

자동 주석 갱신 기법을 이용한 멀티미디어 데이터베이스 시스템

이근왕*, 오택환*, 조경모**
*청운대학교 멀티미디어학과
**한국폴리텍IV 대전대학 멀티미디어과
e-mail:kwlee@chungwoon.ac.kr

A Multimedia Database System using Method of Automatic Annotation Update

Keun-Wang Lee*, Taek-Hwan Oh*, Kyung-Mo Cho**
*Dept of Multimedia Science, Chungwoon University
**Daejeon College Multimedia Science of Korean PolytechnicIV

요 약

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 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.

1. 서론

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[1].

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 특

징기반 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 분류할 수 있다. 하지만, 이 두 가지 비디오 데이터에 대한 내용기반 검색은 모두 단점을 가지고 있다. 주석기반 검색은 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.

2. 관련연구

2.1 주석기반 비디오 검색 기법

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.

2.2 기존 연구의 문제점

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.

따라서, 비디오 데이터의 주석기반 검색을 병행하여 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 분석한다. 그림 1은 제안하는 시스템의 전체 구조를 나타낸다.

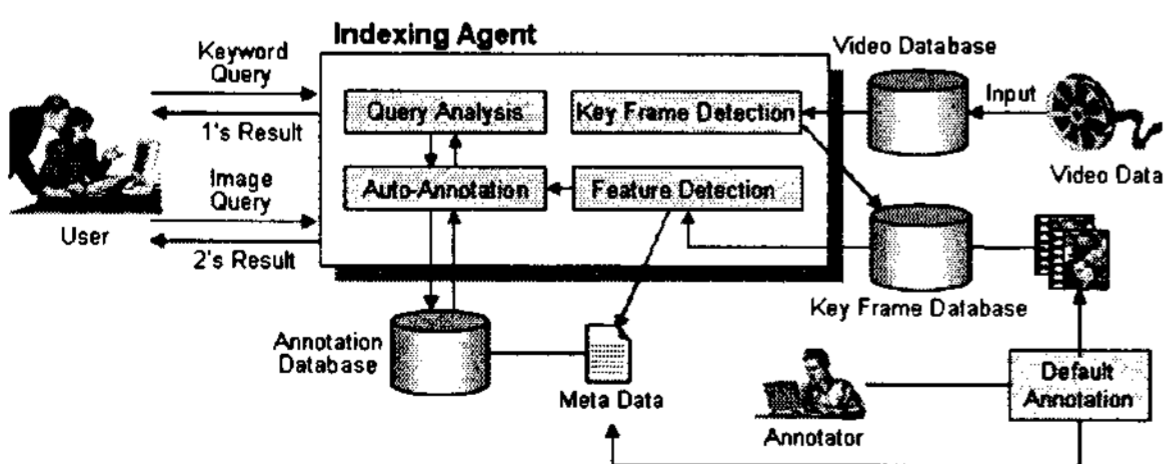


그림 1. 시스템 구조

추출된 키 프레임은 주석자에 의해 가장 기본적인 하나의 단어로 주석이 입력되고 각 키 프레임과 기

본적인 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.

3.2 자동 주석처리 기법

단어 한개 이상으로 구성된 사용자 질의가 입력되면 키워드가 추출되고, 추출된 사용자 키워드에 의해 해당 키워드가 포함된 키 프레임들을 검색한다. 추출된 사용자 키워드는 그림 2와 같이 실 키워드와 잠재적 키워드로 분류된 후, 키 프레임의 주석 정보와 비교하여 주석 정보에 있는 키워드에서 정확히 매칭된 키워드는 동일 키워드로 정확히 매칭되지 않은 키워드는 비동일 키워드로 분류한다.

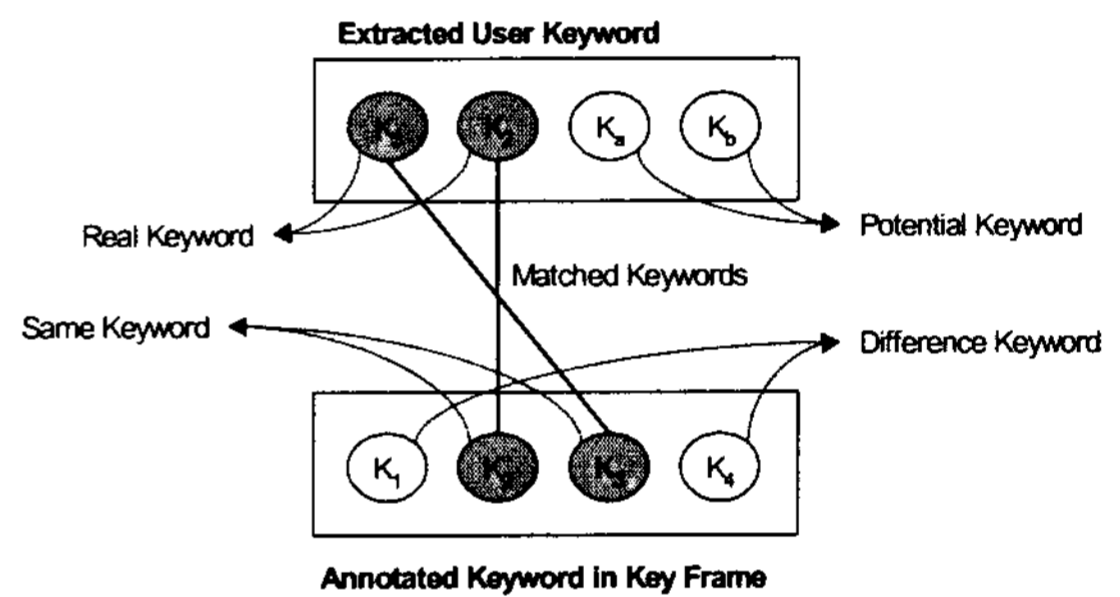


그림 2. 키워드 분류

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 값을 계산한다.

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

$$W_{Keyword_new} = W_{Keyword_old} + \frac{1}{N_{Kframe_SK}} \quad (1)$$

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

또한, 키 프레임에서의 주석 키워드가 비동일 키

워드(difference keyword)일 경우 새로운 의미 가중치는 다음과 같이 계산한다.

$$W_{Keyword_new} = W_{Keyword_old} - \frac{1}{N_{Kframe_SK}} \quad (2)$$

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 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.

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 식 1과 식 2에 의해 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.

4. 시스템 구현 및 실험 평가

4.1 시스템 구현

제안하는 시스템은 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.

사용자는 주석 기반 검색에 의해 몇 개의 검색어를 입력한다. 사용자에게 의해 입력된 키워드는 DB에 저장되어 있는 키 프레임의 주석과 비교하여 매칭되는 키 프레임을 찾아 사용자에게 디스플레이한다. 디스플레이 된 키 프레임 리스트 중에 사용자가 특정 키 프레임을 선택하게 되면 그 키 프레임은 의미 가중치 계산에 의해 새로운 가중치 값을 갖게 된다. 1차적인 주석기반 검색이 완료되면 사용자에게 결과로 키 프레임의 리스트가 디스플레이되고 사용자가 원하는 이미지를 선택한 후 특징 기반 검색을 수행할 수 있다.

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 3의 재현율(Recall)에 의한 성능평가 척도

를 이용하여 실험을 하였다. 기존 연구에서의 비디오 검색 시스템의 일치 검색을 기반으로 하는 검색의 적합성은 재현율로 검색의 성능을 측정하고 있다 [11].

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

$C_{Total, right}$ 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 주석기반 검색을 혼합한 검색을 통한 재현율을 계산하기 이전에 키 프레임의 질의 이미지를 이용하여 주석기반 검색을 각 분할 그룹별로 실험해 보았다. 표 3은 각 영역을 그룹화한 후 각 그룹별로 질의 이미지를 이용한 특징기반 검색만을 실행하였을 때 재현율을 나타낸 것이다.

표 3. 검색 영역별 특징기반 검색의 재현율

영역	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	50.23	34.25	41.25	35.35	43.05	30.41	32.50	40.61	51.25	45.35
2	51.85	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	30.05	36.82	35.65	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	45.29	40.31	42.39	39.44	46.25	41.41	45.29	40.31
7	34.82	44.65	42.82	42.25	42.22	41.87	41.05	41.65	42.82	35.25
8	30.23	35.21	31.98	49.31	38.07	39.61	35.45	32.11	38.98	30.31
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	40.29	45.11	46.15	45.27	42.24	40.87
평균	40.69	40.06	41.34	41.52	41.45	41.54	41.67	42.80	43.54	42.23

표 3의 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 장면에서 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 재현

율이 높게 나타남을 보였다.

5. 결론

본 논문에서는 대용량의 비디오 데이터에 대한 주석기반 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.

실험 결과에서 나타나듯이 제안한 시스템을 통해 비디오 데이터의 의미기반 장면 검색의 정확도를 높일 수 있었다.

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 있다.

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.