

# MPEG-7을 이용한 유동적인 DS의 설계 및 구현

이용남<sup>o</sup>, 고재진, 최기호  
광운대학교, KETI, 광운대학교

## Design and Implementation of Flexible DS based on MPEG7

YongNam Lee, JaJin Ko, KiHo Choi  
Kwangwoon Univ. , KETI , Kwangwoon.  
E-mail : yonalee@korea.com, jaejini@keti.re.kr, khchoi@gwu.ac.kr

### 요 약

기존의 내용기반 검색 시스템에서는 사용자의 개입이 고려되지 않은 채 Low level 특징만을 이용한 연구가 진행되어 왔다. 본 논문에서는 내용 기반 기술 인터페이스인 MPEG-7을 이용하여 주관적인 비디오 구조 재구성 기술과 기술을 보장하는 유동적인 DS를 설계하고 구현하였다.

### 1. 서론

기존의 검색 시스템은 사용자에게 “End-user만의 역할”만을 기대한다. 이런 이유로 검색시스템을 만든 사람과 사용자의 표현의 차이로 인해서 사용자는 원하는 결과를 정확히 얻어내기가 어렵다. 또한 WWW 상에서는 아직도 텍스트 정보에 대한 검색만이 가능한 상태여서 사용자가 원하는 시청각 정보의 실질적인 내용에 대한 검색은 제공되지 않고 있다. 다만 멀티미디어 데이터 베이스에서 영상의 색상, 질감, 물체의 모양에 대한 정보 등의 하위레벨 특징을 이용한 검색이 가능하다. 또한 현재 대부분의 검색프로그램은 HTML을 이용한 문서기반의 정보를 검색하고 있기 때문에 영상, 동영상 등을 참조할 수 있으나 영상, 동영상 내에 설명을 덧붙일 수는 없다. 더구나 텍스트 설명 자체도 설명을 만드는 사람과 검색하는 사람과의 관점 차이로 인해 검색의 효율성이 떨어질 수 있기 때문에 더욱이 멀티미디어 자료에 대한 내용기반 검색은 필요하다. 이러한 문제점을 해결하고

자 MPEG위원회에서는 “Multimedia description Interface”로 MPEG-7을 만들어 세계적 표준안으로 제정하려 한다. 이것은 오디오, 비디오 정보의 기술과 색인, 접근을 가능하게 하는 표준안으로, 다양한 방법으로 기술되던 멀티미디어 데이터에 대한 기술방법을 통일함으로써 사용자로 하여금 보다 쉽게 멀티미디어 데이터를 활용하고자 하는 것에 그 목적이 있다 하겠다. 본 논문에서는 MPEG7을 이용하여 그동안 검색 시스템에서 제외된 사용자에게 중심으로 스스로 비디오 구조를 재구성하게 하고 그 구조를 디스크립션하게 함으로 해서 보다 “개인적인”이고 “유동적인” 검색 시스템을 구현할 수 있도록 한다.

본 연구에서는 기존의 내용기반 검색에서 이용된 Low level 특징을 이용하여 장면을 분할하는데 사용하고, 유동적인 DS(Description Scheme)을 이용하여 Structural DS를 재구성한다. 이를 참조하여 사용자의 주관적인 기술 방법에 의해 High level 내용을 기술해서 사용자의 요구에 맞는 비디오 검색 시스템을 구현한다.

## 2. 관련연구

### 2.1 MPEG-7

다양한 형태의 멀티미디어 정보의 디스크립션 즉 표현을 표준화하는 것으로 이 디스크립션은 콘텐츠 자체와 연결되어 사용자가 관심있는 멀티미디어 자료를 바르고 효율적으로 찾을 수 있게 한다. 표준은 디스크립션 또는 특징을 자동으로 추출하는 것은 제외하며 디스크립션을 사용하는 검색 엔진이나 기타 다른 프로그램도 표준의 대상이 아니다. 다시 말하면 MPEG7은 다양한 형태의 멀티미디어 정보를 기술하기 위해 사용되는 디스크립터 셋(Descriptor Set)을 결정하는 것이다. MPEG7은 디스크립터와 그들 간의 관계에 대한 다른 D(Descriptor)와 DS(Description Schemes)은 사용자가 원하는 자료를 빠르고 효율적으로 찾아줄 수 있도록 콘텐츠 자체에 연결한다. 또한 이러한 관계를 기술하는 언어로써 DDL을 표준화한다.

#### 2.1.1 MPEG-7의 Terminology

MPEG7을 구성하는 요소 가운데 가장 중요한 것은 다음과 같다.

첫째로 D(Descriptor)는 멀티미디어 콘텐츠가 갖는 가장 기본적인 특성이 하위레벨특징의 표현을 기술한다. 한가지 주의할 점은 모든 디스크립터가 자동으로 특징을 추출하지는 않는다는 것이다.

두 번째로 DS(Description Scheme)은 D의 구조적 조합으로, 이 구조는 문서를 annotate하는데 사용될 수 있고, 또 몇몇 특징의 조합을 통하여 보다 상위 레벨의 개념을 잘 표현할 수 있게 해 준다. DS는 오디오 비주얼 데이터의 일반적인 특성과 함께 응용 영역 세부적인 특성을 함께 수용할 수 있도록 설계되어 있으며, 일반적인 DS는 멀티미디어 데이터의 변하지 않는 메타데이터를 기술하는데, 이들 메타데이터는 제작, 생성, 사용과 관리 등과 관련된 정보들이다. 오디오 비주얼 데이터 콘텐츠의 특정 DS는 이들 신호적 특성이나 특징, 모델 및 시맨틱을 표현하

는 방법을 제공한다. 다른 DS들로는 이들 정보를 이용하여 효과적으로 네비게이션하고 접근할 수 있는 방법을 제공한다.

세 번째로 DDL(Description Definition Language)는 사용자가 생성한 DS나 D에 대한 멀티미디어 표현의 기초를 제공한다. DDL은 DS나 D를 합칠 수 있는 구문 규칙을 제공한다. MPEG-7 Requirements 문서에 의하며, DDL은 새로운 DS나 D를 만들 수 있는 언어이며 기존의 DS를 확장 변형 할 수 있게 한다고 정의되어 있다. DDL은 UML과 같은 모델링 언어가 아니라 DS, D로 오디오 비디오 데이터를 모델링한 결과를 표현하는 Schema언어이다.

DDL은 MPEG7 DDL 요구사항을 만족해야 한다. DDL은 DS들의 element들과 DS들간의 spatial, temporal structural, and conceptual 관계를 표현할 수 있어야 한다. 하나 또는 그 이상의 디스크립션과 이것이 기술하는 데이터들 간의 링크나 레퍼런스에 대한 풍부한 모델을 제공해야 한다.

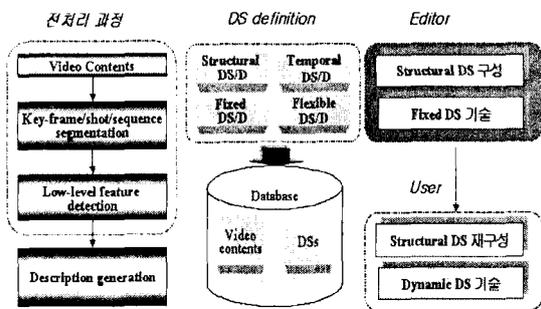
### 2.1 내용기반검색

내용기반 검색의 의미는 이미지, 비디오, 오디오 등 멀티미디어 자료를 검색할 때 자료에 추가된 주석을 이용하여 검색하는 텍스트 기반 검색과는 달리, 멀티미디어 자료를 분석한 특징 값을 이용함으로써 검색하고자 하는 내용에 유사한 멀티미디어 자료를 검색하는 기술을 말한다.[1] 현재 텍스트 검색 엔진 시장에는 6700여개의 서버를 가진 구글(google)이라는 성능 좋은 엔진이 등장했지만, 멀티미디어 검색을 하고자 할 때는 텍스트 설명을 만드는 사람과 검색하는 사람과의 관점 차이로 검색의 효율이 떨어질 수 있기 때문에 기존의 키워드 기반의 정보 검색보다 사용자가 원하는 정보의 내용에 기반한 내용기반검색이 요구된다. 동영상내 객체의 움직임을 찾아내는 VideoQ 시스템에서는 사용자가 움직이는 스케치(animated sketch)질의를 한다. 하지만 이 시스템에서 사용자는 이미 비디오 검색 시스템 제공자에 의해 결정되어진 시퀀스(sequence) 결과만을 보게 된다.

본 논문에서는 프레임웍으로 제공된 MPEG7의 핵심인 DDL을 자동으로 완성해주는 시스템을 구현한다.

### 3. 제안된 시스템

본 논문에서 구축하고자 하는 비디오 검색 시스템은 기존의 컬러 정보를 이용해 장면전환을 검출한다. 이렇게 하위레벨 특징을 이용하는 것은 기존의 내용기반검색에서 취했던 방식으로 자동화 단계에 있으나 인간의 인지적인 능력이 고려되지 않고, 그림이 가지고 있는 특징만을 이용한다. 더구나 가장 많이 활용되는 칼라 히스토그램의 경우 데이터베이스 양이 증가할수록 유사도 계산에 따른 검색시간이 증가하고 검색 효율이 떨어지는 단점이 지적되어 항상 보완이 요구되곤 한다. 이러한 단점을 보완하기 위해서 사용자의 디스크립션이 가능하도록 비디오 구조를 재구성하고, 사용자의 개입을 허락한다.



[그림 1] 제안된 비디오 검색시스템

#### 3.1 전처리 과정

장면분할 하는데 가장 많이 쓰여지는 방법으로 컬러 히스토그램을 이용하는데, 이미 지적한 바와 같은 단점에도 불구하고 프레임표현의 가장 중요 요소인 칼라 분포율을 이용하며, 조명변화, 크기변화 등에 민감하지 않다는 장점을 충분히 활용한다. 또한 대부분의 서로 다른 프레임은 뚜렷이 구별되는 칼라

히스토그램을 갖는 것도 칼라 히스토그램이 많이 사용되는 이유이다. 컬러 히스토그램은 다음 식 3.1과 같이 나타낼 수 있다. 여기서  $H_i$ 는 화상 안에서 컬러  $I$ 의 전체 칼라 히스토그램을 나타내며,  $n$ 은 칼라 버킷의 수이다. 칼라 히스토그램의 차이를 구하는 식은 일반적으로 3.2가 사용된다.

$$H_i = \sum_{i=1}^n h_i \quad \text{식 3.1}$$

$$d(I, I_{...}) = \sum_{j=1}^n |H_i(j) - H_{...}(j)| \quad \text{식 3.2}$$

동영상 정보를 이용하는데 있어 위와 같은 컬러 히스토그램을 이용하여 자동으로 논리적 정보인 샷을 구별해 내며, 샷은 시간과 공간 안에서 연속적인 행동을 표현하는 연속된 프레임에 지칭한다. 따라서 샷을 구분함으로써 프레임에 포함된 정보의 내용 변화를 알아낼 수 있다. 이렇게 해서 물리적으로 키프레임, 샷, 시퀀스로 구별한다.

#### 3.2 DDL Generate

기존의 내용기반 검색에서 하위레벨 특징의 추출은 자동화되어 있다. 그러나 표현방법이 통일된 MPEG7에서는 이러한 하위레벨 뿐 아니라 상위레벨 내용까지 기술할 수 있으므로 본 논문에서는 추출된 특징 및 특징 값. 그리고 다음에 소개한 네 개의 DS를 연결하여 DDL의 자동화시스템을 구현한다.

#### 3.3 제안된 DS

본 논문에서는 기존의 컬러(color), 텍스처(texture), 모션(motion) 등의 하위레벨 특징에 초점을 두었던 내용기반 검색을[6][8] 개선해 인간의 개념적인(perceptual) 기술에 의한 검색 시스템을 구현한다. 본 논문에서 제안한 네 개의 DS는 다음과 같다.

##### 3.3.1 Structural DS

일반적으로 장면분할 작업을 통하여 동영상을 구

조화하는데[7] 특히, 장면전환이 이루어진 프레임은 대표 프레임으로 상아 키프레임(key-frame)으로 지정한다. [1] 장면 전환에는 칼라에 대한 분포율을 정보로 사용하고, 계산이 빠르며, 카메라의 위치 변화, 조명, 크기 변화에 민감하지 않은 방법으로 칼라히스토그램을 이용한다. [7] 칼라히스토그램은 식(1)과 같이 나타내며, 이들의 차이를 구하는 식은 식(2)를 주로 사용하며, 칼라공간 별로 따로 계산하기 위해 식(3)을 사용했다.

$$H_i = \sum_{i=1}^n h_i \quad \text{식(1)}$$

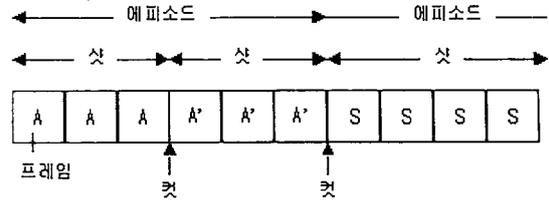
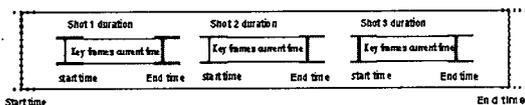
$$d(I_{i,j}, I_{i,k}) = \sum_{j=1}^n |H_i(j) - H_{i,k}(j)| \quad \text{식(2)}$$

$$d(I_{i,j}, I_{i,k}) = \sum_{j=1}^n (|H_i(j) - H_{i,k}(j)| + |H_j(i) - H_{i,k}(i)| + |H_k(i) - H_{i,k}(i)|) \quad \text{식(3)}$$

$H_i$ 는 화상 안에서 칼라  $i$ 의 전체 칼라 히스토그램을 나타내며,  $n$ 은 칼라 버킷의 수이다. 장면전환이 검출되는 임계값은 반복실험을 통해 48%이상의 오차를 장면전환으로 검출하며, 키프레임을 선택한다. 또한 비디오 편집자 또는 프로듀서에 의해, 시간과 줄거리 중심의 시퀀스(sequence)로 나누고, 그 안에서 다시 이벤트 중심으로 샷(shot)을 구별해 낸다. 각각은 고유한 인덱스로 정렬한다.

### 3.3.2 Temporal DS

멀티미디어 데이터는 그 자체가 시간의 순서로써 표현되어진다. 따라서 사건의 시간적 순서를 확실히 하여 사건과 스토리를 동기화 시켜 전개시킨다. 그림 2 에서처럼 시간의 흐름에 의한 비디오 데이터의 특성을 고려하여 Structural DS에서 결정된 키프레임의 현재시간(current time)과 프로듀서 및 사용자에 의해서 선택된 샷과 시퀀스의 기간(duration)으로 동기화 시켜 검색시 디스플레이한다.



[그림 2] Temporal DS와 Structural DS의 동기화

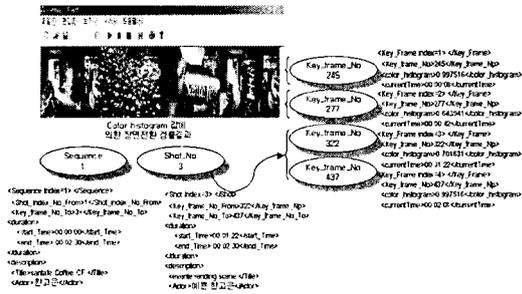
### 3.3.3 Union DS

Union DS는 처음 AVI파일을 저장하고 그 위치를 등록시키며, 컨텐츠에 대한 기본적인 정보들을 디스크립션 하기 위해 구성된다. 이 정보는 제작자가 직접 만들며, 사용자나 편집자 모두에게 공개된다. 그러나 이 정보는 수정이 불가능하고 다만 Read-only 정보로만 사용된다. 여기에 특별히 제작자에 대한 내용이 들어감으로 해서 제작자를 보호하는 기능도 함께 한다. 전체 구성은 비디오 파일의 위치, 파일의 이름, 파일의 위치 등의 Structural 한 논리적 정보들과 Actor등의 메타 정보들이다. 이러한 정보들은 제작과정에서 만들어지며, 사용자는 임의로 수정할 수는 대신 주어진 정보를 이용해 사용자 디스크립션을 할 때 도움을 얻거나 인용이 가능하다.

### 3.3.4 Fixed DS

Fixed DS는 DDL에 의해 DS가 만들어진 후 어플리케이션을 통해 편집자에 의해서 기술되는 Descriptor의 Value를 저장한다. 이 같은 정보는 기본적인 Union Description Scheme에서 제공되는 정보 이외에 추가적인 정보를 별도로 관리하게 된다 이는 편집자가 제작자와 다를 경우 제작자에 의해 의도된 Description과 다른 의도의 Description이 필요하게 될 경우 이용할 수 있으며, 각각에 대한 저작권자를 명시함으로써 저작권에 대한 분쟁 해결에 이용될 수 있다. 또한 서로 같은 경우라 할 지라도 사용자에게 제공될 필요가 없는 내용들을 기술함으로써 보다 효율적인 비디오 관리가

가능하다. 비디오 검색 시스템을 만드는 사람(프로듀서, 편집자 등)의 입장에서, 제목, 부제목, 날짜, 저작권자, 형식, 요약 기술(Summary Description-제작일자, 배우, 제작 스튜디오, 제작시간) 등의 D를 기술한다. 이 DS와 D는 사용자나 프로듀서 등에게 검색어로 사용된다. 본 논문에서는 제목, 부제목, 날짜, 제작자(저작권 소유자), 포맷, 주연배우 등을 엘리먼트로 삼아 기술한다. 이것은 시퀀스 처리에서 얻어진 시퀀스의 인덱스를 중심으로 디스크립션 된다.



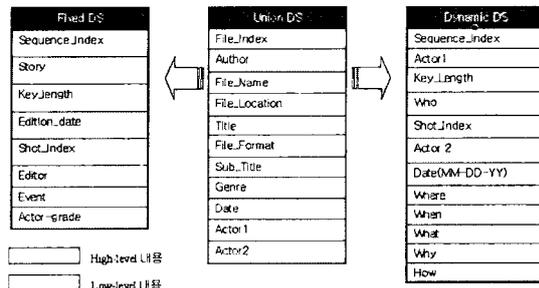
[그림 3] Structural DS의 요소들 간의 관계

### 3.3.5 Fixed DS

본 논문에서 핵심적인 역할을 수행하게 될 Dynamic Description Scheme은 제작자나 편집자에 의해서 제공된 Union Description Scheme과 Fixed Description Scheme을 검색에 이용하지 않을 경우나 혹은 사용자의 의도에 맞는 검색이 어려운 경우 등을 해결하기 위한 목적으로 제안되었다. 사용자는 제작자에 의해 주어진 Union DS를 검색에 그대로 이용하기 보단 사용자 본인의 감성과 취향에 맞는 Description이 가능해짐으로 해서 보다 Personal한 검색기반을 만들 수 있다. 이와 같은 제안은 기존의 내용기반검색에서 배제되어 최하위선에서 End User로만 참여했던 것을 개선해주는 것이다. 또한 이것은 시퀀스 처리에서 본인의 선택에 의해 인덱싱 된 시퀀스 정보를 보고 디스크립션 한다는 장점을 가지고 있다.

### 4.2 DS의 정의

Low level 내용과 High level 내용을 유기적으로 연결하여 사용자의 기술에 의해 검색하고자 하는 것이 본 논문의 목적이므로 이러한 DS들 사이는 다음과 같다.



[그림 4] DS들 사이의 관계

## 4. 구현

### 4.1 전처리

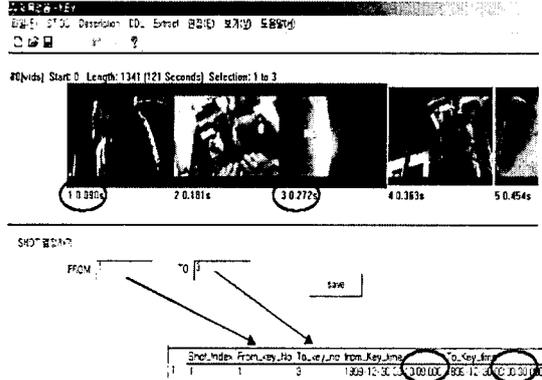
칼라 히스토그램을 사용하여 키 프레임을 검출해 내고 시간으로 동기화 시킨다. 가장 기본적인 논리적 단위인 키 프레임을 보고 샷과 시퀀스를 구성하게 하며, 이것은 편집자나 사용자에 의해 결정되어진다. [그림 3]의 결과는 키 프레임을 검출해 내고 이것을 샷과 시퀀스로 구성했을 경우의 관계에 대해 설명하고 있다.

### 4.4 Flexible DS

로우레벨 특징에 의한 키 프레임 재구성해 시퀀스를 완성해서 Structural DS를 기술하고, 이벤트 중심의 샷 디스크립션과 줄거리 중심의 시퀀스 디스크립션을 기술할 수 있도록 지원한다

이것은 Fixed DS에게도 공통적으로 지원이 되나 사용자의 입장에서 기존 검색 시스템을 사용하는 입장에서 비디오 구조를 재구성할 수 있고, 이를 이용하는 기술이 가능하다는 점에서 사용자에게는 Flexible한 비디오 관리 시스템 및 검색 시스템 구성이 가능하다.

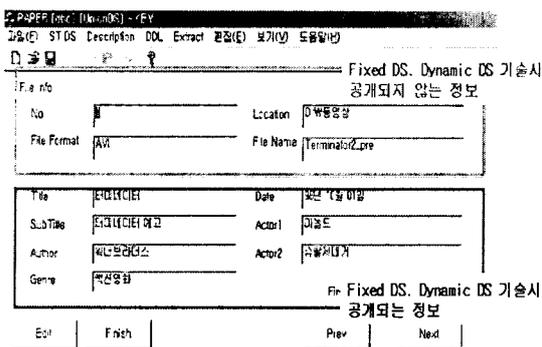
다음 [그림 5]의 구현결과 1과 2는 실제로 터미네이터 2의 예고편을 가지고 실제 응용한 것을 보여준다.



[그림 5] 구현 결과

#### 4.5 디스크립션

위와 같이 Structural DS과 Temporal DS를 통해 Low level 내용을 기술한 후 사용자에게 High level 내용을 기술할 수 있도록 디스크립션의 기회를 제공한다. [그림 6]은 이와 같은 구현 결과를 보여준다. Fixed DS와 Dynamic DS기술시 공개되지 않는 정보는 키프레임 추출시 이미 결정되거나 혹은 디스크립션되는 기본적인 정보들이다.



[그림 6] 디스크립션 구현

#### 5. 결론 및 고찰

기존의 검색 시스템에서 사용자의 적극적인 개입을 유도해 개인적인 비디오 관리 및 검색 시스템을 구현할 수 있다. 이는 디지털 방송등에서 PVR 어플리케이션을 가능하게 할 수 있다.

DDL 완성 자동화 시스템은 MPEG-7 내용기반 검색을 위한 자동화에 기여할 것이다. 다양한 하위레벨 특징과 상위레벨 특징을 이용해 정확하고 빠른 내용기반 검색을 제공할 수 있는 시스템을 구현할 필요가 있다. 또한 멀티미디어 데이터 크기와 처리는 엄청난 속도를 요구하기 때문에, 근본적인 멀티미디어 저장 및 관리도구에 대한 연구가 진행되어야 할 것이고 멀티미디어를 위한 형식(format), 저장소(store), OS에 관한 연구가 활발히 이루어져 멀티미디어 처리에 대한 연구가 병행되어야 할 것이다.

- [1] V.Gay and B.Dervella, MHEGAM-A Multimedia Messaging System, IEEE Multimedia, Oct.-Dec. 1997, pp.22-29
- [2] ITU-T Recommendation H.245 : Control of communication between Visual Telephone Systems and Terminal Equipment, <http://www.itu.int/>, 1998
- [3] ITU-T Recommendation H.323 : Packet Based Multimedia Communication Systems, <http://www.itu.int/>, January 1993
- [4] IETE RFC 1889 RTP : A Transport Protocol for Real-Time Application, January 1996
- [5] IETE RFC 1890 RTCP : RTP Profile for Audio and Video Conferences with Minimal Control January 1996
- [6] R.Steinmetz and K.Nahrstedt, Multimedia: Computing, Communications, and Applications, Prentice Hall, Upper Saddle River, N.J., 1995
- [7] 고영근, 제한적 자원에서의 분할기반 멀티미디어 모델 및 동기화 기법에 관한 연구, 박사학위논문, 연세대학교 컴퓨터학과, 1997