

# MPEG-7 시각 정보 기술자와 텍스트 정보를 이용한 내용 기반 웹 이미지 검색 시스템

박주현<sup>o</sup> 남종호

서강대학교 컴퓨터학과  
{parkjh<sup>o</sup>, jhnang}@sogang.ac.kr

## A Content based Web Image Retrieval System using MPEG-7 Visual Descriptors and Textual Information

Joohyoun Park<sup>o</sup>, Jongho Nang

Dept. of Computer Science, Sogang University

### 요 약

인터넷 기술의 발달과 디지털 카메라와 같은 디지털 미디어 생산 장비의 발달로 WWW에 이미지 데이터의 양이 급격하게 늘어나면서 웹 이미지에 대한 효율적인 검색에 대한 요구가 증가하고 있다. 본 논문에서는 사용자의 다양한 검색 요구를 만족시킬 수 있도록 기존의 텍스트 기반의 검색과 시각 정보 기반의 검색을 병행하여 수행할 수 있는 웹 이미지 검색 시스템을 설계하고 구현한다. 제한한 웹 이미지 검색 시스템은 웹 이미지 수집 및 검색정보 추출 도구, 검색 서버, 그리고 검색 클라이언트로 구성된다. 웹 이미지 수집 및 검색 정보 추출 도구는 웹에서 이미지를 수집하여 이미지가 속해있는 웹 문서 구조를 이용하여 적절한 키워드를 선택하여 시각 정보 기반의 검색을 지원하기 위해 MPEG-7 시각 정보 기술자[1]를 추출한다. 빠른 검색을 위해 추출된 텍스트 정보는 상용 데이터베이스에 저장되며 MPEG-7 시각 정보 기술자는 고차원 데이터 색인 방법인 HBI (Hierarchical Bitmap Index)[2]를 사용하여 색인 정보를 만들어 사용한다. 검색 클라이언트는 사용자가 각 검색 요소에 가중치를 부여하여 검색 할 수 있도록 하며 원하는 검색 결과를 얻을 때까지 반복하여 검색할 수 있는 연관 피드백 과정도 포함한다.

### 1. 서 론

인터넷 기술의 발달로 인해 WWW(World Wide Web)의 규모는 급속히 커져가고 있다. 더욱이 디지털 카메라나 캠코더와 같은 각종 디지털미디어 장비의 발달은 WWW상에서 이미지 데이터의 양을 폭발적으로 증가시키는 결과를 가져왔다. 이러한 대규모의 정보 집합체인 WWW에서 이미지 데이터에 대한 검색 요구는 날로 증가하고 있으며 Google<sup>1)</sup>이나 Yahoo<sup>2)</sup>에서 이미 텍스트 기반의 이미지 검색 서비스를 제공하고 있다. 하지만 이미지 주변에 있는 텍스트 정보를 이용한 이미지 검색은 기존의 웹 문서 검색과 크게 차이가 없으며 이미지 자체가 가지고 있는 시각 정보가 검색에 사용되지 않기 때문에 사용자의 다양한 검색 요구를 만족 시켜 주지 못한다는 문제점을 가지고 있다.

본 논문에서는 동영상, 멀티미디어 검색과 관련하여 진행된 여러 연구 결과를 통합하여 텍스트 정보와 시각 정보를 함께 이용한 내용 기반 웹 이미지 검색 시스템을 설계 및 구현하였다. 본 웹 이미지 검색 시스템은 크게 웹 이미지 수집 및 검색정보 추출 도구, 검색 서버, 검색 클라이언트의 세 부분으로 구성되어 있다. 웹 이미지 수집 및 검색정보 추출 도구는 WWW에서 이미지를 수집하고 수집된 이미지에 대한 키워드를 웹 문서의 이미지 주변 텍스트로부터 추출하는 기능을 수행한다. 이때 키워드 선택의 정확성을 높이기 위해서 웹 문서 구조뿐만 아니라 이미지의 시각 정보를 함께 이용하는 방법을 사용하였다. 그리고 수집된 이미지의 URL, 이미지와 관련된 키워드와 시각 정보는 각각 데이터베이스에

저장되어 검색 엔진에 의해 사용된다. 이미지 검색 엔진은 키워드와 시각 정보를 기반으로 각각 검색을 수행하며 각 검색 결과에 대해 가중치 합을 사용하여 최종 검색 결과를 만들어낸다. 키워드 기반 검색을 수행할 때 워드넷[4]을 사용하여 이미지 간 유사도를 구하고, 시각 정보 기반 검색을 수행할 때 MPEG-7 XM[7]에서 제안된 매칭 함수를 사용하여 이미지 간 유사도를 구한다. 검색 속도 향상을 위해서 텍스트 데이터는 SQL 서버를 이용하여 저장하였으며 MPEG-7 시각 정보 기술자는 HBI를 사용하여 색인하였다. 또한 서버와 클라이언트는 사용자가 검색 결과에 만족하지 못한다면 피드백 과정을 통해 원하는 검색 결과가 나올 때까지 재검색할 수 있도록 연관 피드백을 검색 과정에 포함하고 있다. 제안하는 웹 이미지 검색 시스템은 세분화된 모듈 구조를 통해서 이후 개발되는 각종 이미지 검색 관련 알고리즘들을 테스트할 수 있도록 설계 및 구현한다.

### 2. 웹 이미지 검색 시스템 구조

웹 이미지 검색 시스템은 <그림 1>에서 보이는 바와 같이 크게 웹 이미지 수집 도구, 검색 서버, 그리고 검색 클라이언트의 세 부분으로 구성되어 있다. 각 부분은 독립 프로그램으로 존재하며 소켓을 사용하여 데이터를 주고받는다.

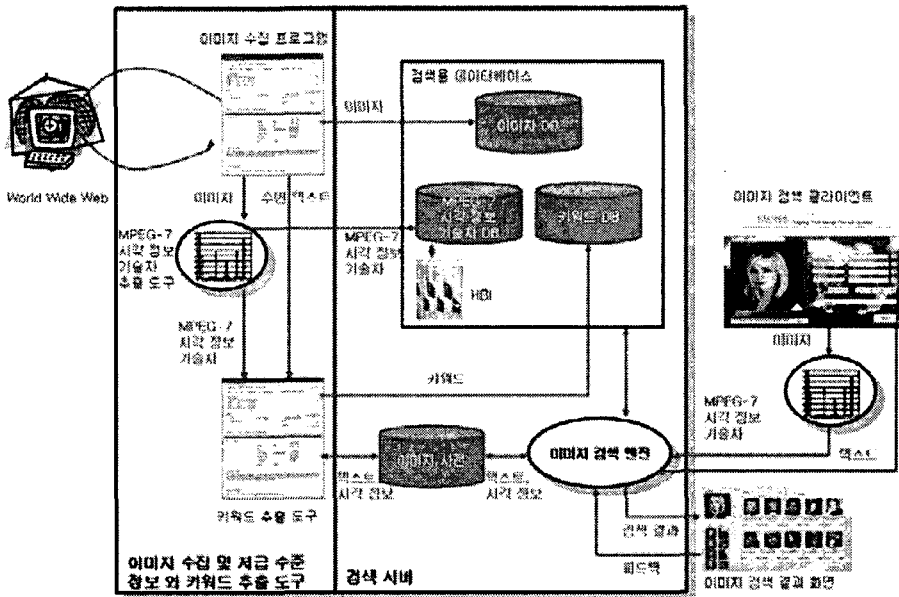
#### 2.1 이미지 수집 및 검색정보 추출 도구

이미지 수집 로봇은 시작 페이지를 사용자로부터 입력 받아 그 페이지 내에서 연결된 페이지들을 BFS (Breadth First Search) 방식으로 탐색하며 이미지들을 수집한다. 이미지가 속해있는 웹 문서의 타이틀, 앵커(Anchor), 주변 텍스트들로부터 선택된 후보 키워드와 이미지로부터 추출된 MPEG-7 시각 정보 기술자인 Dominant Color, Color Layout, Edge Histogram을 사용

1) <http://www.google.co.kr/imghp?hl=ko&tab=wi&q=>

2) <http://kr.image.yahoo.com/>

이 논문은 2006년도 두뇌한국21사업에 의해 지원되었음.



<그림 1> 웹 이미지 검색 시스템 구조

하여 학습된 이미지 사진을 통해 나온 개념 선택 결과를 조합 [3]하여 이미지를 대표할 수 있는 최종 키워드를 선택한다. 선택된 키워드는 텍스트 기반 검색을 위해 사용되기 위하여 데이터베이스에 저장되고 색인 정보가 만들어진다. 또한 추출한 MPEG-7 시각 정보 기술자들 역시 시각 정보 기반의 검색을 위해 서버에 저장되며 빠른 검색을 위해 색인 정보가 만들어진다. 본 논문에서는 고차원 데이터인 MPEG-7 시각 정보 기술자의 색인을 위해서 HBI(Hierarchical Bitmap Indexing)[2]를 사용하였다.

2.2 검색 서버와 클라이언트

검색 서버는 이미지, 시각 정보, 키워드를 저장하고 있는 세 종류의 데이터베이스와 시각정보와 개념간의 관계가 학습되어 있는 이미지 사진, 그리고 검색 엔진으로 구성되어 있다. 검색 엔진은 텍스트 기반 검색과 시각 정보 기반 검색을 모두 수행하여 최종 검색 결과를 만들어낸다. 이미지 검색 과정에 대한 자세한 설명은 3장에서 하도록 한다.

검색 클라이언트는 사용자의 질의를 받아 서버에 전송하고 서버가 전송해주는 검색 결과를 사용자에게 보여주는 기능을 수행한다. 또한 연관 피드백을 위해 Rocchio[5]의 방법을 사용하여 사용자의 피드백을 입력으로 받아 새로운 질의를 만드는 기능을 포함하고 있다.

3. 텍스트 정보와 시각 정보를 이용한 이미지 검색

이미지 검색은 질의 이미지와의 비유사도가  $r$  보다 작은 모든 이미지를 검색하는  $r$ -Range 검색과 질의 이미지와 비유사도가 작은 순서로 순위를 정하여 상위  $k$ 개만 검색하는  $k$ -NN (Nearest Neighbor) 검색으로 나누어 생각해볼 수 있다. 본 시스템은 텍스트 정보와 시각 정보를 기반으로 하여  $r$ -Range 검색과  $k$ -NN 검색을 모두 지원한다. 사용자는 검색 클라이언트를 사용하여 질의  $q$ 를 만들어 서버에게 검색 요청을 할 수 있다. 질의  $q$ 는  $\langle q_1, q_2, q_3, q_4 \rangle$ 로 구

성되며  $q$ 의 원소  $q_1, q_2, q_3, q_4$ 는 각각 키워드와 이미지로부터 추출한 Dominant Color, Color Layout, Edge Histogram을 의미한다. 이 때 사용자는 각 검색 요소인 Dominant Color, Color Layout, Edge Histogram, 그리고 키워드들의 검색 시 반영될 가중치를 설정할 수 있으며  $w = \langle w_1, w_2, w_3, w_4 \rangle$ 로 표현한다. 질의  $q$ 를 받은 서버는 데이터베이스에 저장되어 있는 이미지들의 키워드들과 MPEG-7 시각정보들 간의 비유사도를 계산한다. MPEG-7 시각정보들 간의 비유사도는 XM[7]을 통해서 권고되는 비교 함수를 사용하여 계산한다. 그리고 키워드 간의 비유사도는 워드넷[4]을 사용하여 질의 단어와 대상 이미지를 표현하는 단어 사이의 최단 경로 길이와 이미지 수집 단계에서 키워드 선택 시 계산된 이미지와 키워드의 연관 정도를 곱하여 결정한다. 질의  $q$ 와 임의의 대상 이미지의 키워드와 MPEG-7 시각 정보기술자인  $p$ 와의 최종비유사도  $d(p, q)$ 는 <식 1>과 같이 각 검색 요소 단위로 계산된 비유사도들의 가중치 합으로 결정된다.

$$d(p, q) = \sum_{j=1}^4 w_j \cdot d(p_j, q_j) \quad \text{<식 1>}$$

단, 각 검색 요소간의 비유사도  $d(p_j, q_j)$  ( $1 \leq j \leq 4$ )는 비교 함수에서 계산될 수 있는 최대값과 워드넷에서 표현 가능한 최대 거리로 나누어  $[0, 1]$ 사이의 값으로 정규화하여 사용한다. 결론적으로,  $r$ -Range 검색은 이미지 데이터베이스의 이미지 중 <식 1>을 사용하여 계산된 질의와의 비유사도가  $r$  이하인 이미지 집합을 찾게 되고  $k$ -NN 검색은 질의와의 비유사도를 기준으로 작은 순으로 정렬하여  $k$  번째 이미지까지를 검색 결과로 만들게 된다.

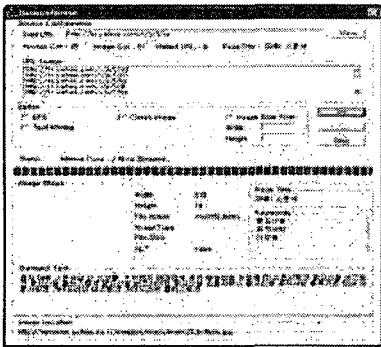
사용자가 원하는 검색 결과를 얻기 위한 연관 피드백은 Rocchio[5]의 방법을 사용하여 수행한다. 즉, 검색 클라이언트에서 사용자가 검색 결과에 대한 만족도를 5단계로 평가하게 한 후 이를 바탕으로 새로운 질의  $q'$ 을 만들어 재검색을 수행하게 되는 것이다.

4. 시스템 구현

본 시스템은 Windows XP 상에서 구현하였으며 시스템의 각 구성요소는 소켓을 사용하여 통신하도록 구현하였다.

4.1 이미지 수집 및 검색정보 추출 도구의 구현

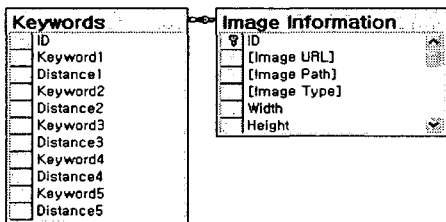
이미지 수집 로봇은 이미지의 타입이나 크기를 제한하여 수집이 가능하도록 구현하였으며 수집된 이미지와 주변 텍스트, 선택된 키워드를 한 눈에 확인할 수 있도록 하였다. 또한 키워드 추출 알고리즘을 다른 알고리즘으로 대체하기 쉽도록 DLL (Dynamic Link Library)로 구현하였다. <그림 2>는 이미지 수집 및 검색 정보 추출 도구 프로그램을 보여준다.



<그림 2> 이미지 수집 및 검색정보 추출 도구

4.2 검색 서버의 구현

검색 속도의 향상을 위해서 이미지 정보와 키워드는 상용 데이터베이스인 MS-SQL을 사용하여 저장하였다. <그림 3>에서 보듯이 확장성을 고려하여 이미지 정보와 키워드는 서로 다른 테이블로 구성되며 ID를 공통 필드로 하여 연결된다. <Keywords>테이블에는 최대 5개의 키워드를 저장할 수 있으며 Distance는 키워드와 이미지와의 연관 정도를 [0,1]사이의 값으로 표현한다. 연관 정도는 이미지 수집 단계에서 계산되며 0에 가까울수록 이미지와 연관성이 큰 키워드임을 의미한다. 하지만 MPEG-7 시각 정보 기술자는 고차원이라는 특성을 가지고 있고 L1 또는 L2 거리를 사용하여 거리가 계산되기 때문에 HBI[2]를 사용하여 색인하고 파일로 저장한다. MPEG-7 시각 정보 기술자 중 Dominant Color는 벡터 공간이나 매트릭 공간에 표현되지 않기 때문에 색인이 불가능하기 때문에 HBI로 색인 가능하도록 벡터 공간으로 변환한 후 사용하였다[6].

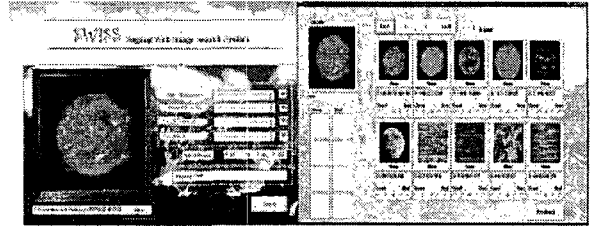


<그림 3> 이미지 정보와 키워드 테이블의 관계도

4.3 검색 클라이언트의 구현

검색 클라이언트는 웹 페이지에서 수행 가능하도록 ActiveX 컨트롤로 구현하였다. <그림 4>의 왼쪽 그림에서 보이는 것처럼 사용자는 질의 이미지를 선택할 수 있고 키워드를 입력할 수 있으며 검색 요소에 대한 가중치를 설정할 수 있다. <그림 4>의 오른쪽 그림은 붉은 태양 이미지와 "sun"이라는 키워드를

질의로 하고 4개의 검색 요소에 대한 가중치를 모두 같게 한 후 검색한 결과화면이다. 결과 화면은 한 페이지에 10개의 검색 결과를 보여주며 5단계의 만족도로서 사용자의 피드백을 받아 재검색을 할 수 있도록 하였다.



<그림 4> 검색 클라이언트의 질의화면과 검색화면

5. 결론

WWW에서의 이미지 데이터의 폭발적인 증가는 웹 이미지 검색 시스템의 필요성을 증대시키고 있다. 하지만 현재 서비스되고 있는 이미지 검색 시스템은 텍스트 기반의 검색만을 제공함으로써 사용자의 다양한 검색 요구를 만족시키지 못하고 있다. 본 웹 이미지 검색 시스템은 이미지 자체의 시각 정보를 키워드 선택 단계에 적용하여 키워드 선택의 정확성을 높였으며 텍스트 기반의 검색과 시각 정보 기반의 검색을 함께 수행하고 연관 피드백 과정을 검색 과정에 포함하여 사용자의 다양한 검색 요구를 반영할 수 있도록 하였다. 또한 검색에 사용되는 정보들의 특색에 맞추어 상용 데이터베이스와 자체 개발한 고차원 데이터 색인 방법인 HBI를 함께 사용하여 검색 속도를 향상시킬 수 있도록 하였다. 아직까지는 고차원의 시각 정보를 검색에 사용하기 때문에 CPU나 메모리와 같은 자원의 사용량이 많아 개인용 이미지 검색 시스템으로서만 활용이 가능하다. 하지만 빠른 검색 알고리즘 개발이나 코드 최적화 작업을 통해서 이러한 문제를 해결한다면 다수 사용자를 위한 웹 이미지 검색 시스템으로서 활용할 수 있을 것이다. 또한, 세분화된 모듈 분리를 통해서 이후 개발되는 검색 알고리즘들을 쉽게 이식할 수 있도록 하였기 때문에 이미지 검색 알고리즘을 위한 테스트베드로서도 활용할 수 있을 것이다.

6. 참고 문헌

[1] ISO/IEC JTC1/SC29/WG11 *Information Technology Multimedia Content Description Interface-Part3: Visual*, 2001.  
 [2] J. Park and J. Nang, "A Hierarchical Bitmap Indexing Method for Content Based Multimedia Retrieval," *Proceedings of EuroIMSA International Conf.*, pp.223-228, 2006.  
 [3] 황지익, *텍스트 정보와 시각 특징 정보를 이용한 효과적인 웹 이미지 캡션 추출 방법*, 서강대학교 석사 학위 논문, 2005.  
 [4] C. Fellbaum, *WordNet: An Electronic Lexical Database*, MIT Press, pp.265-283, 1998.  
 [5] D. Widyantoro, and J. Yen, "Relevant Data Expansion for Learning Concept Drift from Sparsely Labeled Data," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, Issue 3, pp.401-412, 2005.  
 [6] 박주현, 남종호, "인덱싱을 위한 MPEG-7 시각 정보 기술자 분석," *KCC 2005 학술대회 발표집*, 2005.  
 [7] ISO/IEC JTC1/SC29/WG11 *MPEG-7 Visual part of eXperience Model Version 11.0*, 2001.