

키 프레임의 주석과 비교 영역 학습을 이용한 비디오 검색 시스템

이기성*

*호원대학교 컴퓨터학부

e-mail:ygslee@sunny.howon.ac.kr

A Video Retrieval System Using Annotation and Comparison Area Learning of Key-frames

Gi-Sung Lee*

*Dept of Computer Science, Howon University

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. 본 논문에서는 사용자의 키워드 학습과 비교 영역 학습을 이용하여 대용량의 비디오 데이터에 대한 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 자동화된 비디오 검색 시스템을 제안한다.

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 의미기반 검색이 가능하도록 하는 것이 매우 중요하다[2].

현재 비디오 데이터에 대한 내용기반 검색에 대한 연구는 크게 두 가지로 분류할 수 있다. 첫째, 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 분류 할 수 있다. 하지만, 이 두 가지 비디오 데이터에 대한 내용기반 검색은 모두 단점을 가지고 있다.

따라서, 본 논문에서는 비디오 데이터의 키 프레임

에 대한 주석 정보와 비교 영역 학습을 수행하는 인덱싱 에이전트를 이용한 내용기반 비디오 검색 시스템을 제안한다.

2. 관련 연구

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 VisualSEEK는 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 Venus는 비디오의 각 프레임에 나타나는 객체들의 시간관계, 공간관계를

메타데이터로 구축하여 이미지 검색 처리에 이용하였다.

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

3. 의미기반 비디오 검색 시스템

3.1 키워드 학습에 의한 자동 주석 갱신 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 1은 자동 주석처리를 하기 위한 전처리인 키워드 분류에 대해 나타내고 있다.

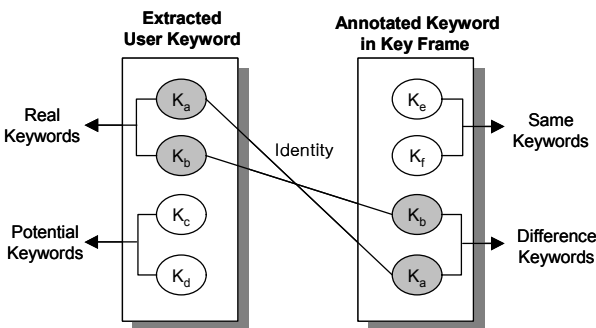


그림 1. 키워드 분류

3.2 키 프레임 특징처리 기법

인덱싱 에이전트는 사용자의 최종 키 프레임 선택에 의한 비교영역을 학습한다. 사용자의 초기 질의에 의해 주석기반 검색이 수행된 후 그 결과인 키 프레임 리스트에서 사용자의 키 프레임 이미지가 이루어지면 데이터베이스 안에 학습되어 저장된 해당 키 프레임 이미지의 멀티분할 색상 히스토그램의 히스토리 값을 이용하여 비교 영역을 선택한다.

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 9개의 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.

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

해당 비교 영역의 색상 차이 가중치는 부여받은 순위 값에 1과의 차를 구하여 이를 비교영역인 멀티 분할된 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.

$$A_{W_new[n]} = Avg[A_{W_old[n]} + W_{diff_color[n]}] \quad (2)$$

(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.

4. 실험 평가

4.1 실험 환경

제안하는 시스템의 실험 평가를 위한 비디오 데이터 도메인은 MPEG 포맷인 약 500600Mbyte 용량의 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. 각각의 비디오 데이터의 최적 비교검색 영역을 추출하기 위해 영화의 각 장르별로 비디오 데이터를 분류한 후 검색 영역을 선택하여 비교 검색해 보았다.

4.2 실험 내용

제안하는 시스템의 검색 적합성을 평가하기 위하여 식 3의 재현율(Recall)과 (식4)의 정확도(Precision)에 의한 성능평가 척도를 이용하여 실험을 하였다.

$$Recall(\%) = \frac{C_{Search, right}}{C_{Total, right}} \times 100 \quad (3)$$

$$Precision(\%) = \frac{C_{Search, right}}{C_{Search, total}} \times 100 \quad (4)$$

$C_{Total, right}$ 합한 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.

4.3 실험 결과

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 2는 시스템의 검색에 대한 재현율을 나타낸 것이다.

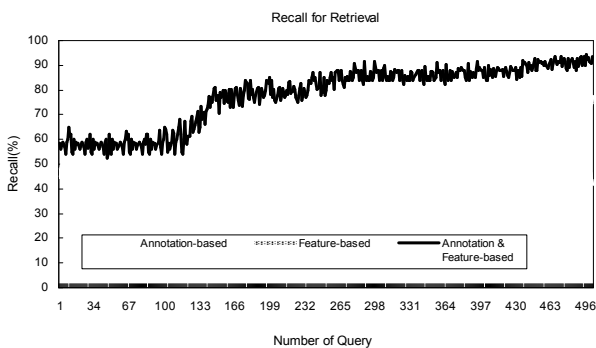


그림 2. 제안 시스템의 검색 재현율

5. 결론

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

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] 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.