

키 프레임의 주석과 비교 영역 학습을 이용한 비디오 검색 시스템의 구현

이근왕[†], 이종희^{**}, 김희숙^{***}

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. 본 논문에서는 사용자의 키워드 학습과 비교 영역 학습을 이용하여 대용량의 비디오 데이터에 대한 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 자동화된 비디오 검색 시스템을 제안한다. 사용자의 기본적인 질의와 질의에 의해 추출된 키 프레임의 이미지를 선택함으로써 에이전트는 추출된 키 프레임의 주석에 대한 의미를 더욱 구체화시킨다. 또한, 사용자에게 의해 선택된 키 프레임은 질의 이미지가 되어 색상 히스토그램 비교기법과 제안하는 비교 영역 학습 기법을 통해 가장 유사한 키 프레임을 검색한다. 설계하고 구현한 시스템은 실험을 통한 성능평가에서 93% 이상의 높은 정확도를 보였다.

Implementation of a Video Retrieval System Using Annotation and Comparison Area Learning of Key-Frames

Keun-Wang Lee[†], Jong-Hee Lee^{**}, Hee-Sook Kim^{***}

ABSTRACT

In order to process video data effectively, it is required that the content information of video data is loaded in database and semantics-based retrieval method can be available for various queries of users. In this paper, we propose a video retrieval system which support semantics retrieval of various users for massive video data by user's keywords and comparison area learning based on automatic agent. 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 becomes a query image and searches the most similar key frame through color histogram comparison and comparison area learning method that proposed. From experiment, the designed and implemented system showed high precision ratio in performance assessment more than 93 percents.

Key words: Semantics-Based Video Retrieval(비디오 검색), User's Keyword Learning(사용자 키워드 학습), Comparison Area Learning(비교영역 학습)

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

※ 교신저자(Corresponding Author) : 이종희, 주소 : 서울시 강남구 역삼동 700-1 수림빌딩 203호(135-080), 전화 : 02)3452-0957, FAX : 02-3452-0956

E-mail : multistar@freechal.com

접수일 : 2004년 6월 11일, 완료일 : 2004년 10월 4일

[†] 중신회원, 청운대학교 멀티미디어학과 교수

(E-mail : kwlee@cwunet.ac.kr)

^{**} 정회원, (주)리테일테크 기술연구소 소장

^{***} 준회원, 아산정보기능대학 멀티미디어과 교수

(E-mail : prima@kopo.or.kr)

※ 본 연구는 2004년도 청운대학교 학술연구조성비 지원에 의하여 수행되었음.

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 문자화되어있지 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[3,11]. 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 검색하

는 방법이다[4,12]. 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 검색으로는 정확한 검색 결과를 얻을 수 없다.

최근 이 두 가지 비디오 데이터 검색 방법을 통합하여 비디오 데이터에 대한 검색 효율을 높이고자 하는 연구가 진행 중에 있으나 개별 사용자들의 서로 다른 주관적인 의미부여로 인해 높은 검색 효율을 기대하기 힘들다. 본 논문에서는 비디오 파일의 키 프레임에 대한 주석기반 검색을 통해 사용자가 비디오 장면의 키프레임에 대한 키워드를 입력하여 검색하면 그 질의의 키워드와 그 결과를 학습하여 학습된 키워드들에 가중치를 부여한 후 검색이 이루어질 때마다 가중치의 값이 재계산되어 가장 최적의 키워드가 유지될 수 있도록 한다. 그리고 특징기반 검색을 통해 비교 영역 학습 기법을 이용하여 사용자의 질의 이미지와 가장 유사한 키 프레임을 결과로 도출한다. 따라서, 본 논문에서는 비디오 데이터의 키 프레임에 대한 주석 정보와 비교 영역 학습을 수행하는 인텍싱 에이전트를 이용한 내용기반 비디오 검색 시스템을 제안한다.

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 샷(shot)의 검출, 샷에 대한 키 프레임 생성과 객체의 움직임 등의 특징 정보를 이용하여 데이터를 검색한다.

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 비디오의 각 프레임에 나타나는 객체들의 시간관계, 공간관계를 메타데이터로 구축하여 이미지 검색 처리에 이용하였다.

특징기반 검색에 있어서 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.

2.2 주석기반 비디오 검색기법

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

노르웨이 공대에서 개발한 VideoSTAR(Video Storage and Retrieval)[9]는 관계형 데이터베이스 모델을 사용한 데이터베이스 시스템으로서 구조화된 비디오 데이터에 대해 인물, 위치, 사건 등의 속성으로 메타데이터를 구성하고, 이를 다시 기본, 1차, 2차 컨텍스트(context)로 구분하여 메타 데이터의 재사용과 공유를 용이하게 하며, 고정된 속성을 통해 사용자가 쉽게 질의를 구성할 수 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 위해서

는 에이전트에 의한 자동화된 메타데이터 생성과 검색 질의와 결과를 학습하여 주석정보로 이용하는 것이 필수적이라 할 수 있다.

3. 의미기반 비디오 검색 시스템

3.1 전체 시스템 구조

새로운 비디오 데이터가 입력되면 시스템의 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. 특징이 분석된 키 프레임들은 키 프레임 데이터베이스에 저장된다. 그림 1은 제안하는 시스템의 전체 구조를 나타낸다.

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.

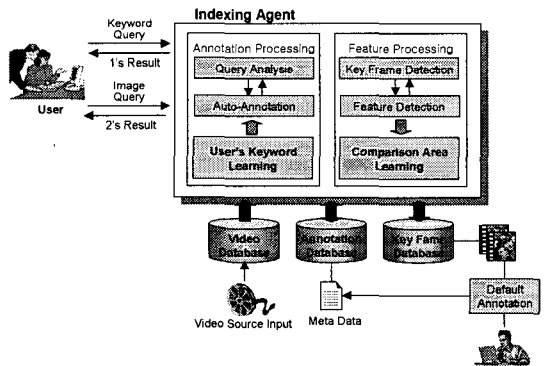


그림 1. 시스템 구조

3.2 키워드 학습에 의한 자동 주석 갱신 기법

단어 한개 이상으로 구성된 사용자 질의가 입력되

면 키워드가 추출되고, 추출된 사용자 키워드에 의해 사용자 키워드가 포함된 키 프레임들을 검색한다. 사용자 키워드는 실 키워드와 잠재적 키워드로 분류된 후, 키 프레임의 주식정보와 비교하여 주식 정보에 있는 키워드에서 정확히 매칭된 키워드는 동일 키워드로 정확히 매칭되지 않은 키워드는 비동일 키워드로 분류한다. 그림 2는 자동 주식처리를 하기 위한 전처리인 키워드 분류에 대해 나타내고 있다.

그림 2에서 각 키워드들의 정의는 표 1과 같다.

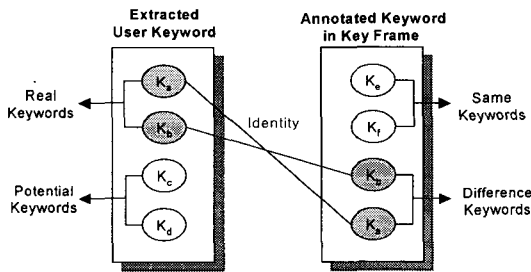


그림 2. 키워드 분류

3.2.1 의미 가중치 계산 알고리즘

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

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

$$W_{Keyword \rightarrow +w} = W_{Keyword \rightarrow old} + \frac{1}{N_{Kframe \rightarrow SK}} \quad (1)$$

$W_{keyword \rightarrow +w}$ 은 주식 키워드에 대한 새로운 의미가

중치이고, $W_{keyword \rightarrow old}$ 은 주식 키워드에 대한 이전의 의미 가중치이며, $N_{Kframe \rightarrow SK}$ 은 동일한 키워드가 포함되어있는 키 프레임의 개수이다.

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

$$W_{Keyword \rightarrow +w} = W_{Keyword \rightarrow old} - \frac{1}{N_{Kframe \rightarrow SK}} \quad (2)$$

사용자들이 질의를 할 때 마다 키 프레임 주식정보인 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 주식 정보에서 삭제한다. 따라서 사용자의 잘못 입력한 키워드나 의미 부여가 잘못된 키워드에 대해 해당 키워드를 삭제시킬 수 있는 결과를 도출할 수 있다.

3.3 키 프레임 특징처리 기법

3.3.1 색상 히스토그램 비교

사용자로부터 입력된 질의 이미지와 비디오 데이터에서 추출된 키 프레임의 이미지에 대한 색상 유사도를 빠르고 정확하게 계산하기 위하여 질의 이미지와 키 프레임의 이미지를 분할한다. 멀티분할 히스토그램 기법을 이용한 색상 유사도 계산 과정은 그림 3과 같다.

키 프레임으로 추출된 $n \times m$ 크기의 이미지는 픽셀을 3×3 영역으로 분할한다. 따라서 분할된 하나의

표 1. 키워드 정의

키워드	키워드 분류	의미
사용자 키워드 (Extracted User Keyword)	실 키워드 (Real Keyword)	사용자가 질의를 위해 질의 입력창에 입력하는 키워드
	잠재적 키워드 (Potential Keyword)	사용자 키워드 중 DB에 있는 주식과 매칭되지 않은 키워드
주식 키워드 (Annotated Keyword)	동일 키워드 (Same Keyword)	DB에 있는 주식 중 사용자 키워드와 매칭된 키워드
	비동일 키워드 (Difference Keyword)	DB에 있는 주식 중 사용자 키워드와 매칭되지 않은 키워드

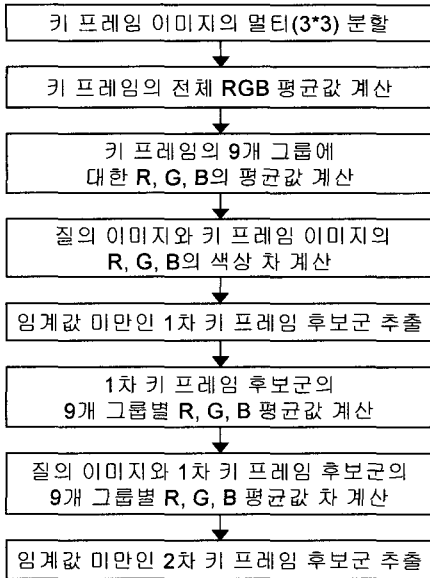


그림 3. 색상 유사도 계산 과정

이미지는 9개의 픽셀 그룹을 가지게 된다. 질의 이미지를 9개의 그룹으로 분할 한 후 질의 이미지 전체에 대한 R, G, B 평균값과 각 그룹의 R, G, B 평균값을 이용하여 멀티분할 색상 히스토그램 기법을 이용하여 질의 이미지와 1차적인 주석기반 검색에 의해 추출된 모든 키 프레임들의 색상 유사도를 계산한다 [13].

3.3.2 비교영역 학습

인덱싱 에이전트는 사용자의 최종 키 프레임 선택에 의한 비교영역을 학습한다. 사용자의 초기 질의에 의해 주석기반 검색이 수행된 후 그 결과인 키 프레

임 리스트에서 사용자의 키 프레임 이미지 선택이 이루어지면 데이터베이스 안에 학습되어 저장된 해당 키 프레임 이미지의 멀티분할 색상 히스토그램의 히스토리 값을 이용하여 비교 영역을 선택한다.

그림 4는 9개의 영역 중 비교대상 영역의 종류를 나타낸 것이다. 인덱싱 에이전트는 분할 영역 중 검색 시간과 정확성을 고려하여 가장 효과적인 비교 검색을 위해 9개의 분할 영역 중 어떤 영역을 선택하여 비교할지 결정한다. 결정된 해당 키 프레임의 비교 영역은 검색이 이루어질 때마다 지속적으로 학습되어 키 프레임의 정보에 저장한다. 최적인 비교 영역 검출을 위해 비교 영역의 색상 차이 가중치를 계산한다. 비교 영역의 색상 차이 가중치란 질의 이미지와 비교 대상이 되는 다른 키 프레임 이미지들과의 각 영역별 색상차이를 가중치로 표현한 것으로, 비교 대상이 되는 9개의 영역에 각각 색상차이 가중치를 부여하여 비교할 때마다 비교된 영역에 대한 가중치를 증가시키기 위한 값이다.

비교 영역의 색상 차이율을 계산하기 위한 색상 차이 가중치 계산은 다음과 같다.

- $W_{diff_color[n]}$: 비교 영역의 색상 차이 가중치
- Rank : 1부터 9까지의 정수
- $R_{[n]}$: 각 영역의 순위 값
- $A_{W_new[n]}$: 현재의 $W_{diff_color[n]}$ 의 누적의 평균
- $A_{W_old[n]}$: 이전의 $W_{diff_color[n]}$ 의 누적의 평균
- β : 비교 영역 추출 임계값 (0.5)
- N : 멀티 분할된 비교 영역 수

질의 이미지와 비교되는 키 프레임 이미지들의 9

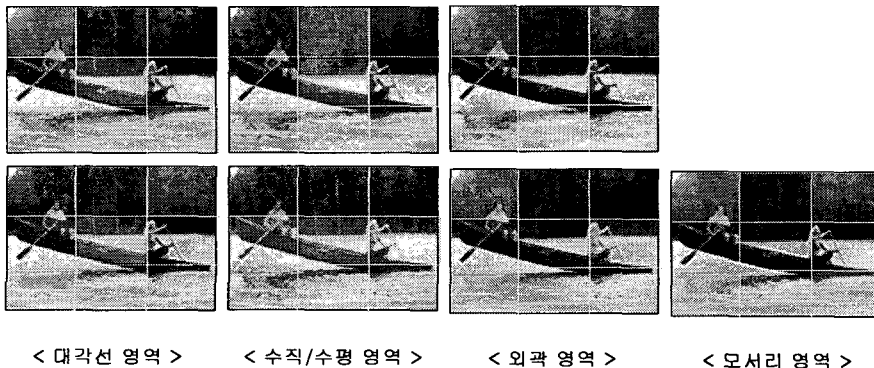


그림 4. 선택 비교대상 영역의 종류

개의 비교 영역별 색상 히스토그램 값의 차를 각각 구한 후 값의 차가 적은 순으로 1부터 9까지의 순위 값을 부여한다. 다음 식에 의해 부여된 순위 값을 이용하여 비교 영역의 색상 차이 가중치를 구한다.

$$W_{diff_color[n]} = 1 - \frac{R_{[n]} - 1}{N} \quad (R_{[n]} | R_{[n]} \in Rank) \quad (3)$$

해당 비교 영역의 색상 차이 가중치는 부여받은 순위 값에 1과의 차를 구하여 이를 비교영역인 멀티분할된 비교 영역 수 N 으로 나눈 후, 이 값을 다시 정수 1과의 차로 구한다. 따라서, 색상 차이 가중치의 값은 0에서 1사이의 실수 값을 갖는다.

색상 차이 가중치를 사용자 검색이 이뤄질 때마다 지속적으로 누적 평균을 구한다. 따라서, 새로운 $W_{diff_color[n]}$ 의 누적 평균인 $A_{W_new[n]}$ 는 이전까지의 누적평균에 현재의 비교 영역 색상 차이 가중치를 합하여 다음과 같이 구한다.

$$A_{W_new[n]} = Avg[A_{W_old[n]} + W_{diff_color[n]}] \quad (4)$$

(단, $A_{W_new[n]} \geq \beta$)

0에서 1까지의 실수 값을 갖게 되는 색상 차이 가중치의 누적 평균값에서 임계값 $\beta(=0.5)$ 이상의 값을 가진 영역만을 선택하여 이를 최적 비교 영역으로 정한다.

색상 차이 가중치의 초기값은 아직 학습된 영역이 없으므로 임의의 영역을 선정해야한다. 이는 색상 기반 검색의 재현율을 측정하여 다른 영역에 비해 높은 재현율을 갖는 임의의 영역 중에서 선택적으로 랜덤한 영역을 초기 비교 영역으로 정의한다. 이렇게 비교영역을 학습하여 최적의 영역만을 비교하는 이유는 전체 영역을 비교하여 특징기반 검색을 하게 되면 유사도 계산에 있어서 계산량 및 계산 시간이 많이 소요되므로 이러한 단점을 줄이기 위함이다.

4. 구현

4.1 키워드 질의 인터페이스

사용자는 주식 기반 검색에 의해 몇 개의 검색어를 입력하여 1차적으로 원하는 장면에 대한 키 프레임을 찾아 수 있다.

사용자에 의해 입력된 키워드는 DB에 저장되어

있는 키 프레임의 주석과 비교하여 매칭되는 키 프레임을 찾아 사용자에게 디스플레이한다. 디스플레이된 키 프레임 리스트 중에 사용자가 특정 키 프레임을 선택하게 되면 그 키 프레임은 의미 가중치 계산에 의해 새로운 가중치 값을 갖게 된다. 1차적인 주석 기반 검색이 완료되면 사용자에게 결과로 키 프레임의 리스트가 디스플레이되고 사용자가 원하는 이미지를 선택한 후 특징 기반 검색을 수행할 수 있다.

사용자의 1차적인 주석기반 검색을 위한 사용자 의 키워드 질의 화면은 그림 5와 같다.



그림 5. 키워드 질의 화면

4.2 주석 데이터베이스

1차적인 주석기반 검색이 완료된 후 키 프레임 이미지를 사용자가 선택하게 되면 선택된 키 프레임 이미지의 의미 가중치가 재 계산된다. 또한, 키워드 매칭 결과 비동일 키워드는 잠재적 키워드로 주석 카테고리에 등록된다. 그림 6은 주석 데이터베이스의 의미 가중치 테이블과 주석 카테고리 테이블을 나타내고 있다.

의미 가중치 테이블에서 'AWeight' 칼럼은 각 주석의 의미 가중치 값을 나타낸다. 'AFeature' 칼럼은 각 키 프레임의 특징 분석 여부를 나타낸 값이다.

4.3 이미지 질의 인터페이스

특징기반 검색을 수행하게 되면 질의 이미지에 대한 멀티 분할 색상 히스토그램 기법에 의해 유사도 계산을 하여 가장 유사한 키 프레임순으로 리스트가

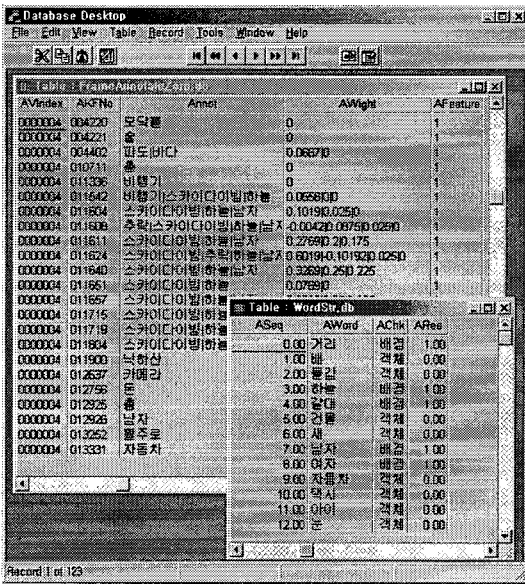


그림 6. 주석 데이터베이스

다시 디스플레이된다. 그림 7은 특징기반 검색을 위한 사용자 인터페이스를 나타내고 있다.

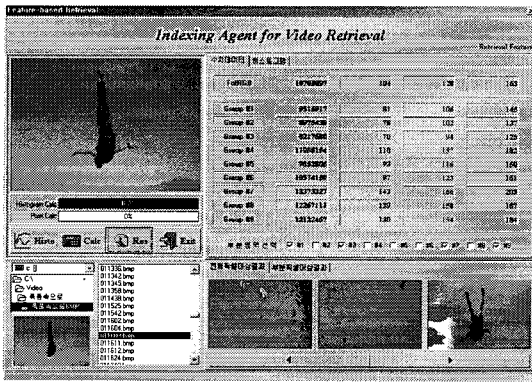


그림 7. 특징기반 검색

5. 실험 평가

5.1 실험 환경

제안하는 시스템의 실험 평가를 위한 비디오 데이터 도메인은 MPEG 포맷인 약 500~600Mbyte 용량의 일반 국내의 영화 비디오 파일을 이용하였다. 실험 데이터 수는 국내영화 14편, 외국영화 16편으로 하여 총 30개의 영화를 대상으로 하였으며, 해당 비디오 클립 파일은 총 37개이다. 총 32,957개의 키 프

레이를 검출하였으며, 연속된 컷에 의한 중복된 내용을 가지고 있는 키 프레임이나 각 개체를 분별할 수 없는 키 프레임을 제외한 5,351개의 키 프레임들에 한 개의 기본 주석을 부여하였다. 각각의 비디오 데이터의 최적 비교검색 영역을 추출하기 위해 영화의 각 장르별로 비디오 데이터를 분류한 후 검색 영역을 선택하여 비교 검색해 보았다.

5.2 실험 내용

제안하는 시스템의 검색 적합성[10]을 평가하기 위하여 식 5의 재현율(Recall)과 식 6의 정확도(Precision)에 의한 성능평가 척도를 이용하여 실험을 하였다.

$$Recall(\%) = \frac{C_{Search, right}}{C_{Total, right}} \times 100 \quad (5)$$

$$Precision(\%) = \frac{C_{Search, right}}{C_{Search, total}} \times 100 \quad (6)$$

$C_{Total, right}$ 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.

5.3 실험 결과

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 2는 각 영역을 그룹화한 후 각 그룹별로 질의 이미지를 이용한 특징기반 검색만을 실행하였을 때 재현율을 나타낸 것이다.

표 2의 재현율의 평균에서 나타나듯이 9개의 영역을 분할한 후 그룹으로 지정하여 칼라 히스토그램

표 2. 검색 영역별 특징기반 검색의 재현율

질의	19 영역 (전체)		3,57 영역 (대각선)		1,59 영역 (대각선)		4,56 영역 (중앙 가로)		1,3,7,9 영역 (모서리)	
	동일	비동일	동일	비동일	동일	비동일	동일	비동일	동일	비동일
1	50.23	34.25	41.25	35.35	43.05	31.41	32.50	36.61	38.25	45.35
2	51.85	40.06	45.61	46.20	44.51	48.65	45.52	45.04	45.61	42.20
3	31.52	25.65	42.21	34.21	38.81	40.36	41.45	45.61	39.21	44.21
4	35.01	27.45	31.05	36.82	35.65	43.48	42.88	47.52	45.05	46.82
5	42.21	41.28	46.31	38.85	45.01	40.91	40.84	45.09	46.31	48.85
6	45.21	49.95	45.29	41.31	43.39	39.44	46.25	41.41	45.29	40.31
7	34.82	44.65	43.82	42.25	42.22	42.87	41.05	41.65	42.82	35.25
8	30.23	35.21	31.98	49.31	38.07	39.61	36.45	40.11	50.98	31.31
9	37.12	47.87	45.65	48.08	44.45	45.52	44.65	43.68	45.65	48.08
10	50.65	56.21	42.24	43.87	41.29	45.11	46.15	45.27	42.24	40.87
평균	40.88	40.26	41.54	41.63	41.65	41.74	41.77	43.20	44.14	42.33

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 것이 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 검색할 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 8은 시스템의 검색에 대한 재현율을 나타낸 것이다.

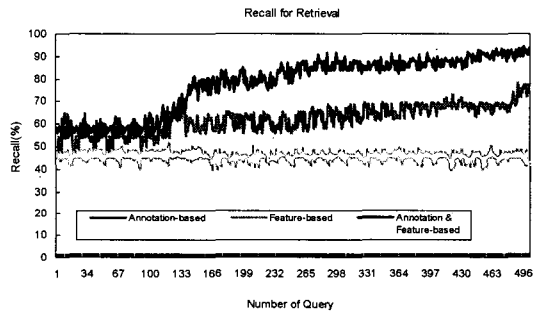


그림 8. 제안 시스템의 검색 재현율

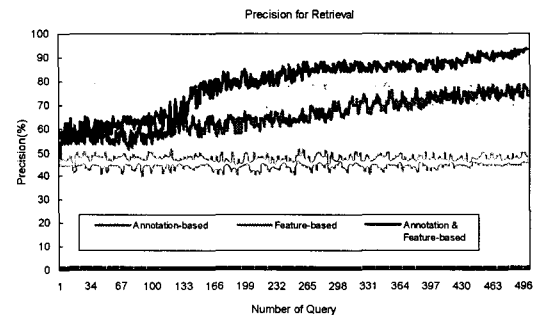


그림 9. 제안 시스템의 검색 정확도

그림 9는 시스템의 검색에 대한 정확도를 나타낸 것이다.

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

표 3. 기존 시스템과 제안 시스템의 처리기능 비교

시스템 \ 항목	주석기반	특징기반	의미기반	인덱싱 자동화	대용량 비디오 처리	초기 특징 추출 속도
OID (Gobeu Univ.)	○	×	×	×	×	빠름
VideoSTAR (Xr. Univ.)	○	×	×	×	×	빠름
Algebraic Video (MIT)	○	×	×	×	×	빠름
Venus (CWU)	×	○	○	×	×	보통
VideoQ (Col. Univ.)	○	○	×	×	×	느림
IVIS (AJ Univ.)	○	○	×	×	×	느림
TIVRON (SNU)	○	○	×	×	×	느림
제안 시스템 (SSU)	○	○	○	○	○	느림

있다.

5.4 기존 시스템과의 비교

제안하는 시스템의 처리와 특징에 대해 기존 시스템과의 비교를 표 3에 요약하였다.

표 3에 나타나듯이 멀티미디어 데이터의 내용기반 검색 시스템인 기존 시스템과의 비교에서 대부분의 기존 시스템들과는 처리기능에 있어 현저한 차이를 보였으며, 특히, 제안하는 시스템과 동일한 방식인 주석 및 특징기반 검색을 동시에 하는 VideoQ 시스템, IVIS 시스템 그리고 TIVRON 시스템은 의미기반 검색을 위한 인덱싱 자동화와 대용량의 비디오 처리가 이루어지지 않는다. 하지만 제안하는 시스템은 인덱싱 자동화 기능을 갖추고 있으며, 대용량 비디오 처리를 지원한다.

6. 결 론

본 논문에서는 약 500~600Mbyte 용량의 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 시스템을 구현하여 검색에 대한 재현율과 정확도를 측정하였다. 사용자의 1차적인 키워드 질의를 지속

적으로 학습하고 질의에 의해 추출된 키 프레임의 이미지를 선택함으로써 에이전트는 추출된 키 프레임의 주석에 대한 의미를 더욱 구체화시킬 수 있다. 또한, 사용자에게 선택된 키 프레임은 색상 히스토그램 기법과 제안하는 영역비교 학습을 이용한 특징기반 검색기법을 통해 모든 비디오 데이터에서 가장 유사한 키 프레임을 비교하여 원하는 장면을 검색할 수 있다.

실험 결과에서 나타나듯이 사용자의 450회 질의부터 90% 이상의 정확도를 보였으며, 500회 질의까지 최고 93.3%의 정확도를 보였다.

참 고 문 헌

- [1] N. Dimitrova, A. Zakhor, and T. Huang, "Applications of Video-content Analysis and Retrieval," IEEE Multimedia, Vol. 9 No. 3 pp. 42-55, 2002.
- [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 and 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] M. Flickner, "Query by Image and Video Content : The QBIC system," IEEE Computer, Vol. 28, No. 9, pp. 23-32, 1995.

[6] J. R. Smith and S. F. Chang, "VisualSEEK : A Fully Automated Content-based Image Query System," ACM Multimedia Conference, Boston, 1996.

[7] T. Kuo and L. P. Chen, "A Content Based Query Language for Video Database," IEEE Multimedia, pp. 209-214, 1996.

[8] S. Adali, "The Advanced Video Information System : Data Structure and Query Processing," ACM Multimedia System, Vol. 4, No. 4, pp. 172-186, 1996.

[9] R. Hjelmsvold, "VideoSTAR-A Database for Video Information Sharing," Ph.D. Thesis, Norwegian Institute of Technology, 1995.

[10] D. Shasha and T. L. Wang, "New Techniques for Best-match Retrieval," ACM TOIS, Vol. 8, No. 2, pp. 140-158, 1990.

[11] 이정봉, 박장춘, "특징 분석과 프랙탈 차원을 이용한 객체 기반 영상검색," 한국멀티미디어학회 논문지, 제7권, 제2권, pp. 173-187, 2004.

[12] 홍성용, 나연목, "멀티미디어 정보처리 : XML을 이용한 지능형 이미지 검색 시스템," 한국멀티미디어학회 논문지, 제7권, 제1호, pp. 132-145. 2004.

[13] 이종희, 오해석, "의미기반 비디오 검색을 위한 인덱싱 에이전트의 설계," 한국정보처리학회 논문지, 제10-B권, 제6호, pp. 687-694, 2003.



이 근 왕

1993년 한밭대학교 전자계산학과 공학사
 1996년 숭실대학교 컴퓨터학과 공학석사
 2000년 숭실대학교 컴퓨터학과 공학박사

2001년~현재 청운대학교 멀티미디어학과 교수

관심분야 : 멀티미디어, 시스템 보안, 이동통신, 원격교육



이 종 희

1998년 한밭대학교 전자계산학과 공학사
 2000년 숭실대학교 컴퓨터학과 공학석사
 2004년 숭실대학교 컴퓨터학과 공학박사

2004년~현재 (주)리테일테크 기술연구소 소장

관심분야 : 멀티미디어, RFID, CRM, S/W 에이전트



김 희 숙

1991년 원광대학교 전자계산공학과 졸업
 1995년 원광대학교 전자계산교육전공 석사
 2003년 원광대학교 컴퓨터공학과 공학박사

2001년~현재 아산정보기능대학 멀티미디어과 교수

관심분야 : 멀티미디어 DB, 멀티미디어 통신, 시스템 보안