

방송 요약을 위한 중요 프레임 및 비 중요 프레임 검출

이웅규, 이재민, 정현중, 송인선, 남종호
서강대학교 컴퓨터공학과

taewoongkyu@naver.com, mangunig@gmail.com, fona@nate.com,

insun0813@naver.com, jhnang@sogang.ac.kr

Exciting and unexciting shot detection in commercial broadcast stream

WoongKyu Lee, JaeMin Lee, HyunJong Jung, InSun Song, JongHo Nang
Univ. Sogang, Computer Engineering

요 약

동영상 데이터에는 요약을 통하여 한눈에 알 수 있도록 하는 기술을 필요로 한다. 그 중 방송용 스트림(stream) 서비스의 경우 그 활용을 더욱 필요로 하고 있다. 여러 채널을 공유하는 TV에서는 각 채널이 무슨 방송을 하고 있는지 중요한 이슈가 된다. 이런 동영상 요약에서 키 프레임(key frame)을 찾는 기술이나 키 프레임과 거리가 먼 프레임을 찾아내는 기술이 필요하다. 이 논문에서는 키 프레임과 비 중요 프레임을 정의하고 그 프레임들을 검출하는 연구에 대하여 소개한다. 비 중요 프레임의 경우 칼라 히스토그램(color histogram)을 분석하여 실제 테스트 이미지들과의 차이점을 분석한다. 키 프레임의 경우 얼굴 검출(face detection)과 샷 변경 검출(shot change detection)의 기술을 이용하여 자동으로 추출하도록 하고 그 성능을 측정하도록 한다.

1. 서 론

스마트 폰의 발전과 스마트 TV의 발전은 수많은 멀티미디어 데이터의 이동 또는 공유를 필요로 하고 있다. 그 중 방송시스템의 발전은 위와 같은 단말기들의 발전을 따라 더욱 가속화 되고 있는 실정이다. 특히 여러 채널을 보유한 TV의 경우, 각 채널이 무슨 방송을 하고 있는지 중요한 이슈가 된다. 이런 동영상 콘텐츠(contents)들은 요약을 통한 정보의 전달을 필요로 한다. 동영상의 요약은 주로 키 프레임들의 집합으로 이루어 지는데, 이런 키 프레임을 선정하는 일은 동영상 요약에서 중요한 부분이 된다. 또한 키 프레임의 선정을 위해서는 비 중요 프레임을 검출하는 일도 중요하다. 본 논문에서는 방송 스트림 내용상 중요한 부분과 관계가 없는 부분을 정의하고 그 부분들을 검출할 수 있는 기술을 고찰하고 실험하여 효율적인 방법을 찾아보도록 한다.

본 논문에서는 크게 두 가지로 정의를 나누고 실험을 나눈다. 한가지는 비 중요 프레임에 대한 내용이고, 나머지 한가지는 키 프레임에 대한 것이다. 물론 그 둘이 완벽히 대치되는 프레임이라면 한가지의 기술만으로도 동영상의 요약은 가능하다. 하지만 프레임들은 아주 많이 존재하기 때문에 다양한 변수가 생길 수 있다. 그래서 여러 가지 기술들을 접목하여 진행하는 것이 효율적일 것이다.

먼저 방송 흐름에 있어서의 중요하지 않은 내용은 크게 본 콘텐츠에 적합하지 않은 내용들을 포함하도록 하였다. 예를 들면, 광고나 콘텐츠 시작 전에 나타나는

예고방송, 방송 시작전의 방송 조정 화면, 방송 신호가 끊어진 후의 상태들을 선정할 수 있다.

방송 요약의 키 프레임을 선정하는 방법은 휴리스틱(heuristic) 하게 생각하도록 하였다. 모든 방송이 그럴지는 않지만, 주로 사람이 클로즈업(close up)되는 상황은 중요한 장면일 가능성이 높다는 것에 착안하여 얼굴이 프레임의 어느 정도 이상 비율을 차지한다면 중요한 장면으로 선정하였다.

앞으로 2 장에서는 서론에 이어서 비 중요 프레임과 키 프레임을 정의하도록 하고, 3 장에서는 두 종류의 프레임을 검출하는 방법에 대하여 기술한다. 4 장에서는 위의 두 종류의 프레임을 검출하는 실험을 통해 결과를 도출한다. 마지막으로 5 장에서는 결론을 내린다.

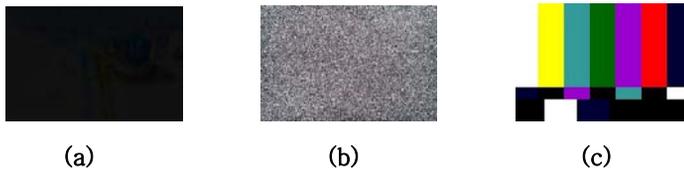
2. 비 중요 프레임 및 키 프레임의 정의

2.1 비 중요 프레임 정의

전체 방송 스트림 중 콘텐츠들의 흐름과 전혀 관계가 없는 프레임을 검출하는 작업이 필요하다는 것은 앞에서 밝혔다. 이런 장면들을 검출하기 위하여 방송의 흐름을 분석해 보았다.

광고의 경우 짧은 시간 동안 내용이 자주 바뀌는 현상을 발견할 수 있었다. 내용이 바뀌는 동안 한 프레임 정도의 까만 화면(black frame)이 나타나는 것을 확인할 수 있었다. 이런 특징은 광고뿐 아니라 방송이 시작할 때나 콘텐츠 중간에도 나타나는 것을 분석을 통해 알 수 있었다. 그러므로, 까만 화면, 또는 극단적인 단색으로 나타나는 화면이 광고 및

방송시작을 알리는 예고 등의 경계를 표현하여 쉽게 제거 할 수 있음을 알 수 있었다.



[그림 1] 정의한 비 중요 프레임의 예시

또한, [그림 1-(C)]처럼 방송이 시작하기 전에 등장하는 방송 조정 화면에 대한 처리도 필요함을 알 수 있다. [그림 1-(a)]와 같은 단색 화면은 방송 콘텐츠와는 무관한 장면임을 쉽게 확인할 수 있다. 마지막으로 방송신호가 끊어진 [그림 1-(b)]와 같은 장면도 방송 콘텐츠와는 무관하다고 할 수 있다.

2.2 키 프레임의 정의

키 프레임의 경우 서론에서 언급했듯이 등장인물의 얼굴이 클로즈업 되는 순간을 동영상 요약에 중요한 순간으로 분류하였다. 물론 여러 다른 상황들 또한 충분히 중요한 상황이 될 수 있지만, 등장인물의 구체적인 상황을 전달하는 프레임이 방송 요약에 가장 중요한 요소라고 초점을 맞추었다.



[그림 2] 정의한 키 프레임의 예시

[그림 2-(a)]에서 나타나는 프레임은 뉴스의 한 장면으로 중요 인물의 얼굴을 클로즈업 한 장면이다. 그리고 [그림 2-(b)]는 드라마의 한 장면으로 대사를 하는 모습을 보여준다. 이런 장면들이 반드시 중요한 장면이 아닐 수도 있지만, 휴리스틱 접근의 가정을 통하여 프레임들을 정의 하였다.

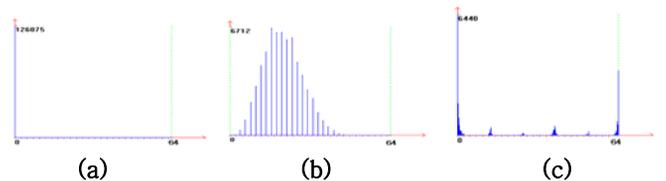
또한 장면이 바뀌면 새로운 내용이 시작하기 때문에 장면이 바뀌는 순간을 검출하여 새로운 등장인물을 찾을 수 있다면 그 장면도 키 프레임이 될 수 있을 것이다.

3. 비 중요 프레임 및 키 프레임의 검출

3.1 비 중요 프레임 검출

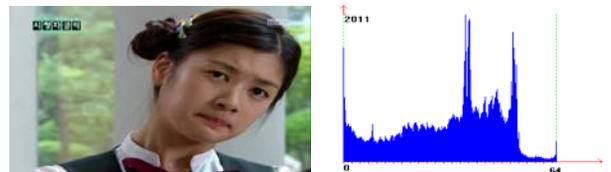
본 논문에서는 위와 같이 방송 콘텐츠와 가장 관계가 없는 장면들을 선정하여 실험의 클래스를 분류하였다. 비 중요 프레임을 검출하기 위하여 프레임들의 칼라

히스토그램[1]을 분석해 보았다. 칼라 히스토그램을 이용한 이유는 [그림 1]의 샘플들을 보고 직관적으로 특징을 뽑을 수 있기 때문이다. [그림 1]의 프레임들을 살펴보면 세 프레임 모두 색에 대한 극단적인 특성이 나타나는 것을 알 수 있다.



[그림 3] 비 중요 프레임 각각의 칼라 히스토그램

[그림 3]은 위에서 선정한 비 중요 프레임들에 대한 각각의 칼라 히스토그램 결과다. 단색의 화면의 경우 맨 위의 그림과 같이 한쪽만 모든 값이 몰려 있는 것을 알 수 있다. 신호가 끊어진 노이즈(noise)화면의 경우 몇 가지 정해진 색상들만 나타나는 것을 확인할 수 있다. 마지막 방송 조정 화면의 경우도 정해진 색상만 데이터가 있음을 알 수 있다.



[그림 4] 평범한 프레임의 칼라 히스토그램

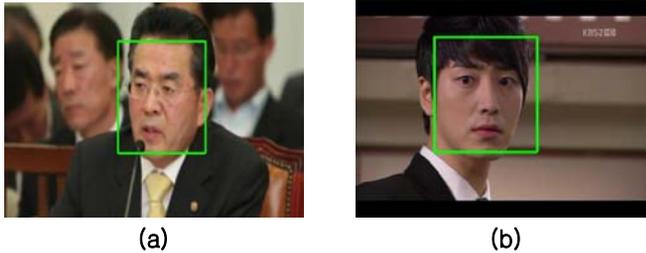
[그림 4]는 평범한 드라마의 한 장면을 칼라 히스토그램으로 분석한 것이다. 거의 모든 색상이 나타나고 있음을 알 수 있다.

위의 데이터들의 분석을 통하여 비 중요 프레임과 평범한 프레임의 차이를 얻을 수 있었다. 비 중요 프레임의 히스토그램에서는 픽셀간의 색 변화가 거의 없거나 인위적인 배열로 인하여 색상이 특정 부분에 몰려 있는 것을 확인할 수 있다. 그래서 칼라 히스토그램상의 데이터가 '0'이 되는 색상들을 많이 찾을 수 있다. 반면, 평범한 프레임의 경우 총천연색으로 구성되어 인접한 픽셀의 색이 같아 보여도 데이터상의 차이가 나기 때문에 데이터 값이 '0'이 되는 부분을 찾아보기 힘들다. 이러한 특징을 이용하면 본 논문에서 정의한 비 중요 프레임들은 쉽고 확실하게 구분할 수 있을 것으로 생각하고 실험을 하였다.

3.2 키 프레임의 검출

본 논문에서 중요 장면을 검출하기 위해 등장인물의 얼굴을 인식하고, 바뀌는 장면을 찾아내는 기술이 필요로 하고 있다. 얼굴을 검출하는 기술은 이미

정확도가 검증된 openCV 라이브러리 중 하나인 Haar-like feature[2]를 이용하여 쉽게 얻을 수 있었다. 그리고 shot change detection 기술은 칼라 히스토그램을 이용하여 장면이 바뀌는 것을 검출할 수 있었다.



[그림 5] 얼굴 검출(face detection)결과

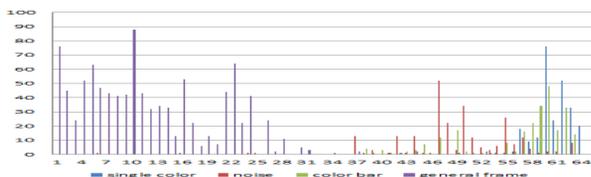
[그림 5]는 Haar-like feature 를 이용한 얼굴 검출의 결과이다. 얼굴 검출은 전체 프레임의 6 분의 1 이상의 얼굴만 검출하도록 설계되어 있다. 그 이유는 수 많은 동영상들을 분석한 결과 클로즈업 되는 화면의 경우 6 분의 1 이상의 얼굴 크기를 보여주었다. 6 분의 1 이상이 될 경우 두 명 이상의 인물이 한 프레임이 나타나는 경우가 많이 나타났다. 그리고 작은 비율의 얼굴을 검출하도록 하면 자막을 검출하는 등 잘못된 검출을 많이 하게 되어 오히려 시스템의 정확도가 낮아졌다.

Shot change detection 의 경우 현재 프레임과 이전 프레임간의 칼라 히스토그램의 차이가 80 퍼센트 이상 되면 다른 장면(shot)이라고 설정하였다. 80 퍼센트의 임계 값은 여러 동영상들을 실험해 본 결과 가장 안정적인 결과를 보였기 때문에 설정하게 되었다.

4. 실험 및 결과

비 중요 프레임의 검출 실험은 각 프레임들을 칼라 히스토그램으로 분석하고 색상이 한번도 나타나지 않은 부분을 카운팅 하여 전체 결과 그래프의 카운팅 된 숫자 위치에 합을 해주는 방식으로 진행 했다. 칼라 히스토그램의 전체 차원은 64 차원이며 [그림 1]의 세 클래스는 250 장의 샘플데이터를 이용하고, 평범한 프레임은 1,000 장을 이용하였다.

[표 1] 비 중요 프레임 실험 결과.

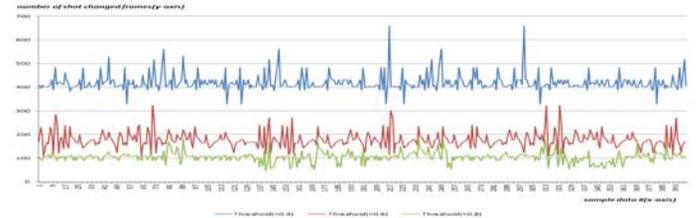


[표 1]의 X축은 '0'이 되는 빈도수를 표현하며, Y축은 X축에 해당하는 수만큼의 프레임의 수를 말한다. [표 1]에서 확인할 수 있듯이 비 중요한 프레임과 정상 프레임 사이에는 확실한 분류가 가능하도록 되어 있다. X

축의 '34'지점을 경계로 1차원의 클래스 분류가 가능한 것을 확인할 수 있다. 본 논문에서 궁극적으로 얻고자 하는 것은 비 중요 프레임을 평범한 프레임들로부터 구분하는 것이기 때문에 원하는 답을 얻을 수 있었다.

키 프레임의 검출 실험은 얼굴 크기가 프레임의 6 분의 1 일 때가 가장 동영상 요약에 적합한 프레임이 된다는 것을 보여주는 실험을 하였다.

[표 2] shot change detection의 임계 값 결정 실험



[표 2]는 임계 값을 임의로 정하여 60 분짜리 동영상 수를 늘려가며 실험했을 때 요약에 가장 적합한 임계 값을 찾기 위한 실험 결과 이다. 추출한 프레임의 수가 적은 것이 80 퍼센트이며, 그 위로 60 퍼센트, 40 퍼센트의 임계 값으로 실험을 진행하였다. 실험 데이터가 증가하여도 평균값은 거의 일정함을 알 수 있다. 결과적으로 프레임간 80 퍼센트의 칼라 히스토그램이 다를 때 가장 적은 프레임 중복을 보이면서 요약에 적합한 모습을 보였다.

5. 결 론

동영상 요약을 위하여 두 가지 가설을 세우고 실험을 하였다. 비록 당장 사용할 수 있는 기술은 아니지만, 방향을 설정하고 목표를 달성하기 위한 초석이 될 것으로 생각한다. 비 중요 프레임의 경우 더 많은 프레임들을 선정하여 새로운 분류 기준을 세우고 진행하는 것이 앞으로의 과제가 될 것이다. 키 프레임의 경우 다른 방송들을 더 많이 분석하여 인물과 관계없는 다른 기준을 세우고 실험을 하는 것이 앞으로의 연구 목표가 될 것이다.

참고문헌

[1] A. Müfit Ferman, A. Murat Tekalp and Rajiv Mehrotra, "Robust Color Histogram Descriptors for Video Segment Retrieval and Identification," IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 11, NO. 5, MAY 2002.
 [2] Min Zuo, Guangping Zeng and Xuyan Tu, "Research and Improvement of Face Detection Algorithm Based on the OpenCV," ICISE 2010 2nd International Conference on, 4-6 Dec. 2010.7