

# 인덱싱 에이전트를 이용한 멀티미디어 데이터 검색시스템

고재운  
콤넥스시스템  
e-mail:gjwmr@naver.com

## A Multimedia Data Search System using Indexing Agent

Jae-Woon Ko  
Connex System Inc.

### 요 약

비디오 데이터를 효율적으로 처리하기 위해서는 비디오 데이터가 가지고 있는 내용에 대한 정보를 데이터베이스에 저장하고 사용자들의 다양한 질의를 처리할 수 있는 의미기반 검색 기법이 요구된다. 기존의 내용기반 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, 사용자에게 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.

### 1. 서론

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]. 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

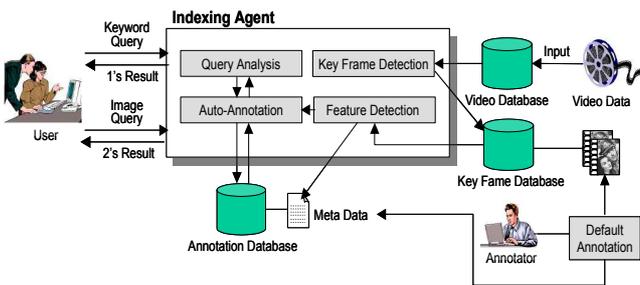
문자로 사용자가 일일이 주석을 부여하여야 하므로 많은 시간과 노력을 필요로 하며, 불필요한 주석의 양이 방대하게 증가할 수 있다는 단점이 있다. 또한, 시스템에서의 주석을 입력하는 주석자 한 사람의 주관적인 의미부여로 인해 많은 사용자의 다양한 검색어로는 정확한 검색 결과를 얻을 수 없다.

따라서, 본 논문에서는 사용자의 질의와 결과를 학습하여 비디오 데이터의 메타데이터를 지속적으로 자동 갱신하는 인택싱 에이전트를 이용한 내용기반 비디오 검색 시스템을 제안한다.

## 2. 인택싱 에이전트

### 2.1 전체 시스템 구조

새로운 비디오 데이터가 입력되면 인택싱 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. 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. (그림 1)은 제안하는 시스템의 전체 구조를 나타낸다.



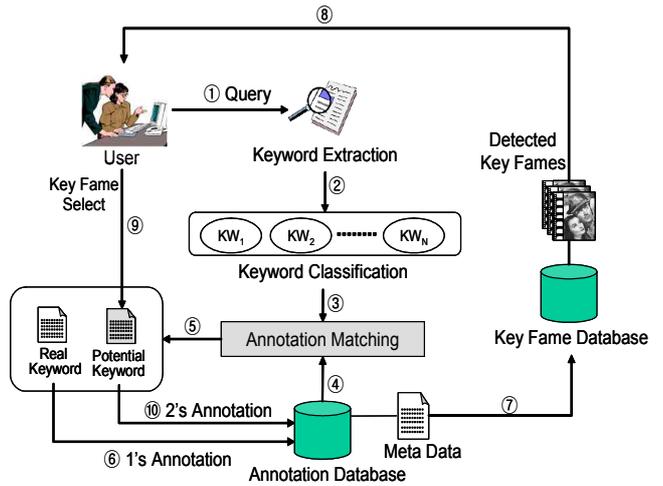
[그림 1] 시스템 구조

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.

### 2.2 자동 주석처리

에이전트는 사용자로부터 비디오 내용 검색에 대한 질의를 받으면 질의어 중 키워드를 추출하여 메

타데이터의 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] 자동 주석처리

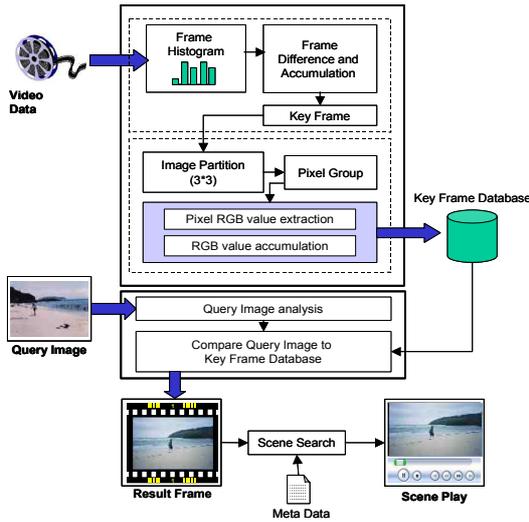
(그림 2)에서 보이는 바와 같이 우선 사용자의 질의가 입력되면 사용자 질의어를 분석하여 키워드를 추출한다. 추출된 사용자 키워드는 주석 데이터베이스 안에 있는 메타데이터의 주석 정보와 매칭시킨다. 매칭시킨 결과 정확히 일치하는 키워드를 주석 정보로 가지고 있는 키 프레임들을 데이터베이스에서 검출하여 사용자에게 전달한다. 또한, 사용자의 질의어 중 주석 정보와 정확히 매칭되지 않는 키워드들은 잠재적 키워드로 정의한다.

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.

### 2.3 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 이미지

는 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 검색

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.

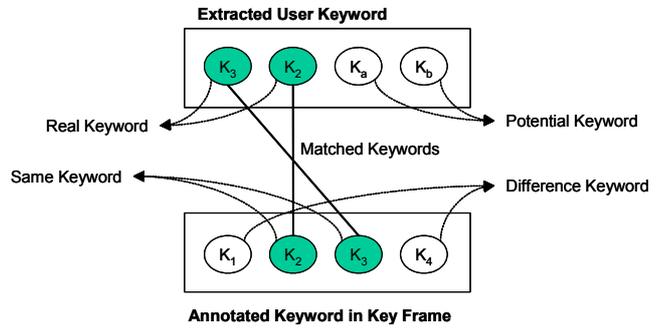
### 3. 인텔싱 기법

#### 3.1 자동 주석처리 기법

단어 한 개 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.

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.



[그림 4] 키워드 분류

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

$$W_{Keyword\_new} = W_{Keyword\_old} + \frac{1}{N_{Kframe\_SK}} \quad (식 1)$$

은 주석 키워드에 대한 새로운 의미 가중치이고, 은 주석 키워드에 대한 이전 의미 가중치이며, 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 때마다 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 1로 하여 의미 가중치가 0의 값을 갖게 되는 키워드는 의미 없는 키워드로 에이전트가 판단하여 해당 키 프레임에 대한 주석 정보에서 삭제한다. 따라서 사용자의 잘못된 입력한 키워드나 의미 부여가 잘못된 키워드에 대해 해당 키워드를 삭제시킬 수 있는 결과를 도출할 수 있다.

#### 4.2 키 프레임 특징처리 기법

사용자로부터 입력된 질의 이미지와 비디오 데이터에서 추출된 키 프레임의 이미지에 대한 색상 유사도를 빠르고 정확하게 계산하기 위하여 질의 이미지와 키 프레임의 이미지를 분할한다. 키 프레임의 전체 이미지에 대한 R, G, B의 평균을 구하는 식은 다음과 같다. 우선, Red의 평균을 구하는 식은 다음과 같다.

$$Avg_{t\_image, red} = \frac{\sum_{i=0}^w \sum_{j=0}^h R_{[i, j]}}{C_t} \quad (\text{식 3})$$

#### 4. 결론

(식 3)에서  $Avg_{t\_image, red}$ 은 키 프레임의 전체 이미지에 대한 Red 값이고,  $R_{[i, j]}$ 은 전체 이미지에서  $i$ 행  $j$ 열에 있는 Red 값이며,  $C_t$ 는 전체 이미지에서 0이 아닌 Red 값을 가지는 픽셀의 개수이다.  $w$ 와  $h$ 는 키 프레임 전체 이미지의 넓이와 높이이다. (식 3)은 키 프레임의 전체 이미지에 대한 Red 값의 평균을 구하는 식이며, Green 값과 Bule 값의 평균도 같은 방법으로 계산한다. 또한, 9개로 분할된 영역에 대한 R, G, B 값의 평균도 동일한 방법으로 구한다.

본 논문에서는 대용량의 비디오 데이터에 대한 주석기반 검색과 특징기반 검색을 이용하여 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 통합된 비디오 내용 검색 시스템을 제안하고 설계하였다.

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.

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.

$$D_{t\_red} = |Avg_{t\_image, red} - Avg_{t\_frame, red}|$$

(단,  $D_{t\_red} < \alpha$ ) (식 4)

(식 4)에서  $D_{t\_red}$ 는 질의 이미지와 키 프레임의 전체 이미지에 대한 Red의 색상 차이이고,  $Avg_{t\_image, red}$ 는 질의 이미지의 전체에 대한 Red의 평균값이며,  $Avg_{t\_frame, red}$ 는 키 프레임의 전체 이미지에 대한 Red의 평균값이다.  $\alpha$ 는 질의 이미지와 키 프레임의 비교 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.

$$ST_{Q=, K\_frame} = |D_{t\_red} + D_{t\_green} + D_{t\_blue}|$$

(식 5)

(식 5)에서  $ST_{Q\_image, K\_frame}$ 은 질의 이미지와 키 프레임 이미지의 유사도이다. 따라서, 각 R, G, B의 차의 절대값은 임계값 보다 작은 값을 1차 후보 프레임으로 선정되고 (식 5)에 의해서 색상 유사도의 값이 작을수록 입력되는 이미지와 유사한 이미지로 계산하여 정렬한다.

#### 참고문헌

- [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] 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.
- [3] 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] 이태동, 김민구, "통합된 비디오 인덱싱 방법을 이용한 내용기반 비디오 데이터베이스의 설계 및 구현", 한국정보과학회 논문지(C), 제7권 제6호 pp.661-683, 2001. 12.