

바다-IV DBMS 에서 색상 특징량을 이용한 내용 기반 이미지 검색

기능의 설계 및 구현

김영균* • 김완석 • 김명준

한국전자통신연구원 컴퓨터소프트웨어기술연구소 인터넷서비스연구부

Design and Implementation of Content-Based Image Retrieval using Color

Feature Distribution in the BADA-IV DBMS

Young-Kyoon Kim, Wan-Seok Kim, Myung-Joon Kim

Internet Service Department, Computer & Software Technology Lab., ETRI

요약

다양하고 방대한 이미지 자료를 효율적으로 저장 관리하고, 또한 효과적인 이미지의 내용 검색을 수행하는 통합 시스템의 필요성이 증가하고 있다. 본 논문에서는 객체지향 멀티미디어 DBMS 인 바다-IV 에서 내용 기반 이미지 검색을 제공할 수 있도록 통합 DBMS 구조를 설계하고, 이미지 자료를 효율적으로 저장, 관리 및 검색할 수 있는 방법을 제시한다. 또한, 이를 이용하여 이미지의 색상 특징에 기반한 내용 검색을 수행하는 시스템을 구현한다. 구현된 바다-IV DBMS 의 이미지 내용 검색은 SQL 을 통한 이미지 검색을 지원하도록 내용 검색 질의어 구문을 지원하고, 그리고 이미지 검색 성능 개선을 위해 2 단계 이미지 검색 알고리즘을 사용한다.

1. 서론

이미지 혹은 동영상 등과 같은 대용량의 멀티미디어 데이터를 초고속으로 전송할 수 있는 인터넷 환경의 구축, 이미지 처리와 컴퓨터 비전(computer vision) 기술 등과 같은 멀티미디어 처리 능력, 그리고 대용량 자료의 효과적인 저장 관리 기술 등을 기반으로 여러 응용들이 엄청난 양의 이미지 정보를 구축하여 유통시키고 있다. 이미지 정보를 생산하는 응용들로 전자도서, 전자 화랑 등과 같은 멀티미디어 정보 검색, 의료 영상 진단, 기상 관측 및 예보, 위성 사진 검색, 전자상거래 등이 있으며, 이러한 응용들은 필수적으로 내용 기반 이미지 검색(content based image retrieval)이 필요하다.

이미지의 내용 검색 기법은 이미지 내용을 사람이 텍스트로 기술하고, 이에 대하여 전문 정보 검색(information retrieval)으로 처리하여 내용 검색을 수행하는 방식과 컴퓨터 비전, 패턴 인식 등과 같은 이미지 처리 기술을 이용하여 이미지가 갖는 고유한 시각적인 특징(visual features) 값을 기반으로 검색하는 방식으로 구분된다[8]. 현재 전자의 방식은 내용 검색의 효과가 낮고, 또한 확장성에 문제점이 있어 다수의 연구들이 후자 방식을 사용하고 있다[2, 3, 4, 5, 6, 9].

컴퓨터 비전 분야에서 이미지의 내용 검색 연구들이 어떻게 가장 효율적인 이미지 특징 값을 추출할 수 있는가에 대한 문제를 중심으로 전개되고 있는 반면에, 이미지 데이터베이스와 이미지 특징 정보를 관리하는 데는 파일 시스템의 사용이나 단순

한 DBMS 활용에 그치고 있다. 이는 이미지 데이터베이스의 규모가 커질 수록 이미지 및 이미지 특징 정보를 일관성있게 유지해야 하는 데이터베이스 관리의 복잡도가 증가하게 되며, 또한 다양한 응용으로의 확장에도 한계가 있다. 이를 해결하기 위한 방안으로 내용 기반 이미지 검색을 강력한 데이터 관리 기능을 제공하는 DBMS 와 통합시킨 시스템들이 개발되고 있다 [3, 4, 6, 7]. 이는 다중 사용자 지원, SQL 을 통한 이미지 검색, 일관성 있는 이미지 특징 정보 관리 등과 같은 통합 데이터 저장 기능과 내용 검색 기능을 지원한다.

현재까지 이미지의 내용 검색 기술에는 모든 응용들에 적용할 수 있고, 내용 검색의 효과가 입증된 이미지 특징 추출 알고리즘 개발에 어려움이 있어, 다수의 연구들이 이미지 데이터베이스의 응용에 종속된 내용 검색 기법을 제시하고 있다. 따라서 다양한 이미지 검색 응용들에서 내용 기반 이미지 검색을 제공하는 DBMS 가 사용되기 위해서는 DBMS 가 다양한 이미지 검색 기법을 쉽게 수용할 수 있는 구조이어야 한다. 객체지향 DBMS 는 확장된 관계형 DBMS 에 비해 응용 개발의 편이성 및 확장성이 뛰어나기 때문에 내용 기반 이미지 검색과 객체지향 DBMS 의 통합이 앞서 언급된 내용 기반 이미지 검색을 제공하는 DBMS 에 비해 효과적이다.

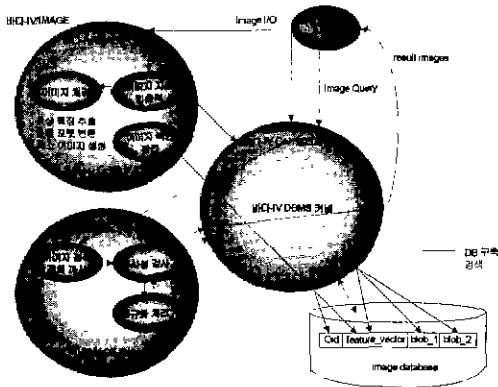
본 논문에서는 객체지향형 멀티미디어 DBMS 인 바다-IV 에서 내용 기반 이미지 검색을 제공하기 위해 DBMS 구조를 설계하고, 그리고 이미지 자료를 효율적으로 저장, 관리하며, 이미지

검색 질의 처리 방법을 제안한다 또한, 제시된 바다-IV DBMS 에서 이미지의 내용 검색 기능을 실현하기 위해서 이미지의 색상 정보를 기반으로 유사한 색상을 갖는 이미지를 탐색하는 내용 검색 기능을 구현한다

논문의 구성은 다음과 같다 먼저, 2장에서 내용 기반 이미지 검색을 지원하는 바다-IV DBMS 의 구조와 이미지 내용 검색을 위한 질의어 구문, 그리고 이미지 검색 질의어 처리에 대해 설명한다. 3 장에서는 바다-IV DBMS 에서 색상 특징량 기반 이미지 검색 구현을 위해 사용한 특징 값 추출 방법 및 검색 알고리즘 등을 설명한다.

2. 이미지 내용 검색을 위한 바다-IV DBMS 구조

색상 내용 기반 이미지 검색을 지원하기 위해서 바다-IV DBMS 는 <그림 1>에서와 같이 이미지 자료를 효율적으로 저장, 관리할 수 있는 바다-IV/IMAGE 클래스 라이브러리와 이미지의 내용 검색을 처리하는 유사도 검색 라이브러리로 이루어진다.



<그림 1> 내용 기반 이미지 검색을 위한 바다-IV DBMS 구조

2.1 이미지 자료형 지원

바다-IV/IMAGE 는 바다-IV DBMS 의 C++ 인터페이스를 기반으로 이미지 자료형을 제공한다. 이미지 자료형 클래스는 이미지의 시각적 특징 값 추출, 축소 이미지 자동 생성, 이미지 포맷 변환 등과 같은 다양한 메소드들을 지원한다.

특정 이미지가 DB 에 저장될 때, <그림 1>에서와 같이 바다-IV/IMAGE 의 특징 추출기, 포맷 변환기, 색인 생성 등과 같은 전처리 과정이 자동으로 수행되어, 내용 검색에 필요한 값들을 생성 관리하게 된다 따라서 이미지 DB 를 구축하려는 사용자는 바다-IV DBMS 의 C++ 인터페이스를 이용하여 이미지 자료형 객체를 생성하고, 이후에 모든 이미지 자료의 입출력은 바다-IV/IMAGE 에서 제공하는 이미지 자료형의 제공 메소드들을 이용하여 된다.

반면에, 이미지의 유사도 검색 라이브러리는 질의로 주어진 이미지와 유사한 이미지들을 DB 에서 검색하기 위한 것으로서, 위의 그림에서와 같이 바다-IV DBMS 의 커널과 밀접합된 구조

이다. 이는 바다-IV DBMS 에서 이미지 검색 질의어가 바다-IV DBMS 커널이 제공하는 SQL SELECT 문을 기반으로 시스템 제공 메소드인 'contains()'를 이용하여 정의되고, 이 시스템 메소드가 바다-IV DBMS 커널의 서버에서 수행되기 때문이다.

2.2 이미지 검색 질의어 처리

바다-IV DBMS 에서 이미지의 내용 검색 질의어를 작성하기 위해서는 이미지 내용 검색을 위한 이미지의 유사도 검색 조건에 해당하는 문자열 질의어(string query)를 contains() 메소드내의 입력 인자로 기술해야 한다. 이를 위해서 아래와 같은 BNF 형식의 검색 질의 구문을 제공한다.

```

<string_query> ::= <feature_description> ',' <threshold> ',' <result>
// threshold: 탐색할 임계 값, result: 탐색 이미지 개수
<feature_description> ::= <feature> '=' ('<values> ')
<feature> ::= <string> // 키워드 : color
<values> ::= <double_number> | <values> ',' <double_number>
// 0.00023, 0.00501, ...
<threshold> & <result> ::= <long_number>
    
```

이미지 검색 질의어 구문의 <threshold>와 <result>를 이용하여 사용자는 DB 에 저장된 이미지들에 대해 최근접 탐색, 범위 탐색, 그리고 K 개 최근접 탐색을 수행시킬 수 있다 사용자가 정의한 내용 기반 이미지 검색 SQL 문은 바다-IV DBMS 의 질의 처리기에서 접수하여, contains() 메소드가 존재하는 경우에 유사도 검색 라이브러리의 유사도 검색 메소드가 호출된다. 먼저, 문자열 질의어로 정의된 질의어를 파싱하여 색상 특징 값을 추출하고, 이를 이용하여 32 절에 언급된 검색 알고리즘에 의해 유사한 이미지 집합을 검색한다 검색된 이미지 객체들은 질의 이미지와 유사한 정도의 순위로 정렬되어, 객체 식별자와 유사도 값 쌍이 최종적으로 질의 처리기에 전달된다 이 때, 유사한 이미지의 유사도 값은 [0, 1] 사이의 값으로 정규화된다. 이는 이미지의 주색 기반 검색과 내용 기반 이미지 검색을 통합시킨 질의를 처리하는 과정에서 통합된 질의 결과로 얻는 이미지들의 최종 유사도 값을 결정하는데 필요하다 통합 검색 질의를 처리한 결과로 얻는 이미지들의 유사도 값 결정은 [5]에 제시된 퍼지 검색 모델을 사용한다.

3. 색상 기반 이미지 검색 구현

2장에 제시된 바다-IV DBMS 에서 색상에 기반한 이미지 유사도 검색을 실현하기 위해서는 색상 모델 및 검색에 이용되는 특징 값의 서술, 그리고 검색 기법이 필요하다. 이를 위해 바다-IV DBMS 에서는 [10]에서 제시한 색상 특징량 분포에 의한 색상 이미지 검색 기법을 사용한다.

3.1 이미지 특징 추출 및 유사도 검색 알고리즘

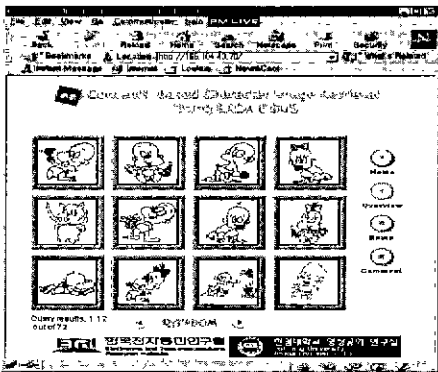
이미지 특징 값 추출을 위해 구현에 이용한 색상 모델이 RGB 모델에 비해 근접한 색상에 대한 처리가 용이한 HSI 모델이다. 따라서 이미지의 색상 특징 값은 Hue, Saturation, Intensity 3 가지 요소에 대한 색상 히스토그램(color histogram)이 된다. 구현에서

는 Hue 를 8 단계, S 와 I 를 각각 4 단계로 나누어 각 bin(bin)에 해당하는 히스토그램 양을 계산하고, 이를 특징 벡터 값으로 사용한다[10].

색상 특징 벡터에 기반하여 이미지를 검색하는 방법에 있어서 질의로 전달된 이미지에서 16 차원의 특징 벡터를 구하고, 이를 데이터베이스 내에 있는 모든 이미지들의 특징 벡터들에 대해 유사도 계산을 수행하는 것은 검색 성능의 저하를 발생시킨다. 이를 극복하기 위해 바다-IV DBMS 의 색상 기반 이미지 검색은 색상 특성이 두드러지게 나타나는 Hue 값의 확률 분포에 기반하여 2 단계 검색을 수행한다. 먼저, 질의 이미지에서 가장 영향이 큰 Hue 값(most salient color feature, MSCF)을 이용하여 이미지의 검색 범위를 축소시킨 후에 한정된 이미지 결과 집합에 포함된 이미지의 16 차원의 특징 벡터와 질의 이미지의 특징 벡터간에 유사도 계산을 수행하여 최종 순위 리스트를 결정한다. 2 단계 이미지 유사도 검색 알고리즘은 다음과 같다.

Algorithm . SimilarityQuery(fv)

```
// Q(fv) : 질의 이미지 색상 특징 벡터 값
// D(fv) : 저장 이미지 색상 특징 벡터 값
1. [Hue 성분의 확률 분포 값 계산]
// Mh : mean of i-th bin of Hue value
// vh : value of i-th bin. fh : frequency of vh
// N : total number of pixels in histogram
FOR i in each bin of Hue values DO
    D(Mh) = (D(vh) * D(fh)) / N where i = 1, ..., 8
END
2. [MSCF 결정 및 1 단계 이미지 검색]
FOR i in each bin of Hue values of query image DO
    MDCFi = | D(Mh) - (Q(vh) * Q(fh)) / N |
MDCF = Max(MDCFi)
END
find a OID set satisfying T1 < MDCF < T2 // T1, T2 : 임계값
3. [2 단계 이미지 유사도 값(Euclidian distance) 계산]
FOR i in each element of feature vector DO
    distance += ((Q(fvi) - D(fvi))2
return (distance)1/2
```



<그림 2> 바다-IV DBMS 의 이미지 검색 GUI 인터페이스

3.2 구현 결과

바다-IV DBMS 의 이미지 내용 검색의 구현에서는 사용자를 위한 GUI 인터페이스로 <그림 2>와 같은 예제 이미지에 의한 검색 인터페이스를 지원하며, 데이터베이스의 이미지 자료는 캐릭터 이미지 총 3000 건을 구축하였다

구현된 2 단계 이미지 검색 기법과 순차 검색으로 수행한 결과를 비교해 볼 때, 약 15 배 정도의 검색 속도가 향상됨을 확인할 수 있었으며, 검색 결과의 효과는 정확율과 평균 재현율을 기반으로 [10]에서와 같은 결과를 얻었다

4 결론

내용 기반 이미지 검색과 DBMS 의 효과적인 통합을 위해 본 논문에서는 바다-IV DBMS 구조를 설계하였으며, SQL 문을 통해서 이미지 검색 질의어가 수행될 수 있도록 내용 검색 질의어 구문을 정의하였다. 그리고 설계된 바다-IV DBMS 의 내용 기반 이미지 검색을 실현하기 위해서 이미지의 색상 특징량 정보를 기반으로 유사한 색상을 갖는 이미지를 검색하는 기능을 구현하였다.

추후에는 바다-IV DBMS 가 색상, 길감, 모양 등과 같은 이미지의 다양한 특징 값을 이용한 내용 검색을 지원하고, 또한 다차원 색인에 기반한 신속한 검색이 지원되도록 시스템을 확장 개발할 예정이다.

5. 참고문헌

- [1] A Gupta and R. Jain, "Visual Information Retrieval," Communications of the ACM. 40(5), pages 69-79, 1997.
- [2] INFORMIX, Excalbur Image DataBlade Module User's Guide. Informix Press, 1997.
- [3] IBM Corporation. DB2 Universal Database Image, Audio, and Video Extenders: Administration and Programming, Ver. 5, Release 2 International Business Machines Corporation, 1998
- [4] M. Flickner et al., "Query by Image and Video Content: The QBIC System," IEEE Computer, 28(9), pages 23-32, 1995
- [5] M. Ortega et al., "Supporting Ranked Boolean Similarity Queries in MARS," IEEE Trans. on Knowledge and Data Eng., 10(6), pages 905-924, 1998.
- [6] Oracle Corporation, Oracle8i Visual Information Retrieval User's Guide and Reference, Release 8.1.5, Oracle Corporation. 1998. Available at <http://technet.oracle.com/doc/inter.815/>
- [7] V. E. Ogle and M. Stonebraker, "Chabot: Retrieval from Relational Database of Images." IEEE Computer, 29(9), pages 40-48, 1995.
- [8] V. N. Gudivada and V. V. Raghavan, "Content-Based Image Retrieval Systems," IEEE Computer. 29(9), pages 18-22. 1995.
- [9] Y. A. Aslandogan and C. T. Yu. "Techniques and Systems for Image and Video Retrieval." IEEE Trans. on Knowledge and Data Engineering. 11(1), pages 56-63. 1999.
- [10] 유광석, 김희율, "칼라 특징량의 확률분포에 의한 내용기반 캐릭터 이미지 검색," 11 회 영상처리 및 이해에 관한 워크샵 발표 논문집, pages 56-60, 1999.