

자동 주석 갱신 및 카테고리라이징 기법을 이용한 의미기반 동영상 검색 시스템

A Semantic-based Video Retrieval System using Design of Automatic Annotation Update and Categorizing

김정재(Jung-Jae Kim)¹⁾ 이창수(Chang-Soo Lee)²⁾

이종희(Jong-Hee Lee)³⁾ 전문석(Moon-Seok jeon)⁴⁾

요 약

비디오 데이터를 효율적으로 처리하기 위해서는 비디오 데이터가 가지고 있는 내용에 대한 정보를 데이터 베이스에 저장하고 사용자들의 다양한 질의를 처리할 수 있는 의미기반 검색 기법이 요구된다. 기존의 내용 기반 비디오 검색 시스템들은 주석기반 검색 또는 특징기반 검색과 같은 단일 방식으로만 검색을 하므로 검색 효율이 낮을 뿐 아니라 완전한 자동 처리가 되지 않아 시스템 관리자나 주석자의 많은 노력을 요구한다. 본 논문에서는 주석기반 검색과 특징기반 검색을 이용하여 대용량의 비디오 데이터에 대한 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 자동화되고 통합된 비디오 의미기반 검색 시스템을 제안한다. 사용자의 기본적인 질의와 질의에 의해 추출된 키 프레임의 이미지를 선택함으로써 에이전트는 추출된 키 프레임의 주석에 대한 의미를 더욱 구체화시킨다. 또한, 사용자에 의해 선택된 키 프레임은 질의 이미지가 되어 제안하는 특징기반 검색기법을 통해 가장 유사한 키 프레임을 검색한다. 따라서 의미기반 검색을 통해 비디오 데이터의 검색의 효율을 높일 수 있도록 시스템을 설계한다.

ABSTRACT

In order to process video data effectively, it is required that the content information of video data is loaded in database and semantic-based retrieval method can be available for various query of users. Currently existent contents-based video retrieval systems search by single method such as annotation-based or feature-based retrieval, and show low search efficiency and requires many efforts of system administrator or annotator form less perfect automatic processing. In this paper, we propose semantic-based video retrieval system which support semantic retrieval of various users by feature-based retrieval and annotation-based retrieval of massive video data. By user's fundamental query and selection of image for key frame that extracted from query, the agent gives the detail shape for annotation of extracted key frame. Also, key frame selected by user become query image and searches the most similar key frame through feature based retrieval method that propose. Therefore, we design the system that can heighten retrieval efficiency of video data through semantic-based retrieval.

키워드 : 동영상 검색(Video Retrieval), 인덱싱 에이전트(Indexing Agent), 의미기반 검색 (Semantic-based Retrieval)

-
- 1) 정회원 : 숭실대학교 컴퓨터학과 박사과정
 - 2) 정회원 : 숭실대학교 컴퓨터학과 박사과정
 - 3) 정회원 : 숭실대학교 컴퓨터학과 박사과정
 - 4) 정회원 : 숭실대학교 부교수

논문접수 : 2004. 2. 3.

심사완료 : 2004. 2. 10.

1. 서 론

최근 인터넷의 발전과 더불어 영상 압축 기술과 멀티미디어 콘텐츠 서비스 기술이 빠르게 발전하고 있다. 이에 따라, 비디오 데이터의 정보에 대한 사용자의 다양한 요구가 증가하고 있고 이러한 다양한 사용자의 요구를 충족시키기 위해서는 방대한 비디오 데이터의 효율적인 관리가 필요하다[1]. 비디오 데이터의 효율적인 관리를 위해서는 대용량의 비디오 데이터의 정보를 체계적으로 분류하고 통합하는 기술이 필요하다. 또한, 비디오 데이터에 대한 사용자의 다양한 요구에 따라 원하는 결과를 서비스해 주기 위해서는 비디오 데이터를 효율적으로 검색하고 저장할 수 있어야 된다.

비디오 데이터는 일반적인 텍스트 데이터와 달리 비디오 내에 데이터의 정보가 문자화되어 있지 않아 비디오 데이터에 다양한 정보를 부여하는 것이 쉽지 않다. 따라서 비디오 내에 있는 프레임과 그 프레임들의 키 프레임 및 주석과 같은 부가적인 정보에 의한 내용 기반 검색이 필요하다. 이러한 비디오 데이터의 내용기반 검색을 위해서는 비디오 데이터의 정보를 구조적으로 체계화하고 구체화하여 사용자의 의미 기반 검색이 가능하도록 하는 것이 매우 중요하다 [2].

현재 비디오 데이터에 대한 내용기반 검색에 대한 연구는 크게 두 가지로 분류할 수 있다. 첫 째, 키 프레임에서의 특징을 추출하여 유사성을 이용한 특징기반 검색(feature-based retrieval)이 있으며, 둘째, 키 프레임에 대해 사용자의 주석을 입력하여 저장한 후 이러한 사용자의 주석을 비교 검색하는 주석기반 검색(annotation-based retrieval)으로 크게 두 가지로 분류할 수 있다. 하지만, 이 두 가지 비디오 데이터에 대한 내용기반 검색은 모두 단점을 가지고 있다.

우선, 특징기반 검색은 비디오 데이터 자체로부터 그 데이터가 가지고 있는 Color, Texture,

Region 정보, Spatial Color Distribution 등의 저차원(low-level) 특징 정보들을 추출하여 검색하는 방법이다[3]. 따라서 특징기반 검색 방법은 비디오 자체의 시각적 특징을 추출하여 유사성을 계산하여 비교 검색하는 방식이므로 시각적인 특징을 추출하는 것이 매우 중요하다. 하지만, 무수히 많은 비디오의 특징 정보를 정확하게 추출하기 힘들 뿐 아니라 추출한 특징 정보를 방대한 비디오 데이터에 매칭시켜 검색하기가 쉽지 않다.

그리고 비디오 데이터 검색 방법인 주석기반 검색은 각각의 비디오 데이터에 자동 인식이 어려운 의미 정보를 사용자가 직접 문자로 주석을 부여하여 저장한 후 검색 시 미리 부여된 주석을 이용하여 검색하는 방법이다[4]. 이 방법은 사용자가 직접 비디오를 보면서 내용을 주석으로 처리할 수 있기 때문에 비디오의 내용을 정확하게 표현하고 검색할 수 있다는 장점이 있다. 그러나 각각의 비디오에 대해 문자로 사용자가 일일이 주석을 부여하여야 하므로 많은 시간과 노력을 필요로 하며, 불필요한 주석의 양이 방대하게 증가할 수 있다는 단점이 있다. 또한, 시스템에서의 주석을 입력하는 주석자 한 사람의 주관적인 의미부여로 인해 많은 사용자의 다양한 검색어로는 정확한 검색 결과를 얻을 수 없다.

최근 이 두 가지 비디오 데이터 검색 방법을 통합하여 비디오 데이터에 대한 검색 효율을 높이고자하는 연구가 진행 중에 있으나 개별 사용자들의 서로 다른 주관적인 의미부여로 인해 높은 검색 효율을 기대하기 힘들다. 따라서, 본 논문에서는 사용자의 질의와 결과를 학습하여 비디오 데이터의 메타데이터를 지속적으로 자동 갱신하는 인덱싱 에이전트를 이용한 내용기반 비디오 검색 시스템을 제안한다.

에이전트는 1차적으로 다양한 사용자의 검색 질의와 그 결과값을 지속적으로 학습하고 학습된 내용을 자동으로 주석 처리하도록 하는 주석기반 검색을 수행하고 2차적으로 비디오 데이터에서 키 프레임을 검색하여 검색 결과로 나타난

키 프레임을 이미지화한 후 제안하는 멀티분할 색상비교 기법을 이용하여 데이터베이스 안에 저장되어 있는 모든 비디오 데이터의 키 프레임들과 유사성을 비교하여 가장 유사한 프레임이 포함되어 있는 장면을 재생해 준다. 따라서, 비디오 데이터의 주석기반 검색기법과 특징기반 검색기법을 통합한 인덱싱 에이전트를 통해 자동화하여 비디오 내용 검색 시스템의 검색 결과에 대한 정확도를 높이는 것이 연구의 최종 목적이며, 본 논문에서는 시스템을 제안하고 설계한다.

2. 관련연구

2.1 특징기반 비디오 검색 기법

IBM Almaden 연구소에서 개발한 QBIC(Query by Image and Video Content)[5]는 예제 이미지를 통한 유사도 질의를 하며 사용자 스케치에 의한 질의와 색상 및 질감 패턴에 대해 질의를 지원하는 시스템이다. QBIC은 이미지뿐만 아니라 비디오 데이터도 지원하므로 샷의 검출, 샷에 대한 키 프레임 생성과 객체의 움직임에 이용하여 데이터베이스를 구성한다.

콜롬비아 대학에서 개발된 VisualSEEK[6]는 색상과 공간 질의를 지원하는 이미지 데이터베이스 시스템으로서 이미지의 구분은 색상과 히스토그램과 같은 특성에 의해 이루어지며, 이미지 비교를 위해 이미지의 영역과 색상, 크기 및 공간적 위치들을 추가적으로 사용하는 시스템이다.

대만 청화대학에서 개발한 Venus[7]는 특징기반 검색을 위해 개발된 비디오 데이터베이스 시스템이다. 이미지 처리 기술을 사용하며, 비디오의 각 프레임에 나타나는 객체들의 시간관계, 공간관계를 메타데이터로 구축하였다. Venus는 특징기반 질의 언어인 CVQL(Content-based Video Query Language)을 제공한다.

2.2 주석기반 비디오 검색기법

미국 메릴랜드 대학에서 개발한 AVIS(Advanced Video Information System)[8]는 비디오 내용에 나타나는 객체, 사건 행위 유형에 대한 메타데이터를 정의하고, 이들을 비디오 세그먼트와 연계시킴으로써 효율적인 검색 방법을 제안했다.

노르웨이 공대에서 개발한 VideoSTAR(Video Storage and Retrieval)[9]은 비디오와 메타데이터의 공유와 재사용을 지원하기 위해 관계형 데이터베이스 모델을 사용한 데이터베이스 시스템이다. 메타데이터의 구성은 분할 기법을 통해 비디오데이터를 구조화하고, 구조화된 비디오 데이터에 대해 인물, 위치, 사건 등의 속성으로 메타데이터를 구성하고, 이를 다시 기본, 1차, 2차 컨텍스트로 구분하여 메타 데이터의 재사용과 공유를 용이하게 하며, 고정된 속성을 통해 사용자가 쉽게 질의를 구성할 수 있다.

일본 고베 대학에서 개발한 OVID(Object-oriented Video Information Database)[10]는 비디오 데이터를 모델링하고 검색하기 위해 객체지향 접근방법을 사용한 객체지향 비디오 데이터베이스 시스템이다. OVID는 의미있는 장면을 비디오 객체라는 개념을 도입하고 각 비디오 객체마다 포함된 내용을 설명하는 속성과 속성값을 부여하여 메타데이터를 구성하였으며 비디오 객체를 검색하기 위해 Video SQL이라는 사용자 인터페이스를 제공한다.

2.3 기존 연구의 문제점

특징기반 검색 기법의 경우 우선 키 프레임 추출을 위한 프레임, 샷 그리고 장면의 추출을 위해 비디오 프레임을 분석하여 그 경계를 검출하는 것이 중요하다. 하지만 카메라 기법이나 장면의 변환이 거의 없는 경우 샷의 경계를 정확하게 검출하는 것은 매우 어려운 작업이며, 단일 스토리를 구성하는 장면의 경계 검출은 더욱 힘들다. 기존의 많은 연구에서 비디오 프레임에 대한 색상 히스토그램과 대표 색상을 기반으로 장면과 샷의 경계 검출이 진행되어 왔지만

사용자의 의미기반 검색을 완전히 지원하지는 못한다. 따라서, 비디오 데이터의 특징 정보 분석은 여전히 어렵고 발전시켜야 할 과제로 남아 있다.

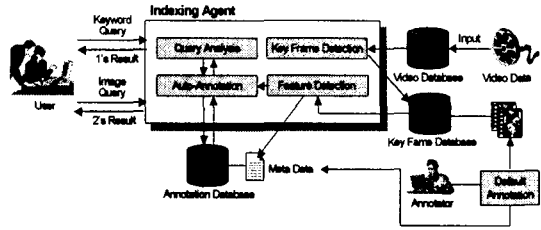
주석기반 검색 기법의 경우 시스템을 구축할 때 주석자의 사물, 위치, 사건 등 단순한 속성 정의와 주관적인 의미 부여로 인해 많은 사용자들의 다양한 검색에 대해 정확한 결과를 도출해 내기 힘들다. 더욱이 방대한 비디오 데이터에서 각 장면에 대한 상세한 주석입력은 주석자의 많은 수고와 시간을 필요로 하므로 방대한 비디오 데이터들을 검색할 수 있는 시스템을 구축하는데 많은 어려움이 따른다.

따라서, 비디오 데이터의 특징기반 검색과 주석기반 검색을 병행하여야 기존의 단일 기법을 사용한 기법보다 검색 효율을 높일 수 있으며, 더욱 체계적인 메타데이터의 생성과 정확한 검색 결과를 위해서는 에이전트에 의한 자동화된 메타데이터 생성과 검색 질의와 결과를 학습하여 주석정보로 이용하는 것이 필수적이라 할 수 있다.

3. 통합된 인덱싱 에이전트의 구조

3.1 전체 시스템 구조

새로운 비디오 데이터가 입력되면 인덱싱 에이전트는 장면전환 검출(Scene Change Detection)을 이용하여 비디오 데이터의 각 장면을 검출하여 분류하고 분류된 장면에서 키 프레임을 추출하여 프레임 전체를 색상 히스토그램기법을 이용하여 특징을 분석한다. (그림 1)은 제안하는 시스템의 전체 구조를 나타낸다.

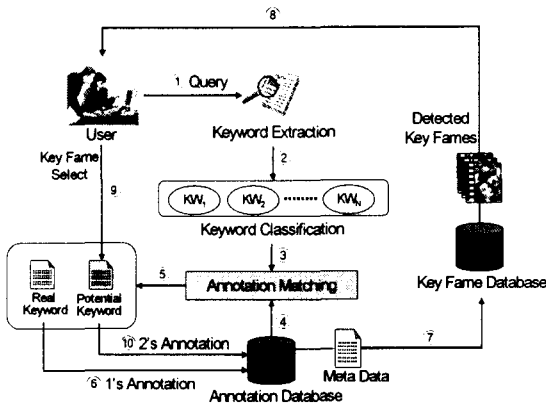


(그림 1) 시스템 구조

추출된 키 프레임은 주석자에 의해 가장 기본적인 하나의 단어로 주석이 입력되고 각 키 프레임과 기본적인 주석이 메타데이터로 저장된다. 사용자의 비디오 의미기반 검색 질의가 전달되면 사용자 질의에서 필요한 키워드를 추출하여 주석 데이터베이스에 있는 프레임 정보와 비교한 후, 유사한 후보 키 프레임 리스트를 생성하여 사용자에게 1차적으로 디스플레이한다. 디스플레이된 후보 키 프레임 리스트 중 사용자가 선택한 특정 키 프레임은 멀티분할 히스토그램 비교법을 이용하여 특징기반 검색을 실시한 후 가장 유사한 키 프레임을 검출하여 사용자에게 2차적으로 디스플레이한다.

3.2 자동 주석처리

에이전트는 사용자로부터 비디오 내용 검색에 대한 질의를 받으면 질의어 중 키워드를 추출하여 메타데이터의 기본적인 주석 정보와 매칭시켜 동일한 키워드를 주석 정보로 가지고 있는 키 프레임을 검출하여 사용자에게 보내준다. 에이전트에 의한 자동 주석처리의 구조는 (그림 2)와 같다.



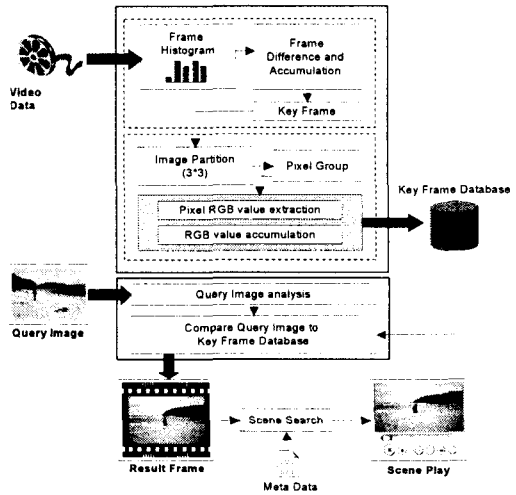
(그림 2) 에이전트의 자동 주석처리

(그림 2)에서 보이는 바와 같이 우선 사용자의 질의가 입력되면 사용자 질의어를 분석하여 키워드를 추출한다. 추출된 사용자 키워드는 주석 데이터베이스 안에 있는 메타데이터의 주석 정보와 매칭시킨다. 매칭시킨 결과 정확히 일치하는 키워드를 주석 정보로 가지고 있는 키 프레임들을 데이터베이스에서 검출하여 사용자에게 전달한다. 또한, 사용자의 질의어 중 주석 정보와 정확히 매칭되지 않는 키워드들은 잠재적 키워드로 정의한다.

사용자는 에이전트가 보내준 사용자 질의에 의한 결과인 검출된 후보 키 프레임들을 이미지화된 예제 리스트 형태로 받게 되고 유사성의 내림차순으로 디스플레이된다. 키 프레임 리스트들 중에 사용자가 원하는 키 프레임을 선택하게 되면 선택된 키 프레임은 다시 2차적인 질의 이미지가 되고 해당 키 프레임의 주석 정보에 잠재적 키워드를 포함시켜 키 프레임에 대한 의미를 구체화시킨다. 또한, 사용자의 질의에 의한 자동 주석처리가 이루어질 때 마다 개별 사용자들의 질의에서 추출된 키워드와 선택된 예제 이미지의 키워드를 매칭시켜 정확히 일치하는 키워드는 유사 가중치를 높여주고 일치하지 않는 키워드는 유사 가중치를 낮추어 줌으로써 해당 키 프레임에 대한 주석의 의미를 더욱 구체화시킨다.

3.3 키 프레임 추출 및 특징 검출

에이전트에 의해 사용자에게 제시된 키 프레임 리스트 중에서 사용자로부터 선택된 키 프레임은 특징기반 검색을 위한 질의 이미지가 된다. 질의 이미지는 제안하는 멀티분할 히스토그램 비교법에 의해 키 프레임 데이터베이스에 있는 모든 비디오 데이터의 키 프레임들과 비교하여 유사도를 계산한 후 유사성이 높은 순으로 사용자에게 키 프레임을 디스플레이한다. 제안하는 프레임의 특징기반 검색에 대한 구조는 (그림 3)과 같다.



(그림 3) 키 프레임 추출 및 특징 검출

비디오 데이터에 대한 특징기반 검색은 우선 분석하고자 하는 비디오 데이터로부터 컬러 히스토그램 방법을 이용하여 장면 전환을 검출하여 각 장면(Scene)을 추출한 후 각 장면의 키 프레임을 추출한다. 각 장면에서 추출한 키 프레임 크기를 3*3으로 영역을 분할한 후 9개의 픽셀 그룹으로 분류하고 전체 픽셀과 분할되어진 영역에 대한 RGB값의 평균을 구하여 키 프레임 데이터베이스에 저장한다. 3*3 영역으로 이미지를 분할한 후 그룹화하여 인덱싱하는 이유는 시간대비 유사도 계산을 했을 때 멀티분할 중 가장 높은 효율을 보였기 때문이다. 주석기반 검색 결과로부터 선택되어진 사용자의

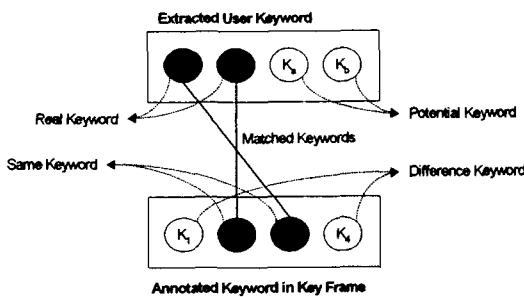
질의 이미지는 질의 분석기를 통해 키 프레임과 동일하게 3*3 이미지 분할과 RGB 평균값을 계산한다. 질의 이미지의 픽셀 그룹에 대하여 모든 계산이 끝나면 키 프레임 데이터베이스에서 가장 유사한 값을 가지는 항목을 선택하게 되고 최종적으로 비디오 데이터의 가장 유사한 키 프레임으로 이동한 후 사용자의 질의 장면을 플래이하게 된다.

4. 인덱싱 기법

4.1 자동 주석처리 기법

단어 한개 이상으로 구성된 사용자 질의가 입력되면 키워드가 추출되고, 추출된 사용자 키워드에 의해 사용자 키워드가 포함된 키 프레임들을 검색한다. 사용자 키워드는 실 키워드와 잠재적 키워드로 분류된 후, 키 프레임의 주석정보와 비교하여 주석 정보에 있는 키워드에서 정확히 매칭된 키워드는 동일 키워드로 정확히 매칭되지 않은 키워드는 비동일 키워드로 분류한다.

(그림 4)는 자동 주석처리를 하기 위한 전처리인 키워드 분류에 대해 나타내고 있다.



(그림 4) 키워드 분류

에이전트는 동일 키워드를 포함하고 있는 키 프레임들을 추출하여 사용자에게 키 프레임 리스트를 디스플레이하고 키 프레임 리스트 중 사용자가 특정 키 프레임을 선택하게 되면 특정 키 프레임이 가지고 있는 각 키워드들에 대한 의미 가중치 계산을 하게 된다.

우선, 키 프레임에서의 주석 키워드가 동일 키워드(same keyword)일 경우 새로운 의미 가중치는 다음과 같이 계산한다.

$$W_{keyword_new} = W_{keyword_old} + \frac{1}{N_{Kframe_SK}} \quad (식 1)$$

$W_{keyword_new}$ 은 주석 키워드에 대한 새로운 의미 가중치이고, $W_{keyword_old}$ 은 주석 키워드에 대한 이전 의미 가중치이며, N_{Kframe_SK} 은 동일한 키워드가 포함되어있는 키 프레임의 개수이다.

또한, 키 프레임에서의 주석 키워드가 비동일 키워드(difference keyword)일 경우 새로운 의미 가중치는 다음과 같이 계산한다.

$$W_{keyword_new} = W_{keyword_old} - \frac{1}{N_{Kframe_SK}} \quad (식 2)$$

사용자들이 질의를 할 때 마다 키 프레임 주석정보인 키워드가 지속적으로 갱신된다. 동일 키워드를 갖는 키 프레임이 계속해서 선택될 경우 의미 가중치가 점차적으로 증가하여 해당 키워드에 의한 의미가 더욱 명확해지는 결과를 도출할 수 있으며, 비동일 키워드인 경우 가중치가 점차적으로 감소하여 의미가 불명확한 키워드로 인식하게 된다. 의미 가중치의 초기값은 0으로 하여 의미 가중치가 -1의 값을 갖게 되는 키워드는 의미 없는 키워드로 에이전트가 판단하여 해당 키 프레임에 대한 주석 정보에서 삭제한다. 따라서 사용자의 잘못 입력한 키워드나 의미 부여가 잘못된 키워드에 대해 해당 키워드를 삭제시킬 수 있는 결과를 도출할 수 있다.

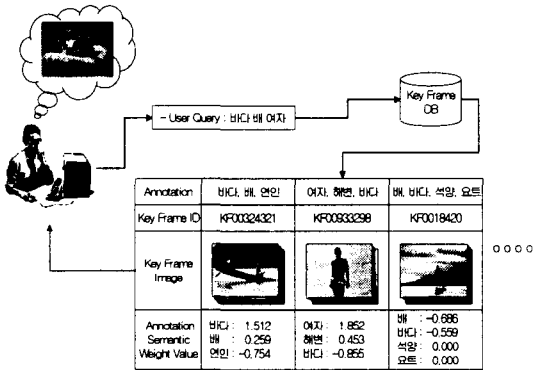
4.2 자동 주석처리 시나리오

사용자들이 질의를 할 때마다 키 프레임 주석 정보인 키워드가 지속적으로 갱신된다. 동일 키워드를 갖는 키 프레임이 계속해서 선택될 경우 의미 가중치가 점차적으로 증가하여 해당 키워드에 의해 의미가 더욱 명확해지는 결과를 도출할 수 있으며, 상이한 키워드인 경우 가중치가 점차적으로 감소하여 의미가 불명확한 키워드로

인식하게 된다.

의미 가중치의 초기 기본값은 0으로 하며, 의미 가중치가 -1 이하의 값을 갖게 되는 키워드는 의미 없는 키워드로 에이전트가 판단하여 해당 키 프레임에 대한 주석 정보에서 삭제된다. 따라서 사용자의 잘못 입력한 키워드나 의미가 잘못된 키워드에 대해 해당 키워드를 삭제시킬 수 있는 결과를 도출할 수 있다.

예를 들어, 사용자가 바다에 여자가 배를 타고 있는 장면이 포함되어있는 영화를 찾기 위해 (그림 5)와 같이 '바다', '배', '여자'라는 키워드로 질의를 했다고 가정해보자.



(그림 5) 키 프레임 주석 및 의미 가중치

인덱싱 에이전트는 데이터베이스 안에 저장되어있는 키 프레임 메타데이터에서 3가지 키워드와 가장 유사한 키 프레임들을 찾는다.

<표 1> 키 프레임 선택 후 키워드 분류

선택된 키 프레임	키워드 분류	키워드
 KF00324321	실 키워드 (Real Keyword)	바다, 배, 여자
	동일 키워드 (Same Keyword)	바다, 배
	상이한 키워드 (Difference Keyword)	연인
	잠재적 키워드 (Potential Keyword)	여자

각 키 프레임들은 기본 주석 이외의 인덱싱 에이전트에 의해 자동으로 갱신된 1개 이상의

주석을 가지고 있다. 예를 들어, 사용자가 '바다', '배', '여자'라는 키워드로 질의한 결과 (그림 5)와 같이 여러 개의 키 프레임들이 검색되었고 그 중 사용자가 키 프레임 ID가 'KF00324321'인 키 프레임 이미지를 선택했다면, <표 1>과 같은 키워드 분류가 시작된다. 키워드 분류가 이루어지면 인덱싱 에이전트에 의해 의미 가중치 자동 갱신을 위한 동일 키워드와 상이한 키워드의 가중치 계산이 (식 1)과 (식 2)에 의해 <표 2>와 같이 이루어진다.

<표 2> 자동 주석 갱신 결과

선택된 키 프레임	주석	주석을 포함한 키워드 수	의미 가중치	
			이전 가중치	새 가중치
 KF00324321	바다	23	1.512	1.555
	배	56	0.259	0.276
	연인	89	-0.754	-0.765
	여자			0

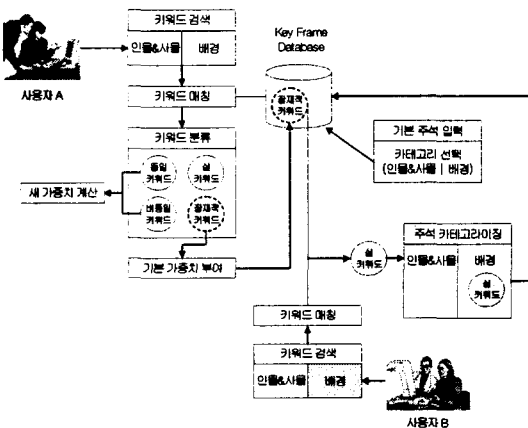
<표 2>에 나타나듯이 실 키워드인 '바다', '배', '여자' 중에 '바다'와 '배'의 의미 가중치는 제안하는 동일 키워드에 대한 의미 가중치 계산에 의해 가중치 값이 증가했으며, '연인'은 상이한 키워드에 대한 의미 가중치 계산에 의해 가중치 값이 감소한 것을 알 수 있다. 또한, 잠재적 키워드인 '여자'는 해당 키 프레임의 실 키워드로 주석이 등록되면서 기본 가중치 값인 0으로 세팅된다. 이러한 방법으로 사용자로부터 키워드 질의에 의해 검색되어지는 키 프레임들의 의미 가중치 값이 지속적으로 자동 갱신되어 더욱 객관적이고 의미적인 주석을 키 프레임들이 갖게 된다.

앞에서 기술한 1차적인 주석기반 검색으로 인덱싱 에이전트에 의해 검색되어진 키 프레임을 다시 멀티분할 색상 히스토그램 기법을 이용하여 2차적인 특징기반 검색을 수행하게 된다.

4.3 주석 카테고리라이징

사용자의 키워드 질의에 있어서 정확하고 상세한 검색을 위해 검색 방식을 두 가지로 분류

한다. 인물 및 사물 중심의 검색 방식과 배경 중심의 검색 방식을 제안한다. 사건 검색 방식을 배제시킨 이유는 주석자의 주관적인 주석이 입력될 가능성이 많고, 사용자의 키워드 검색과정에서 사용자가 입력한 키워드가 데이터베이스의 주석과 일치하지 않을 확률이 매우 높으며, 사건에 대한 단어들의 시소러스를 만들기도 매우 힘들기 때문이다. 따라서, 본 논문에서는 주석에 대한 카테고리를 인물 및 사건과 배경 두 가지로 분류하여 저장하고 검색하도록 제안한다. (그림 6)은 주석에 대한 카테고리라이징에 대한 구조이다.



(그림 6) 주석 카테고리라이징

우선, 키 프레임에 대한 주석자의 기본 주석이 입력될 때 해당 키 프레임의 주석 카테고리를 선택하여 저장하게 된다. 즉, 인물 및 사물에 대한 주석인지 배경에 대한 주석인지를 선택하여 저장한다.

사용자 A가 카테고리별 키워드 검색을 하게 되면 데이터베이스에 카테고리별로 분류된 주석과 키워드 매칭을 통해 키워드를 분류하게 된다. 동일 키워드와 상이한 키워드는 새로운 의미가 중치를 계산하게 되고, 잠재적 키워드는 기본가중치를 부여받고 데이터베이스에 저장된다. 처음 입력된 잠재적 키워드는 일단 카테고리로 분류하지 않고 데이터베이스에 저장한다. 이후

에 또 다른 사용자 B가 키워드 검색을 할 때 사용자 B가 카테고리별로 입력한 키워드 중에 사용자 A에 의해 등록된 잠재적 키워드와 정확히 매칭되면 실 키워드로 등록되며, 주석 카테고리라이징에 의해 사용자 B가 입력한 카테고리로 그 실 키워드가 카테고리를 갖게 된다. (그림 6)에서는 사용자 A의 잠재적 키워드가 사용자 B에 의해 배경 카테고리를 갖는 실 키워드로 변경된 과정을 나타내고 있다.

4.4 키 프레임 특징처리 기법

본 논문에서는 키 프레임 검출을 위한 전처리인 장면 분할을 위해 칼라 히스토그램을 사용하였다. 기존의 장면 분할 기법 중에서 히스토그램을 사용한 방법은 비교적 정확하게 장면 전환을 검출한다. 사용자로부터 입력된 질의 이미지와 비디오 데이터에서 추출된 키 프레임의 이미지에 대한 색상 유사도를 빠르고 정확하게 계산하기 위하여 질의 이미지와 키 프레임의 이미지를 분할한다.

질의 이미지와 키 프레임의 이미지에 대한 분할과 유사도 비교를 위한 색상값의 평균값 계산은 서로 동일하므로 본 논문에서는 키 프레임 이미지에 대한 이미지 분할과 색상 평균값 계산만을 언급한다. 키 프레임으로 추출된 n×m 크기의 이미지는 픽셀을 3×3 영역으로 분할한다. 따라서 분할된 하나의 이미지는 9개의 픽셀그룹을 가지게 된다.

키 프레임의 전체 이미지에 대한 R, G, B의 평균을 구하는 식은 다음과 같다.

$$A_{t,image.red} = \frac{\sum_{i=0}^w \sum_{j=0}^h R_{i,j}}{C_t} \quad (식 3)$$

$R_{i,j}$ 은 전체 이미지에서 i행 j열에 있는 Red 값이며, C_t 는 전체 이미지에서 0이 아닌 Red 값을 가지는 픽셀의 개수이다. (식 3)는 키 프레임의 전체 이미지에 대한 Red 값의 평균을 구하는 식이며, Green 값과 Blue 값의 평균도 같은 방법으로 계산한다. 또한, 9개로 분할된 영역에

대한 R, G, B 값의 평균도 동일한 방법으로 구한다.

입력되어진 질의 이미지를 데이터베이스에 저장되어 있는 키 프레임과 비교하여 가장 유사한 색상정보를 가지고 있는 프레임을 찾기 위하여 (식 4)를 이용하여 색상의 차이값이 임계값 미만인 1차 유사 키 프레임 후보군을 추출한다. 질의 이미지와 키 프레임의 전체 이미지에 대한 Red의 색상 차를 다음과 같이 구한다.

$$D_{t, red} = |A_{t, image, red} - A_{t, frame, red}| \quad (\text{단, } D_{red} < \alpha) \quad (\text{식 4})$$

$A_{t, image, red}$ 는 질의 이미지의 전체에 대한 Red의 평균값이고, $A_{t, frame, red}$ 는 키 프레임의 전체 이미지에 대한 Red의 평균값이며, α 는 임계값이다. 임계값 1차 후보 프레임의 수를 결정하는 값이므로 유동적으로 할 수 있으나 1~255중에 5로 입력했을 때 유사한 키 프레임들을 찾는 것으로 나타났다. Green과 Blue의 색상차도 (식 4)와 동일한 방법으로 구한다.

1차 유사 키 프레임 후보군으로 선택된 이미지는 다음 식을 이용하여 입력된 이미지와 유사도 계산을 하게 된다.

$$S_{total, Q, image, K, frame} = |D_{t, red} + D_{t, green} + D_{t, blue}| \quad (\text{식 5})$$

따라서, 각 R, G, B의 차의 절대값은 임계값 α 보다 작은 값을 1차 후보 프레임으로 선정되고 (식 5)에 의해서 색상 유사도의 값이 작을수록 입력되는 이미지와 유사한 이미지로 계산하여 정렬한다.

1차 후보군으로 선택된 키 프레임들은 색상의 분포가 비슷하지만 입력되어진 질의 이미지와 서로 전혀 유사하지 않은 이미지 일 수 있다. 따라서 1차 후보군에 대하여 (식 7)과 같이 9개의 분할 그룹에 대한 R, G, B 평균값의 차를 이용하여 2차 키 프레임 후보군을 생성한다. (식 6)은 질의 이미지와 1차 후보군의 이미지들

과 1번 그룹의 Red 값이 차를 구하는 식이다.

$$D_{g1, red} = |A_{g1, image, red} - A_{g1, frame, red}| \quad (\text{단, } D_{red} < \alpha) \quad (\text{식 6})$$

$A_{g1, image, red}$ 은 질의 이미지에서 그룹1의 Red에 대한 평균값이며, $A_{g1, frame, red}$ 은 1차 키 프레임 후보군에서 그룹1의 Red에 대한 평균값이다.

동일한 방법으로 9개의 영역에 대한 R, G, B 값의 차를 구하며, 위에서 언급한 이미지 전체에 대한 유사도를 구하는 방식과 같이 각 분할 영역에 대한 유사도를 구하여 가장 유사도가 높은 순으로 사용자에게 디스플레이한다.

5. 실험 평가

5.1 프로토타입 구현

제안하는 시스템의 실험 평가를 위한 시스템 프로토타입은 Visual C++ 6.0과 Delphi 6.0을 이용하여 구현하였으며, 실험은 Intel(R) P-IV CPU 2.4GHz, 512M RAM의 PC에서 MS-Windows 2000 Pro. Server 운영체제하에서 실험하였다.

5.1.1 키워드 질의 인터페이스

사용자의 1차적인 주석기반 검색을 위한 인터페이스는 (그림 7)과 같다.



(그림 7) 주석기반 검색 화면

사용자는 몇 개의 검색어를 입력하여 1차적으로 원하는 장면에 대한 키 프레임을 찾을 수 있다.

1차적인 주석기반 검색이 완료되면 사용자에게 결과로 키 프레임의 리스트가 디스플레이되고 사용자가 원하는 이미지를 선택한 후 특징기반 검색을 수행할 수 있다.

5.1.2 이미지 질의 인터페이스

특징기반 검색을 수행하게 되면 질의 이미지에 대한 멀티 분할 색상 히스토그램 기법에 의해 유사도 계산을 하여 가장 유사한 키 프레임 순으로 리스트가 다시 디스플레이된다. (그림 8)은 특징기반 검색을 위한 사용자 인터페이스를 나타내고 있다.



(그림 8) 특징기반 검색 화면

5.2 실험 평가

제안하는 시스템의 실험 평가를 위한 비디오 데이터 도메인은 MPEG 포맷인 일반 국내의 영화 비디오 파일을 이용하였다. 실험 데이터 수는 국내영화 10편, 외국영화 10편으로 하여 총 20개의 영화를 대상으로 하였으며, 해당 비디오 클립 파일은 총 32개이다. 총 28,641개의 키 프레임을 검출하였으며, 연속된 컷에 의한 중복된

내용을 가지고 있는 키 프레임이나 각 개체를 분별할 수 없는 키 프레임을 제외한 7,258개의 키 프레임들에 한 개의 기본 주석을 부여하였다.

제안하는 시스템의 검색 적합성[10]을 평가하기 위하여 (식 7)의 재현율(Recall)과 (식 8)의 정확율(Precision)에 의한 성능평가 척도를 이용하여 실험을 하였다. 기존 연구에서의 비디오 검색 시스템의 일치 검색을 기반으로 하는 검색의 적합성은 재현율과 정확율로 검색의 성능을 측정하고 있다[11].

$$Recall(\%) = \frac{C_{Search, right}}{C_{Total, right}} \times 100 \quad (\text{식 } 7)$$

$$Precision(\%) = \frac{C_{Search, right}}{C_{Search, total}} \times 100 \quad (\text{식 } 8)$$

$C_{Total, right}$ 는 전체 키 프레임 중 사용자 질의에 적합한 키 프레임의 수이고, $C_{Search, right}$ 는 검색된 키 프레임 중 사용자 질의에 적합한 키 프레임의 수이며, $C_{Search, total}$ 은 질의에 의해 검색된 전체 키 프레임의 수이다.

재현율은 전체 키 프레임을 대상으로 하였을 때 질의 이미지와 연관된 키 프레임들 중에서 질의 결과로 검색되어진 키 프레임이 질의 이미지와 얼마나 연관성이 있는지를 나타내는 값이고, 정확도는 검색된 키 프레임을 대상으로 하였을 때 검색된 모든 키 프레임들 중에서 질의 이미지와 연관된 키 프레임이 얼마나 되는지를 나타내는 값이다.

우선, 시스템의 주석기반 검색과 특징기반 검색을 혼합한 검색을 통한 재현율과 정확도를 계산하기 이전에 키 프레임의 질의 이미지를 이용하여 주석기반 검색을 배제한 특징기반 검색을 각 분할 그룹별로 실험해 보았다. <표 3>은 각 영역을 그룹화한 후 각 그룹별로 질의 이미지를 이용한 특징기반 검색만을 실험하였을 때 재현율을 나타낸 것이다.

<표 3> 검색 영역별 특징기반 검색의 재현율

영역 번호	1~9 영역 (전체)		3,5,7 영역 (대각선)		1,5,9 영역 (대각선)		4,5,6 영역 (중앙 가로)		1,3,7,9 영역 (모서리)	
	동일	비동일	동일	비동일	동일	비동일	동일	비동일	동일	비동일
1	50.21	34.25	41.21	35.35	43.04	30.41	32.50	40.61	51.25	45.35
2	51.84	40.06	45.61	46.20	44.51	48.65	45.52	55.04	45.61	42.20
3	30.52	25.65	42.21	34.21	38.81	40.36	41.45	35.61	32.21	44.21
4	34.01	26.45	30.01	36.82	35.61	43.48	42.88	47.52	45.05	46.82
5	42.21	40.28	46.31	38.85	45.01	40.91	40.84	45.09	46.31	48.85
6	45.21	49.95	45.21	40.31	42.31	39.44	46.25	41.41	45.29	40.31
7	34.81	44.65	42.81	42.25	42.21	41.87	41.05	41.65	42.82	35.25
8	30.21	35.21	31.91	49.31	38.01	39.61	35.45	32.11	38.98	30.31
9	37.11	47.87	45.61	48.08	44.41	45.52	44.65	43.68	45.65	48.08
10	50.61	56.21	42.21	43.87	40.21	45.11	46.15	45.27	42.24	40.87
평균	40.61	40.06	41.31	41.52	41.41	41.54	41.67	42.80	43.54	42.23

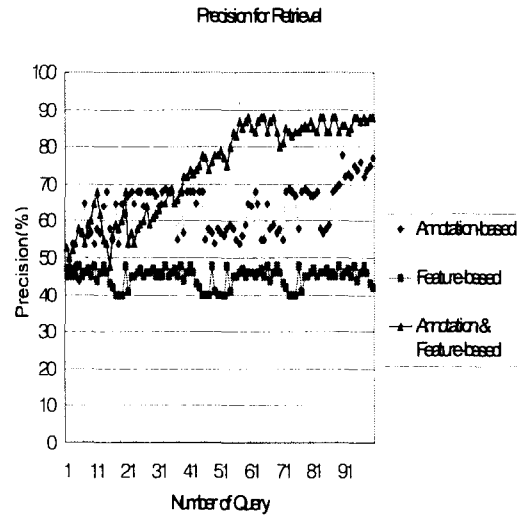
<표 3>의 재현율의 평균에서 나타나듯이 9개의 영역을 분할한 후 그룹으로 지정하여 칼라 히스토그램 분석을 하였을 때, 동일한 비디오 데이터 내에서 유사한 키 프레임을 검출하는 것은 1, 3, 7, 9번 영역을 비교하였을 때 가장 좋은 검색 결과를 보였으며, 동일하지 않은 비디오 데이터 내에서 유사한 키 프레임을 검출하는 것은 4, 5, 6번 영역을 비교하였을 때 가장 좋은 검색 결과를 보였다.

이는 동일한 비디오 데이터에서는 영화의 특성상 하나의 장면에 유사한 컷이 여러 개 있을 수 있으며, 키 프레임에 대한 색상 비교를 하였을 때 전체 이미지에서 중앙에 위치한 영역보다 상하의 모서리 부분의 이미지가 상대적으로 변화가 작기 때문이다. 반대로, 비동일 비디오 데이터에서는 영화의 특성상 각 영화마다 전체적인 배경과 조명이 서로 많이 다르므로 모서리 영역보다는 중앙의 가로 영역이 재현율이 높게 나타남을 보였다.

따라서, 제안하는 시스템에서는 1차 주석기반 검색이 완료된 후 질의 이미지를 통한 2차 특징기반 검색에서는 비동일 비디오 데이터에서 검

색할 경우는 4, 5, 6 영역을 동일 비디오 데이터에서 검색할 경우는 1, 3, 7, 9 영역을 비교 대상으로 하여 유사도 계산을 실행함으로써 계산량 및 계산 시간을 줄여 시스템 부하를 줄이도록 하였다.

제안하는 시스템의 전체적인 검색에 대한 정확도를 계산하기 위해 500회 이상의 사용자 질의를 실행하여 제안한 시스템의 검색에 대한 실험을 하였다. (그림 9)는 시스템의 검색에 대한 정확도를 나타낸 것이다.



(그림 9) 제안 시스템의 검색 정확도

(그림 9)에서 보이는 바와 같이 검색에 대한 정확도가 높게 나타났으며, 질의어가 증가할수록 에이전트의 질의어 학습과 주석정보의 갱신으로 인해 검색에 대한 정확도가 계속 증가하는 것을 알 수 있다.

6. 결론

본 논문에서는 대용량의 비디오 데이터에 대한 주석기반 검색과 특징기반 검색을 이용하여

사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 통합된 비디오 내용 검색 시스템을 제안하고 설계하였다. 또한, 실험을 위해 프로토타입을 구현하여 검색에 대한 정확도를 측정하였다. 사용자의 1차적인 키워드 질의와 질의에 의해 추출된 키 프레임의 이미지를 선택함으로써 에이전트는 추출된 키 프레임의 주석에 대한 의미를 더욱 구체화시킬 수 있다. 또한, 사용자에게 선택된 키 프레임은 제안하는 RGB 칼라 히스토그램 기법을 이용한 특징기반 검색기법을 통해 모든 비디오 데이터에서 가장 유사한 키 프레임을 비교하여 원하는 장면을 검색할 수 있다.

실험 결과에서 나타나듯이 사용자에게 대한 질의어와 키 프레임 선택에 의해 지속적인 에이전트의 키워드 학습과 그에 따른 주석의 자동 갱신 및 주석기반 검색에 의해 검출된 키 프레임 중에서의 2차적인 특징기반 검색으로 인해 비디오 데이터의 의미기반 장면 검색의 정확도를 높일 수 있었으며, 사용자의 질의어 시험결과 90%에 가까운 정확도를 보였다.

앞으로 제안된 알고리즘 외에 개체 인식 알고리즘을 추가하여 더욱 유사한 키 프레임을 검색할 수 있도록 함으로써 검색에 대한 완성도를 높일 수 있는 연구가 진행 중에 있으며, 제안된 시스템의 완전한 구현이 요구된다.

참 고 문 헌

- [1] Sibel Adali, et. al., "The Advanced Video Information System : data structure and query processing," *Multimedia System*, pp.172-186, 1996.
- [2] N. Dimitrova, A. Zakhor and T. Huang, "Applications of video-content analysis and retrieval", *IEEE Multimedia*, Vol.9, No.3, pp.42-55, 2002.
- [3] C. W. Ngo, T. C. Pong, H. J. Zhang, "Clustering and retrieval of video shots through temporal slices analysis," *IEEE Trans on Multimedia*, Vol.04, No.04, pp.446-458, 2002.
- [4] M. S. Kankanhalli and T. S. Chua, "Video modeling using strata-based annotation," *IEEE Multimedia*, Vol.7, No.1, pp.68-74, 2000.
- [5] Myron Flickner and et. al, "Query by Image and Video Content : The QBIC system," *IEEE Computer*, Vol. 28, No. 9, 1995.
- [6] J. R. Smith and S. F. Chang, "VisualSEEK : a fully automated content-based image query system," *ACM Multimedia, Boston*, 1996.
- [7] Tony C. T. Kuo and Arbee L. P. Chen, "A Content Based Query Language for Video Database," *IEEE M.M.* '96, pp. 209-214, 1996.
- [8] Sibel Adali, et. al., "the Advanced Video Information System : data structures and query processing," *Multimedia System*, pp. 172-186, 1996.
- [9] R. Hjelsvold, "VideoSTAR-A Database for Video Information Sharing," Ph.D. Thesis, Norwegian Institute of Technology, 1995.
- [10] Eitetsu Oomoto, Katsumi Tanaka, "OVID : Design and Implementation of a Video Object Database System," *IEEE TKDE*, Vol. 5, No. 4, pp. 629-643, 1993.
- [11] D. Shasha and T.L. Wang, "New Techniques for Best-match Retrieval," *ACM TOIS*, Vol. 8, No. 2, pp.140-158, 1990.
- [12] G. Salton and M. J. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.
- [13] 이태동, 김민구, "통합된 비디오 인덱싱 방법을 이용한 내용기반 비디오 데이터베이스의 설계 및 구현", *한국정보과학회 논문지(C)*, 제7권 제6호 pp.661-683, 2001. 12.
- [14] 윤미희, 조동욱, "비디오의 의미검색과 예제기반 장면검색을 위한 비디오 검색시스템", *한국정보처*

리학회 논문지(B), 제9-B권 제1호 pp.105-112,
2002. 2.

- [15] 남윤영, 황인준, "XMARS : XML 기반 멀티미디어 주석 및 검색 시스템", 한국정보처리학회 논문지(B), 제9-B권, 제5호, pp.541-548, 2002. 10.

김 정 재



1999년 영동대학교 컴퓨터공학과(공학사)
2001년 숭실대학교 컴퓨터학과(공학석사)
2003년 숭실대학교 컴퓨터학과(박사수료)

관심분야 : EC, CRM, S/W Agent, PKI

1989년~1991년 New Mexico State University
Physical Science Lab 책임 연구원

1991년~현재 숭실대학교 부교수
관심분야 전자상거래 보안, 인터넷 보안, 멀티미디어 보안, 인증 시스템

이 창 수



1999년 한서대학교 컴퓨터과학과
(이학사)
2002년 숭실대학교 컴퓨터학과(공학석사)
2004년 숭실대학교 컴퓨터학과(박사수료)

관심분야 : DRM, Image processing, PKI

이 중 희



1998년 한밭대학교 전자계산학과(공학사)
2000년 숭실대학교 컴퓨터학과(공학석사)
2002년 숭실대학교 컴퓨터학과(박사수료)

관심분야 : Multimedia, EC, Agent

전 문 석



1981년 숭실대학교 전자계산학과
(공학사)
1986년 University of Maryland
Computer Science(공학석사)
1989년 University of Maryland
Computer Science(공학박사)
1989년 3월~7월 Morgan State

University 조교수