

# 주석 기반 검색 기법을 이용한 멀티미디어 데이터베이스 시스템

조경모\*

\*한국폴리텍IV 대전대학 멀티미디어과

e-mail:mo0519@naver.com

## A Multimedia Database System using Method of annotation-based retrieval

Kyung-Mo Cho\*

\*Daejeon College Multimedia Science of Korean PolytechnicIV

### 요 약

본 논문에서는 특징기반 검색을 이용하여 대용량의 비디오 데이터에 대한 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 자동화되고 통합된 비디오 의미기반 검색 시스템을 제안한다. 사용자의 기본적인 질의를 분석하고 질의에 의해 추출된 키 프레임의 이미지를 사용자가 선택함으로써 인덱싱 에이전트는 추출된 키 프레임의 주석에 대한 의미를 더욱 구체화시킨다. 또한, 사용자에게 의해 선택된 키 프레임은 특징기반 검색의 질의 이미지가 되고 인덱싱 에이전트는 제안하는 다중 분할 칼라 히스토그램 기법을 통해 질의 이미지와 데이터베이스의 키 프레임들을 비교한 후 가장 유사한 키 프레임 이미지를 검색하여 사용자에게 디스플레이한다. 제안하여 구현된 시스템은 현저히 향상된 성능을 보였다.

### 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[1,2].

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)[3]이 있으며, 둘째, 키 프레임에 대해 사용자의 주석을 입력하여 저장한 후 이러한 사용자의 주석을 비교 검색하는 주석기반 검색(annotation-based retrieval)[4]으로 크게 두 가지로 분류 할 수 있다. 하지만, 이 두 가지 비디오 데이터에 대한 내용기반 검색은 모두 단점을 가지고 있다. 특징기반 검색은 비디오 자체의 시각적 특징을 추출하여 유사성을 계산하여 비교 검

색하는 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.

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. 따라서, 본 논문에서는 사용자의 질의와 결과를 학습하여 비디오 데이터의 메타데이터를 지속적으로 자동 갱신하는 인덱싱 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 프

레이 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.

## 2.2. 기존 연구의 문제점

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 칼라 히스토그램과 대표 색상을 기반으로 장면과 샷의 경계 검출이 진행되어 왔지만 사용자의 의미기반 검색을 완전히 지원하지는 못한다.

따라서, 비디오 데이터의 특징기반 검색을 이용하여 에이전트에 의한 자동화된 인덱싱에 의한 데이터베이스 시스템의 구축이 필수적이라 할 수 있다.

## 3. 제안 시스템


### 3.1. 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 칼라 히스토그램 기법을 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. 디스플레이된 후보 키 프레임 리스트 중 사용자가 선택한 특정 키 프레임은 다중 분할 히

스토그램 비교법을 이용하여 특징기반 검색을 실시한 후 가장 유사한 키 프레임을 검출하여 사용자에게 2차적으로 디스플레이한다.

[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
	비동일 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 3과 같이 여러 개의 키 프레임들이 검색되었고 그 중 사용자가 키 프레임 ID가 'KF00324321'인 키 프레임 이미지를 선택했다면, 표 1과 같은 키워드 분류가 시작된다. 키워드 분류가 이루어지면 인덱싱 에이전트에 주석 자동 갱신을 위한 동일 키워드와 비동일 키워드의 가중치 계산이 표 2와 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 2에 나타나듯이 실 키워드인 '바다', '배', '여자' 중에 '바다'와 '배'의 의미 가중치는 제안하는 동일 키워드에 대한 의미 가중치 계산에 의해 값이 증가했으며, '연인'은 비동일 키워드에 대한 의미 가중치 계산에 의해 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.

### 3.2. 키 프레임 특징처리 기법

검색되어진 키 프레임을 다시 다중 분할 칼라 히스토그램 기법을 이용하여 2차적인 특징기반 검색을

수행하게 된다. 사용자로부터 입력된 질의 이미지와 비디오 데이터에서 추출된 키 프레임의 이미지에 대한 색상 유사도를 빠르고 정확하게 계산하기 위하여 질의 이미지와 키 프레임의 이미지를 3\*3 영역으로 분할한다. 즉, 분할된 하나의 이미지는 9개의 픽셀 그룹을 가지게 된다. 질의 이미지와 키 프레임의 이미지에 대한 분할과 유사도 비교를 위한 색상의 평균값 계산은 서로 동일하므로 본 논문에서는 키 프레임 이미지에 대한 이미지 분할과 색상 평균값 계산만을 언급한다.

키 프레임의 전체 이미지에 대한 R, G, B 색상의 평균값을 구하는 식은 다음과 같다.

$$A_{t\_image,red} = \frac{\sum_{i=0}^w \sum_{j=0}^h R_{[i,j]}}{C_t} \quad (1)$$

$R_{[i,j]}$ 은 전체 이미지에서  $i$ 행  $j$ 열에 있는 Red 값이며,  $C_t$ 는 전체 이미지에서 0이 아닌 Red 값을 가지는 픽셀의 개수이다. 식 3은 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 색상 정보를 가지고 있는 프레임을 찾기 위하여 식 2를 이용하여 색상의 차이값이 임계값 미만인 1차 유사 키 프레임 후보군을 추출한다. 질의 이미지와 키 프레임의 전체 이미지에 대한 Red의 색상 차는 다음과 같이 구한다.

$$D_{t\_red} = |A_{t\_image,red} - A_{t\_frame,red}| \quad (\text{단, } D_{red} < \alpha) \quad (2)$$

$A_{t\_image,red}$ 는 질의 이미지의 전체에 대한 Red의 평균값이고,  $A_{t\_frame,red}$ 는 키 프레임의 전체 이미지에 대한 Red의 평균값이며,  $\alpha$ 는 임계값이다. 임계값은 1차 칼라 히스토그램 기법의 결과로서 후보 프레임의 수를 결정하는 값이며 1~255중에 5로 입력했을 때 유사한 키 프레임들을 찾은 것으로 나타났다. Green과 Blue의 색상 차도 식 2와 동일한 방법으로 구한다.

1차 유사 키 프레임 후보군으로 선택된 이미지는 식 3를 이용하여 입력된 이미지와 유사도 계산을 하게 된다.

$$S_{total_{Q\_image,K\_frame}} = |D_{t\_red} + D_{t\_green} + D_{t\_blue}| \quad (3)$$

따라서, 식 2에서 질의 이미지와 키 프레임 전체 이미지의 각 R, G, B 색상 차의 절대값은 임계값  $\alpha$ 보다 작은 값을 1차 후보 프레임으로 선정되고 식 3에 의해서 색상 유사도의 값이 작을수록 입력되는 이미지와 더욱 유사한 이미지로 판단한다.

1차 후보군으로 선택된 키 프레임들은 색상의 분포가 비슷하지만 입력되어진 질의 이미지와 서로 전혀 유사하지 않은 이미지 일 수도 있다. 따라서 1차 후보군에 대하여 식 4와 같이 9개의 분할 그룹에 대한 R, G, B 평균값의 차를 이용하여 2차 키 프레임 후보군을 생성한다. 식 4는 질의 이미지의 1번 영역과 1차 후보군의 이미지들의 1번 영역 그룹의 Red 값 평균의 차를 구하는 식이다.

$$D_{g1\_red} = |A_{g1\_image,red} - A_{g1\_frame,red}| \quad (\text{단, } D_{red} < \alpha) \quad (4)$$

$A_{g1\_image,red}$ 은 질의 이미지에서 영역1의 Red에 대한 평균값이며,  $A_{g1\_frame,red}$ 은 1차 키 프레임 후보군에서 영역 1그룹의 Red에 대한 평균값이다.

상기한 방법으로 9개의 영역에 대한 R, G, B 값의 차를 구한다.

#### 4. 시스템 구현 및 실험 평가

##### 4.1. 이미지 질의 인터페이스

특징기반 검색을 수행하게 되면 질의 이미지에 대한 다중 분할 칼라 히스토그램 기법에 의해 유사도 계산을 하여 가장 유사한 키 프레임순으로 리스트가 다시 디스플레이된다.

##### 4.2. 실험 평가

제안하는 시스템의 실험 평가를 위한 비디오 데이터 도메인은 MPEG 포맷인 일반 국내외 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 프레임을 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.

제안하는 시스템의 검색 적합성[10]을 평가하기 위하여 식 5의 정확율(Precision)에 의한 성능평가 척도를 이용하여 실험을 하였다. 기존 연구에서의

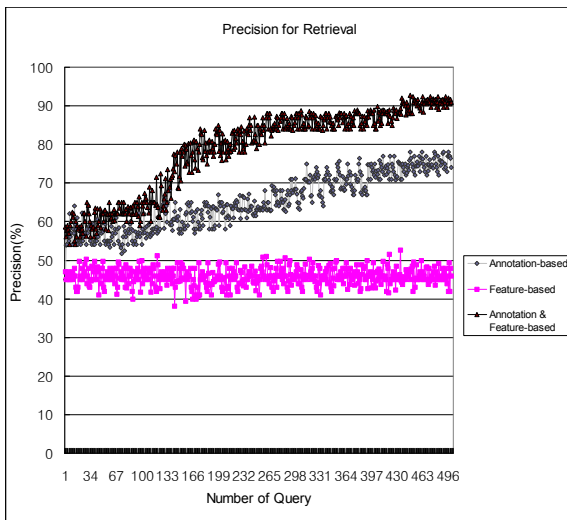
비디오 검색 시스템의 일치 검색을 기반으로 하는 검색의 적합성은 정확율로 검색의 성능을 측정하고 있다[11].

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

$C_{Total, right}$ 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.

우선, 시스템의 특징기반 검색을 혼합한 검색을 통한 재현율과 정확도를 계산하기 이전에 키 프레임의 질의 이미지를 이용하여 주석기반 검색을 배제한 특징기반 검색을 각 분할 그룹별로 실험해 보았다.

제안하는 시스템의 전체적인 검색에 대한 정확도를 계산하기 위해 500회 이상의 사용자 질의를 실행하여 제안한 시스템의 검색에 대한 실험을 하였다. 그림 1은 시스템의 검색에 대한 정확도를 나타낸 것이다.



[그림 1] 제안 시스템의 검색 정확도

### 5. 결론

본 논문에서는 대용량의 비디오 데이터에 대한 특징기반 검색을 이용하여 사용자의 다양한 의미검색을 지원하는 멀티미디어 데이터베이스 시스템을 제안하였다. 사용자의 1차적인 키워드 질의를 에이전트가 자동으로 학습하고 키워드 분류와 의미 가중치 값 계산에 의해 의미 있는 주석으로 자동 등록 및 갱신된다. 또한, 키워드 검색 후 사용자에게 선택된 키 프레임은 제안하는 다중 분할 칼라 히스토그램

기법의 유사도 측정을 통해 기존의 기법보다 시스템 부하를 줄이고 검색 시간을 단축하여 가장 유사한 키 프레임을 추출한다.

실험 결과에서 나타나듯이 제안한 시스템을 통해 비디오 데이터의 의미기반 장면 검색의 정확도를 높일 수 있었으며, 사용자의 질의어 시험결과 450회 질의부터 90% 이상의 정확도를 보였으며, 500회 질의까지 최고(최고 92.7%)의 정확도를 보였다.

### 참고문헌

- [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] D. Shasha and T.L. Wang, "New Techniques for Best-match Retrieval," *ACM TOIS*, Vol. 8, No. 2, pp.140-158, 1990.
- [11] G. Salton and M. J. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.