, 조명의 변화 등 색 정보에 민감하다는 단점을 갖고 있다. 이를 개선하고자 움직임 정보를 기반으로 하는 방법이 연구되었다. 이 방법은 시간적으로 연속된 프레임에서 움직임 방향과 거리 정

본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음. [10044615, 클라우드 기반 개방형 소셜 방송미디어 콘텐츠 융합 생성, 편집 및 재생을 위한 미디어 제작 및 전송 시스템 개발]

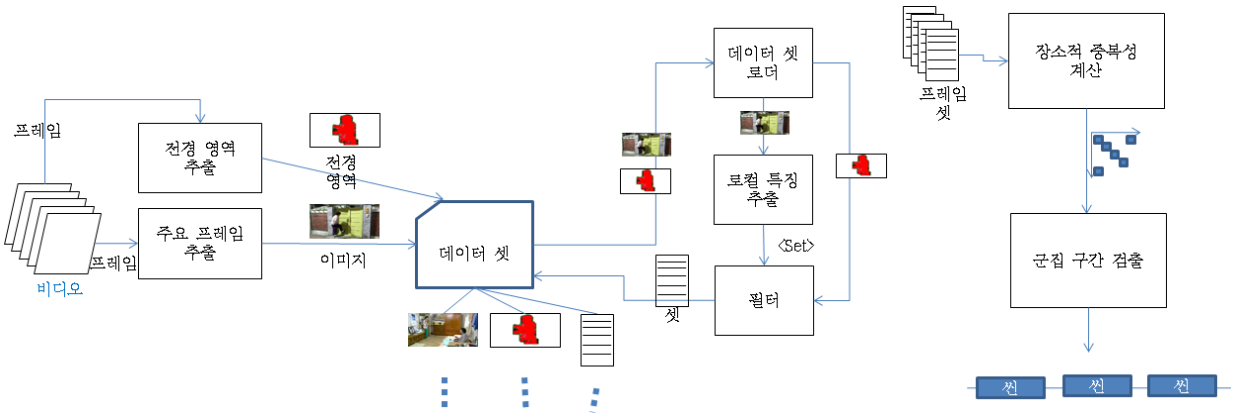

그림 2. 장소기반 동영상 세그먼테이션 시스템 구조도

보를 이용하여 전경영역과 배경영역을 분리한다. 따라서 배경의 위치가 바뀌더라도 분리가 가능하다는 장점을 갖고 있다. 한편, 모델 기반의 방법은 전경영역에 대한 모델을 학습하여 모델을 만들고 이를 이용하여 전경과 배경을 분리하는 방법이다. 예를 들어 전경이 사람이 될 경우 사람의 머리, 팔, 다리 그리고 몸의 형태를 학습하여 모델을 만들고 이를 분리해내는 것이다. 하지만 이 방법은 학습된 모델만을 분리할 수 있는 단점이 있다. 따라서 본 논문에서는 배경의 위치가 바뀌는 특성을 갖는 시리즈 동영상의 전경과 배경을 분리하기 위하여 움직임 정보를 기반으로 하는 방법을 사용한다.

2.2 군집화 검출 방법

일반적으로 동영상 내에서의 세그먼테이션을 위해서 유사도 공간 매트릭스를 생성한다. 유사도 공간 매트릭스는 세그먼테이션 구간에서의 군집성을 갖고 있다는 특징이 있다. 본 논문에서는 이러한 특성을 바탕으로 그림3과 같이 하프-변환(Hough-Transform)[6]을 통한 유사도 공간상에서의 네모 검출 방식을 사용한다.

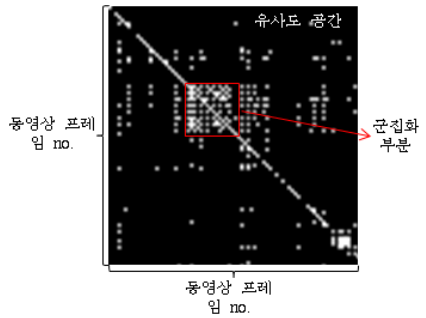

그림 3. 하프-변환을 통한 군집화 검출

3. 동영상 세그먼테이션

본 논문에서는 그림2에서 표시한 바와 같이 시스템을 구성한다. 제일먼저 주요 프레임을 대상으로 전경과 배경을 분리하고, 동시에 로컬 패치를 추출한다. 다음에 전경영역을 제거한 로컬 패치들을 이용하여 장소적 중복성을 계산하고 이를 이용하여 동일 장소 구간을 검출한다.

3.1 로컬 특징 추출

동영상에서 장소 정보를 얻기 위해선 장소라는 배경 앞에서 영향을 주는 전경들이 있다. 특히 시리즈 동영상의 경우 사람이 전경이 되는 것이 대부분이다. 이러한 전경은 장소와 다른 움직임 벡터를 갖고 있는 특징이 있다. 따라서 주요 프레임과 다음 프레임에 대해서 옵티컬 플로우[7]를 이용하여 움직임 벡터가 커지는 영역을 검출하여 이 영역은 전경이라고 정의한다. 동시에 SIFT를 이용하여 주요 프레임의 특징점들을 추출한 후 전경 영역을 제외한 특징점들을 추출한다. 그림4는 위의 과정을 보여준다.

그림 4. 배경 중심의 로컬 패치 추출 과정

3.2 장소적 중복성 계산

주요 프레임간의 장소적 중복성을 계산하기 위하여 두 주요 프레임의 로컬 패치들의 교집합을 구한다. 교집합을 구하는 과정은 먼저 두 이미지의 로컬패치들 간의 매칭 페어 중 유사도가 임계치보다 낮은 페어들을 선별하고 선별된 페어들에 대하여 기울기의 분포를 통해 공간적 일관성을 계산한 후 마지막으로 교집합의 이미지 좌표계에서의 영역이 충분히 넓은지를 판단하기 위하여 블록 분할을 통하여 페어가 포함된 블록 수의 비율을 계산한다. 위와 같은 과정을 그림5에서 보여주고 있다.

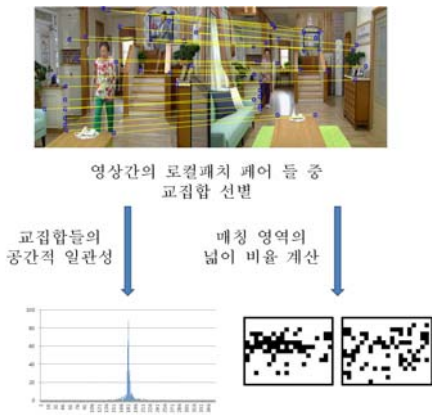

그림 5. 영상간의 장소 중복성 연산

3.3 동일 장소 구간 검출

모든 주요 프레임들 간의 장소적 중복성을 3.2에서 제시한 방법을 통해 계산하여 유사도 매트릭스를 생성한다. 같은 장소의 구간은 그림 6과 같이 대각선 대칭인 군집을 형성하게 되므로 군집 검출을 통해 동일 장소 구간을 검출한다.

그림 6. 동일 장소 구간 검출 예시

4. 실험 결과

본 논문에서는 성능평가를 위해 장소 구간의 개수 검출이 된 것들 중 성공한 것과 성공이 되지 않은 것으로 구분을 해서 표 1과 같이 표현하였다.

표 1. 검출결과

횟수	장소 구간의 개수	검출 성공	잘못된 검출 수
1	52	22	10
2	36	12	18
3	35	15	17
4	28	11	15
5	35	17	15
6	27	11	13
7	39	11	10
8	42	19	17
9	25	11	13
10	33	18	15

위 결과에서 볼 수 있듯이 제안한 방법은 성능이 높지 않음을 볼 수 있다. 이는 분석결과 그림7에서 보이는 것과 같이 전경영역의 추출 정확도가 높지 않아 전경영역에서의 로컬 특징점도 중복성 계산에 들어가게 되어 성능이 떨어지게 되는 현상이 생겼다. 또한 중복성 매트릭스의 군집화 검출을 하는 과정에서 검출 정확도의 영향을 받아 성능에 영향을 주게 되는 현상이 발생했다.

그림 7. 잘못된 전경영역 추출

5. 결론 및 향후 연구 방향

본 논문에서는 장소를 기반으로 하여 동영상을 세그먼테이션 하는 방법을 제안하였다. 기존 연구에서는 칼라정보나 질감 등의 사용자와 컴퓨터간의 의미적 관점차가 있는 내용을 기반으로 하는 반면에 제안 방법은 사용자가 직접적으로 의미가 있다고 판단되는 장소를 기반으로 하여 의미적 관점차를 줄였다는 것에 의의가 있다.
그러나 실험결과에서 보이듯이 성능 개선을 위해 전경영역의 추출 방법 개선, 중복성 매트릭스상의 군집화 검출 방법을 개선하고 처리 속도를 올리는 방법이 연구되어야 한다.

6. 참고 문헌

[1] LOWE, David G. “Object recognition from local scale-invariant features. In: Computer vision”, The proceedings of the seventh IEEE international conference on. Ieee, pp. 1150-1157, 1999

[2] C. Stauffer, W. Grimson, “Adaptive background mixture models for real-time tracking”, in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 246-252, 1999

[3] P. Kumar, S. Ranganath, W. Huang, “Queue based fast background modelling and fast hysteresis thresholding for better foreground segmentation”, The 2003 Joint Conference of the Fourth ICICS and PCM, Vol. 2, pp. 15-18, 2003.

[4] J. Badenas, J. M. Sanchiz, F. Pla, “Motion-based segmentation and region tracking in image sequences”, Pattern Recognition, Vol. 34, No. 3, pp. 661-670, 2001.

[5] H. Luo, Eleftheriadis, A., “Model-based segmentation and tracking of head-and-shoulder video objects for real time multimedia services”, IEEE Transactions on Multimedia, Vol. 5, No. 3, pp. 379-389, 2003.

[6] Illingworth, John, and J. Kittler. “A survey of the Hough transform,“ *Computer vision, graphics, and image processing.*, Vol. 44. No.1, pp.87-116, 1988.

[7] HORN, Berthold K., SCHUNCK, Brian G. “Determining optical flow.” In Technical Symposium East. International Society for Optics and Photonics, p. 319-331. 1981.