

인터랙티브 스마트 TV 적용을 위한 사운드트랙 검색 시스템

류상현 조재만 김형국

광운대학교 전파공학과

rshfly@nate.com, coolshut@hotmail.com, hkim@kw.ac.kr

Soundtrack Search System for Interactive-Smart-Television

Ryu, Sang-Hyeon Cho, Jea-Man Kim, Hyoung-Gook

Dept. of Radio Science and Engineering Kwangwoon University

요약

본 논문에서는 인터랙티브 스마트 TV 적용을 위한 사운드트랙 검색 시스템을 제안한다. 제안하는 시스템은 동영상을 오디오와 비디오특징을 구분한 후, 각 오디오와 비디오 신호를 분석한다. 비디오 신호의 분석은 MPEG-2 비디오 인코더로부터 영상의 장면전환과 시작과 끝 위치를 검출하고, 오디오 신호의 분석은 AC-3 오디오 인코더로부터 오디오 특징을 추출한 후, 오디오 정보의 비트 벡터를 추출하여 데이터베이스를 생성한다. 생성된 데이터베이스와 사용자가 북마크를 하여 요청한 쿼리와 비교를 통하여 오디오 특징정보가 유사한 부분의 장면을 검색하고, 검색된 장면을 사용자에게 제공한다. 제안된 시스템의 성능 측정을 위해서 뉴스, 패널리톤, 음악방송, 광고, 드라마 등 50개 TV 방송 프로그램의 데이터베이스를 이용해서 정확성을 측정하였다.

1. 서론

사용자가 원하는 비디오를 제공하는 서비스(VOD: Video on Demand)는 IPTV를 성공으로 이끈 중요한 기술 중 하나이다. VOD 시스템의 다운로드와 스트리밍은 언제 어디서든 사용자가 원하는 다양한 종류의 비디오를 제공할 수 있게 해주었다. TV 프로그램에서 시청자가 원하는 부분을 다시 시청하길 희망하는 수요가 증가함에 따라서, 시청자가 원하는 비디오 콘텐츠를 검색해서 찾아내는 기술의 개발이 활발하게 이루어지고 있다. 자동으로 비디오를 북마크 하는 기능은 비디오 세그먼트의 한 부분을 검색하는 기술과 TV프로그램에서 시청자의 흥미를 끄는 특정한 부분을 표시해주는 기술을 사용한다. 또한, 위와 같은 IPTV의 기능들은 다른 사용자들과 비디오를 공유할 수 있게 해주고, 검색기록은 추후 검색을 위해서 저장된다.

비디오를 북마크 하는 시스템을 개발하는데 중요한 요소는 시스템의 간단함, 검색의 정확도, 빠른 검색속도이다. 본 논문에서는 인터랙티브 스마트 TV에 적용을 위한 빠르고 정확한 비디오 검색 시스템을 제안한다. 제안된 비디오 검색 시스템은 고성능 사운드트랙 식별 시스템과 함께 수행된다.

2. 전체적인 시스템구조

제안된 시스템의 목적은 시청자들이 쉽고 빠르게 북마크 된 비디오 세그먼트를 검색하는 기능을 제공하는 것이다. 제안된 시스템은 비디오 콘텐츠의 사운드트랙을 식별하는 방법으로 구현된다. 그림1은 제안된 시스템의 기본 개념도이다. 데이터베이스 서버에서 사용자가 북마크한 비디오 세그먼트의 사운드트랙으로부터 오디오 특징 정보를 추출한다. 그리고 오디오 특징 정보는 TV 프로그램에서 사용자가 원하는 비디오 구간을 검색하는데 사용된다. 제안하는 시스템에는 데이

터베이스 생성과 쿼리 식별 두가지단계로 구분이 된다.

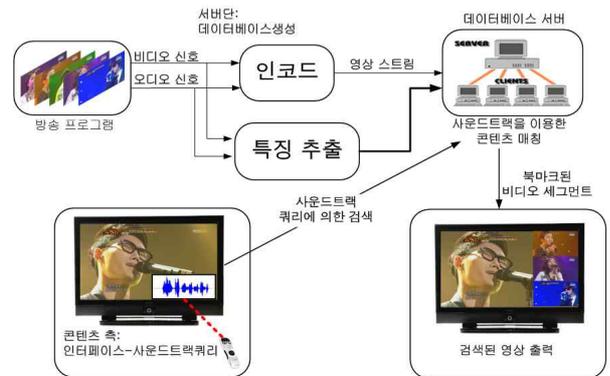


그림 1. 제안된 시스템의 기본 개념도

그림2는 제안하는 검색 시스템의 블록도이다. 블록도는 크게 색인화와 검색 두가지 단계로 나눌수 있다. 색인화란, 데이터베이스 생성을 의미하며 검색은 쿼리 식별을 포함하는 사용자 인터페이스에 제공되는 영상의 검색을 의미한다. 제안한 시스템은 입력된 비디오 스트림의 비디오와 오디오 신호 분석에 바탕을 두고 있다.

가. 데이터베이스 생성

데이터베이스를 생성할 때, 제안한 시스템은 오디오와 비디오 특징을 구분해서 분석하는데, 모든 방송 비디오 콘텐츠의 비디오와 오디오 신호는 MPEG-2 비디오 인코더와 AC-3 오디오 인코더로 인코딩되어 오디오의 특징정보와 비디오 영상의 특징을 추출한다. 추출된 오디오 특징 정보와 찾아진 비디오 영상의 특징 정보가 데이터베이스 서버에 저장된다.

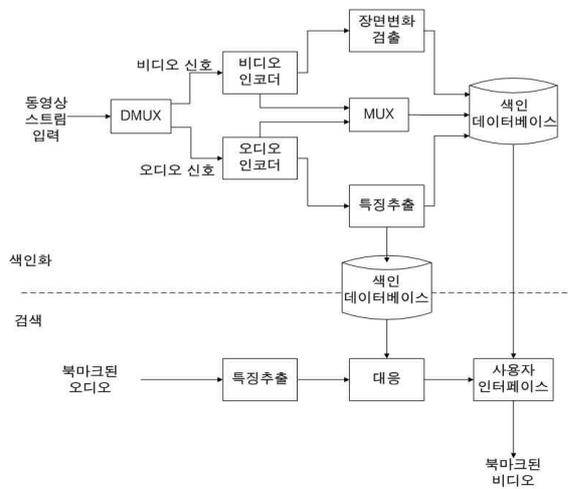


그림 2. 검색 시스템의 블록도

비디오 영상의 특징정보 추출은, MPEG-2 비디오 인코더로부터 RGB색상 히스토그램을 바탕으로 장면전환시간과 키 프레임이 결정된다. RGB색상 공간에서 색의 히스토그램간의 차이를 장면전환을 검출하는데 사용한다. 이와 마찬가지로, AC-3 오디오 인코더로부터 오디오 특징 정보와 비트 벡터가 추출 된다. 비트벡터의 추출은 MDCT(Modified Discrete Cosine Transform)계수를 Logarithmic하게 추출하여 LMDCT(Logarithmic MDCT)계수를 추출한다. 추출한 LMDCT계수를 Low pass filter에 통과시킨다. Low pass filter를 통과한 LMDCT계수를 양자화하여 비트벡터를 추출한다. 양자화는 아래 식(1)과 같다.

$$fbv(i) = \begin{cases} 1, & AMS(i) > AMS(i+1) \\ 0, & otherwise \end{cases} \quad (1)$$

$AMS(i)$ 는 Low pass filter를 통과한 LMDCT계수이다.

나. 쿼리 식별과 검색

쿼리 식별에서 사용자에게 의해 북마크 된 사운드트랙은 데이터베이스의 쿼리를 요청하고 데이터베이스 서버에 저장된 정보들과 비교한다. 북마크 된 사운드트랙과 일치하는 부분이 찾아지면, 해당하는 영상의 시작과 끝의 위치가 데이터베이스 서버에서 검색된다.



그림 3 비디오검색의 순서

제안하는 시스템은 사용자가 비디오 콘텐츠를 시청할 때, 원하는 장면을 선택해서 시청하는 것이 가능하다. 제안된 시스템은 그림2와 같이 북마크 된 오디오 세그먼트를 식별하기위해서 오디오 특징 정보를 사용한다. 또한, 쿼리 식별은 북마크 된 쿼리 사운드트랙으로부터 오디오 특징 정보를 추출한다. 북마크 된 특징 정보는 검색을 위해서 데이터베이스 서버로 전송된다. 전송된 오디오 특징정보와 유사한 구간들이 검색결과로 결정되어, 검색결과 구간들에 대한 영상이 사용자에게 보여진다.

3. 실험결과

본 논문에서 설명한 시스템의 정확성을 측정하기 위해 뉴스, 공개 토론, 음악 방송, 광고와 드라마를 포함하는 50개의 한국어 방송 TV프로그램을 이용하여 북마크된 오디오의 검출 정확도를 측정하였다. 측정된 실험 결과는 98.7%의 정확도를 나타내었다.

4. 결론

본 논문에서는 인터랙티브 스마트 TV 적용을 위한 사운드 트랙 검색시스템을 제안하였고, 제안된 시스템의 성능 측정을 위해서 TV 방송 프로그램의 데이터베이스를 이용해서 검색속도와 정확성을 측정하였다. 향후 사용자 제작 콘텐츠(UCC) 데이터베이스를 이용한 사운드 트랙검색시스템에 대한 연구를 시행할 계획이다.

감사의 글

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구 사업 지원을 받아 수행된 것임(2011-0004311).

참고 문헌

- [1] M. Flink, M. Covell, and S. Baluja, "Social-and interactive-television applications based on real-time ambient-audio identification," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, Jul. 2002.
- [2] P. Cano, E. Batlle, T. Kalker and J. Haitsma, "A review of algorithms for audio fingerprinting", *Proc. of International Workshop on Multimedia Signal Processing*, 2002.
- [3] Y. Ke, D. Hoiem, and R. Sukthakar, "Computer vision for music identification," *Proc. of Computer Vision and Pattern Recognition*, 2005.
- [4] Y. Y. Shi, and X. Zuan, H.-G. Kim, and K.-W. Eom, "A robust music retrieval system," *Proc. 120th Convention of Audio Engineering Society*, pp. 151-158, May, 2006.
- [5] Jurgen Herre, Oliver Hellmich, and Markus Cremer, "Scalable robust audio fingerprinting using MPEG-7 content description", *Proc. of IEEE Workshop on Multimedia Signal Processing*, pp. 165~168, 2002.
- [6] Jaap Haitsma, and T. Kalker, "A Highly Robust Audio Fingerprinting System", *Proc. of International Conf. on Music Information Retrieval*, pp. 14~17, 2002.