

MPEG-7 기반 비디오/이미지 검색 시스템(VIRS)

이 재 호[†] · 김 형 준[†] · 김 회 울^{††}

요 약

멀티미디어데이터의 급격한 양적 팽창은 원하는 데이터를 빠르고 정확하게 검색해야 한다는 새로운 과제를 안겨주었다. 이러한 효율적 검색을 위해서 가장 중요한 기반이 되는 것이 바로 데이터의 적절한 표현이다. 2001년 국제 표준으로 제정된 MPEG-7은 바로 이러한 이유로 멀티미디어 데이터의 표현에 대한 표준화를 다루고 있다. 그러나 MPEG-7의 내용은 표준의 특성상 포함하는 범위가 방대하고 실제 검색시스템을 구축하려는 이들에게 다루기 힘든 것이 사실이다. 이에 본 논문에서는 MPEG-7에 제시되어있는 표준 중 비주얼 기술자들만을 이용하여 간단한 검색시스템을 구축하는 방법에 대하여 제시하고 그 검색 결과를 도시하였다. 또한, 개발된 시스템인 MPEG-7 VIRS(Video/Image Retrieval System)의 검색 결과를 통하여 각 비주얼 기술자를 이용한 검색과 다중 기술자들의 조합을 이용한 검색간의 결과를 분석하였으며 앞으로 MPEG-7을 이용한 검색 시스템이 나아갈 방향에 대한 간단한 제시를 하고 있다.

MPEG-7 based Video/Image Retrieval System (VIRS)

Jae-Ho Lee[†] · Hyung-Joon Kim[†] · Whoi-Yul Kim^{††}

ABSTRACT

An increasing in quantity of multimedia data brought a new problem that expected data should be retrieved fast and exactly. The adequate representation is a key element for the efficient retrieval. For this reason, MPEG-7 standard was established for description of multimedia data in 2001. However, the content of the standard is massive and the approach method is not clear for real application system yet, because of properties of MPEG-7 standard that has to include a lot of potential cases. In this paper, we suggested implementation scheme of retrieval system with using of only visual descriptors and presented the performance results of developed system. From the result of developed system, MPEG-7 VIRS (Video/Image Retrieval System), we analyzed the retrieval results between using individual descriptor and using multiple descriptors, and showed a layout for real application system.

키워드 : MPEG-7 비주얼(MPEG-7 Visual), 내용기반 검색(Content-Based Retrieval), 멀티미디어 검색(Multimedia Searching), 비주얼 기술자(Visual Descriptor)

1. 서 론

최근 다양한 디지털 기기들의 발달과 이들의 대중적 보급으로 디지털 이미지나 비디오의 제작이나 저장 등의 활용이 일반화 되고 있다. 또한 인터넷의 전 세계적인 확산은 멀티미디어의 양적 팽창을 돕는데 큰 역할을 하고 있다. 멀티미디어 데이터의 양적/질적 팽창은 검색기술의 필요성을 한층 부각시키는 결과를 가져오게 되었다. 이러한 이유로 전세계적으로 원하는 데이터를 쉽고 정확하게 찾으려는 연구들에 대한 활발한 진행이 이루어지고 있는 상황이다. 효율적인 검색을 위하여 반드시 이루어져야 할 기반적 요소가 바로 데이터의 적절한 표현이라 하겠다. MPEG-7은 그 정식 명칭인 "멀티미디어 콘텐츠 표현에 관한 인터페이스"

에서 알 수 있듯이 검색을 위한 콘텐츠 표현의 규격을 만드는 데 그 목적을 두고 있다[1].

MPEG-7의 표준화와 맞물려 MPEG-7의 효율성과 이를 이용한 검색 시스템들에 대한 연구가 종종 발표되어 왔다. R. B. Johnson은 그의 논문에서 멀티미디어와 MPEG-7의 관계에 대하여 설명하고 있다[2]. 각 비주얼 기술자들에 대한 연구로 S. Jeannin 등은 Motion Activity와 Motion Trajectory 기술자를 이용한 응용 시나리오를 보였으며[3], M. Bober 는 Contour Shape 기술자와 어플리케이션에 대하여 설명하였고[4], D. S. Messing등은 컬러와 지역 공간 정보를 이용한 검색 응용에 대하여 제시하였다[5]. 그러나 이러한 논문들은 각 저자들의 표준에 제안한 기술자들에 대하여 중점적으로 다루어지고 있으며, 활용에 대한 예시도 제안한 기술자에 한정되어 있다. MPEG-7 기술을 이용한 실제 시스템 개발에 관련된 연구들도 최근 발표되고 있다.

[†] 준 회 원 : 한양대학교 대학원 전자전기컴퓨터공학부

^{††} 종신회원 : 한양대학교 전자전기공학부 교수

논문접수 : 2003년 5월 20일, 심사완료 : 2003년 8월 22일

T. Echigo 등 IBM과 NTT의 연구원들은 모바일 폰에서의 개인 휴대폰 전송에 적합하게 요약한 비디오 전송에 관한 연구를 제시하였고[6], N. Fatemi 등은 표준에 근거하여 뉴스 동영상의 표현에 관한 모델구축을 이용한 검색시스템을 보였으며[7], T. Ebrahimi 등은 근래에 표준의 활용성이 증가할 것으로 예상되고 있는 응용분야인 물체추적, 가상현실 또는 가상 디스플레이 등의 분야를 위한 MPEG-7 카메라에 대하여 소개하였다[8]. 이들은 MPEG-7이 활용될 수 있는 각각의 응용분야에 대한 좋은 제안들을 하고 있으나, 표준의 어떤 항목을 사용하여 어떻게 실제적으로 시스템 개발할 지에 대한 내용은 부족하다. 이는 현재 MPEG-7 표준이 실제 사용을 위한 표준화 프로파일링 단계에 머물러 있어 예상 가능한 각각의 응용분야를 위한 적합한 구조가 정의되어 있지 않기 때문이다.

이에 본 논문에서는 하나의 기술자만을 사용한 시스템이 아닌 표준에 제안되어 있는 대부분의 비주얼 기술자를 사용하여 검색시스템을 제작하는 과정과 그의 결과를 제시하려 한다. 또한 개발된 검색 시스템인 MPEG-7 VIRS(Video/Image Retrieval System)의 성능 평가를 위하여 조합된 비주얼 기술자들을 통한 검색 결과를 분석하고 MPEG-7의 실제 검색 시스템의 개발에 필요한 요소들을 점검해보려 한다.

본 논문의 구성은 2장에서는 현재까지 개발되어진 대표적인 내용기반 검색 시스템들에 대하여 언급하였고 3장에서는 시스템 개발에 사용된 비주얼 기술자에 대한 설명을 하고 있다. 또한 4장에서는 본 연구에서 개발된 MPEG-7 VIRS의 대한 설명을 하고 있으며, 시스템의 결과와 이를 분석한 내용이 5장에 담겨 있다. 마지막 6장은 결론으로 이루어져 있다.

2. 내용기반 검색 시스템

본 장에서는 현재까지 여러 연구기관이나 대학 등에서 개발된 대표적인 내용기반 검색 시스템들에 대하여 간략히 설명하고자 한다. 이는 검색 시스템 개발자들과 연구자들에게 실제 개발된 검색 시스템에 대한 정보를 공유함으로써 개발되어질 검색 시스템에 대한 길잡이를 주고자 그 목적이 있다.

2.1 VisualSEEK 시스템

Columbia University at New York 에서 개발한 최초의 종합적인 검색 시스템으로 이미지 영역 칼라, 모양, 공간적 위치 및 키워드에 의한 검색을 제공한다. 자동으로 이미지들을 웹에서 획득하여 적당한 카테고리 분류하고 이들은 알맞은 키워드로 저장된다. 검색을 할 때 분류된 카테고리

안에서만 검색하도록 정의되어 있으며, 피드백 기능을 제공한다. 현재 WebSEEK, VidoEQ(사용자의 특정 움직임 질의에 의한 비디오 검색), MetaSEEK(웹에서의 이미지 검색을 위한 meta 검색 엔진) 등의 시스템으로 보여지고 있다[9].

2.2 MARS 시스템

Illinois 대학에서 만든 이미지 검색 시스템으로 사용자 피드백 기능을 강화한 시스템이다. 사용자의 피드백은 각 특징에 대한 가중치 값을 조절하는데 사용되며, 이러한 가중치 조절을 통하여 각기 다른 방식의 유사도 측정이 이용된다[10].

2.3 Blobworld 시스템

종합적이고 복합적인 컬러영역을 검색 기술을 응용하여 University of California at Berkley 에서 만든 시스템이다. 칼라처리 기술에 기반을 둔 검색 시스템으로 영역별 검색 기술도 제공한다[11].

2.4 QBIC system

IBM에서 개발한 이미지 검색 시스템으로 가장 널리 알려진 제품이다. 키워드에 의한 검색 및 칼라, 질감, 모양 등의 특징 조합을 이용한 검색을 지원한다. 현재 흑백 이미지의 검색과 비디오 스토리보드 생성 등의 추가 기능이 첨부되어 있다. 검색 질의 방식으로는 팔레트에서의 선택, 예제 기반 검색, 모양 그려서 검색 등이 제공되고 있다[12].

2.5 Photobook 시스템

MIT에서 개발한 초반 콘텐츠 기반 이미지 검색의 대표적인 시스템으로 모양, 질감 등 여러 종류의 특징을 사용하여 이미지를 검색한다. 이 시스템의 특징은 사용자가 적절한 특징을 선택하여 검색에 사용할 수 있다는 것이다[13].

2.6 Virage 시스템

이 시스템은 쿼리 인터페이스나 특정 이미지들을 위한 추가적 모듈에 대한 확장이 용이하도록 제작되었다. 현재 Altavista의 AV Photo Finder에서 사용중이며, 콘텐츠 유사도를 이용한 이미지 검색에 사용되고 있다. 또한 Video-logger 제품에는 비디오 데이터의 관리에 관한 기술들이 추가되어 있다[14].

위에 열거한 시스템 이외에도 Synapse[15] Surfimage[16], Netra[17], Informedia[18], PicHunter[19] 등 최근 들어 내용 기반 검색 시스템의 개발이 전 세계적으로 활발히 진행되고 지속적으로 그 결과물들이 발표되고 있다. 이러한 검색 시스템들에 대한 검색 성능의 평가는 정확한 판단 기준을 정하기가 어렵다. 이는 검색 결과를 객관적으로 증명할

만한 어떤 척도가 미비하기 때문이다. 사용자의 검색 결과물에 대한 피드백을 이용하여 시스템의 가중치를 재조정하는 방식에 대한 연구가 계속적으로 수행되는 이유는 검색 결과의 평가가 위한 이렇게 사용자마다 다르게 나타날 수 있기 때문이다[20].

현재까지 개발된 여러 검색 시스템들은 데이터를 표현하는데 각각의 특징 값들을 사용하여 검색을 수행하였다. 그러나 이러한 데이터 표현의 불일치성은 하나의 데이터에 대해서도 서로 다른 표현으로 나타내게 되므로, 데이터 표현에 대한 불필요한 중복성을 유발시킨다. 이러한 표현에 대한 이질감을 극복하여 동일한 데이터 표현으로 나타내게 하고자 하는 것이 바로 MPEG-7 표준화의 목적이라 하겠다.

3. MPEG-7 비주얼 기술자

2001년 제정된 MPEG-7 표준 버전 1은 크게 7개의 파트로 구성되어 있다[21]. <표 1>에서 MPEG-7 표준의 구성과 그 내용을 보인다.

<표 1> MPEG-7 표준의 구성

MPEG-7 표준	내용
Part-1 : System	콘텐츠와 기술자들간의 동기화를 위한 효율적인 전송과 저장에 관한 부분, 관리와 지적 재산권 보호 등도 포함되어있다.
Part-2 : DDL (Description Definition Language)	새로운 기술자 스키마를 표현하기 위한 언어
Part-3 : Visual	비주얼 정보의 표현과 표현 스키마에 대한 부분
Part-4 : Audio	오디오 정보의 표현과 표현 스키마에 관한 부분
Part-5 : MDS (Generic Entities and Multimedia Description Schemes)	일반적인 멀티미디어 표현과 표현 스키마에 관한 부분
Part-6 : Reference Software	표준에 포함되어 있는 소프트웨어에 관한 부분
Part-7 : Conformance Testing	MPEG-7 표현과 터미널에 대한 적합성 테스트에 대한 부분

MPEG-7 표준 문서를 접하게 되면 대부분의 많은 표준에서도 그렇듯이 그 내용이 방대하고 초보자가 접하기에는 한계가 있다는 것을 알 수 있다. 이는 표준에 특성상 가능한 많은 분야를 포함해야 하기 때문에 발생한 것이다. 이러한 방대한 내용은 실제 시스템 개발자로 하여금 혼란을 야기할 수도 있다. 현재 표준화 회의에서는 실제 사용에 관한 가이드를 주기 위하여 표준화의 내용에는 포함되어 있지만 Part-8로 'Extraction and Use of MPEG-7 Descriptions'를 두고 있다. 또한 프로파일링 작업등을 통하여 실제 어플리케이션에 이용될 형태 등에 대하여 논의하고 있는 실정이다.

본 논문에서는 표준의 구성 중 Part-3인 비주얼 기술자들만을 이용하여 검색 시스템을 구현하는 것을 목적으로 하였다. 이는 비주얼 기술자가 기술자 생성시에 별도의 입력 절차 없이 자동으로 표현될 수 있으며 이로 인해 시스템을 간단하게 구현할 수 있기 때문이다. 모든 비주얼 기술자들에 대한 자세한 설명은[21, 22]에서 찾을 수 있다. 본 장에서는 개발된 검색 시스템인 MPEG-7 VIRS에 사용한 비주얼 기술자들에 대하여 간략히 설명하도록 하겠다.

비주얼 기술자는 크게 각 기술자들의 표현의 기반을 제공하는 기본 구성(Basic Structure)과 칼라, 텍스처, 모양, 움직임, 영역 그리고 얼굴 등의 형태로 구성되어 있다. 본 개발에서는 하나의 기술자가 아닌 대부분의 비주얼 기술자의 사용을 목적으로 하고 있다. 본 개발에 사용된 기술자들을 열거하면 <표 2>와 같다.

<표 2> MPEG-7 VIRS에 사용된 비주얼 기술자의 종류

칼라 기술자(Color Descriptor)	Dominant Color, Color Layout, Color Structure
질감 기술자(Texture Descriptor)	Edge Histogram, Homogenous Texture, Texture Browsing
움직임 기술자(Motion Descriptor)	Motion Activity, Camera Motion, Motion Trajectory
모양 기술자(Shape Descriptor)	Region Shape

본 개발에서는 초기에 모든 분야에 걸쳐 표준에 열거되어 있는 기술자들을 전부 사용하려고 하였으나 몇몇의 비주얼 기술자들은 본 개발에는 사용되지 않았다. 이는 그 성능이 월등하지 않다는 것은 아니라 저자의 주관적인 판단에 따라 본 시스템 구현에 적합하지 않거나 기술자 추출이 용이하지 않은 것들은 배제하였다. 그러면 이제 MPEG-7 VIRS에 사용한 각 기술자들에 대한 특성들만 간단히 설명하도록 하겠다.

3.1 Dominant Color Descriptor

본 칼라 기술자는 적은 수의 칼라로 이미지를 표현하는데 적합하다. 최대 8가지의 칼라로 이미지를 표현할 수 있으며 각 칼라는 이미지에서 점유하는 퍼센트로 정의 되어진다.

3.2 Color Layout Descriptor

본 칼라 기술자의 목적은 공간적 정보와 칼라 정보를 동시에 이용하려는데 있다. 즉 이미지를 가로 8, 세로 8 크기의 영상으로 변형하고 각 영역의 칼라 정보를 구하는 것이다. 실제 저장되는 값은 YCbCr 계수 값들로 정의 되어져 있다.

3.3 Color Structure Descriptor

본 칼라 기술자는 칼라 히스토그램과 유사한 성질을 가진다. Structuring element 에 따라 칼라의 구조적 특성이 정의된다는 것이 차이점이라 하겠다.

3.4 Homogeneous Texture Descriptor

본 질감 기술자는 이미지의 정규성 즉, 방향성, 밀집성, 패턴의 정규성 등을 표현한다. 질감을 표현하기 위하여 주파수 레이아웃에서 에너지와 에너지 분산 등이 계산되어진다.

3.5 Texture Browsing Descriptor

본 질감 기술자는 다섯 개의 요소로 구성되어 있는데 인간의 질감에 대한 인식방식과 같이 정규성, 밀집성, 방향성들에 대한 정의를 두고 있다.

3.6 Edge Histogram Descriptor

본 질감 기술자는 다섯 종류의(방향성 타입 4, 비방향성 타입 1) 에지 타입에 대하여 질감에 대한 표현을 정의하고 이를 이용하여 계산 되어진다.

3.7 Motion Activity Descriptor

본 움직임 기술자는 동영상에서의 움직임 정도를 기술하기 위하여 정의되었으며 다섯 개의 요소로 구성되어 있다.

3.8 Camera Motion Descriptor

본 움직임 기술자는 카메라의 움직임을 panning, zooming 등의 분류를 통하여 3차원적으로 기술하기 위하여 정의되었으며, 각각의 움직임에 대한 정도를 표현하고 있다.

3.9 Motion Trajectory Descriptor

본 움직임 기술자는 객체기반 검색에 주요하게 사용될 수 있도록 객체의 움직임을 위치에 대한 리스트를 기술하고 있다.

3.10 Region Shape Descriptor

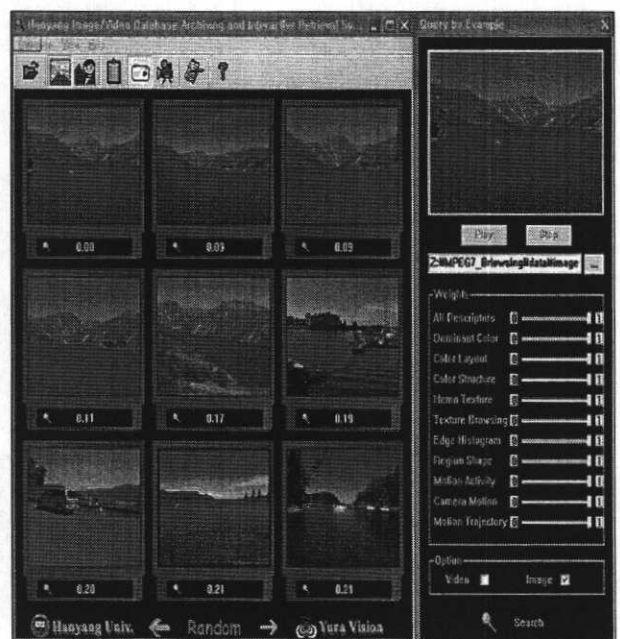
본 모양 기술자는 객체의 영역이 가지는 특징에 대하여 기술하고 있으며, 이미지의 사소한 변형에도 강인한 검색 능력을 보일 수 있도록 정의되어 있다.

4. 검색 시스템 구현

앞장에서 언급한 것과 같이 본 연구에서는 열 개의 비주얼 기술자 들을 이용하여 검색 시스템을 개발하였다. 이는 기존의 검색 시스템이 하나의 카테고리에 속하는 하나 혹

은 두어 개의 기술자들로 구성되어 있었다는 것과 비교할 만하다. 물론 2장에서 열거했다시피 이전에도 내용 기반 영상 검색에 관한 연구와 다른 종류의 특징을 결합하여 사용한 한 검색 시스템들이 그 중요성과 필요성이 인식됨에 따라 많이 개발되어 왔다. 또한, 이러한 꾸준한 연구와 개발이 MPEG-7의 표준의 제정에 많은 도움을 줄 수 있었다. 그러나, 표준이 제정되고 공표된 현재는 새로운 MPEG-7을 이용한 검색 시스템의 개발이 필요하다. 이에 본 연구는 MPEG-7 표준이 실제 검색 시스템에서 어떻게 사용되어질 수 있는지에 대한 예시와 그 검색 결과를 보이며 한다. 본 시스템의 특징 추출과정과 유사도 측정은 MPEG-7표준의 XM 문서와 소프트웨어를 참조하여 제작하였다[23, 24].

MPEG-7 VIRS 시스템에 사용한 최초의 데이터 베이스는 크게 칼라이미지와 비디오 클립 그리고 객체 이미지로 구성되어 있다. 이중 칼라 이미지데이터는 표준화 과정 중 칼라 기술자들의 성능평가를 위하여 사용하였던 CCD(Common Color Dataset) 5466장을 사용하였고, 객체이미지는 모양 기술자를 위하여 사용했던 데이터 셋에 칼라를 입혀서 사용하였다. 또한 동영상 클립은 표준화에 사용되었던 클립과 일반 영화나 인터넷에서 가져온 클립들로 구성되어 있다. 이중 새로 획득한 동영상에서는 동일한 장면의 그룹을 만들기 위하여 비디오 클립들은 하나의 동일한 샷에서 클립의 길이를 다르게 하거나 클립의 시작시점 조절하여 동일한 그룹의 비디오 클립을 제작하였다. 객체 이미지와 비디오 클립의 데이터 수는 각각 3917장과 1241개로 이루어져 있다.



(그림 1) MPEG-7 VIRS의 QBE 인터페이스와 검색 결과

개발된 시스템은 두 가지 방식의 검색 질의 방법을 제공한다. 하나는 샘플 영상을 이용하여 비슷한 이미지를 검색하는 QBE(Query By Example) 방법과 또 하나는 직접 사용자가 그림을 그려서 질의하는 QBD(Query by Draw) 방식이다. 아래 그림은 개발된 시스템의 QBE 방식의 인터페이스와 그 결과를 보여주고 있다.

QBE 방식은 왼쪽의 창에서 찾으려는 영상을 선택하여 검색 명령을 수행시키면, 오른쪽 상단에 자신이 질의한 영상이 보이며 왼쪽의 창에는 이에 대한 검색 결과들을 보이는 형식으로 작동한다. 결과 영상 아래에 보이는 수치들은 질의 영상과의 유사도를 나타내며 좌측 상단부터 우측으로 정렬 되어진 결과를 보인다. QBE 방식에서는 칼라 이미지의 경우 칼라 기술자들과 질감 기술자들의 여섯 개의 기술자들이 검색에 사용되었다. 이는 움직임 정보가 정지 영상에서는 없기 때문이다. 비디오 클립의 검색에서는 시스템에 적용된 열 개의 비주얼 기술자가 모두 사용되었고, 모양 기술자에 사용된 이미지에 색을 입힌 객체 이미지의 검색에서는 정지 영상과 동일한 수의 기술자들이 검색에 사용되었다. 또한 시스템 우측에서 보이는 것과 같이 각 기술자들의 가중치를 조정할 수 있도록 하여 검색 결과를 검색하는 사용자의 용도나 목적에 맞게 조절할 수 있도록 제작하였다. 즉, 사용자가 카메라의 움직임이 오른쪽으로 움직인 동영상 클립을 찾기를 원한다면 카메라 모션 기술자만을 사용하던가 가중치를 높게 선택할 수 있도록 하여 시스템의 응용성을 높였다. 또한 샘플 영상을 이용한 검색시, 이미 데이터베이스에 비주얼 기술자가 추출되어 있는 영상뿐 아니라 사용자가 임의의 영상을 선택하여 이 영상에서 비주얼 기술자를 자동으로 추출하고 검색하는 기능도 제공하고 있다.

QBE 방식은 이미 찾으려는 데이터에 대한 샘플영상이 존재할 경우 가장 손쉬운 방법을 제공한다. 그러나 샘플 영상이 없거나 찾기가 힘든 경우 이러한 QBE 방식은 무용지물이 되기 쉽다. 그러므로 본 시스템에서는 직접 사용자의 입력을 받아 검색을 하는 QBD 방식을 지원하여 이러한 경우를 대비하였다. QBD 방식은 팔레트에서 색상을 선택하고 각각의 그림 도구들을 이용하여 사용자가 직접 그림을 그린 후 이를 이용하여 질의하는 방식이다. QBD 방식은 상대적으로 QBE 방식보다 그 결과가 좋지 않게 나타난다. 그러나, 이러한 QBD 방식은 일차적으로 찾으려는 그림에 대하여 한번 검색하고 이러한 결과를 이용하여 다시 QBE 방식으로 검색을 하는 데에 유용하게 사용될 수 있을 것이다. 이러한 QBD 방식에서는 컬러 레이아웃 기술자만을 사용하여 구성하였다. 다음 그림은 QBD 방식의 인터페이스와 그 결과 영상이다.

현재 MPEG-7 VIRS는 인터넷 상에서 그 검색 결과를 확인할 수 있도록 <http://mpeg7.hanyang.ac.kr>에서 웹 서비스를 제공하고 있다. 직접 사용자의 이미지를 업로드하여 데이터 베이스에서 유사 이미지를 검색하는 기능도 제공되고 있다. 또한 새로운 데이터의 추가를 위하여 MPEG-7 VIRS는 웹에서 이미지들을 자동으로 수집하고 이를 분석하여 분류하는 필터링 기능이 추가 되어 있으며, 이를 이용하여 계속적으로 데이터베이스의 용량을 늘리고 있다. 이러한 필터링 기능도 MPEG-7 비주얼 기술자 들을 이용하여 제작되었다.



(그림 2) MPEG-7 VIRS의 QBD 인터페이스와 검색 결과

5. 결과 분석과 향후 과제

본 시스템에서는 여러 개의 MPEG-7 비주얼 기술자를 사용하여 동영상 및 이미지에 대한 검색 결과를 유도하였다. 이는 표준화 작업을 통해 각각 하나의 기술자의 성능만을 측정할 결과만을 알고 있는 상태에서 유익한 정보를 제공할 것이다. 그러나, 우리는 결과에 대한 평가를 할 때 아주 주의해야 할 것이 있다. 그것은 결코 데이터베이스와 검색 결과의 상관관계가 작지 않다는 것이다. 이는 어떠한 데이터 베이스를 사용하느냐에 따라 그 결과가 확연히 다르게 나타날 수 있다는 것이다. 또 생각해 보아야 할 것은 현재 사용하고 있는 데이터 셋의 그룹핑이 실제 존재하는 수 많은 영상들에 대한 완전한 대표적 그룹핑인가에 대하여서도 생각해 봐야 한다. 또한 그렇게 그룹핑한 것이 모든 사람들에게 동의될 수 있을 것인가에 대한 것도 고려해 봐야 한다. 이러한 그룹핑에 대한 문제는 보는 사람마다 시각의

차이가 있기 때문에 완전한 결론을 내리는 것은 무리이다. 그렇기 때문에 각각의 사람의 성향에 따라 시스템의 가중치가 결정되도록 하는 피드백 방식의 검색 시스템이 개발되고 있는 것이다[10, 19, 20]. MPEG-7 에서도 이러한 사용자의 성향에 따라 검색 방식을 결정하는 파트가 MDS파트에 포함되어 있다[27].

그래서, 본 논문에서는 모든 기술자들의 조합에 대한 성능 평가는 자체하려 한다. 이는 앞에서 설명하였듯이 각각 칼라와 질감 모양과 그리고 움직임 기술자가 표준화를 제정할 때 사용한 데이터 셋이 다르기 때문이다. 기술자들의 조합들 통한 검색 결과의 분석을 위하여, 본 논문에서는 가장 실제의 영상들과 상관도가 높다고 판단되는 CCD에 대하여 각각의 기술자들에 대한 성능을 평가하고 이를 도시하였다. CCD이미지 5446장은 총 50개의 그룹으로 구성되어 있으며 이에 대한 내용분석과 각 그룹의 이미지들에 대한 설명은 MPEG-7 문서에 자세히 기록 되어 있다[25]. 성능 평가에 사용한 기술자들도 칼라에 대하여만 제한을 두었다. 이는 질감과 모양 그리고 움직임 기술자에 대한 CCD 데이터 셋에서의 검색 결과 평가는 데이터 셋이 이들 기술자들에 대하여 그룹핑이 되어있지 않기 때문에 검색 결과에 대한 평가가 난해하기 때문이다.

또한 칼라기술자들만의 조합을 통한 검색 결과를 얻을 때에도 고려해야 할 사항이 있다. 이는 MPEG-7의 각 칼라 기술자를 표현하는 빈의 숫자가 다르다는 것이다. 여기에 유사도의 범위가 각 기술자마다 달라서 범위에 대한 정규화 작업을 반드시 해야 한다는 것이다. 본 논문에서는 각각의 기술자에 대하여 유사도 차이의 최소값과 최대값을 각각 0과 1에 대응시켜 정규화를 유도하였다. 식 (1)에 정규화한 값을 얻는 과정을 보인다.

$$D_{i,j}^{normalized} = \frac{D_{i,j} - D_{min}}{D_{max} - D_{min}} \quad (1)$$

위의 식 (1)에서 $D_{i,j}$ 는 이미지 i, j 에서의 유사도 값을 나타내며, D_{min}, D_{max} 는 각각의 결과에서 나타나는 최소와 최대의 유사도 값을 의미한다.

정규화하는 방법에 대하여는 많은 이론이 존재하나 본 논문에서는 가장 보편적으로 쉽게 접근할 수 있는 방법을 선택하였다. 이는 본 시스템이 각 기술자에 대한 가중치를 조정할 수 있다는 점을 이용한 것이다. 즉, 정규화의 과정과 가중치의 방식을 접목하여 CCD 데이터셋에 대한 검색 결과의 성능 향상을 보이는 칼라 기술자들의 조합을 계산하는 것이다. 본 논문의 목적은 현재의 시스템에서 완벽한 조화를 이루는 최고의 조합을 찾으려는 것이 아니라, 개발자들에게 MPEG-7 검색 시스템 구현과 검색 성능 분석의

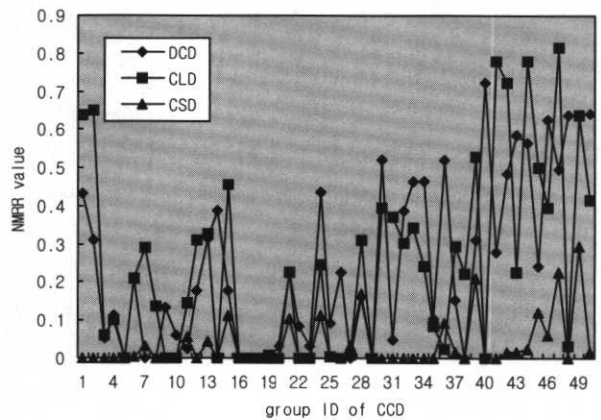
하나의 지침을 주는 데에 그 의미가 있다고 하겠다. 기존의 다른 검색 시스템에서 사용되었던 피드백을 이용한 검색 성능의 향상 및 신경망이나 각종 학습 방법을 이용한 최상의 특징 조합값의 구현 등에 대한 내용은 본 논문에서는 포함하지 않고 있다. 식 (2)에는 가중치를 사용한 최종 유사도 계산 과정을 보인다.

$$D_{i,j}^{new} = \sum_k^N w_k D_{i,j}^{normalized} \quad (2)$$

위의 식 (2)에서 w_k 는 k 번째 기술자의 가중치를 뜻하며, N 은 기술자들의 조합에 사용된 기술자의 개수를 나타낸다. 그리고 각 조합의 가중치의 합은 $\sum_k^N = 1$ 이다.

각 기술자의 가중치의 조정은 0부터 1까지 열 개의 단계로 구분되어 있다. 이 가중치는 시스템 인터페이스에서 슬라이드로 조정 가능하다. 두 개의 기술자를 쓸 경우 두 가중치 조절 슬라이드를 모두 1로 놓았다고 하더라도 실제 각 가중치는 1/2이 된다.

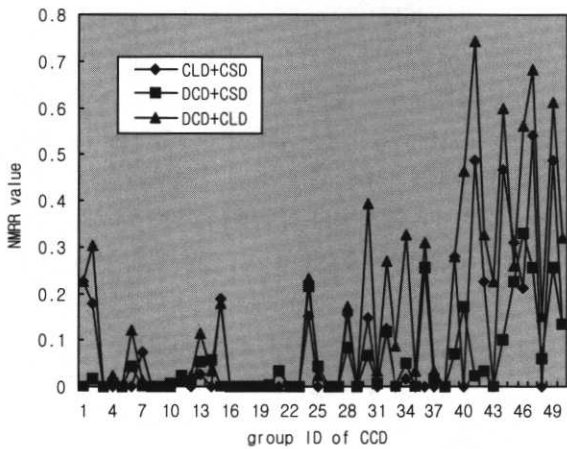
시스템 검색 결과를 위한 측정 척도로 ANMRR(Average Normalized Modified Retrieval Rank)를 사용한다. ANMRR은 MPEG-7에서 칼라 부분에 제안된 기술자들의 검색 효율을 평가하기 위한 평가 기준이다. ANMRR은 대표 질의 영상의 질의로부터 검색된 영상들 중, 질의한 영상과 같은 그룹에 속하는 영상의 개수와 그 영상들의 순위(rank)를 이용하여 검색효율을 평가하게 된다. ANMRR 값은 0에 가까울수록 질의한 영상과 같은 그룹의 모든 영상을 높은 순위에 잘 맞게 찾은 것을 의미한다. ANMRR의 자세한 계산식은[26]에서 찾을 수 있다. 아래 그림은 각 칼라 기술자 하나만을 이용한 경우의 검색 결과를 보여주고 있다.



(그림 2) 각각의 칼라 기술자를 이용한 검색 결과 : DCD(Dominant Color Descriptor), CLD(Color Layout Descriptor), CSD(Color Structure Descriptor)

결과에서 ANMRR의 값은 Color Layout의 경우 0.245, Dominant Color의 경우 0.252로 그 결과 값이 표준에서 나와 있는 다른 문서들의 결과와 비슷하게 산출되었다. 그러나, color Structure의 경우 ANMRR이 0.035로 다른 문서의 0.066 보다 약간 더 잘 나온 것을 볼 수 있다. 이는 쿼타이제이션 혹은 비트 표현 등의 차이 때문에 발생하는 문제이다. 그러나 오차 값이 범위가 확연히 크지 않으므로 본 시스템에서의 각 기술자의 예 대한 성능 측정이 올바르게 나왔다고 판단하였다.

이제 이러한 각 기술자들의 조합을 이용하여 검색 결과를 확인하여 보았다. 아래 그림은 두 개의 조합만을 이용하여 검색한 결과를 도시한 것이다. 이때의 가중치의 조정은 두 개의 기술자에 모두 1로 두었다.



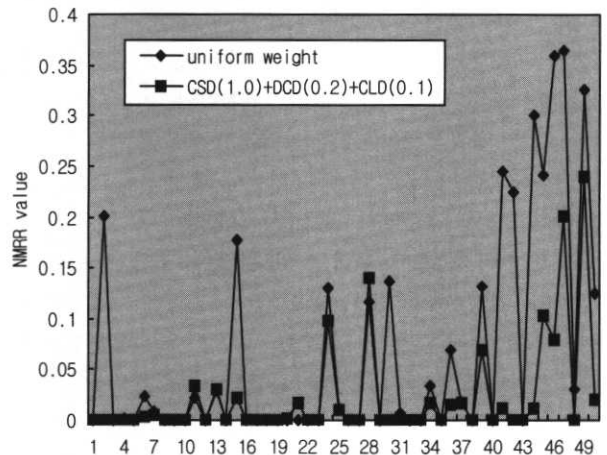
(그림 3) 두 개의 칼라 기술자의 조합을 이용한 검색 결과들

위의 두 개의 칼라 기술자를 조합하여 검색한 ANMRR 결과는 CLD + CSD의 경우 0.089, DCD + CSD의 경우 0.556, 그리고 DCD + CLD의 경우 0.163로 나타남을 알 수 있었다. 이러한 결과는 각각의 직관적으로 생각하기에 두 개의 기술자가 하나의 기술자보다 검색 결과가 좋으리라는 예상과 어긋남을 알 수 있다. Dominant Color와 Color Layout의 경우 두 기술자의 조합이 각각의 기술자만을 사용하였을 때 보다 확연히 좋은 것을 보인다. 그러나, Color Structure와 다른 칼라 기술자와의 조합은 오히려 Color Structure 하나만을 사용하였을 때의 결과 보다 좋지 않은 검색 성능을 보이고 있다.

이러한 기술자들의 조합에 있어서 몇 가지 간과하지 말아야 할 문제점들이 있다. 그 첫째는 각 기술자의 유사도의 분포가 가우시안 분포처럼 정규분포를 보이지 않는다는 것이다. 이는 간단한 최소값과 최대값을 통한 정규화는 어떤 영상에서 각 기술자에 따라 유사도 차이 값의 변화가 완전히 달라질 수 있다는 것이다. 두 번째로 기술자를 표현하는

변의 숫자가 모두 다르다는 것이다. 이는 가중치를 지금처럼 절반씩 준 경우 어느 하나에 기술자에게는 많은 영향을 미칠 수도 있다는 것이다. 마지막으로 각 기술자들에 대하여 서로 중복적인 성분이 존재한다는 것이다. 이는 칼라 기술자의 경우 완벽히 독립적인 성질의 서로 다른 기술자가 되지 않는다는 것이다.

그러나, 이러한 사항과 결과에 따라 MPEG-7 칼라 기술자의 조합은 좋은 결과를 낼 수 없다는 결론을 내리는 것은 성급한 판단이다. 왜냐하면 각 기술자마다 지향하는 용도가 다르기 때문이다. 즉, Dominant Color 기술자의 경우 적은 수의 대표 색깔만으로 유사한 영상의 검색을 시행하려는 목표에 따라 만들어졌으며, 컬러 레이아웃의 경우 영역의 정보를 추가하여 접근한 기술자이고, Color Structure의 경우 전체 영상의 칼라 분포의 구조를 감안하여 제안된 것이기 때문이다.



(그림4) 세 개의 칼라 기술자에 균등 가중치를 주었을 때와 가장 좋은 검색 결과를 가지는 가중치(CSD(1.0)+DCD(0.2)+CLD(0.1))를 주었을 때에 대한 검색 결과

본 논문에서는 CCD에 대하여 MPEG-7 기술자 중 하나의 기술자가 아닌 다수의 기술자를 이용하여 영상 검색에 적합한 조합을 유도하려고 시도하였다. 이러한 시도는 실제 검색 시스템의 개발에 있어, 기존의 연구에서 접근한 하나의 기술자에 대한 검색 결과는 다양한 시스템 개발에 충분한 정보를 제공하지 못하기 때문이다. 그러므로, 객관적으로 입증된 하나의 데이터 셋에 대하여 다수의 기술자를 이용한 검색 결과를 유도하고, 이를 통한 다수 기술자 조합의 성질을 연구함으로써 향후 MPEG-7 검색 시스템 개발에 기여하고자 함이다. 이에 본 논문에서는 사용된 세 개의 칼라 기술자를 모두 사용한다는 전제 하에 가중치를 조절해가면서 현재의 조건하에서 가장 좋은 결과를 가질 때의 검색 결과 값을 구해 보았다. 위의 그림은 세 기술자

의 가중치를 동일하게 1로 설정하였을 때와 가중치의 조정으로 가장 좋은 검색 결과를 나타냈을 때의 결과를 도시하였다.

세 개의 칼라 기술자를 모두 사용한다는 조건하에 구할 수 있는 조합의 수는 가중치의 조정 단계에 따라 무수한 경우의 수를 발생할 수도 있다. 이에 본 논문에서는 가중치 조절 범위에 각 단계를 0.1로 제한하여 실험하였다. 세 가지의 칼라기술자를 전부 사용한다는 조건으로부터 가중치의 조절 단계 0.1부터 1까지의 열 개의 단계가 된다. 여기에서 동일하게 가중치를 두었을 때는 하나의 의미만 가지므로 총 1331가지의 경우가 된다. 각 기술자에 동일한 가중치를 두었을 때의 ANMRR 값은 0.067로 나왔다. 그리고 가장 좋은 결과 값을 가질 때의 조합은 CSD : DCD : CLD 가 각각 1.0 : 0.2 : 0.1의 가중치로 책정되었을 때이며 그때의 ANMRR 값은 0.023였다. 이는 Color Structure만 사용한 검색 결과인 0.035 보다 더 좋은 결과를 보이는 것을 알 수 있었다.

현재 개발된 시스템인 MPEG-7 VIRS를 통하여 시스템이 포함하고 있는 모든 기술자를 이용하여 정지영상 및 동영상 검색 결과도 측정하고 있다. 그러나 본 시스템에서 사용하고 있는 동영상 클립의 데이터 베이스에 대한 객관적인 그룹핑이 검증되지 않은 상태이므로 본 논문에서는 그 값을 수치적으로 표현하고 있지 않다. 또한 정지영상의 경우에도 질감 기술자를 포함한 검색 결과가 그룹핑한 데이터 베이스에서 나온 검색 성능 수치는 적으나 질감의 특성을 고려하여 결과 이미지를 판단해 보면 결과로 보여주는 수치만으로는 좋지 않다는 평가를 내릴 수 없음을 알 수 있다. 즉 각 기술자들은 그 특성과 표현하는 빈의 수에 따라 각각의 용도가 변경될 수 있는 여건이 충분히 있다는 것이다. 예를 들어 Dominant Color와 같이 적은 빈수만으로 표현하고 있는 기술자를 대용량 데이터 베이스에서 원하는 영상만을 필터링 하는 경우에 사용하고, 그렇게 필터링된 이미지에 대하여 Color Structure를 이용하여 검색하는 방법도 하나의 어플리케이션이 될 수 있는 것이다.

MPEG-7에는 3장에서 언급한 것처럼 많은 비디오, 오디오, MDS 및 시스템 등 많은 파트로 구성되어 있으며, 각 파트 또한 그 내용이 상당한 것이 사실이다. 이러한 내용은 자칫 MPEG-7을 처음 접하는 시스템 개발자나 연구자들에게 있어 감당하기 힘든 과제로 주어질 수 있다. 그러나, MPEG-7에서 포함하고 있는 요소는 모든 항목이 필수요소가 아니며, 이러한 표준의 내용 중 표준의 기본적인 요구사항 외에 어떠한 추가 요소를 사용하여 검색 시스템을 구현할지는 개발자나 연구자의 몫이라 할 수 있다. 현재까지 MPEG-7 표준 회의에서도 정확한 프로파일링 작

업이 마무리 되지 않은 상태에서 실제 검색 시스템의 개발은 그 프로토타입을 보인다는 점에서 좋은 시도라 할 수 있다. 향후 과제로는 MPEG-7 VIRS의 특징 값 조합에 대한 학습 방법을 통한 개선책 연구와 검색 성능 향상을 위한 사용자 피드백 기능부가 등 기존의 내용기반 검색 시스템에서 연구되었던 항목들이 추가적으로 MPEG-7 기술자들에 대하여 수행 될 것이다.

6. 결 론

본 논문에서는 MPEG-7의 비주얼 기술자들만을 이용한 이미지 및 동영상을 위한 검색 시스템인 MPEG-7 VIRS에 대하여 구성과 그 검색 성능을 보였다. 이는 비주얼 기술자들만을 이용하여 사람의 개입 없이 자동으로 검색 시스템을 구축할 수 있는 MPEG-7 검색 시스템을 위한 하나의 프로토타입을 보인 것이다. 이러한 시스템의 접근 방법은 실제 모든 비주얼 데이터에 대하여 일일이 사람이 그 내용을 입력하기엔 역부족이라는 점을 감안할 때, 양적으로 급팽창하는 멀티미디어 데이터에 대한 검색 시스템의 좋은 제안이라 하겠다. MPEG-7 VIRS의 성능평가를 위하여 공인된 데이터셋인 CCD를 사용하였고 이 경우 단순 조합만을 이용하여 평가하였을 경우, 가장 알맞은 복합 기술자의 조합은 CSD : DCD : CLD 각각 1.0 : 0.2 : 0.1인 경우임을 확인하였다. 본 논문은 앞으로 MPEG-7을 이용하여 검색 시스템을 구축하려는 이들에게 하나의 방향 제시를 함으로써 MPEG-7 표준의 실제 응용으로의 접근에 기여할 것으로 예측된다.

참 고 문 헌

- [1] S. F. Chang, "Overview of the MPEG-7 Standard," IEEE Trans. On Circuits and Systems for Video Technology, Vol.11, No.6, pp.799-695, June, 2001.
- [2] R. B. Johnson, "Multimedia databases and MPEG-7," Electronics & Communication Engineering Journal, issue 3, Vol.13, pp.98-99, June, 2001.
- [3] S. Jeannin and A. Divakaran, "MPEG-7 visual motion descriptors," IEEE Transactions on Circuits and Systems for Video Technology, Vol.11, No.6, pp.720-724, June, 2001.
- [4] M. Bober, "MPEG-7 visual shape descriptors," IEEE Trans. on Circuits and Systems for Video Technology, Vol.11, No. 6, pp.716-719, June, 2001.
- [5] D. S. Messing, P. van Beek and J. H. Errico, "The MPEG-7 colour structure descriptor : image description using colour and local spatial information," Proc. International Con-

- ference on Image Processing, Vol.1, pp.670-673, 2001.
- [6] T. Echigo, K. Masumitsu, M. Teraguchi, M. Etoh and S. Sekihuchi, "Personalized delivery of digest video managed on MPEG-7," *Information Technology : Proc. Coding and Computing*, pp.216-220, 2001.
- [7] N. Fatemi and O. A. Khaled, "Indexing and retrieval of TV news programs based on MPEG-7," *Proc. International Conference on Consumer Electronics*, pp.360-361, 2001.
- [8] T. Ebrahimi, Y. Abdeljaoued, R. M. Figueras i Ventura and O. Divorra Escoda, "MPEG-7 camera," *Proc. International Conference on Image Processing, Vol.3*, pp.600-603, 2001.
- [9] J. R. Smith and S.-F. Chang, "VisualSEEK : A Fully Automated Content-Based Query System," *Proc. ACM Multimedia*, pp.87-98, 1996. <http://disney.ctr.columbia.edu/WebSEEK>.
- [10] S. Mehrotra, Y. Rui, M. Ortega-Binderberger and T. S. Huang, "Supporting Content-Based Queries over Images in MARS," *Proc. IEEE Int'l Conf. on Multimedia Computing and Systems*, pp.632-633, June, 1997. <http://jadzia.ifi.uiuc.edu : 8001/>.
- [11] M. Thomas, C. Carson and J. M. Hellerstein, "Creating a Customized Access Method for Blobworld," *Proc. 16th International Conference on Data Engineering*, pp.82-82, 2000, <http://elib.cs.berkeley.edu/photos/blobworld>.
- [12] W. Niblack, et al, "Updates to the QBIC system," *Proc. SPIE on Storage and Retrieval for Image and Video Databases, Vol.6*, pp.150-161, 1998, <http://www.qbic.almaden.ibm.com/>.
- [13] Pentland, R. W. Picard and S. Scarlo, "Photobook : Tools for Content-Based Manipulation of image Databases," *Proc. SPIE on Storage and Retrieval for Image and Video Databases II, Vol.2, Issue 185*, pp.34-47, 1994, <http://whitechapel.media.mit.edu/vismod/demos/photobook>.
- [14] A. Gupta, et al "The Virage image search engine : an open framework for image management," *Proc. SPIE on Storage and Retrieval for Image and Video Databases Vol.4*, pp.76-87, 1996, www.virage.com.
- [15] S. Ravela and R. Manmatha, "On computing global similarity in images," *Proc. IEEE Workshop on Applications of Computer Vision (WACV98)*, Princeton, NJ, pp.82-87, 1998, <http://cowarie.cs.umass.edu/~demo/>.
- [16] C. Nastar, "Surfimage : a flexible content-based image retrieval system," *Proc. ACM Multimedia '98*, Bristol, UK, 1998, <http://www-syntim.inria.fr/htbin/syntim/surfimage/surfimage.cgi>.
- [17] W. Y. Ma and B. S. Manjunath, "NeTra : a toolbox for navigating large image databases," *Proc. International Conference on Image Processing, Vol.1*, pp.568-571, 1997, <http://vivaldi.ece.ucsb.edu/Netra>.
- [18] M. G. Christel, R. B. Allen and E. Rasmussen, "Multimedia abstractions for a digital video library" *Proc. ACM Digital Libraries '97*, pp.21-29, 1997. <http://www.islip.com/fprod.htm>.
- [19] I. J. Cox, M.L. Miller, T. P. Minka, T. V. Pappathomas and P. N. Yianilos, "The Bayesian Image Retrieval System, PicHunter : Theory, Implementation and Psychophysical Experiments," *IEEE Trans. Image Processing, Vol.9, No.1*, pp.20-37, 2000.
- [20] S. Kulkarni, B. Verma, P. Sharma and H. Selvaraj, "Content Based Image Retrieval Using a Neuro-Fuzzy Technique," *Proc. IEEE Int'l Joint Conf. on Neural Networks*, pp.846-850, July, 1999.
- [21] B. S. Manjunath, et al., "Introduction to MPEG-7," John Wiley & Sons Ltd., West Sussex, England, 2002.
- [22] ISO/IEC 15938-3, "Multimedia Content Description Interface-Part 3 : Visual," version 1, 2001.
- [23] A. Yamada, M. Pickering, S. Jeannin, L. Cieplinski, J. R. Ohm and M. C. Kim, "MPEG-7 Visual part of eXperimentation Model Version 10.0," *ISO/IEC JTC1/SC29/WG11/N4063*, Singapore, March 2001.
- [24] http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html, MPEG-7 Experimental Model Software.
- [25] D. Zier and J-R. Ohm "Common Datasets and Queries in MPEG-7 Color Core Experiments," *ISO/IEC JTC1/SC29/WG11/M5060*, Melbourne, October, 1999.
- [26] P. Ndjiki-Nya et al., "Subjective Evaluation of the MPEG-7 Retrieval Accuracy Measure (ANMRR)," *ISO/IEC JTC/SC29/WG11/M6029*, Geneva, May, 2000.
- [27] ISO/IEC 15938-3, "Multimedia Content Description Interface-Part 5 : MDS," version 1, 2001.



이 재 호

e-mail: jhlee@vision.hanyang.ac.kr

1999년 한양대학교 전자전통전파공학군
(공학사)

2001년 한양대학교 전자전기공학부 대학원
(공학석사)

2001년~2001년 일본 NHK 방문연구원

2001년~현재 한양대학교 전자전기 컴퓨터공학부 대학원 박사

과정

관심분야 : 얼굴 검출 및 인식, 내용기반 검색, 실시간 동영상 처리



김형준

e-mail: khjoon@vision.hanyang.ac.kr

1999년 한양대학교 전자전통전파공학군
(공학사)

2001년 한양대학교 전자전기공학부 대학원
(공학석사)

2001년~현재 한양대학교 전자전기컴퓨터
공학부 대학원 박사과정

관심분야 : 내용기반 검색, 오브젝트 트래킹



김희울

e-mail: wykim@hanyang.ac.kr

1980년 한양대학교 전자공학과(공학사)

1983년 Pennsylvania State University

전자공학과(공학석사)

1989년 Purdue University, West Lafayette,

전자공학과(공학박사)

1989년~1994년 Erick Johnson School of Engineering and Com-
puter Science at the University of Texas at Dallas
조교수

1994년~현재 한양대학교 전자전기공학부 교수

관심분야 : 컴퓨터 비전, 머신 비전, 내용기반 이미지 검색,
MPEG-7