

효과적인 비디오 Scene/Shot 단위 내용 Tagging을 위한 시스템1)

이호준^o 김윤후 최장원 낭종호

서강대학교 컴퓨터공학과

hughlee89@gmail.com, yhookim89@gmail.com, persevere7@naver.com, jhnang@sogang.ac.kr

Efficient Browsing for Video Scene/Shot Contents Tagging

HoJun Lee^o YoonHoo Kim JangWon Choi JongHo Nang
Sogang University Department of Computer Science and Engineering

요 약

최근 빠르게 확산되고 있는 멀티미디어 콘텐츠에 대한 효과적 관리 및 활용의 필요성이 커지고 있다. 특히 비디오 콘텐츠의 효과적 관리와 활용을 위해 정보 관계를 기반으로 한 시멘틱 검색 기술을 도입하고 이를 위해 데이터베이스의 일종인 온톨로지를 구축한다. 효과적이고 유용한 온톨로지 구축은 멀티미디어 콘텐츠 관리 및 활용에 있어 가장 기본이자 핵심이라 할 수 있다. 현재 사용되고 있는 비디오 전체에 대한 Tagging으로는 온톨로지를 효과적으로 구축할 수 없다. 영상의 세부적인 Scene과 Shot에 대해 정보를 수집하기 어렵고 직접 Tagging이 가능하더라도 비효율적인 측면이 있다. 따라서 우리는 효과적으로 온톨로지를 구축하기 위해 비디오 콘텐츠 Tagging에 대한 비효율적인 측면을 개선해 효율적으로 Scene과 Shot 단위의 Tagging이 가능한 시스템을 제안한다. 다양한 장르의 비디오 콘텐츠를 분석하여 현재 비디오 콘텐츠에 적합한 UI(User Interface)를 갖춘 시스템을 설계하였다. Scene과 Shot에 따른 효과적인 분류를 통해 비디오의 구조를 쉽고 입체적으로 파악할 수 있으며 사용자의 입장을 고려한 효율적 태그 입력 및 분석이 가능하도록 했다. 수집된 정보를 토대로 정확한 데이터베이스를 구성해 직관성, 편의성, 신뢰성을 모두 갖춘 시스템을 제안한다.

1. 서 론

최근 디지털 매체의 발달과 함께 이를 활용할 수 있는 멀티미디어 콘텐츠들이 빠르게 확산되고 있다. 특히 비디오와 같은 동적인 멀티미디어의 콘텐츠들의 사용이 많아짐에 따라 비디오 정보에 대한 관리의 필요성이 커지고 있다. 효율적인 정보관리를 위해서는 텍스트 중심의 단순한 검색 보다는 정보들의 관계를 기반으로 하는 시멘틱 검색 기술이 필요하다[1]. 따라서 사용자의 Tagging 입력을 바탕으로 비디오 콘텐츠들의 관계와 내용 정보를 가지는 데이터베이스 형태의 일종인 온톨로지를 구축해 시멘틱 검색을 활용할 수 있게 해야 한다[2].

온톨로지 구축을 위해서는 각 콘텐츠들마다 Tagging된 정보들이 필요하며, Tagging된 정보들이 많을수록 정보들 간의 관계를 잘 유추할 수 있다. 그러므로 많은 정보를 얻기 위해서는 비디오 전체내용 보다는 장면에 대한 Tagging이 필요하다. 따라서 비디오의 기본 구성 단위인 Scene/Shot단위의 Tagging은 온톨로지를 보다 효과적으로 구축하는데 필수적이다. 비디오 전체보다 작은 단위의 Tagging을 함으로써 특정한 장면만 보고 싶어 하는

사용자에게 비디오 전체영상 검색결과 대신 장면 단위의 비디오 콘텐츠를 제공한다. 이를 활용해 수많은 콘텐츠들 속에서 빠르고 정확하게 원하는 비디오 콘텐츠를 접할 수 있다. 본 논문에서 제안할 시스템은 Tagging tool의 데이터베이스를 구축하고, 직관성과 편의성이 있는 UI(User Interface)를 제공하여 Shot/Scene단위의 장면 Tagging을 가능하게 한다. 이를 통해 현재 우리가 접할 수 있는 Tagging tool들보다 세부적인 Tagging을 가능하도록 하여 최종적으로 온톨로지 구축 도구의 일종으로 사용할 수 있도록 했다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 조사에 대해 소개하고, 3장에서는 현재의 콘텐츠들의 분석과 본 논문에서 제안한 시스템 방법에 대해 설명하며 4장에서는 결론 및 향후 연구로 맺는다.

2. 관련조사

2.1 현재 비디오 Tagging 시스템

현재 비디오 Tagging을 하는 시스템은 두 가지로 나누어진다. 사용자가 비디오 전체에 대한 메타정보를 작성하는 방법과 비디오를 시청하며 Shot단위의 정보를 순간마다 입력하는 방법이다.

1) 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]

2.1.1 비디오 전체에 대한 Tagging 시스템

미국의 YouTube, 한국의 엠앤케스트와 태그스토리가 비디오 전체에 대한 Tagging 방식을 사용하는데 이 방식은 비디오 전체에 대한 정보를 입력 할 수 있지만 Shot/Scene 단위의 내용에 대한 Tagging이 불가능하기 때문에 장면위주의 검색에는 한계가 있다. BeeDeo.com에서 Scene에 직접 태그를 붙이는 방식은 Scene단위의 검색이 가능하다는 장점이 있지만 비디오의 처음부터 마지막까지 모두 시청해야하는 단점이 존재한다.

2.1.2 Shot 단위에 대한 Tagging 시스템

현재 대표적인 Shot단위의 Tagging이 가능한 tool로는 IBM사에서 제작한 비디오AnnEx annotation tool 이다. 그러나 tool은 Scene단위의 Tagging이 불가능하고 Tagging을 하는데 있어서 자유로운 Tagging보다는 주어진 항목을 선택하는 방식이어서, 다양한 상황에 대한 정확한 Tagging이 불가능하다. 그리고 Shot change detector의 정확도가 100%가 아님에도 불구하고, 수동으로 수정할 수 있는 tool을 제공해 주지 않는 것도 정확도를 낮춘다. 따라서 본 논문에서는 지금까지 비디오 Tagging 방식의 비효율적인 측면들을 개선해 Shot/Scene 단위의 장면 Tagging을 가능하게 해 비디오 시멘틱 검색을 위한 온톨로지를 보다 효과적으로 구축할 수 있음 보였다[3].

3. 제안 시스템

본 논문에서 제안하는 시스템의 주된 목적은 Shot/Scene 단위의 세부적인 Tagging을 얼마나 효율적으로 하느냐 하는 것이다. 이를 위해서는 UI(User Interface)가 중요한데, 이는 현재 비디오 콘텐츠 분석을 통해 Scene과 Shot에 대한 특성이 분명해 졌을 때 가능할 것이다. 그래서 본 논문에서는 비디오 콘텐츠 분석을 통한 Scene과 Shot의 일반적인 특성을 바탕으로 효율적인 시스템을 제안한다.

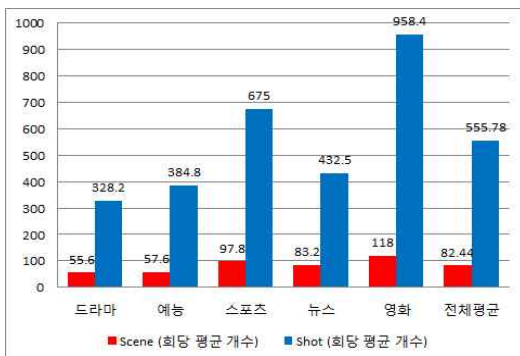


그림 1. 장르별 비디오 콘텐츠 분석

3.1 비디오 콘텐츠 분석

대표적인 비디오 콘텐츠를 장르별(드라마, 예능, 스포츠, 뉴스, 영화)로 나누어, 5가지의 장르별 콘텐츠를 각각 5회 분량씩 분석하였다. 영상의 가장 기본 단위인 Scene과 Shot의 평균 개수를 토대로 분석한 결과는 다음과 같다.

표 1. Scene에 따른 Shot의 평균 개수

장르	Scene에 따른 Shot의 평균 개수
드라마	5.9
예능	6.7
스포츠	6.9
뉴스	5.2
영화	8.1
전체 평균	6.7

그림 1에서 보면 장르에 따라 Scene과 Shot의 평균 개수는 모두 다르나 모든 장르에서 Scene에 비해 Shot이 상대적으로 많은 비율을 나타냄을 알 수 있다. 이는 영상의 기본 단위와도 관련이 깊다. 영상의 단위를 나눌 때 Scene이라는 단위 속에 Shot이라는 단위가 포함되어 있기 때문이다. 표 1에서 보면 비디오 콘텐츠의 장르는 다르나 Scene안에 속한 Shot의 평균 개수는 적게는 5개에서 많게는 8개로 어느 정도 일정한 범위 내에서 제작됨을 알 수 있다. 이러한 분석 결과를 토대로 우리는 사용자가 Scene과 Shot별로 Tagging을 쉽게 할 수 있도록 하는 UI 구성을 중심으로 다음과 같은 시스템을 제안한다.

3.2 시스템 설계

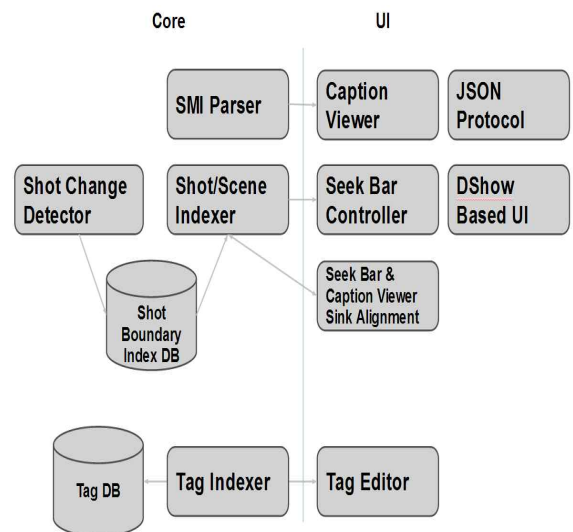


그림 2. 시스템 구조도

Scene/Shot Change Detector를 이용해 영상에 대한 Scene과 Shot의 범위를 나누고 이를 인덱싱 하여 데이터베이스에 저장한다. 인덱스 정보를 통해 Scene/Shot단위의 화면 출력 및 Tagging할 수 있는 tool을 제공한다. 이를 활용하여 Data 입력 시, 입력 결과가 데이터베이스에 저장되는 구조를 가지고 있다. 이러한 기본 구조를 바탕으로 Tagging tool에 있어 다음과 같은 주안점을 두고 제작하였다.

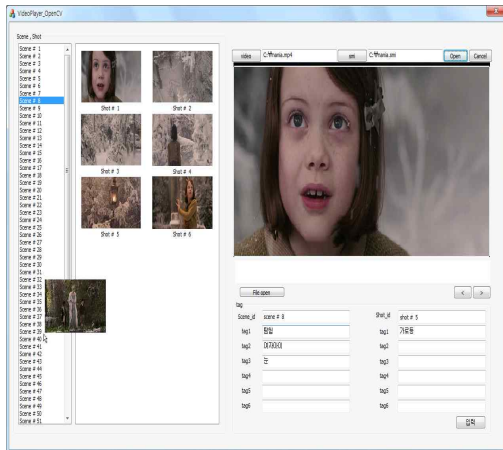


그림 3. Tagging tool

3.2.1 직관성

비디오 콘텐츠의 전체적인 구조를 한눈에 알 수 있도록 Scene은 리스트 형식으로 Shot은 이미지 형태로 나타낸다. Scene의 경우 그림1에서 볼 수 있듯이 그 수가 매우 많고, 여러 개의 Shot으로 구성되어 있기 때문에 하나의 frame 이미지로 각 Shot들의 내용을 나타낼 수 없다. 이러한 이유로 직관적 브라우징을 위해서 이를 리스트 형식으로 나타내었다. 대신 리스트의 마우스오버(mouse over)를 통한 하위 Shot들을 미리 볼 수 있는 Thumbnail을 제공한다. Shot의 경우, 표 1에서 알 수 있듯이 대부분의 Scene에 따른 Shot의 개수가 일정 범위 내에서 제작되고, 또한 하나의 frame 이미지로 Shot에 대한 정보를 대표해 줄 수 있으므로 이를 이미지로 구성한다.

3.2.2 편의성

사용자의 편의성을 도모하기 위해 ¹⁾Shot 이미지 선택 시 해당 영상으로 이동, ²⁾디렉터를 이용한 Scene과 Shot의 자동 검출 및 인덱싱 기능, ³⁾Scene/Shot change detection algorithm을 보완하기 위한 기능을 제공한다.

¹⁾영상의 특성상 Shot의 이미지만을 통해 정확한 정보를 알 수 없으므로 Shot 이미지를 선택하여 해당하는 영상으로 이동 및 재생 할 수 있는 기능을 제공한다.

²⁾비디오 콘텐츠는 그림 1에서 보듯이 상당히 많은 양의 Scene과 Shot을 가진다. 이를 사용자가 직접 검출하는 시간을 줄이기 위해 디렉터를 활용한 자동 검출 및 해당하는 Scene과 Shot의 인덱싱 기능을 제공한다[4].

³⁾Scene/Shot change detection algorithm을 이용하여 자동 검출한 정보가 정확하지 않을 경우, frame단위의 이동 tool을 제공하여, 사용자가 편리하게 직접 인덱스 정보를

수정할 수 있는 기능을 제공한다.

3.2.3 데이터베이스 구축

데이터베이스 구축을 통해 사용자가 저장한 Tagging Data를 검색할 수 있으며, Scene과 Shot의 고유 인덱스를 통해 이를 용이하게 한다. 또한 Data의 위치 정보 파악 및 해석을 가능하게 하며 수정을 통한 보완이 가능하다. 지속적인 정보 보완을 통해 좀 더 완벽한 Data를 생성하여 이는 신뢰도 있는 데이터베이스 구축과 직결 된다[5].

4. 결론 및 향후 연구

본 논문에서는 비디오 시맨틱 검색을 위해 Shot/Scene 단위로 Tagging이 가능하며 비디오 내용의 구조를 직관적으로 볼 수 있는 tool을 제공함으로써 불완전한 Shot Change Detector에 대해 비디오 인덱싱을 보다 정확하게 찾을 수 있도록 했다. 본 논문에서 제안한 시스템을 바탕으로 사용자들이 Shot/Scene을 일일이 찾는 수고를 덜 수 있고 세부적인 Tagging을 효율적으로 할 수 있도록 했다. 향후에는 방대한 멀티미디어 콘텐츠들 속에서 사용자들은 시맨틱 웹 검색을 통해 원하는 비디오 콘텐츠 정보들을 정확하고 신속하게 접근 가능할 것이다. 또 Shot/Scene 단위로 콘텐츠에 접근함으로써 사용자가 기존의 방송 콘텐츠나 자체 콘텐츠를 활용한 UEC(User Edited Contents)를 생성하는 등 다양한 사용자의 needs를 만족시킬 수 있다. 하지만 본 논문에서는 비디오 Shot/Scene 단위의 메타정보 입력에 있어 자동 Tagging이 되지 않는다는 한계가 있다. 향후 연구에서는 자동/반자동 Tagging이 가능하도록 하는 방법과 Tagging 키워드 추출을 위한 동영상 분석 알고리즘을 연구할 계획이다.

참고문헌

- [1] KLAMMA, Ralf, SPANIOL, Marc; RENZEL, Dominik, "Community-Aware Semantic Multimedia Tagging-From Folksonomies to Commsonomies," *Proceedings of I-Media*, pp. 163-171, 2007.
- [2] 임해철, "멀티미디어 콘텐츠의 강화된 의미 검색을 위한 온톨로지 설계," *인터넷정보학회논문지*, 13권 1호, pp. 107-115, 2012.
- [3] SMITH, John R, LUGEON, Blaise, "Visual annotation tool for multimedia content description in Information Technologies 2000," *International Society for Optics and Photonics*, pp. 49-59, 2000.
- [4] LEFÈVRE, Sébastien; VINCENT, Nicole, "Efficient and robust Shot change detection," *Journal of Real-Time Image Processing*, Vol.2 No.1, pp. 23-34, 2007.
- [5] 이태동, 김민구, "통합된 비디오 인덱싱 방법을 이용한 내용기반 비디오 데이터베이스의 설계 및 구현," *정보과학회논문지*, 7권 6 호, 2001.