

## MPEG-7 기술자를 이용한 영상 검색 시스템 구현

이 희 경, 정 용 주, 윤 정 현, \*강 경 욱, 노 용 만  
한국정보통신대학원대학교 영상 & 비디오 시스템 연구실, \*한국전자통신연구원  
전화 : 042-866-6189 / 핸드폰 : 011-9352-8587

### Implementation of Image Retrieval System Using MPEG-7 Descriptors

Hee-Kyung Lee, Yong-Ju Jung, Jeong-Hyun Yoon, Kyung-OK Kang, Yong Man Ro  
Image & Video System Lab ,  
Information & Communication Graduate University  
E-mail : lkh95@icu.ac.kr

#### Abstract

In this paper, a multimedia database retrieval system is proposed using MPEG-7 meta data. Multimedia content based retrieval system is implemented with the MPEG-7 meta data extraction and matching technique. MPEG-7 descriptors and descriptor schemes are stored into the database with other meta data. When a query image is given, the descriptors and descriptor schemes of the query image are extracted and compared with the descriptors and descriptor schemes in the database. Finally, images having more similarity are retrieved.

리, 검색 및 인식의 필요성에 의하여 진행중인 MPEG-7 표준화 과정에서는 영상의 내용 특징에 기반한 다양한 검색 알고리즘들이 개발되고 있다. 그리고 영상 데이터를 기존의 웹 검색 엔진들이 제공하던 문자 기반 검색 뿐 아니라 이미지 기반 검색이 가능하도록 하는 검색 엔진의 개발이 중요한 화두로 제시되고 있다.

본 논문에서는 MPEG-7 표준 모델 중 영상에 대한 질감, 색 기술자를 기반으로 하는 통합 검색 모델을 개발하고, 이를 이용하여 인터넷에서 이미지 기반 검색이 가능하도록 하는 웹 기반 이미지 검색 엔진을 구현하였다.

#### 1. 서론

현재 컴퓨터와 인터넷의 발달과 더불어 멀티미디어 데이터는 기하급수적으로 증가하고 있으며, 특히 전세계 누구나 접근이 가능한 인터넷상의 멀티미디어 데이터는 그 수를 다 헤아릴 수 없을 정도다.

이러한 방대한 양의 멀티미디어 데이터의 효율적 관

#### 2. 본론

##### 2.1 MPEG-7

MPEG-7은 MPEG에 의해 개발된 ISO/IEC표준으로, 공식적으로 '멀티미디어 내용 기술 인터페이스'라 불린다. MPEG-7은 다양한 종류의 멀티미디어 정보를 기술하는 방법에 관한 표준이며, 각 기술자들은 멀티

미디어 데이터가 갖는 내용에 기반을 두고 만들어졌다. MPEG-7은 메타 정보를 멀티미디어 데이터 안에 삽입할 수 있는 방법을 제공함으로써 보다 빠르고 효율적인 내용 기반 검색이 가능하게 했다. MPEG-7은 멀티미디어 정보 기술자인 D(descriptor) 뿐 아니라 기술자들간의 관계 표현을 위한 구조인 DS(description schemes), DS를 정의하는 데 사용하는 DDL(description definition language)에 관한 표준화 방안[1][2]도 마련 중이다.

2.2. MPEG-7기반 데이터베이스 시스템

MPEG-7이 표준화 중인 대표적인 기술자에는 질감, 모양, 색, 움직임 등이 있다. 본 논문에서는 그 중 질감과 색 기술자를 함께 사용하여 이미지 검색 시스템을 구현하였으며, 시스템 구조는 그림 1과 같다.

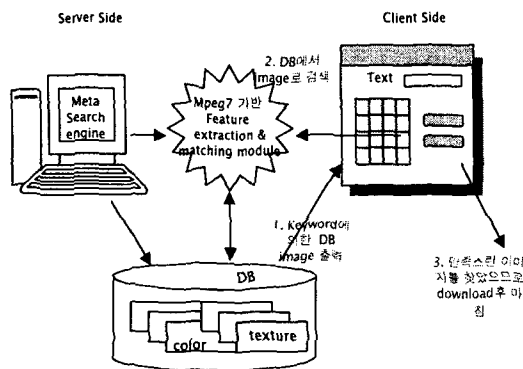


그림 1. MPEG-7 기술자 기반 검색 시스템 구조

데이터베이스 시스템을 구성 할 이미지들은 멀티미디어 데이터의 보고라 할 수 있는 인터넷을 통해서 수집한다. 인터넷상에 존재하는 멀티미디어 데이터들은 그 양의 방대함과 종류의 다양함 때문에 오래 전부터 빠르고 효율적인 검색 엔진을 필요로 했다. 따라서 비록 문자 기반 검색이기는 하지만 멀티미디어 데이터에 대한 상용 검색 엔진들의 검색률은 비교적 괜찮은 편이다. 그러므로 우리는 기존 5개의 상용 검색 엔진을 이용하여 이미지를 찾는 meta search engine을 구현하고 이를 통하여 필요로 하는 이미지를 얻는다.

먼저, 시스템 관리자는 웹에 구현된 관리자용 인터페이스를 이용하여 찾고자 하는 이미지의 keyword를 입력한다. 그러면 meta search engine은 관리자가 입력한 keyword를 5개의 상용 검색 엔진에 전송하여 해당 이미지를 찾는다. 그러나 상용 검색 엔진들을 이용하여 수집한 이미지들 중에는 해당 keyword와 상관이

없는 이미지나, 관리자가 보기에 데이터베이스에 넣지 말아야할 이미지도 있을 것이다. 그러므로 관리자가 개입하여 데이터베이스에 저장하고자 하는 이미지를 선택할 수 있도록 하는 방법이 필요하다. 구현된 meta search engine에서는 찾아진 모든 이미지들을 체크박스과 함께 웹 페이지에 출력하여, 관리자가 데이터베이스에 저장할 이미지를 선택할 수 있도록 했다.

선택된 이미지들을 MPEG-7 기술자 기반 DB에 저장하기 위해서는 이미지 뿐 아니라 해당 이미지의 질감 및 색 기술자 정보가 있어야 한다. 그림1에서 MPEG-7 기반 feature extraction 및 matching 모듈이 바로 입력 이미지에서 질감 및 색 기술자를 추출하거나 기술자들간의 유사도를 측정하는 루틴이다. MPEG-7 기반 feature extraction 및 matching 모듈의 알고리즘과 기능은 다음절에서 자세히 설명하기로 하겠다.

표 1. DB 데이터 필드

필드 명	데이터 형
Keyword	문자열
Filename	문자열
Position	문자열
Color_feature_1	실수형(Double)
Color_feature_64	
Texture_feature_1	실수형(Double)
Texture_feature_62	

MPEG-7 기반 feature extraction 및 matching 모듈에서 질감 및 색 기술자가 추출되면, 추출된 두 기술자와 함께 이미지 파일의 이름, 파일의 위치, keyword 그리고 URL등의 메타 데이터들이 데이터베이스에 저장된다. 데이터베이스의 자세한 구조는 표1과 같다.

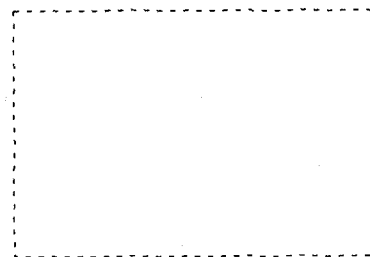


그림 2. 사용자 인터페이스 화면

일반 사용자들은 역시 웹에 구현된 그림 2의 사용자 인터페이스를 이용해서 찾고자 하는 이미지의 keyword나, 찾고자 하는 이미지와 비슷한 특성을 갖는 이미지를 질의 이미지로 입력한다. 전자의 경우에는 이미 데이터베이스에 keyword가 저장되어 있으므로, 사용자

가 입력한 keyword와의 단순 string 비교로 원하는 이미지를 찾을 수 있다. 그러나 후자의 경우에는 로컬 호스트에 있는 gif나 jpg 이미지를 질의 이미지로써 서버에 업로드하고, MPEG-7 기반 feature extraction 모듈을 통해 질의 이미지의 질감 및 색 기술자를 추출한다. 그리고 마지막으로 추출된 두 기술자와 데이터베이스에 저장되어 있던 기술자들간의 유사도 측정을 통해 가장 비슷한 16개의 이미지를 순서대로 출력해 준다. 또 무작위로 보여지는 16개의 데이터베이스 이미지들 중에서 하나를 택해 검색을 수행할 수도 있다.

이어지는 절에서는 앞에서 자주 언급된 MPEG-7 질감 및 색 기술자 extraction 알고리즘 및 matching 알고리즘에 대해 설명하겠다.

### 2.3. 질감 기술자

영상의 질감 정보는 옷감의 무늬, 벽, 타일 등의 균질 또는 비균질의 패턴을 말하는 것으로 색상, 물체 모양, 객체 움직임 등과 함께 영상의 내용 정보를 기술하는 중요한 시각적인 특징의 하나로 현재 MPEG-7의 표준안으로 진행 중에 있는 기술이다. 현재의 WD(Working Draft)에 있는 기술은 영상에 대하여 각 좌표계의 주파수 공간으로 변환을 수행한 뒤에 주파수 공간을 5x6으로 나누어 각 채널에서의 에너지와 에너지의 분산을 구하여 그 정보를 영상의 질감 정보로 사용한다.

각 채널에서 에너지 30개와 한 영상의 평균값과 분산의 값이 검색에 수행되는 32개의 Base Layer로 표현할 수 있다. 62개의 특징을 가진다.

이 질감 기술자는 다음과 같이 표시된다. 우선 기본층(Base Layer)으로는 수식 1과 같이 32개의 특징으로 구성되고

$$TD = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}] \quad (1)$$

그리고 확장층(enhancement layer)으로 는 수식 2와 같이 구성된다.

$$TD = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}] \quad (2)$$

여기서 TD는 질감 기술자(Texture descriptor)라 하고  $f_{DC}$  는 평균,  $f_{SD}$  는 분산값, 그리고  $e$  는 각 채널에서의 에너지  $d$  는 채널에서 에너지의 분산을 의미한다.

### 2.4. 칼라 히스토그램 기술자

칼라 히스토그램 기술자는 특정 칼라 공간의 세분화된 영역에 포함되는 픽셀들의 수를 카운트하여, 그 수를 기술자로 사용하는 방법이다[3][4]. 여기서 픽셀의 수를 세는 기준이 되는 세분화 칼라 공간을 빈이라 한다.

칼라 히스토그램 기술자는 칼라 공간, 칼라 양자화, 히스토그램과 같은 기술자들의 결합으로 이루어지는데, 이들 기본적 요소들의 유연하고 일반적인 정의에 의해 매우 유연하고, 넓은 범위의 응용에 적합한 기술자가 될 수 있다.

대표적인 칼라 공간에는 RGB, YCrCb, HSV, HMMD가 있다. RGB 색 공간에서 각 색들은 빨강, 초록 그리고 파랑의 기본 스펙트럼으로 나타난다. 이 모델은 직교 좌표 시스템에 기초한다. YCbCr 색 공간은 MPEG-1 과 MPEG-2에서 사용되어지는 색 공간으로 사람의 눈이 chrominance 성분에는 덜 민감하고 luminance 성분에 매우 민감하다는 이론에 근거하여 영상전송을 위한 압축시 Y 축은 넓게, Cb, Cr 축은 좁게 표현하기 위해 고안된 색 공간이다. HMMD는 Hue, Max, Min, Diff, Sum의 다섯 가지 축에 의한 색 공간으로 여기서 Hue는 색조, Max는 흑색의 함유량, Min은 흰색의 함유량, Diff는 색의 순수도, 그리고 Sum은 색의 밝기를 나타낸다.

칼라 양자화에도 uniform, subspace\_uniform, subspace\_nonuniform, lookup\_table을 사용하는 방법 등이 있다. 어떤 양자화 방법을 쓰느냐에 따라 세분화 칼라 공간인, 빈의 수가 바뀌며 그에 따라 히스토그램의 레이어아웃이 결정된다. 이미지들의 전체적인 칼라 특성을 256개 빈을 갖는 히스토그램으로 매우 정확한 표현을 할 수도 있지만 64개의 빈을 이용하는 적은 수의 경우에도 충분한 정확도를 얻을 수 있다.

칼라 히스토그램의 특징 추출 단계는 다음의 모듈로 나타낼 수 있다. 여기서 quantized\_color\_index 함수는 pixel의 칼라 값이 칼라 양자화로 재구성된 칼라 빈들 중에서 어떤 빈에 속하는지를 알려주는 함수이다.

```

numpix = 0;
for ( i=0 : i<NumberBins : i++ ) HistogramValue[i] = 0;
for ( all_pixels_of_a_visual_item ) {
    i = quantized_color_index ( pixel_value );
    HistogramValue[i]++;
    numpix++;
}
    
```

유사도 측정을 위한 distance는 여러가지가 있는데 본문에 사용된 것은 L1 distance이다. L1 distance는 다음의 식으로 계산되어 진다.

$$d(CD_{query}, CD_{Database}) = \text{distance}(CD_{query}, CD_{Database}) = \sum_k |CD_{query}(k) - CD_{Database}(k)| \quad (3)$$

여기서  $CD_{query}(k)$  는 질의 이미지에 대한 히스토그램 벡터 값이고  $CD_{Database}(k)$  은 데이터베이스에 있는 이미지의 히스토그램 벡터 값이다.

#### 2.4. 두 기술자를 이용한 matching 알고리즘

그림 3은 질의 이미지와 데이터베이스 이미지들간의 유사도 측정 과정을 보여준다. 만약 질의 이미지가 MPEG-7 기술자 기반 검색 시스템에 입력되면, MPEG-7 기반 feature extraction 및 matching 모듈은 질의 이미지에서 질감과 칼라 히스토그램 기술자를 뽑는다. 뽑아진 두 기술자들은 다음과 같이 표현될 수 있다.

$$MD = TD + CD \quad (4)$$

여기서 MD는 MPEG-7 질감 기술자와 칼라 히스토그램 기술자를 함께 의미한다. 우리는 이 논문에서 독립적으로 추출되고 비교되던 두 기술자를 함께 사용함으로써 자연 영상에 대해 상호 보완적인 효과를 얻을 수 있다. 그러므로 조합된 두 기술자들간의 유사도 측정은 다음의 식에 의해 이루어진다.

$$\text{merged distance } (MD_{query}, MD_{Database}) = \sum_k \left| \frac{TD_{query}(k) - TD_{Database}(k)}{\alpha(k)} \right| + w \times \sum_k |CD_{query}(k) - CD_{Database}(k)| \quad (5)$$

### 3. 실험 및 결론

그림 4에서 구현된 시스템을 이용한 이미지 검색 결과를 보여주고 있다. 질의 이미지인 좌측 상단의 이미지와 유사한 질감과 색을 가진 이미지들을 잘 찾고 있음을 볼 수 있다.

구현된 시스템은 구현 성능 뿐 아니라 다음 두 가지 측면에서 이점을 갖는다. 첫째, 사용자 인터페이스가 웹상에 구현됨으로써 일반 사용자들이 쉽게 시스템에 접근할 수 있다. 둘째, 관리자는 필요시 meta search engine을 이용해서 데이터베이스 정보를 쉽게 보충할 수 있으므로 검색 엔진의 생명이라 할 수 있는 최신 정보 구축이 용이하다.

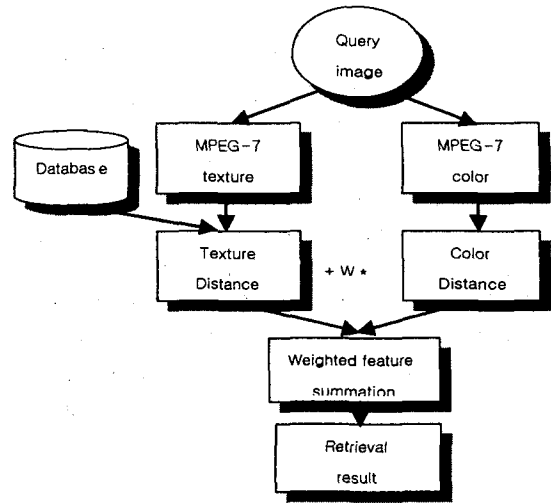


그림 3. 두 기술자를 이용한 matching 알고리즘

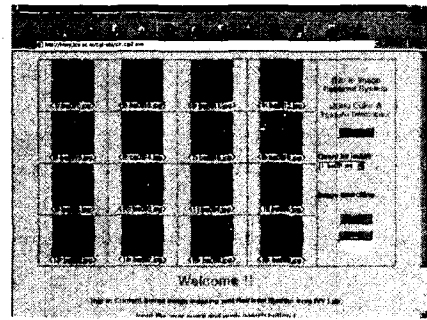


그림 4. 이미지 검색 결과 화면

### 참고문헌

- [1] MPEG-7 Visual part of eXperimentation Model Version 7.0 , *ISO/MPEG*, n3521, Beijing, July, 2000
- [2] MPEG-7 Visual Working Draft 3.0 , *ISO/MPEG*, n3399, Geneva, June, 2000
- [3] M. J. Swain and D. H. Ballard. Color indexing, *International Journal of Computer Vision*, 7(11):11-32,1991
- [4] F. Idris and S. Panchanathan, Review of Image and Video indexing Techniques , *Journal of Visual Communication and Image Representation.*, Vol. 8, No. 2, pp. 146- 166, June 1997