

스마트 TV의 골프동영상 썸네일 생성을 위한 오디오기반 경계영역 검출 기법

*최희민 **이진호 ***김형국

광운대학교 전자공학과

*kikimaro@naver.com

**jhlee0506@kw.ac.kr

***hkim@kw.ac.kr

Thumbnail Generation of Golf Videos Using Audio-Based Boundary Detection for Smart TV

*Choi, Hee-Min **Lee, Jin-Ho ***Kim, Hyoung-Gook

Dept. of Radio Science and Engineering, Kwangwoon University

요약

본 논문에서는 스마트 TV 시청시에 녹화하는 골프 동영상에서 오디오기반의 경계영역 검출을 이용하여 썸네일을 고속으로 생성하는 방법을 제안한다. 제안된 방법은 녹화되는 골프동영상의 인코딩된 오디오 정보로부터 추출된 MDCT계수를 이용하여 온셋 구간 검출 및 오디오 세그멘테이션을 수행함으로써 골프 동영상에 6개의 오디오 클래스로 자동 분할한다. 분할된 오디오 세그먼트와 상응하는 비디오 프레임을 맵핑하여 골프 동영상의 썸네일을 생성한다. 제안된 오디오기반 경계영역 검출 방법의 성능 측정 결과, 97.4%의 Recall과 96.85%의 Precision의 우수한 분류 성능을 나타내었다.

1. 서론

디지털 비디오 레코더(DVR)을 통해 저장된 미디어 콘텐츠를 이용하는 사용자가 증가함에 따라 저장된 미디어 콘텐츠를 효율적으로 시청할 수 있는 타임 시프팅과 같은 연구가 대두되고 있다. 이는 비교적 재생시간이 긴 스포츠 동영상을 시청할 때 필요하며, 일반적으로 스포츠 동영상에 대해 썸네일을 구성함으로써 타임 시프팅을 구현하고, 이로 인해 사용자에게 편의를 제공한다. 이러한 DVR 기능과 타임 시프팅 기능이 스마트 TV에 적용되고 있다.

스포츠 비디오의 중요 이벤트 검출 및 썸네일 형성을 위한 연구는 다양한 방향으로 진행되어 왔다. 이벤트 검출 및 썸네일 형성을 위해 멀티모달 기반 검출 방법이 연구[1]되어져 왔으며, 오프라인 학습 기반에 적응적 모델을 결합한 연구[2]도 진행되어져 왔다.

본 논문에서는 스마트 TV 시청시에 녹화하는 골프 동영상에서 고속 썸네일 추출을 위한 오디오기반의 경계영역 검출 방법을 제안한다. 제안된 방법은 골프 동영상내에서 연산량이 작은 오디오 신호분석을 이용하여 경계영역을 검출하며 영상신호의 상응하는 프레임을 맵핑하여 썸네일을 생성함으로써, 연산량과 복잡도 측면에서 이점을 지니고 있다.

본 논문에 구성은 다음과 같다. 2장에서는 제안된 방법에 대해 설명하고, 3장에서는 설명된 방법을 통한 실험결과를 제시한다. 그리고 마지막 4장에서는 결론을 맺는다.

2. 골프 동영상 썸네일 형성을 위한 전체 시스템

그림 1은 본 논문에서 제안한 스마트 TV의 썸네일 자동생성 및 사용자 인터페이스가 연계된 블록도를 나타낸다. TV 수신기를 통해 입력되는 동영상은 사용자 선호도에 맞게 자동으로 녹화된다. 녹화된 동영상은 썸네일 추출을 통해 오디오 스트림을 6개의 오디오 클래스로

자동 분할한다. 6개의 오디오 클래스는 스튜디오 음성(STD), 음악(SOM), 관중의 환호성(APP), 필드 음성(SPC), 기타 잡음(SIL), 그리고 스윙(SWN) 사운드 등으로 구성된다. 추출된 썸네일은 썸네일 저장소에 전달되고, 썸네일 추출을 마친 골프 동영상은 미디어 콘텐츠 저장소에 저장된다. 그리고 사용자가 해당 골프 동영상을 시청하고자 할 때 그림 1의 하단 사진과 같이 검색 수행 사용자 인터페이스에 썸네일을 형성함으로써 사용자에게 편의를 제공한다.

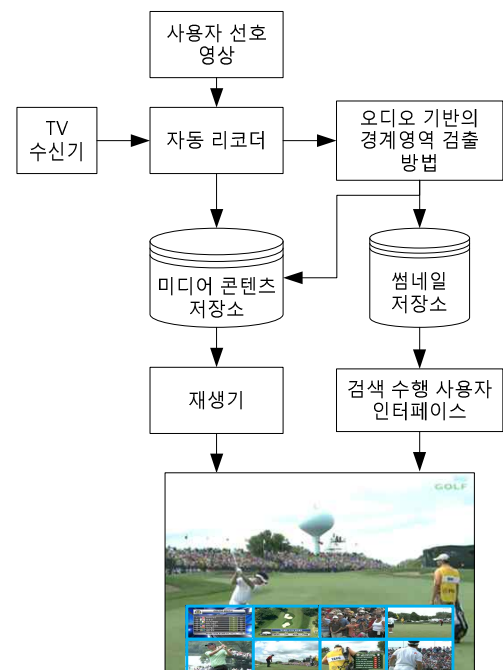


그림 1. 제안하는 시스템의 기본 개념

3. 오디오기반 경계영역 검출 및 썸네일 추출

그림 2는 골프 동영상에서 오디오기반 경계영역 검출 및 썸네일 추출 방법에 대한 블록도를 나타낸다. 녹화되는 골프 동영상은 비디오 신호와 오디오 신호로 분리되어 각각 비디오 인코더(H.264/AVC)와 오디오 인코더(AC-3)로 입력된다. 인코딩된 오디오 신호에서 Modified Discrete Cosine Transform (MDCT) 계수의 오디오 특징값을 추출한다. 추출한 MDCT는 스윙 사운드 검출부로 입력되어 골프 동영상에서 스윙 구간(SWN)을 검출한다[1]. 그리고 오디오 세그먼트이전에 입력되어 SWN을 제외한 STD, SOM, APP, SPC, 그리고 SIL 구간을 검출한다[2]. 스윙 사운드 검출부와 오디오 세그먼트이전에서 검출된 SWN과 APP를 이용하여 오디오기반 하이라이트 구간을 검출한다. 검출된 오디오기반 하이라이트 구간과 분할된 오디오 세그먼트 구간에 상응하는 인코딩된 비디오의 프레임과 맵핑함으로써 썸네일 구간을 결정짓는다.

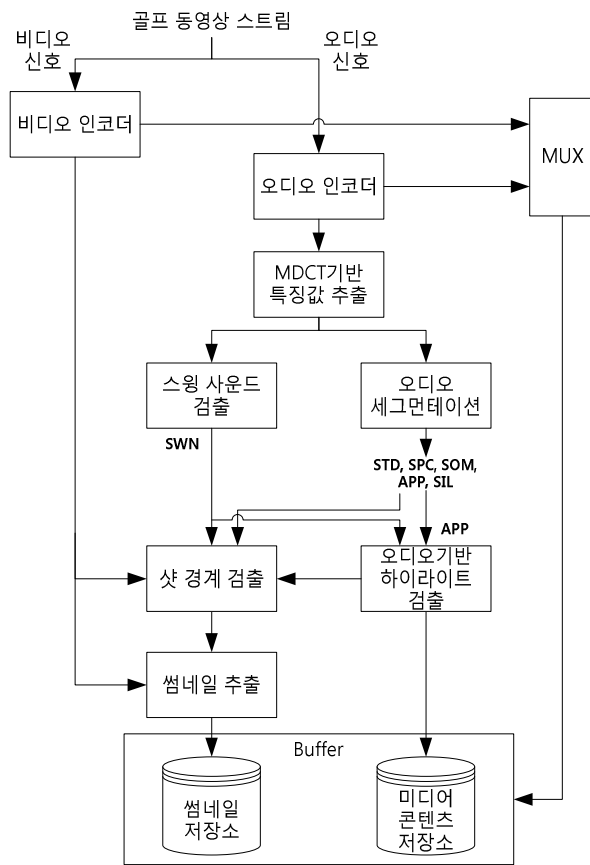


그림 2. 오디오기반 시멘틱 경계 검출 및 썸네일 추출

4. 실험 결과

본 논문에서 제안한 방법의 성능을 측정하기 위해 200분짜리 100개의 골프 동영상 실험군을 이용하였다. 다음 표 1은 골프 동영상 실험군을 이용한 오디오 세그먼트 분류 정확도를 나타낸다.

표 1에서 보면, STD / SOM / SWN / APP에 대해 평균 97.4%의 Recall과 96.85%의 Precision의 우수한 성능을 제시함을 확인할 수 있다.

BASICSs	Recall		Precision	
	Mean	95% CI	Mean	95% CI
STD	100%	100%, 100%	97.5%	91.8%, 103.3%
SOM	92.4%	91.9%, 102.8%	93.7%	90.6%, 101.8%
SWN	100%	100%, 100%	98.6%	95.4%, 105.2%
APP	97.2%	94.2%, 100.1%	97.6%	92.8%, 102.4%

표 1. 골프 오디오 세그먼트별 경계분류 정확도

5. 결론

본 논문에서는 스마트 TV에 녹화되는 골프 동영상으로부터 고속 썸네일 추출을 위한 오디오기반 경계영역 검출 방법을 제안하였다. 제안된 방법은 동영상의 오디오 신호를 이용하여 고속으로 썸네일을 형성할 수 있기 때문에, 스마트 TV와 같은 임베디드 환경에서 효과적으로 썸네일을 추출할 수 있다.

감사의 글

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

참고문헌

- [1] D. A. Sadlier, and N. E. O'Connor, "Event detection in field sports video using audio-visual features and a support vector Machine," *IEEE Transactions on Circuits and Systems for Video Technology*, pp.1225-1233, 2005.
- [2] J. Wu, X. Hua, J. Li, B. Zhang, and H. J. Zhang, "An Online Learning Framework for Sports Video View Classification," *Proceedings of Pacific Rim Conference on Multimedia*, pp.289-297, 2004.
- [3] 서광덕, 김형국, "골프 동영상에서의 Mel-Filter Bank를 이용한 스윙구간 검출 알고리즘," *한국멀티미디어학회 춘계학술대회*, Vol.14, No. 1, p.5, 2011.
- [4] 남상순, 김형국, "멀티 모달 정보 분석을 이용한 TV골프 방송 프로그램에서의 이벤트 검출 및 요약," *한국방송공학회 학술대회*, p.3-407, 2009.