일반논문-04-09-2-06

MPEG-7 표준에 따른 내용기반 비디오 검색 시스템

김 형 준[a], 김 회 율[a]‡

Content-based Video Indexing and Retrieval System using MPEG-7 Standard

Hyoung-Joon Kim[a] and Whoi-Yul Kim[a]‡

요 약

본 논문에서는 비디오의 효율적인 검색과 관리를 위해 MPEG-7 표준에 따른 내용기반 비디오 검색 시스템을 제안한다. 제안된 시스템은 비디오 DB 구축을 위한 인덱싱 모듈과 웹을 통한 비디오 검색 모듈로 구성되며 검색 모듈에서는 다양한 질의 방법을 지원한다. 비디오 인덱싱 모듈은 관리자가 입력한 키워드, 인덱싱 모듈이 자동으로 추출한 등장 인물 정보와 MPEG-7 비주얼 서술자와 같은 메타데이터를 서버에 저장한다. 일반 사용자는 웹을 통해 검색 모듈에 접근하며 키워드, 얼굴, 예제 및 스케치 질의와 같은 다양한 질의 방법을 통해 원하는 비디오를 검색할 수 있다. 이러한 비디오 검색 시스템을 구성하기 위해서 본 논문에서는 효율적인 비디오 인덱싱을 위한 장면 전환 검출 방법으로 ATC(Adaptive Twin Comparison)와 사용자 편의성을 위한 개선된 내용기반 질의 방법으로 QBME(Query By Modified Example)를 제안한다. 실험에서 제안된 장면 전환 검출 방법이 기존의 방법보다 우수함을 보였고, 제안된 질의 방법을 통해 기존의 질의 방법인 QBE(Query By Example)나 QBS(Query By Sketch) 보다 사용자에게 검색의 편의성을 제공할 수 있음을 보였다.

Abstract

In this paper, we propose a content-based video indexing and retrieval system using MPEG-7 standard to retrieve and manage videos efficiently. The proposed system consists of video indexing module for a video DB and video retrieval module to allow various query methods on a web environment. Video indexing module stores metadata such as manually typed in keywords, automatically recognized character names, and MPEG-7 visual descriptors extracted by indexing module into a DB in a sever side. A user can access to retrieval module by a web and retrieve desired videos through various query methods like keywords, faces, example and sketch. For this retrieval system, we propose ATC(Adaptive Twin Comparison) as a cut detection method for efficient video indexing and QBME(Query By Modified Example) as an improved content-based query method for the convenience of users. Experimental results show that the proposed ATC method detects cuts well and the proposed QBME method provides the conveniences better than existing query methods such as QBE(Query By Example) and QBS(Query By Sketch).

Keywords : MPEG-7, video retrieval, cut detection, query by sketch

a) 한양대학교 전자통신전파공학과
Division of Electrical and Computer Engineering, Hanyang Univ.
※ 본 연구는 한양대학교 교내 연구 특성화 연구팀 공모 사업과 KBS 기술연구소의 지원을 받아 수행되었습니다.

I. 서 론

최근 컴퓨터와 통신 기술의 급속한 발달로 인해 인터넷

을 통한 멀티미디어 데이터의 전송이 보편화되어 정보 통신망 사용자들이 시간과 장소에 상관없이 다양한 정보에 접근할 수 있게 되었으며 이용할 수 있는 멀티미디어 정보의 양도 폭발적으로 증가하게 되었다. 하지만 정보량이 늘어날수록 원하는 정보를 찾기는 더욱 어려워지고 멀티미디어 정보의 검색, 저장, 관리 기술에 대한 요구가 늘어나면서 대용량 비디오 데이터베이스에서 원하는 정보를 빠른 시간 내에 검색할 수 있는 검색 기법이 필요하게 되었다. 이를 위해 기존의 텍스트 기반 정보 검색과는 달리 멀티미디어 컨텐츠의 내용을 기반으로 하는 다양한 멀티미디어 검색 방법이 제안되었다[1].

내용기반 검색을 위해 QBIC, VisualSEEk, Photobook, Virage, Cypress, CVEPS, JACOB 등 다양한 검색 시스템이 제안되었다. QBIC은 IBM에서 개발한 이미지 검색 시스템으로 키워드에 의한 검색 및 칼라, 질감, 모양 등의 조합을 이용하여 검색을 지원하고[2], VisualSEEk는 지역적인 특징 질의 및 사용자 피드백을 위한 히스토그램 정제 등의 기능을 제공한다[3]. Photobook은 사용자가 자신만의 내용 분석 기능과 학습을 통해 사용자 피드백에 기반한 특징 선택 기능을 제공한다[4]. Virage 시스템은 질의 인터페이스나 특정 이미지들을 위한 추가적 모듈에 대한 확장이 용이하도록 제작되었다[5]. 또한 VideoLogger 제품에는 비디오 데이터의 관리에 관한 기술들이 추가되어 있다. Cypress는 예를 들어 해변을 표현하기 위해 노란색으로 태양을 그리고 베이지색으로 모래를, 파란색으로 바다를 표현하는 것과 같은 스케치 질의 방법을 제공한다[6]. CVEPS와 JACOB는 비디오에서 장면 전환을 검출하고, 키프레임이나 물체에 기반하여 인덱싱 및 검색이 이루어진다. CVEPS는 또한 압축 도메인에서 비디오 에디팅 기능을 제공하며[7], JACOB는 신경망을 이용하여 샷 검출을 수행한다[8]. 이러한 기존의 검색 시스템들은 멀티미디어 콘텐츠를 표현하기 위한 메타데이터의 형식이 서로 다르기 때문에 동일한 멀티미디어 콘텐츠라 하더라도 시스템 상호간의 호환이 되지 않고 각각의 시스템은 다른 형식의 메타데이터를 저장하기 위해 중복적인 메타데이터 DB를 필요로 하는 문제가 있다. 따라서 멀티미디어 콘텐츠 표현에 관한 표준으로서 MPEG-7이 제정되었고[9], 이를 이용한 검색 시스템으로 MPEG-7 VIRS가 제안되었다[10].

현재 웹에서 비디오 검색 서비스를 제공하는 기능은 키워드 질의에 기반하는 경우가 일반적이고 일부 테스트베드 시스템들이 내용기반 검색 방법을 제공한다. 이에 반해 본 논문에서 제안하는 통합형 비디오 검색 시스템은 키워드, 등장 인물, 예제 및 스케치 질의 방법 중에서 하나 혹은 둘 이상을 조합하여 비디오를 검색하는 기능을 제공한다. 구체적으로는 1) 자동으로 추출되는 MPEG-7 비주얼 서술자들을 이용하여 내용기반 비디오 검색 기능을 제공하고, 2) 비디오에 등장하는 인물을 자동 인식하여 의미기반 비디오 검색 기능을 제공하고, 3) 비디오로부터 추출된 메타데이터는 MPEG-7 국제 표준에 따라 표현되어 타 검색시스템 혹은 다른 MPEG-7 어플리케이션과 상호 호환하여 쉽게 시스템의 확장과 다양한 응용을 기대하였으며, 4) 웹 기반에서 다양한 형태의 질의 방법을 통해 비디오 검색 서비스를 제공하는 통합형 비디오 검색 시스템을 제안한다.

제안된 비디오 검색 시스템은 비디오 DB 구축을 위한 비디오 인덱싱 모듈과 웹을 통한 비디오 검색 모듈로 구성된다. 비디오 인덱싱 모듈은 입력된 비디오로부터 메타데이터를 자동으로 추출하고 이를 MPEG-7 국제 표준 규격에 따라 서술하여 서버에 저장한다. 효과적인 비디오 인덱싱을 위해서는 정확한 장면 전환 검출이 먼저 이루어져야 하며, 급격한 장면 전환 뿐만 아니라 점진적인 장면 전환에도 강인하게 장면 전환을 검출할 수 있어야 한다. 이러한 장면 전환 검출을 위해서 본 논문에서는 기존의 적응적 임계값[11]과 twin comparison 방법[12]을 기반으로 하여 점진적인 장면 전환 구간이 갖는 특성을 이용한 ATC(Adaptive Twin Comparison) 방법을 제안한다. 비디오 검색 모듈에서는 네 가지 형태의 질의 방법으로서 키워드, 얼굴, 예제 및 스케치 질의 방법을 지원하도록 구성되어 있다. 특히 예제 및 스케치 질의와 같은 내용기반 검색을 위해서 QBME(Query By Modified Example)를 제안한다. 기존의 예제 질의 방법(QBE)에서는 사용자가 검색을 원하는 영상과 유사한 영상을 가지고 있어야 하고, 반면에 스케치 질의 방법(QBS)에서는 사용자가 생각하는 영상을 표현하는데 어려움이 있어 극히 단순한 형태의 스케치 형태를 제외하고는 그 활용도에 한계가 있다. 반면에 QBME는 사용자가 주어진 영상을 단순화 시킴으로써 쉽고 편리하게 유사한 영상으로 수정 하거나 새로운 영상을 구성할 수 있는 방법을 제공한다.

본 논문의 구성은 다음과 같다. 2장에서 구현된 비디오 검색 시스템을 소개하고, 3장과 4장에서 각각 장면 전환 검출 방법과 QBME를 설명한다. 5장에서 장면 전환 검출 방법과 QBME에 대한 실험 결과를 보이고 6장에서 결론을 맺는다.

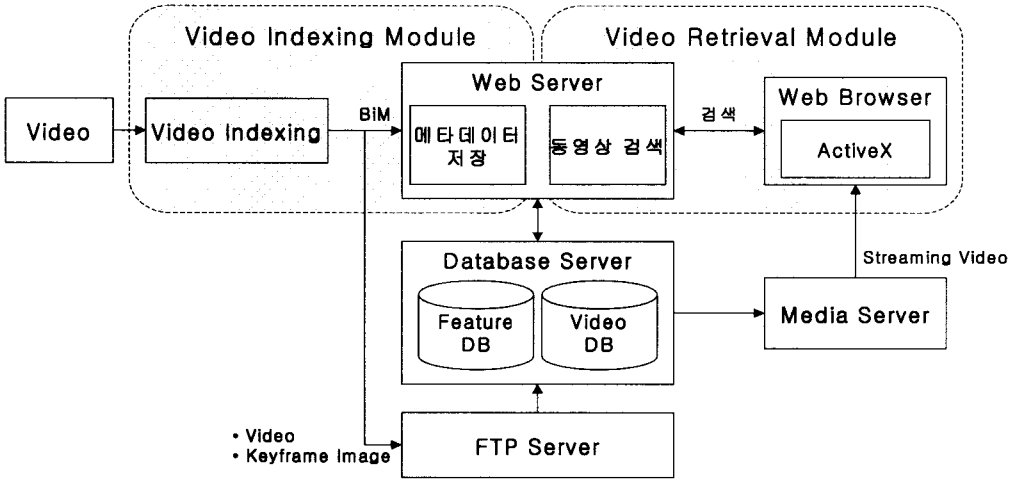


그림 1. MPEG-7 기반 통합형 비디오 검색 시스템의 전체 구성도
Fig. 1. The overall structure of an integrated video retrieval system based on MPEG-7 standard

Ⅱ. 통합형 비디오 검색 시스템

구현된 비디오 검색 시스템은 그림 1과 같이 비디오 인덱싱 모듈과 비디오 검색 모듈로 구성되어있다. 관리자는 새로운 비디오를 인덱싱 클라이언트를 이용하여 비디오 메타데이터와 비디오를 서버에 전송하여 데이터베이스에 추가한다. 일반 사용자는 웹 브라우저를 이용하여 서버에 접속한 후, 키워드, 얼굴, 예제 및 스케치 질의 방법을 이용하여 자신이 원하는 비디오를 검색한다.

전체 시스템은 기능적으로 네 개의 서버, 즉 웹 서버, FTP 서버, 스트리밍 서버, 데이터베이스 서버로 구성되어 있다. FTP 서버는 인덱싱 과정에서 비디오를 서버로 전송하기 위한 것으로, 스트리밍 서버와 같은 곳에 위치한다. 이 모든 것은 한 개의 PC에 모두 구성되어 있으나 성능 향상을 위해서 여러 개의 PC로 분산되어 설치될 수 있다. OS로는 Windows 2000 Server를 사용하였고, 웹 서버 및 FTP 서버로는 IIS 5.0, 스트리밍 서버로는 Windows Media Server, 데이터베이스는 MSSQL 2000을 사용하였다.

1. 비디오 인덱싱 모듈

비디오 인덱싱 모듈은 비디오 검색에 필요한 특징값 추출을 위한 것으로 그림 2와 같은 일련의 과정을 갖는다. 특

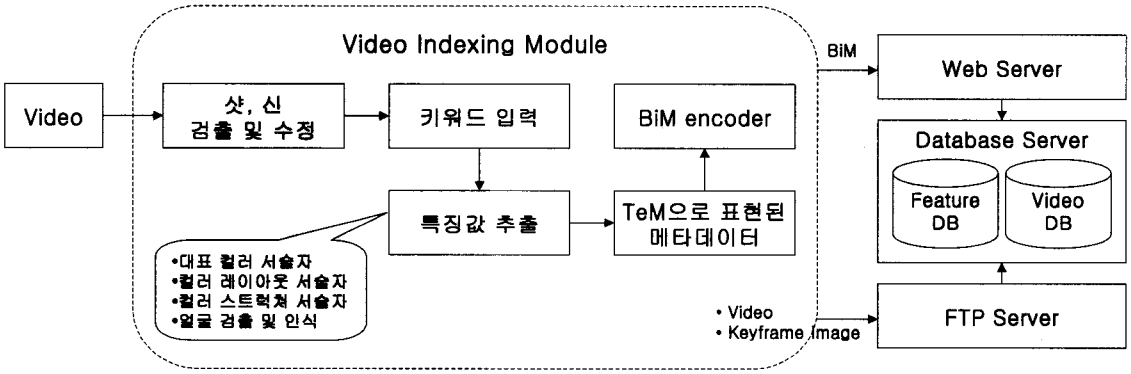


그림 2. 비디오 인덱싱 과정
Fig. 2. The video indexing process

히 비디오의 경우에는 샷이나 신과 같은 구조적인 분할이 우선 필요하며, 분할된 각 샷에서 정의된 키프레임으로부터 비주얼 특징값을 추출하여 데이터베이스에 그 정보를 저장한다.

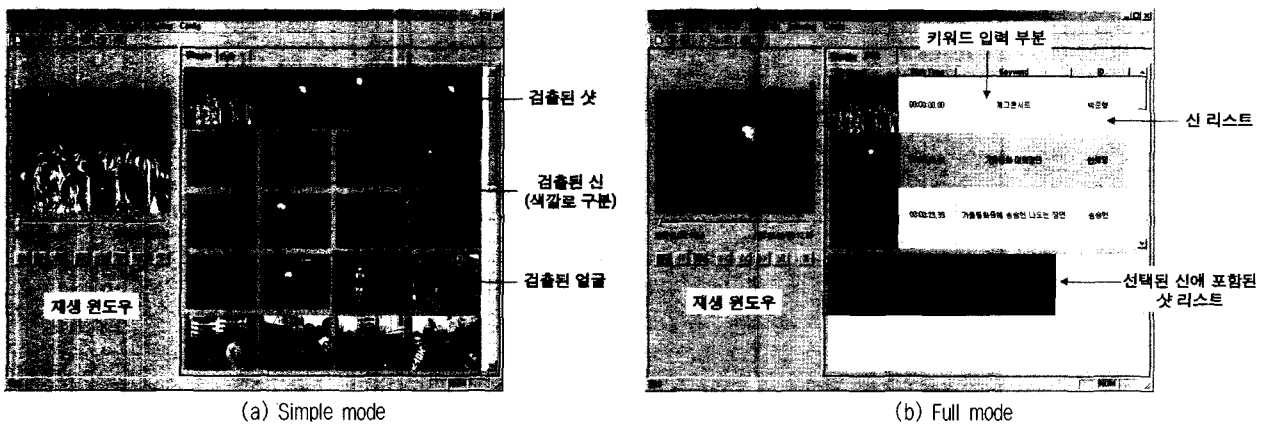
3장에서 소개하는 장면 전환 검출 방법을 이용하여 비디오가 샷 단위로 구분이 이루어지면 각 샷의 첫 번째 프레임을 키프레임으로 선정하고 Lee가 제안한 방법[13]과 같이 시간 연속성을 고려하여 신을 판단한다. 판단된 샷과 신 정보가 잘못된 경우 관리자의 수정을 거친 후, 키워드 검색을 위해 각 신에 대해 키워드 입력을 한다. 내용기반 검색에 필요한 특징값으로는 MPEG-7에 정의된 대표 컬러 서술자, 컬러 레이아웃 서술자, 컬러 스트럭쳐 서술자를 이용하였고[14], 각 샷의 키프레임으로부터 이들 특징값을 추출한다. 또한 비디오에 등장하는 인물 정보를 이용하여 검색에 사용할 수 있도록 하기 위해 SVM을 이용한 얼굴 검출[15]과 DCT/LDA를 이용한 얼굴 인식[16]을 적용하여 등장 인물에 대한 정보를 추출한다. 기본적으로 등장 인물 분석은 비디오의 모든 프레임에 대해 수행되지만, 인덱싱 시간을 줄이기 위해 키프레임에 대해서만 수행되어질 수 있다.

위와 같은 과정을 거쳐서 생성된 비디오 메타데이터는 MPEG-7 MDS[17]에 정의된 스키마에 따라서 XML로 표현된다. XML로 표현된 메타데이터는 텍스트 형식(TeM: Text format for MPEG-7)을 가지며 한 시간 길이의 비디오에 대해서 키프레임 수에 따라 적게는 2MByte에서 많게는 4Mbyte 크기를 갖는다. 이를 서버로 전송하는데 있어서 전송 시간을 줄이기 위해 BiM Encoder를 이용하여 이진 파일(BiM: Binary format for MPEG-7)로 변환하여 압축한다. 여기서 사용된 BiM Encoder는 EXPWAY사의 BiM XM 소프트웨어를 사용하였다[18][19]. 최종적으로 BiM으로 표현된 메타데이터는 HTTP를 통해 서버로 전송되고, 비디오 데이터와 키프레임 이미지는 FTP를 통해 전송되어 데이터베이스에 저장된다.

구현된 비디오 인덱싱 프로그램은 그림 3과 같이 simple mode와 full mode로 구성된 인터페이스를 갖는다. Simple mode에서는 신과 샷의 구성을 확인할 수 있으며, 특정 신에 속한 샷들은 같은 색으로 표현되어서 각각의 신의 구성을 쉽게 알 수 있다. Full mode는 각 신에 속한 등장 인물의 이름과 샷의 구성을 보여주고 사용자는 신 단위로 키워드를 입력할 수 있다.

2. 비디오 검색 모듈

비디오 검색 모듈은 인덱싱 과정에서 입력된 키워드, 등장 인물 정보, MPEG-7 비주얼 서술자 등을 이용하여 키워드, 얼굴, 예제 및 스케치 질의 방법을 지원하도록 구성되어 있다. 키워드 질의 검색은 인덱싱 과정에서 관리자가 입력한 키워드 정보를 검색 대상으로 하여, 질의한 키워드가 포함되어 있는 신을 검색해서 보여준다. 얼굴 질의 검색에서는 직접 이름을 입력하거나 혹은 영상에서 자동으로 검출된 얼굴에 대해 얼굴 인식을 하고, 그 인식된 얼굴의 이름을 질의할 수 있다. 데이터베이스에는 각 샷에 등장하



(a) Simple mode (b) Full mode

그림 3. 비디오 인덱싱 프로그램 인터페이스
Fig. 3. Interfaces of the video indexing program: (a) simple mode, (b) full mode

는 인물의 이름이 저장되어 있고, 이를 대상으로 얼굴 검색이 이루어진다. 예제 질의 검색은 MPEG-7 XM[14]에 정의된 대표 컬러 서술자, 컬러 레이아웃 서술자, 컬러 스트럭쳐 서술자의 각 유사도에 대한 가중합을 이용한다. 즉, 사용자가 질의한 영상에서 추출된 특징값과 데이터베이스에 들어있는 샷의 키프레임의 특징값에 대해 식 (1)과 같은 유사도 계산식을 이용하여 유사한 샷을 검색하여 사용자에게 보여준다.

$$D_{total}=\sum_i w_i D_i \quad (1)$$

where $i \in$ {대표 컬러, 컬러 레이아웃, 컬러 스트럭쳐 서술자}

여기서, D_i는 대표 컬러 서술자, 컬러 레이아웃 서술자, 컬러 스트럭쳐 서술자에 대한 유사도를 의미한다. 각 서술자의 유사도는 그 범위가 서로 다르기 때문에, 각 서술자의 유사도는 최대값과 최소값을 이용해서 0에서 1 사이로 정규화하였다. 가중치 w_i는 대표 컬러 : 컬러 레이아웃 : 컬러 스트럭쳐 서술자 각각 1.0 : 0.2 : 0.1을 사용하였다[10]. 마지막으로 스케치 질의 검색은 예제 질의 검색과 같은 과정으로 이루어지나, 스케치 질의 영상을 만드는 과정에서 4장에서 소개하는 QBME 방법을 통해 사용자는 쉽게 원하는 질의 영상을 만들 수 있도록 한다.

비디오 검색을 위한 사용자 인터페이스는 웹 브라우저에서 동작할 수 있도록 ActiveX로 구현되어 현재 웹에서 비디오 검색 서비스를 제공하고 있다[20]. 사용자는 Internet Explorer를 이용하여 웹 서버에 접속 후 그림 4와 같은 인터페이스를 통해 원하는 비디오를 검색할 수 있다. 그림 4

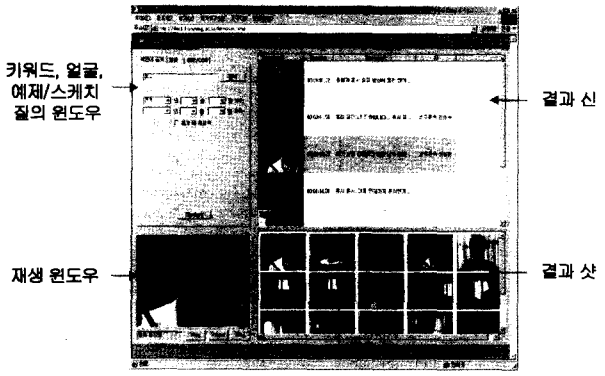
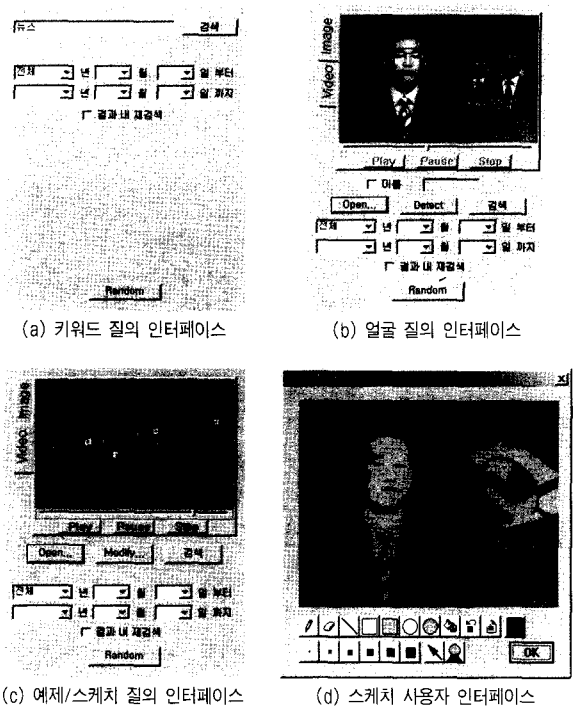


그림 4. 비디오 검색 클라이언트 인터페이스
Fig. 4. Interface of the client for video retrieval



(a) 키워드 질의 인터페이스 (b) 얼굴 질의 인터페이스

(c) 예제/스케치 질의 인터페이스 (d) 스케치 사용자 인터페이스

그림 5. 질의 인터페이스
Fig. 5. Query interfaces: (a) query by keyword, (b) query by face, (c) query by example/sketch, (d) user interface for sketch

의 왼쪽 상단 윈도우는 그림 5와 같은 인터페이스를 갖는 질의 윈도우이다. 사용자는 각 질의 인터페이스를 통해 키워드, 얼굴, 예제 및 스케치 질의를 할 수 있다. 그림 4의 오른쪽은 검색 결과를 보여주며 상단에는 신을, 하단에는 신에 포함된 샷의 키프레임을 보여준다. 사용자는 검색된 샷이나 신을 선택하여 서버로부터 스트리밍되는 비디오를 왼쪽 하단의 재생 윈도우를 통해 그 내용을 확인할 수 있다.

본 시스템에서 사용된 예제/스케치 질의 검색 방법은 키프레임 만을 비교하는 정지 영상 검색 방법이 사용되었고, 객체 혹은 움직임 정보 등은 사용되지 않았다. 비디오에서 객체에 대한 정보는 매우 중요한 의미를 갖지만 객체에 대한 정보를 추출하기 위해서는 우선 객체 검출이 이루어져야 한다. 그러나 아직까지 완벽하게 자동으로 객체를 검출하는 방법은 없으며, 객체를 기반으로 비디오를 검색하기 위해서는 사람이 수동으로 객체를 검출해서 정보를 DB에 저장해야 한다. 이러한 객체의 수동 검출 대신에 본 시스템에서는 자동 얼굴 검출과 인식 기술을 이용하여 등장 인물

검색이 가능하도록 하였다. 따라서 본 시스템은 사람의 개입 없이 자동으로 비디오 DB를 구축할 수 있는 편의성을 제공한다. 또한 키워드의 경우에는 뉴스와 같이 스크립트가 제공되는 경우 쉽게 해결할 수 있으며 음성 인식 및 문자 인식과 같은 기술이 적용되면 키워드 입력도 자동으로 이루어질 수 있다.

Ⅲ. 장면 전환 검출 방법

비디오 인덱싱 모듈에서는 우선 장면 전환 검출이 먼저 이루어지며, 장면 전환에는 급격한 장면 전환과 점진적인 장면 전환이 있다. 본 논문에서는 두 가지 장면 전환을 함께 검출하기 위해 적응적 임계값[11]과 twin comparison 방법[12]을 기반으로 하여 점진적인 장면 전환 구간이 갖는 특징값 분포 특성을 이용한 장면 전환 검출 방법을 제안한다. Twin comparison 방법은 두 개의 임계값을 이용하여 급격한 장면 전환 뿐만 아니라 점진적인 장면 전환도 검출하기 위해 제안된 방법이다. 그러나 이 방법은 임계값의 선택에 따라서 장면 전환 검출 성능이 영향을 받기 때문에 본 논문에서는 임계값을 적응적으로 선택함으로써 이 문제를 해결하도록 한다. 또한 점진적인 장면 전환 구간에서 인접한 프레임 간의 특징값의 비유사도 분포 특성을 이용하여 더 정확한 장면 전환 검출을 한다.

본 논문에서는 연속된 두 프레임 (f_i, f_{i-1}) 간의 비유사도(D)를 계산하기 위해서 HSV 컬러 히스토그램(H)을 특징값으로 사용하였고, 다음과 같은 식을 이용해서 비유사도를 계산하였다.

$$D(f_i, f_{i-1}) = d(i) = \sum_{j}^{\text{histogram bin}} |H_i(j) - H_{i-1}(j)| \qquad (2)$$

1. 적응적 임계값과 twin comparison

장면 전환 검출을 위한 기본적인 방법은 연속한 프레임 간의 특징값의 차이가 미리 설정된 임계값 이상이면 장면 전환이 발생한 것으로 판단하는 것이다. 그러나 임계값에 기반한 장면 전환 검출 방법은 사용된 임계값에 따라 결과가 크게 좌우되는 문제가 있어서, Yusoff는 적응적 임계값을 제안하였다[11]. 적응적 임계값은 현재 프레임의 앞, 뒤로 길이 N인 로컬 윈도우를 정의하고, 윈도우 내에서 연속한 프레임 간 특징값의 비유사도의 평균에 비례하도록 적응적으로 임계값을 결정하는 방법이다. 즉, 식 (3)과 같이 윈도우 내의 비유사도 평균 μ_N에 미리 주어진 임계값 T_p를 곱한 값을 임계값으로 결정하고, 현재 프레임과 이전 프레임의 특징값의 비유사도가 적응적 임계값 m_r 보다 크면 장면 전환이 발생한 것으로 판단한다. 그러나 실제로 장면 전환 검출을 위한 적응적 임계값의 적용은 현재 프레임과 이전 프레임의 특징값의 비유사도를 μ_N으로 나눈값과 주어진 임계값 T_p을 비교하여 임계값보다 크면 장면 전환이 발생한 것으로 판단한다.

$$m_r = T_p \mu_N \qquad (3)$$

점진적인 장면 전환은 급격한 장면 전환에 비해 그림 6과 같이 상대적으로 낮은 특징값이 여러 프레임에 걸쳐 지속적으로 발생하는 특징이 있다. 따라서, 임계값이 너무 높게 설정되면 점진적인 장면 전환을 찾아낼 수 없고, 반대로 임계값이 너무 낮게 설정되면 하나의 점진적 장면 전환에서 너무 많은 장면 전환을 검출하는 문제가 발생한다. 이러한 문제점을 해결하기 위해 twin comparison 방법[12]이 소개되었다. 이 방법은 두 개의 임계값을 설정하여 특징값의 비유사도가 높은 임계값 이상이면 급격한 장면 전환으로 판단하고, 낮은 임계값 이상이면 낮은 임계값 이상의 값을 갖는 구간에서 비유사도를 누적하여 이 누적한 값이

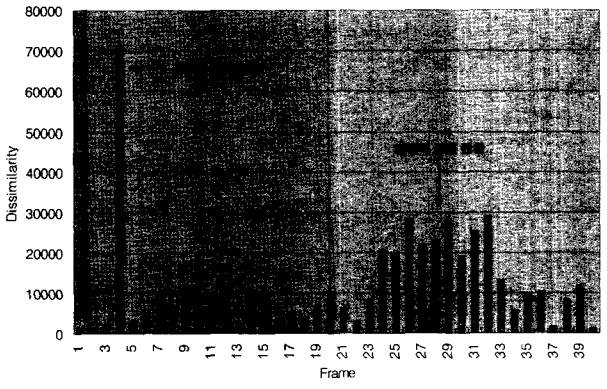


그림 6. 점진적인 장면 전환의 특징
Fig. 6. A characteristic of gradual transition

주어진 임계값보다 클 경우에 점진적인 장면 전환으로 판단한다.

Twin comparison 방법은 점진적인 장면 전환 구간에서 비유사도의 변화에 기반하여 점진적인 장면 전환 검출의 기본적인 방법을 제시한다. 하지만, 임계값이 정확하게 선택되지 못했을 경우에는 장면 전환 구간을 정확하게 찾아내지 못하는 문제가 있다. 또한 큰 물체의 이동이나 장면 전환 구간 내에서 연속한 프레임이 매우 유사하여 비유사도가 일시적으로 임계값 아래로 떨어지는 등의 순간적인 변화에 민감하게 반응한다.

2. ATC (Adaptive Twin Comparison)

본 논문에서는 장면 전환 검출을 위해 적응적 임계값과 twin comparison 방법을 병합하고 점진적인 장면 전환 구간의 특성을 이용하는 ATC 방법을 제안한다. Twin comparison 방법에서 임계값 선택의 문제점을 적응적 임계값을 이용하여 해결하고, 점진적인 장면 전환이 시작되는 부분과 끝 부분에서 연속 프레임간의 특징값의 비유사도가 그림 7 (a)와 같은 모양으로 변화하는 특성을 이용하여 점진적인 장면 전환을 검출하도록 한다.

점진적인 장면 전환의 시작부분에서는 그림 7의 (b)와 같은 특징값 비유사도의 변화가 일어나기 때문에, 현재 프레임을 중심으로 길이 $2N$인 로컬 윈도우에서 오른쪽 부분의 비유사도 평균은 왼쪽 부분의 비유사도 평균보다 크다. 장면 전환의 끝 부분에서는 마찬가지로 그림 7의 (c)와 같은 변화가 발생한다. 따라서, 다음과 같은 조건으로 장면 전환의 구간을 판단할 수 있다.

$$\begin{aligned} &\text{Start of transition if: } \alpha\mu_{left} < \mu_{right} \\ &\text{End of transition if: } \mu_{left} > \alpha\mu_{right} \\ &\text{where, } \mu_{left} = \frac{1}{N}\sum_{i=0}^{N-1} d(i), \quad \mu_{right} = \frac{1}{N}\sum_{i=N+1}^{2N} d(i) \end{aligned} \tag{4}$$

$d(i)$는 식 (2)에서 정의된 i 번째 프레임과 $i-1$번째 프레임의 비유사도를 의미하고, α는 상수로서 본 논문에서는 실험을 통해 1.4를 선택하였다. μ_{left}와 μ_{right}는 각각 윈도우 내에서 현재 프레임을 기준으로 왼쪽 부분과 오른쪽 부분에 대한 비유사도의 평균이다. 식 (4)의 시작점과 끝점 조건을 동시에 만족하는 구간에 대해서 특징값의 비유사도를 누적한 값이 주어진 임계값보다 클 경우 점진적인 장면 전환이 발생하였다고 판단한다. 이때 사용된 임계값은 장면 전환 시작 부분에서 특징값의 평균을 이용하여 적응적으로 선택되었다. 이러한 장면 전환 구간 판단 방법은 윈도우 내의 평균값을 이용하기 때문에 특징값의 일시적인 변화에 덜 민감하게 동작한다.

Ⅳ. QBME (Query By Modified Example)

내용기반 영상 검색에서 사용되는 대표적인 질의 방법으로는 사용자가 원하는 영상과 유사한 영상을 질의 영상으로 입력하는 QBE(Query By Example) 방법과 사용자가 직접 원하는 영상을 스케치하여 질의하는 QBS(Query By Sketch) 방법이 있다[21]. QBE 방법의 경우 질의한 영상에

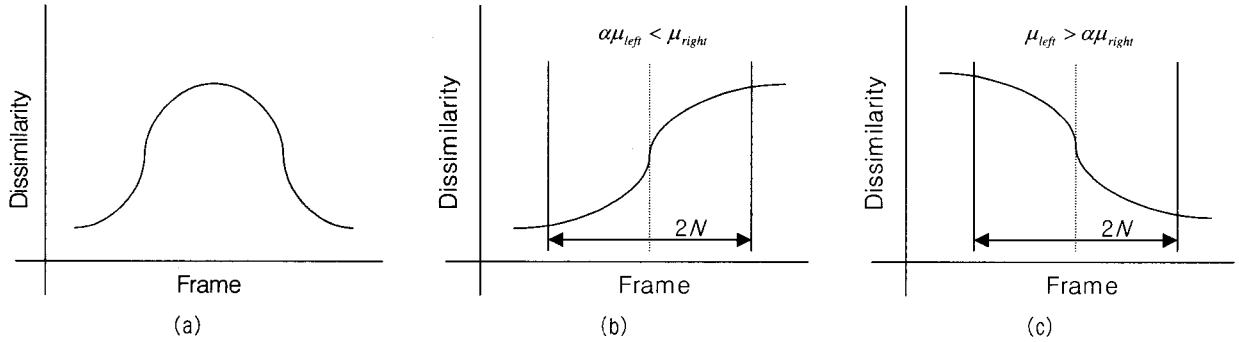


그림 7. (a) 점진적 장면 전환, (b) 점진적 장면 전환 시작 지점, (c) 점진적 장면 전환 끝 지점에서 특징값의 비유사도 변화
Fig. 7. Distributions of dissimilarities in (a) gradual transition, (b) start point and (c) end point of gradual transition

대해 유사한 영상을 찾아주는 방식이므로 사용자는 원하는 영상과 같은, 또는 매우 유사한 영상을 가지고 있어야 한다는 제약이 있다. QBS 방법의 경우, 특히 영상이 복잡할수록 찾고자 하는 영상을 시각화 하는데 어려움이 있고, 사용자의 스케치 능력이 요구되는 문제가 있다. 즉, 다른 영상과 구분되는 특징의 적절한 표현이 어렵고, 색상 및 질감 선택, 전체 영상을 그려야 하는 등의 불편함이 있다.

본 논문에서는 기존의 QBE와 QBS의 문제점을 해결하기 위해 QBME를 제안한다. 제안된 방법은 사용자가 입력한 유사 영상을 기반으로 단순화한 영상을 만들어 밑그림을 제공하고 이를 간단한 수정 과정을 거쳐 원하는 질의 영상을 생성하도록 한다. 이 때 입력 영상 중 찾고자 하는 영상과 비슷한 영역은 색상과 위치 등을 그대로 질의에 사용하고 일부분만 원하는 내용으로 수정하여 사용하기 때문에 스케치에 대한 부담감을 줄여준다. 또한 단순화된 영상을 이용하므로 색상 채우기 등의 수정 작업이 몇 가지의 단순한 조작으로 가능하여 스케치의 편리성을 향상시킬 수 있다.

1. 영상의 단순화

스케치를 위한 밑그림은 입력 영상을 몇 개의 대표색 만으로 표현함으로써 생성된다. 대표색의 추출은 MPEG-7의 대표 컬러 서술자를 이용하였으며, 입력 영상에서 각 픽셀을 원래 색상과 가장 유사한 대표색으로 대체하여 영상을 단순화한다. 이때 비슷한 색상의 영역이 과분할 되는 것을 방지하기 위하여 전처리를 하였다. 그림 8은 입력 영상을 단순화하여 스케치를 위한 밑그림 생성 과정을 보여준다.

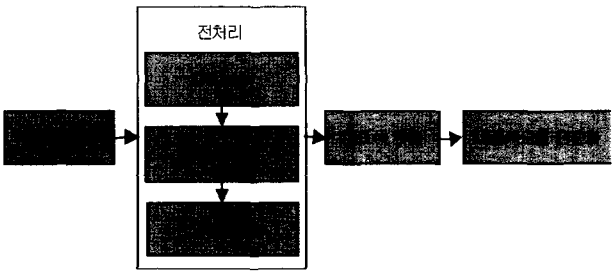
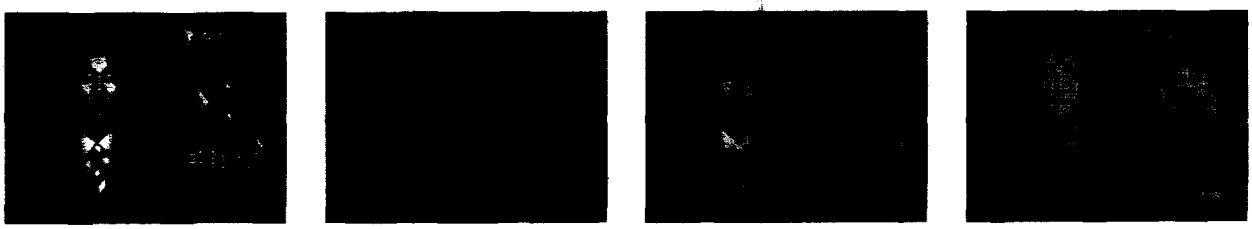


그림 8. 입력 영상의 단순화 과정
Fig. 8. The simplification process of an input image

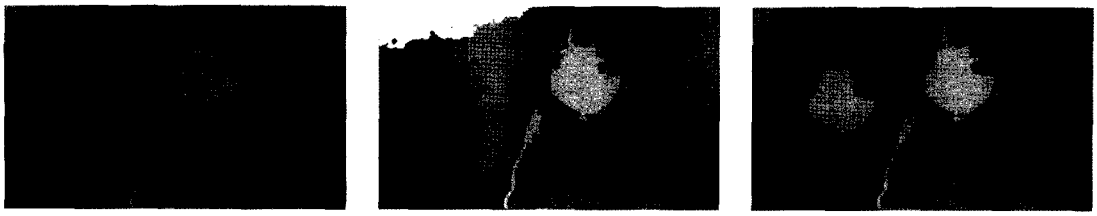
전처리 과정의 핵심으로는 watershed 알고리즘을 사용하였다[22][23]. Watershed 알고리즘은 영상을 각 픽셀의 그레디언트 값에 따라 여러 개의 영역으로 분할한다. 일반적으로 watershed 알고리즘을 적용한 결과 영상은 수많은 영역으로 분할되므로 이에 대한 전후처리가 필요하다. 본 논문에서는 잡음에 의한 영역의 과분할을 방지하기 위해 비등방성 확산(Anisotrofic diffusion)[24]을 전처리로 적용하였고 watershed 알고리즘 적용 후 각 영역의 그레이 값과 크기에 따라 유사한 속성을 가진 인접 영역들간의 영역 병합 과정을 수행하였다[25][26].

위와 같은 전처리 과정을 거친 후의 영상에서 추출된 대표색을 이용하여 영상을 표현한다. 대표색은 MPEG-7의 대표 컬러 서술자를 이용하였으며 이것은 입력된 영상의 색상을 LUV 색상 공간으로 변환한 후 GLA(Generalized Lloyd Algorithm)를 적용하여 최대 8개의 영상의 대표색을 추출한다[14]. 그림 9의 (b)는 원본 영상을 전처리 과정 없이 대표색으로 표현한 것이며 (c)는 전처리 결과, (d)는 전처리 후 대표색으로 표현한 것이다. (d)가 (b)에 비해 불필요하게 분할되는 영역이 줄어들어 사용자가 수정하기 위한 밑그림으로 더 적합한 것을 볼 수 있다.



(a) 원본 영상 (b) 전처리 없이 단순화 (c) 전처리 결과 (d) 전처리 후 단순화

그림 9. 입력 영상의 단순화 결과
Fig. 9. The results of simplification: (a) original image, (b) simplified image without preprocessing, (c) preprocessed image, (d) simplified image after preprocessing



(a) 단순화된 영상 (b) 색상 변경 (c) 객체 복사

그림 10. 단순화된 영상의 수정
Fig. 10. Image modification of a simplified image:(a) simplified image, (b) color modification, (c) objects copy

2. 영상 수정

사용자는 그림 10과 같이 입력 영상으로부터 단순화된 영상을 수정하여 질의를 위한 새로운 영상으로 만들 수 있다. 즉, 단순화된 영상 그림 10 (a)를 밑그림으로 이용하여 색상 변환, 객체의 추가 및 제거 등의 단순한 조작을 통해 (b)나 (c)와 같이 원하는 영상으로 수정한 후 질의한다. 이 때 입력 영상 중 찾고자 하는 영상과 비슷한 영역은 색상과 위치 등을 그대로 질의에 사용하고 일부분만 원하는 내용으로 수정하여 사용하기 때문에 스케치에 대한 부담감을 줄여준다. 또한 단순화된 영상은 이미 색상에 기반해서 각 영역으로 분할되어 있으므로 (c)와 같이 객체 단위의 조작이 가능한 장점이 있다.

V. 실험 결과

1. 장면 전환 검출

장면 전환 검출 성능을 평가하기 위해 영화 예고편과 뮤직 비디오를 실험 영상으로 선택하였다. 이들은 빠른 움직임을 많이 포함하고 다양한 특수효과가 사용되어 장면 전환 검출에 어려움이 많지만, 다수의 점진적인 장면 전환을 포함하고 있기 때문에 실험 영상으로 선택되었다. 사용한 특징량으로는 HSV 컬러 히스토그램을 사용하였으며, 영상의 작은 변화에 덜 민감하도록 히스토그램의 각 성분이 4 bit로 양자화 되었다. 성능 평가를 위해 ATC 방법과 twin comparison 방법[12]을 이용하여 점진적인 장면 전환 검출 성능을 비교하였고, 모두 로컬 윈도우 길이를 41로 고정하였다. 급격한 장면 전환에 대해서는 모두 만족스러운 검출 성능을 보였기 때문에 실험에서는 제외하였다.

표 1은 실험에 사용된 비디오에 대한 각 길이와 사람이 판단한 장면 전환의 개수 등의 정보를 나타내고 있고 그림 11에는 실험 결과가 나타나있다. 실험결과에서 recall과 precision은 아래와 같이 정의된다.

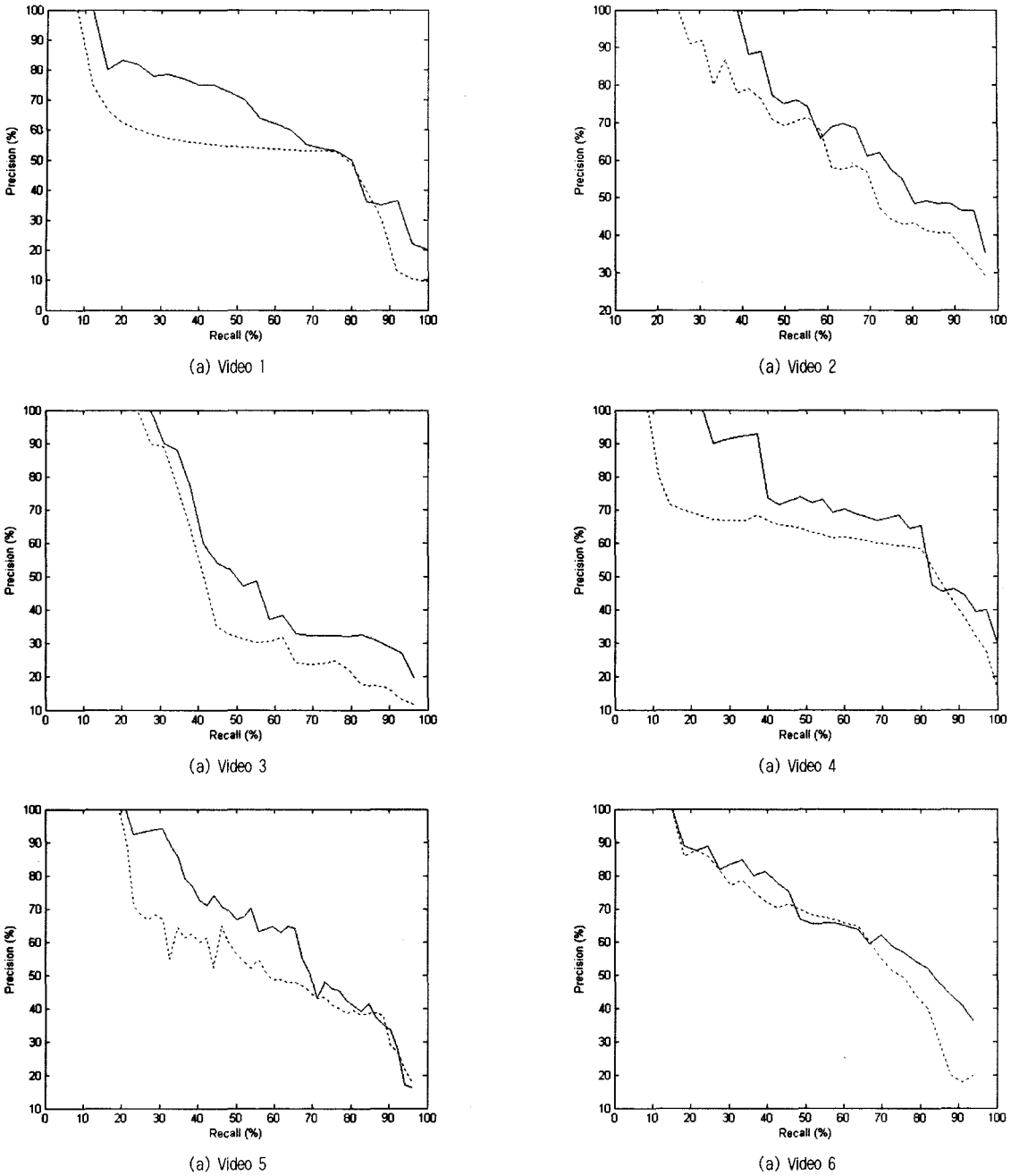
$$recall = \frac{N_c}{N_c + N_m} \times 100\%$$
$$precision = \frac{N_c}{N_c + N_f} \times 100\% \quad (5)$$

N_c는 검출된 장면 전환 중 올바른 장면 전환의 개수이며, N_m은 검출하지 못한 장면 전환의 개수, N_f는 잘못 검출된 장면 전환의 개수를 나타낸다.

표 1. 실험에 사용된 비디오
Table 1. Videos used in the experiments

Video	Time (sec)	No. of abrupt change	No. of gradual transitions	Genre
1	166	52	25	Music video
2	180	43	36	Music video
3	171	13	29	Music video
4	148	34	35	Movie trailer
5	175	31	52	Movie trailer
6	124	58	33	Movie trailer

실험 결과 점진적인 장면 전환 검출에 대해 ATC 방법이 twin comparison 방법 보다 전반적으로 우수한 성능을 보여주었다. 실험 결과에서 전체적으로 낮은 성능을 보이는 것은 비디오 특성상 조명 변화 및 특수 효과 등이 많이 사용되고 이를 점진적인 장면 전환으로 잘못 검출하는 경우가 많이 발생하기 때문이다. 또한 점진적인 장면 전환이 매우 짧은 시간 내에 이루어지는 경우가 많이 발생하여 로컬



(a) Video 1 (a) Video 2

(a) Video 3 (a) Video 4

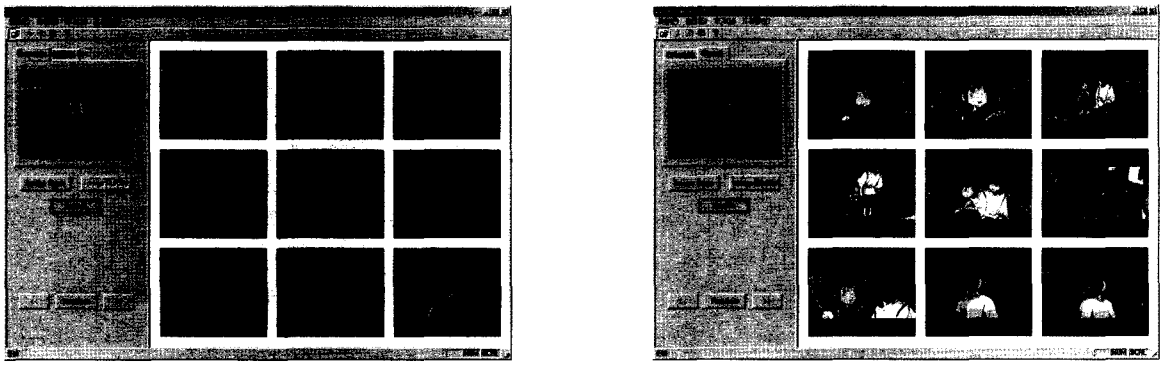
(a) Video 5 (a) Video 6

그림 11. 점진적인 장면 전환 검출 결과 (실선: ATC 방법, 점선: twin comparison 방법)
Fig. 11. The results of cut detection for gradual transitions (solid line: ATC method, dotted line: twin comparison method)

윈도우의 크기를 어떻게 결정하는지에 따라서 그 성능의 변화가 매우 심하였다. 그럼에도 불구하고, 그림 11과 같이 본 논문에서 제안하는 ATC 방법이 twin comparison 방법보다 좋은 검출 결과를 보여주고 있다.

2. QBME

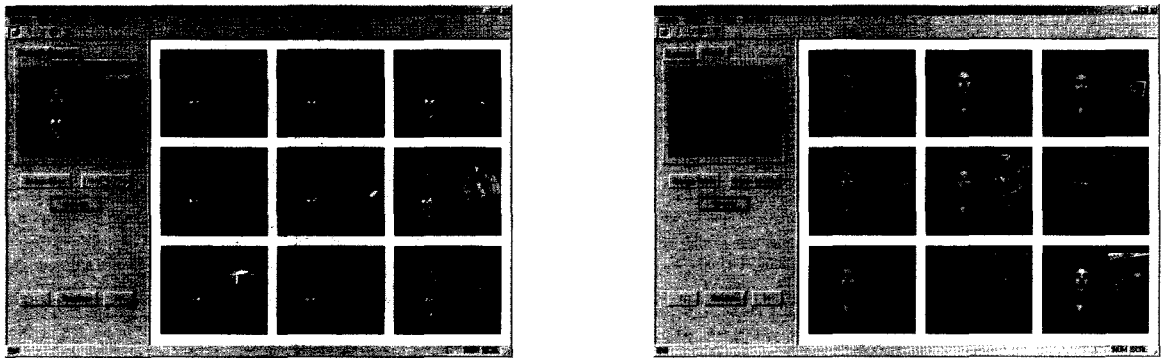
QBME 질의를 위해 MPEG-7의 컬러 서술자인 대표 컬러 서술자와 컬러 레이아웃 서술자를 사용하여 영상의 특징값을 추출하고 이를 영상의 유사도 비교에 사용하였다. 데이터베이스는 자연 영상, 영화, 애니메이션, 뉴스 등의 다양한 분야에서 추출한 3600여장의 영상으로 구성되었으며 그림 12, 13, 14는 검색 결과를 보여준다. 화면의 왼쪽은 사용자가 질



(a) QBE 검색 결과 (b) QBME 검색 결과

그림 12. 배경의 색상 수정 후의 검색 결과

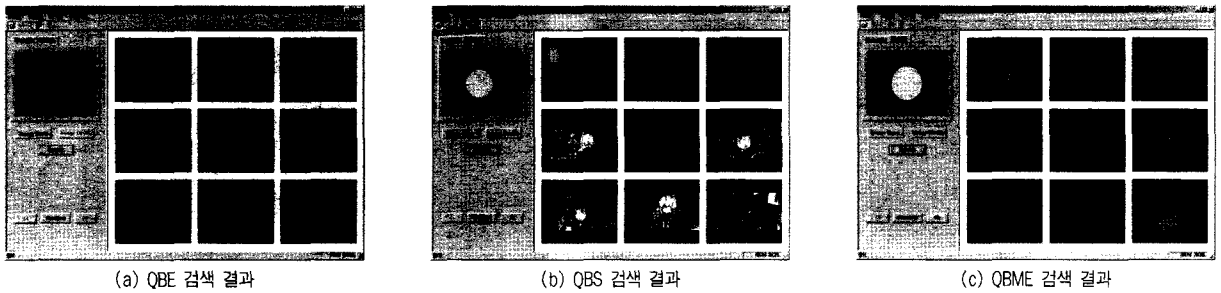
Fig. 12. The result of retrieval after modifying the background color: (a) QBE, (b) QBME



(a) QBE 검색 결과 (b) QBME 검색 결과

그림 13. 객체의 색상 수정 후의 검색 결과

Fig. 13. The result of retrieval after modifying the object color: (a) QBE, (b) QBME



(a) QBE 검색 결과 (b) QBS 검색 결과 (c) QBME 검색 결과

그림 14. 객체 추가 후의 검색 결과

Fig. 14. The result of retrieval after adding an object: (a) QBE, (b) QBS, (c) QBME

의한 영상을 보여주며 오른쪽 9개의 그림은 그 결과를 유사도에 따라 순서대로 보여준다. 그림 12, 13의 (a)는 원본 영상을 질의한 QBE 검색 결과이며 (b)는 원본 영상을 수정한 후 검색한 QBME 검색 결과이다. 그림 14의 (a)는 원본 영상을 질의한 QBE 검색 결과, (b)는 QBS 검색 결과, (c)는 원본 영상을 수정한 후 검색한 QBME 검색 결과이다.

그림 12는 사용자가 가진 영상의 등장 인물이 다른 배경에서 나타난 장면을 검색하고자 할 때의 예이다. 그림 12의 (b)는 입력영상을 단순화 시킨 후 배경색상을 원하는 색상으로 변경하여 검색한 결과를 보여준다. 그림 12의 (b)에서 질의한 스케치 영상은 녹색 배경에 3개의 영역으로 이루어진 인물을 포함하는 간단한 영상이지만, 밑그림 없이 스케치 질의 영상을 만들기 위해서는 배경뿐만 아니라 인물의 각 영역의 모양을 그리고 색칠하기 위한 여러 번의 조작이 필요하다. 반면에 밑그림을 기반으로 스케치를 하면 단순히 배경 색상만을 수정함으로써 원하는 영상을 쉽게 만들 수 있다.

그림 13은 비슷한 배경에서 다른 인물이 등장하는 장면을 검색하고자 할 경우의 예이다. 그림 13의 (b)는 사용자가 찾고자 하는 영상과 비슷한 배경을 가진 영상을 가지고 있을 때 입력 영상을 단순화 한 후 찾고자 하는 인물의 의상 색상으로 변경하여 검색한 결과이다. 이 경우에는 그림 12와는 다르게 배경이 여러 영역으로 이루어져서 사용자가 이를 표현하기는 어렵지만, 주어진 밑그림으로부터 얻은 배경을 그대로 이용하고 인물만을 수정함으로써 쉽게 새로운 스케치 질의 영상을 얻을 수 있다.

그림 14는 비슷한 배경을 가지고 있을 때 원하는 내용을 추가하여 검색한 결과이다. 사용자가 풀밭 영상을 가지고 있고 풀밭에 있는 민들레 영상을 찾고자 한다면 가지고 있는 풀밭 영상을 단순화 한 후 그림 14의 (c)와 같이 노란색의 원을 추가하여 검색할 수 있다. 그림 14의 (b)는 밑그림이 없을 경우 (c)와 같은 결과를 얻고자 스케치를 통하여 질의 영상을 만들어 질의한 결과이다. (b)의 결과와 같이 간단한 스케치만으로 원하는 영상을 검색하는 것이 쉽지 않은 것을 알 수 있다.

실험 결과와 같이 제안하는 QBME 방법은 사용자가 검색을 원하는 영상과 같은 혹은 매우 유사한 영상을 가지고 있어야 하는 QBE의 제약을 보완할 수 있다. 또한 밑그림을 사용하므로 스케치를 통하여 만드는 경우보다 손쉽고 빠르게 질의 영상을 얻을 수 있으며, 단순화된 영상을 밑그림으로 사용하므로 색상 변경 등의 수정 과정이 용이한 장점이 있다.

VI. 결론

본 논문에서는 MPEG-7 국제 표준 규격을 따르면서 키워드, 얼굴, 예제/스케치 등의 다양한 질의 방법을 제공하는 통합형 비디오 검색 시스템을 구현하였고, 비디오 인덱싱에 필요한 강인한 장면 전환 검출 방법인 ATC 방법과 내용기반 검색을 위한 개선된 질의 방법인 QBME를 제안하였다. ATC 방법은 적응적 임계값과 twin comparison 방법을 병합함으로써 급격한 장면 전환 뿐만 아니라 점진적 장면 전환까지 검출할 수 있었다. QBME는 입력 영상을 단순화하여 밑그림을 제공함으로써 사용자 편의성을 제공하고 기존의 스케치 질의 방법의 문제점을 개선하였다. 실험에서 제안된 ATC 방법이 점진적인 장면 전환에 대해 기존의 twin comparison 방법보다 우수함을 보였고, QBME를 이용함으로써 기존의 QBE와 QBS의 단점을 보완할 수 있음을 보였다.

지금까지 다양한 비디오 검색 시스템이 제안되었으나, 이들 시스템은 특정 질의 방법만을 제공하는 경우가 많았다. 그러나 본 논문에서 구현된 검색 시스템은 키워드 및 등장인물, 예제/스케치 질의와 같은 다양한 질의 방법 통해 검색의 편의성을 제공한다. 또한 MPEG-7 표준에 의해 비디오 메타데이터를 표현함으로써 향후 다양한 MPEG-7 어플리케이션이 제안되면 이들과의 데이터 상호 호환을 통해 보다 더 쉽고 빠르게 비디오 DB 구축이 가능하고 다양한 응용을 기대할 수 있다. 그러나 지속적인 데이터의 증가에 따라서 검색 속도가 느려지는 문제점이 발생할 수 있으므로 이를 해결하기 위해서 유사 비디오에 대한 클러스터링과 같은 검색 속도 향상을 위한 향후 연구가 필요하다.

참 고 문 헌

[1] Y. Alp Aslandogan and Clememt T. Yu, "Techniques and Systems for Image and Video Retrieval," IEEE Trans. Knowledge and Data Engineering, Vol. 11, No. 1, pp. 56-63, 1999. 1.

[2] 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.

[3] J. R, Smith and S.-F. Chang, "VisualSEEk: A Fully Automated Content-Based Image Query System," Proc. ACM Multimedia, pp. 87-98, 1996.

[4] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databaes," Proc. SPIE on Storage and Retrieval of Image and Video Databaes II, Vol. 2, Issue 185, pp. 34-47, 1994.

[5] 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.
[6] V. E. Ogle and M. Stonebraker, "Chabot: Retrieval from a Relational Database of Image," Computer, Vol. 28, No. 9, 1995.
[7] S.-F. Chang, J. R. Smith, and J. Meng, "Efficient Techniques for Feature-Based Image/Video Access and Manipulation," Proc. 33rd Ann. Clinic on Library Applications of Data Processing Image Access and Retrieval, 1996.
[8] M. La Cascia and E. Ardizzone, "JACOB: Just A Content-Based Query System for Video Databases," Proc. ICASSP, 1996.
[9] "Overview of the MPEG-7 Standard," ISO/IEC JTC1/SC29/WG11 N4031, 2001. 3.
[10] 이재호, 김형준, 김회율, "MPEG-7 기반 비디오/이미지 검색 시스템 (VIRS)," 정보처리학회논문지, 제10-B권, 제5호, pp. 543-552, 2003. 8.
[11] Y. Yusoff, W. Christmas, and J. Kittler, "Video Shot Cut Detection Using Adaptive Thresholding," British Machine Vision Conference, pp. 362-372, 2000.
[12] S. W. Smoliar and H. J. Zhang, "Content-Based Video Indexing and Retrieval," IEEE Multimedia, Vol. 1, pp. 6272, 1994.
[13] J.-H. Lee, G.-G. Lee, and W.-Y. Kim, "Automatic Video Summarizing Tool using MPEG-7 Descriptors for Personal Video Recorder," IEEE Trans. Consumer Electronics, Vol. 49, No. 3, pp. 742-749, 2003. 8.
[14] "MPEG-7 visual part of experimentation Model Version 10.0," ISO/IEC JTC1/SC29/WG11, N4063, 2001. 3.
[15] 박현선, 김경수, 김회정, 정병희, 하명환, 김회율, "Integer DCT와 SVM을 이용한 실시간 얼굴 검출," 대한전자공학회 하계학술대회, Vol. 26, No. 1, pp. 2112-2115, 2003. 7.
[16] 이혼진, 김형준, 김회정, 하명환, 정병희, 김회율, "DCT/LDA를 이용한 얼굴 인식의 성능 향상," 제16회 신호처리합동학술대회, Vol. 16, No. 1, pp. 854-857, 2003. 9.
[17] "Text of 15938-5 FCD Information Technology Multimedia Content Description Interface Part 5: Multimedia Description Schemes," ISO/IEC JTC1/SC29/WG11, N3966, 2001. 3.
[18] http://www.expway.com
[19] "Information Technology Multimedia Content Description Interface Part 6: Reference Software," ISO/IEC JTC1/SC29/WG11, N4475, 2001. 10.
[20] http://duck.hanyang.ac.kr/browser.html
[21] J. Assfalg, A. Del Bimbo, and P. Pala, "Image Retrieval by Positive and Negative Examples," ICPR, Vol. 4, pp.4267-4270, 2000. 9.
[22] L. Vincent and P. Sollie, "Watershed in digital spaces: An efficient algorithm based on immersion simulations," IEEE Trans. PAMI, Vol. 13, No. 6, pp. 583-598, 1991. 6.
[23] P. De Smet and D. De Vleeschauwer, "Performance and Scalability of a highly optimized rainfalling watershed algorithm," CISST 98, pp. 266-273, 1998. 7.
[24] P. Perona and J. Malik, "Scale Space and Edge Detection Using Anisotropic Diffusion," IEEE Trans. PAMI, Vol. 12, No. 7, pp. 629-639, 1990. 7.
[25] L. Najman and M. Schmitt, "Geodesic Saliency of Watershed Contours and Hierarchical Segmentation," IEEE Trans. PAMI, Vol. 18, No. 12, pp.1163-1173, 1996. 12.
[26] H. Gao, W. Siu, and C. Hou, "Improved Techniques for Automatic Image Segmentation," IEEE Trans. Circuits and Systems for Video Technology, Vol. 11, No. 12, pp. 1273-1280, 2001. 12.

저 자 소 개

김 형 준

- 1999년 : 한양대학교 전자전자통신전파공학과 졸업 (공학사)
- 2001년 : 한양대학교 대학원 전자통신전파공학과 졸업 (공학석사)
- 2003년 ~ 현재 : 한양대학교 대학원 전자통신전파공학과 박사과정
- 주관심분야 : 패턴 인식, 얼굴 인식, 멀티미디어 검색 등

김 회 율

- 1980년 : 한양대학교 전자공학과 졸업 (공학사)
- 1983년 : Pennsylvania State University 전기공학과 졸업 (공학석사)
- 1989년 : Purdue University 전기공학과 졸업 (공학박사)
- 1989년 9월 ~ 1994년 2월 : University of Texas 조교수
- 1994년 ~ 현재 : 한양대학교 전자전기컴퓨터 공학부 정교수
- 주관심분야 : 영상처리, 컴퓨터비전, 패턴 인식, 머신비전, MPEG-7 등