

# 멀티미디어 콘텐츠의 지능형 선택/검색 시스템 구현

이중설, 이윤주, 박우철, 정하중, 조위덕  
전자부품연구원 인터넷미디어연구센터  
{jslee, yilee0618, wcpark, chunghj, chowd}@keti.re.kr

Korea Electronic Technology Institute

Jong-Sul Lee<sup>0</sup>, Yun-Ju lee, Woo-Chool Park, Ha-Joong Chung, We-Duke Cho

Internetmedia Research Center, KETI

## 요 약

멀티미디어 콘텐츠의 지능형 선택/검색 시스템(MISS: Multimedia Content Intelligent Selection/Search)은 콘텐츠를 공급하는 서버에 다량의 멀티미디어 콘텐츠들이 존재하며, 이 콘텐츠들 중에서 원하는 것을 검색, 선택하는 시스템이다. 지능적 검색, 선택기능을 갖는 MISS 시스템은 인터넷 및 네트워크상에 연결된 시스템들간의 맞춤형 서비스 구현에 필요한 핵심이며, 모든 종류의 멀티미디어 콘텐츠에 적용 가능하다. 현재 WWW 서비스 경우는 정보를 찾기 위하여 웹상에서 문서를 찾아주는 텍스트 기반 정보검색기술이 사용되고 있는데, 점점 우리가 접하는 정보의 형태는 텍스트와 함께 화상, 음성, 동영상 등의 멀티미디어화 및 디지털화하고 있다. 사용자들에게는 멀티미디어 데이터를 효과적으로 찾아야 하는 필요성이 증가하고 이에 따라 방대한 양의 분산된 멀티미디어 데이터를 처리할 수 있는 색인 및 검색 도구의 요구가 커지게 되었다. MISS 시스템은 WWW 서비스의 요구에도 적용될 수 있다. MISS 시스템은 다량의 동영상 콘텐츠중에서 특정 배우, 감독등의 여러 가지 검색 조건으로 콘텐츠를 검색/선택할 수 있고, 하나의 동영상 콘텐츠 내에서 특정 Video Segment를 검색할 수 있다. 본 MISS 시스템은 동영상에 대한 Search/Query를 위한 DS 구조로써 MPEG-7의 User preference metadata를 이용하였다.

## 1. 서 론

현재의 인터넷 시대는 네트워크의 발달과 함께 점점 더 많은 오디오-비디오 정보들이 늘어나고 있다. 이러한 멀티미디어 정보들은 사진, 비디오(동영상), 그래픽, 3차원 모델, 소리, 목소리 등 다양한 형태로 표현되고 있다. 이러한 멀티미디어 정보들은 우리 주변에서 많이 찾아 볼 수 있으며 디지털화 되어 사용되고 있다. 이러한 멀티미디어 정보의 증가와 함께 사람들은 이들 멀티미디어 정보를 생성하고, 가공하고, 교환하며 필요한 정보를 검색하고자 하는 요구가 늘어났다.

이러한 예로는 surveillance, intelligent vision, smart camera 등의 image understanding과 speech-to-text, picture-to-speech, speech-to-picture 등의 미디어 변환 뿐만 아니라, 사용자가 관심 있는 멀티미디어 정보를 빠르고 쉽게 찾아주는 멀티미디어 정보검색과 찾아진 결과를 사용자의 요구나 선호도에 맞도록 걸러주는 filtering 기능 등이 있다.

MPEG-7은 이러한 기능들을 지원할 수 있도록 멀티미디어 콘텐츠에 대한 표현 및 기술 방법을 표준화한다. MPEG-7은 어느 특정 응용 어플리케이션을 염두에 두고 있는 것이 아니라 MPEG-7에서 표준화하는 요소들이 가능한 많은 응용분야에 쓰일 수 있도록 하는데 목표를 두고 있다.

디지털방송과 전자도서관, 전자상거래 등에서 멀티미디어 데이터를 효과적으로 저장, 전송, 그리고 검색하기 위한 차세대 표준인 MPEG-7에 많은 기술들이 제안되었다. MPEG-7은 MPEG 그룹에서 ISO/IEC 표준으로 제안한 멀티미디어 콘텐츠 표현에 대한 인터페이스이다. 다시 말해 멀티미디어 콘텐츠가 갖는 내용을 해석하여

다른 컴퓨터에서 사용자가 찾고자 하는 내용을 검색할 수 있게 해준다. 여기서, MPEG-7은 어떤 특정 시스템을 위한 것이 아니라 광범위한 응용을 염두에 둔 요소들을 표준화하는데 있다.

## 2. MISS 시스템의 Search/Query/Filter 알고리즘

MPEG-7에서는 무수히 많은 Description Scheme(DS)에 대해서 정의하고 있으며, 해당 DS를 통해서 동영상에 대한 검색을 수행 할 수 있다.

본 MISS 시스템은 동영상의 classification DS를 통해서 유용한 검색이 가능한데 본 MISS 시스템에서는 저장된 동영상에 대한 메타데이터 DS의 검색을 위한 Keyframe, Publisher, Creator, Keyword, Title에 대한 검색을 수행한다.

그림 1과 같이 Client에서 검색할 DS name과 DS의 값을 전송하면 Search/Query 모듈은 User의 Search/Query 값을 가지고 콘텐츠 메타데이터들을 검색하여 원하는 콘텐츠가 존재하는지 확인한다. 이때 각 DS에 대한 Filter를 통해 해당 DS에 대한 검색을 실시하며 콘텐츠가 검색이 되면 콘텐츠의 저장 위치 및 전달 방법 등에 대한 정보를 사용자에게 전달한다.

## 3. MISS 시스템의 Search/Query/Filtering API 설계 및 구현

다음은 MISS 시스템에서 구현된 Creator, Publisher, Title, Keyword, KeyFrame 등에 대한 Search/Query API 이다.

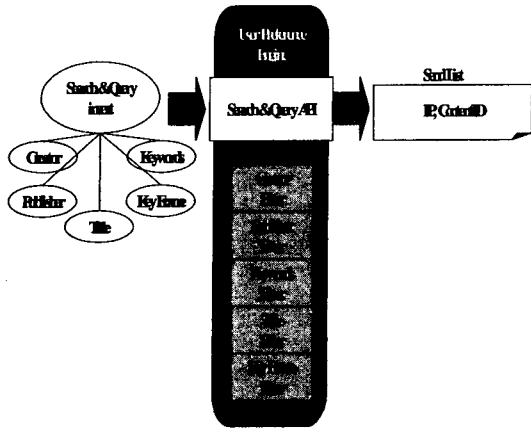


그림 1 MISS 시스템의 DS에 대한 Search & Query 구조

| 함수명   | 기능  |
|---|---|
| public Vector SearchTitle(String value)     | value에 입력되어진 DS 값과 일치하거나 부분 포함한 Title에 대한 검색을 실시한다.     |
| public Vector SearchPublisher(String value) | value에 입력되어진 DS 값과 일치하거나 부분 포함한 Publisher에 대한 검색을 실시한다. |
| public Vector SearchCreator(String value)   | value에 입력되어진 DS 값과 일치하거나 부분 포함한 Creator에 대한 검색을 실시한다.   |
| public Vector SearchKeyFrame(String value)  | value에 입력되어진 DS 값과 일치하거나 부분 포함한 KeyFrame에 대한 검색을 실시한다.  |
| public Vector SearchKeyword(String value)   | value에 입력되어진 DS 값과 일치하거나 부분 포함한 Keyword에 대한 검색을 실시한다.   |

표 1. MISS 시스템에서 Search/Query API

### 3.1 픽셀 밝기 차이에 의한 장면 전환 검출

영상의 픽셀 밝기(intensity)를 차이 측정 측도에 사용한다. 만약 프레임간의 픽셀 밝기의 차이가 임계치보다 크다면 서로 다른 셋에 존재하는 프레임으로 판단한다. 차이 측정 측도는 다음과 같다.

$D_i(x,y)$ 는  $i$ 번째 프레임에서의 차이 측도,  $F_i(x,y)$ 는  $i$ 번째 프레임에서의 밝기값,  $M, N$ 은 프레임의 크기를 나타낸다. 여기에서 프레임에 대한 임계치  $t$ 는 각 픽셀  $(x,y)$  위치에서의 밝기 차이에 적용시키는 임계치이다. 즉 픽셀의 밝기차가  $t$  이상인 픽셀들은 연속된 두 프레임이 서로 다른 셋에 존재할 확률을 높이는 데 기여하는 픽셀로 인식한다.

$$D_i(x,y) = \begin{cases} 1 & \text{if } |F_{i-1}(x,y) - F_i(x,y)| > t \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\sum_{x,y=1}^{M,N} D_i(x,y)}{M \cdot N} \times 100 > T$$

반대로 밝기차가  $t$ 보다 작으면 동일한 셋 내부의 두 프레임의 확률을 높이는 데 기여하는 픽셀이다.  $D_i(x,y)$ 에 의해 두 프레임 간의 차분 영상에 대한 이진화가 수행된다. 임계치  $T$ 는 이진화된 영상에서  $D_i(x,y)$ 가 1인 픽셀들의 개수를 조사하기 위해 설정되는 임계치이다. 따라서 전체 프레임 크기에 대한 1 값을 갖는 픽셀들이 차지하는 비율이  $T$  이상인 경우에 실제 컷이 발생한 지점으로 결정한다.

픽셀 밝기 차이에 의한 장면 전환 검출 방법은 대체적으로 움직임(motion)이 거의 없는 동영상 시퀀스에 대해서는 셋 경계를 잘 검출해 내지만 카메라나 물체의 움직임에 민감하게 반응한다는 문제점이 있다. 즉 셋 내부에서 카메라나 물체의 움직임 영향으로 발생한 픽셀차에 의한 방법은 차이 측정 측도가 간단하고 구현하기가 쉬우나 점진적으로 검출할 만한 정교한 처리를 하지 못하는 문제가 있다.

### 3.2 히스토그램 비교법에 의한 장면 전환 검출

히스토그램 비교법은 장면 전환 검출 방법의 대표적인 방법으로 영상의 밝기(intensity)나 컬러(color)의 확률 분포도인 히스토그램을 구성한 후 히스토그램간의 차이를 구하는 방법이다. 히스토그램 비교법에 사용되는 차이 측정 측도는 다음과 같다.

$D_i$ 는 장면 전환 검출을 위한 차이 측정 측도이고  $j$ 는 픽셀의 밝기 구간(0-255)을 나타내고  $G$ 는 최대의 밝기값을 나타낸다.  $H_{i-1}$ 은  $i-1$ 번째 프레임의 밝기 히스토그램이고

$$D_{i(c)} = \sum_{j=0}^G |H_{i(c)}(j) - H_{i-1(c)}(j)| > T$$

$$D_i = \sum_{c \in \{R,G,B\}} D_{i(c)}$$

$H_i$ 는  $i$ 번째 프레임의 히스토그램에 해당된다. 따라서 두 히스토그램간의 차이의 합이 임계치  $T$  이상인 경우가 셋 경계 지점이다. 그리고 컬러 히스토그램을 이용할 경우에는 식 3.2.2에서와 같이 각각 RGB 별로 히스토그램을 구성하여 차이를 계산한다.

히스토그램 비교법은 픽셀 밝기 차이에 의한 검출법에서 문제가 되었던 움직임에 민감하다는 단점을 어느 정도 극복하고 있다. 즉 히스토그램은 영상의 전체적인 분위기를 파악하는 것으로 영상을 이해하는데 있어서 영상 내에 물체나 카메라의 움직임이 존재하는 경우라도 밝기 분포나 색상 분포는 거의 일정하기 때문이다.

히스토그램 비교법은 가장 많이 사용되고 있는 방법이고 다양한 차이 측정 측도를 설계하여 응용에 따라 적용시킬 수 있으나 공통적으로 몇 가지 문제점을 가지고 있다. 첫째로 영상

에서 차지하는 물체의 비율이 적고 셋 간에 배경의 변화가 없을 경우 또는 서로 다른 셋에 존재하는 프레임이라도 비슷한 밝기 분포를 갖게 되는 경우에는 히스토그램 비교법으로 셋 경계를 검출할 수 없다. 이 경우에는 셋 경계 지점을 대부분 놓치게 된다. 또한 큰 물체가 갑자기 등장하는 장면이나 밝기 변화가 급격하게 발생하는 장면에서는 동일한 셋 내부의 프레임인데도 불구하고 셋 경계로 잘못 판단하는 문제가 발생한다.

MISS 시스템에서는 위 두가지 알고리즘을 적용한 키프레임 추출 API를 이용하여 장면전환에 대한 키프레임을 생성하고 이때 각 키프레임에 대한 TimePoint값을 추출하여 AV컨텐츠의 세그먼트에 대한 Script를 생성하는데 사용한다.

#### 4. MISS 시스템의 사용자 인터페이스

본 MISS 시스템에서 클라이언트 사용자 인터페이스는 컨텐츠 메타데이터를 이용하여 컨텐츠를 검색하고, 선택한 동영상 컨텐츠를 재생할 수 있는 사용자 환경을 제공한다. 검색 결과는 리스트 형태로 제공되며, 컨텐츠 데이터와 컨텐츠에 대한 메타데이터 정보를 사용자 상호작용에 의해 동영상과 이미지, 텍스트의 다양한 형태로 화면에 디스플레이 하고, 사용자 이벤트에 의한 실시간 처리를 제공하여야 한다. 클라이언트 사용자 인터페이스는 자바 플랫폼의 그래픽 사용자 인터페이스 개발로 다양한 단말 플랫폼에서의 연동을 지원한다.

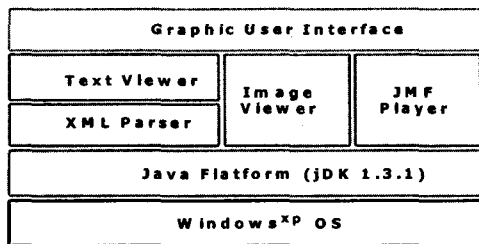


그림 2. MISS 시스템의 클라이언트 구조도

다음 그림 3은 서버의 컨텐츠 데이터 베이스와 클라이언트의 사용자간에 컨텐츠 검색 요구와 컨텐츠 데이터 흐름을 나타낸 서버-클라이언트 구조도이다.

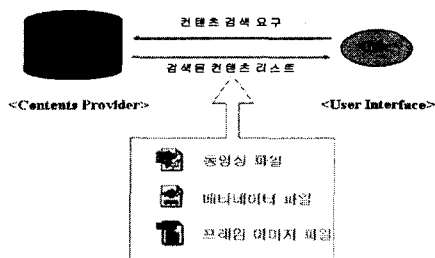


그림 3. MISS 시스템의 서버-클라이언트 구성도

사용자 컨텐츠 검색 결과에 따른 컨텐츠 리스트와 각 해당 컨텐츠와 메타데이터 정보를 각각 동영상과 이미지, 텍스트로 화면에 디스플레이 한다. 사용자의 컨텐츠 선호도에 맞게 추출, 저장된 컨텐츠를 보여주는 컨텐츠 리스트와 동영상 플레이어, 이미지 뷰, 텍스트 뷰로 구성된다. 사용자가 컨텐츠 리스트에 나타난 컨텐츠 중에서 원하는 컨텐츠 바에 버튼을 클릭하면, 해당 컨텐츠의 동영상이 플레이되고, 컨텐츠에 관련된 이미지와 텍스트 정보가 디스플레이 된다.

동영상 플레이어는 재생, 정지, 일시정지, 종료의 기능을 지원한다. 다음은 검색 분류 항목을 장르로, 검색어를 스포츠로 선택한 컨텐츠 검색 결과 화면의 예이다.

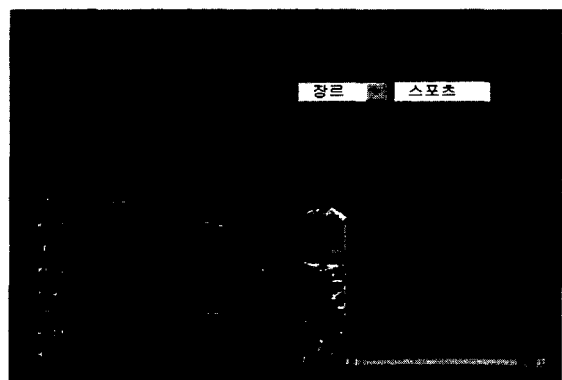


그림 4. MISS 시스템의 컨텐츠 디스플레이

#### 5. 결론

본 MISS 시스템은 동영상에 대한 Search/Query를 위한 DS 구조로써 MPEG-7의 User preference metadata를 이용하였으며, 동영상에 대한 키 프레임 정보들, 즉 메타데이터를 같이 제공함으로써 사용자로 하여금 전체 영상을 다 재생시키지 않고도 사용자가 관심있는 컨텐츠를 선택, 재생 할 수 있는 기능을 제공하였다. 향후과제로는 사용자의 보다 다양한 쿼리를 반영할 수 있는 연구가 필요하다.

#### 6. 참고 문헌

[1] XML Schema version 1.0 <http://www.w3.org/XML/scheme>  
 [2] Text of ISO/IEC CD 15938-5 Information Technology - Multimedia content description Interface - Part 5 Multimedia Description Schemes, N3705, La Baule, France, Oct. 2000.  
 [3] The TV-Anytime Forum <http://www.tv-anytime.org>  
 [4] MPEG-7 Multimedia Description Schemes XM(v5.0), La Baule, France, Oct. 2000.  
 [5] 박우철, 이종철, 이윤주, 이석필, 조위덕, "컨텐츠 메타데이터와 사용자 선호도에 의한 지능화 EPG 서비스 시스템", 한국통신학회 하계 학술대회 2002. 7.