

대용량 멀티미디어 데이터의 효율적인 검색엔진 설계

이광형*, 민소연*

*서일대학 인터넷정보과

*서일대학 정보통신과

요 약

본 논문에서는 대용량 멀티미디어 데이터에 대한 사용자의 다양한 의미검색을 지원하는 비디오 검색 시스템의 설계를 제안한다. 제안하는 시스템은 주석기반검색과 특징기반 검색을 각각의 에이전트를 통하여 자동으로 처리하였다. 먼저 주석기반검색은 사용자의 검색어를 입력하게 되면 가중치를 적용하여 의미를 더욱 구체화 하여 오류율을 최소화 하였으며, 특징기반검색은 주석기반검색에서 선택된 키프레임에 의해 데이터베이스의 영상들과 유사도를 검사하여 검색하였다. 시스템의 구현결과 기본시스템보다 0.5%의 재현율의 향상과 97.8%의 정확률을 나타내었다.

1. 서론

최근 멀티미디어 정보의 양이 매우 빠른 속도로 증가함에 따라 비디오 데이터에 대한 다양한 검색은 매우 중요한 이슈가 되고 있다. 비디오 데이터를 효율적으로 처리하기 위해서는 비디오 데이터가 가지고 있는 의미 있는 내용을 추출할 수 있는 의미기반 검색 기법이 요구된다.

지금까지 연구되어온 멀티미디어 데이터의 검색은 주석기반검색, 특징기반검색, 주석과 특징기반검색의 통합검색 시스템이 있다. 이러한 시스템들은 검색데이터의 생성을 위해 관리자의 많은 노력과 시간을 요구하고 특징추출을 위한 복잡한 계산을 요구하며, 생성된 데이터는 변화되지 않는 정적인 검색을 수행하는 단점이 있다.

본 논문에서는 대용량의 비디오 데이터에 대한 사용자의 다양한 의미검색을 지원하는 의미기반 비디오 검색 시스템을 제안한다. 제안하는 시스템은 주석기반 검색과 특징기반 검색을 통합한 에이전트를 이용하여 비디오 데이터의 내용 정보 추출 및 검색 과정을 자동으로 처리한다.

우선, 제안하는 인덱싱 에이전트는 사용자의 기본적인 질의를 분석하여 실 키워드를 추출한 후 의미 가중치와 의존 가중치 계산을 이용하여 키 프레임의 주석에 대한 의미를 더욱 구체화시킨다.

또한, 주석기반 검색에서 사용자에 의해 선택된 키 프레임은 특징기반 검색의 질의 이미지가 되고 에이전트는 제안하는 전체영역 및 중심영역 색상분포와

이진영상 히스토그램 기법을 통해 질의 이미지와 데이터베이스의 키 프레임들을 비교한 후 가장 유사한 키 프레임 이미지를 검색하여 사용자에게 디스플레이 한다. 따라서, 사용자는 비디오 데이터에 대한 검색을 반복할수록 에이전트의 학습에 의해 더욱 정확한 검색 결과를 얻을 수 있다.

설계하고 구현한 시스템은 실험을 통한 성능 평가에서 비디오 데이터의 장면 검색에 대하여 기존 시스템 보다 1.5%의 재현율의 증가를 보였고, 정확률은 94.2%로 약 2.0%의 증가를 보였다.

2. 기존연구

2.1 비디오데이터의 특징

비디오 데이터의 한 장면(scene)에서 검출된 컷(cut)과 컷 사이의 샷(shot)은 색상, 텍스처(texture), 모양, 경계(edge)등으로 표현을 하게 된다. 이러한 비디오 데이터의 키 프레임 이미지에서 각 표현방법을 이용하여 색상 정보와 모양 정보를 사용하여 비디오 데이터의 특징과 의미 정보를 추출한다.

2.2 색상정보의 표현

색상을 표현하는 칼라모델은 사용되는 용도에 따라 다양하다. 가장 일반적으로 사용되는 하드웨어 기준 모델들은 칼라 모니터와 많은 종류의 색채 비디오 카메라를 위한 RGB(Red, Green, Blue)모델, 칼라 프린터를 위한 CMY(Cyan, Magenta, Yellow)모델, 칼라 텔레비전 방송의 표준인 YIQ(Luminosity, In-phase, Quadrature)모델, 이미지 처리를 위해 가

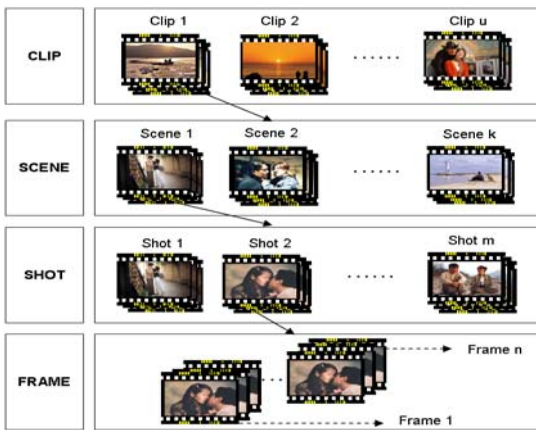
장 많이 사용되는 HSI(Hue, Saturation, Intensity)모델, HSV(Hue, Saturation, Value)모델 등이 있다[1].

2.3 비디오 인덱싱

비디오 인덱싱은 대용량의 비정형적인 비디오 데이터를 보다 계층적이고 구조적인 형태로 재구성하기 위한 기술이다. 비정형적인 비디오 데이터를 물리적인 샷으로 나누고, 각 샷을 가장 잘 나타낼 수 있는 대표적인 프레임 즉, 키 프레임을 추출하여 각 샷들 간의 유사성을 통하여 의미적으로 유사한 샷들을 하나로 묶어 장면을 구성함으로써 사용자에게 보다 편리한 브라우징 환경을 제공할 수 있다.

일반적으로 비디오 데이터는 프레임(frame), 샷(shot), 장면(scene), 컷(cut) 그리고 클립(clip)등의 요소로 구성되어 있다. 비디오 데이터의 구성 요소들과 그들의 관계는 [그림 1]과 같다.

프레임이란 비디오 데이터를 구성하는 최소 단위로서 필름 한 장에 해당하는 하나의 영상을 표현하는 용어이다. 샷은 하나의 카메라로 촬영한 연속적인 영상들을 표현한 것이며, 한 샷 내에서는 영상이 끊어지지 않고 연결되어 있으며, 비디오 데이터 분할의 기본 단위로 사용된다.

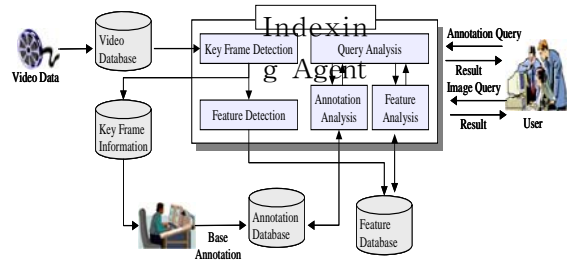


[그림 1] 비디오 데이터의 구성

3. 제안시스템

3.1 제안시스템 개요

비디오 데이터가 입력되면 시스템은 컷 검출(cut detection)을 통하여 비디오 데이터의 각 장면에서 컷을 검출하고 검출된 컷에서 키 프레임을 추출하여 각 프레임에 대해 이진영상변환과 중심영역과 전체영역에 대한 특징을 분석하는 시스템으로 제안하는 시스템의 전체 구조는 [그림 2]와 같다.



[그림 2] 전체 시스템 구조

입력된 비디오에서 추출된 키 프레임은 특징추출기에 의해 전체영역의 특징과 중심영역의 특징을 분석하여 특징데이터베이스에 저장된다. 또한 전체영역을 이진영상으로 변환하여 이진영상의 특징을 추출한 후 특징데이터베이스에 저장된다. 주석데이터베이스에는 추출된 키 프레임의 기본적인 주석이 주석자에 의해 하나의 단어로 입력되어 진다. 사용자로부터 검색하고자 하는 비디오의 주석이 입력되면 사용자 질의에서 필요한 키워드를 추출하여 주석데이터베이스에 있는 키 프레임들의 주석 정보와 유사도 비교를 통하여 검색대상을 추출한다. 검색대상 키 프레임은 사용자로부터 입력된 주석간의 의존가중치에 의해 정렬을 하게 되고 유사한 후보 키 프레임 리스트를 생성하여 사용자에게 1차적으로 디스플레이 한다. 사용자에게 디스플레이된 후보 키 프레임 리스트는 이미지 형태로 보여지게 되고 키 프레임 리스트에서 사용자가 선택한 특정 키 프레임 이미지는 전체영역과 중심영역의 특징정보에 의하여 특징기반 검색을 실시한 후 가장 유사한 키 프레임을 검출하게 된다. 검출된 키 프레임에 대해 이진영상변환과 가로/세로 히스토그램의 이진영상 특징정보를 이용하여 순서를 정렬한 후 사용자에게 2차적으로 디스플레이 한다.

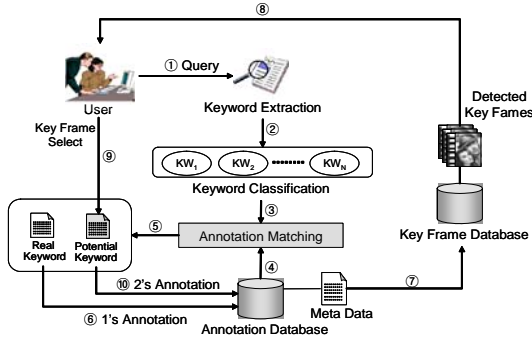
3.2 주석기반 검색구조

에이전트는 사용자로부터 비디오 내용 검색에 대한 질의를 받게 되면 질의어 중 키워드를 추출하여 주석데이터베이스의 주석 정보와 매칭시켜 동일한 키워드를 주석 정보로 가지고 있는 키 프레임을 검출하여 사용자에게 보내어준다.

우선 사용자의 질의가 입력되면 사용자 질의어를 분석하여 키워드를 추출한 후 추출된 사용자 키워드를 주석데이터베이스 안에 있는 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 매칭 되지 않는 키워드들은 잠재적 키워드로 정의한다. 잠재적 키워드로 정의된 키워드는 차후에 다른 사용자의 질의 시 매칭 여부에 의해 실 키워드로 등록 될 수 있다.

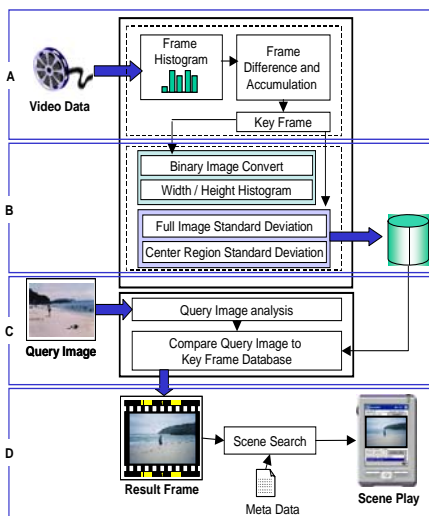
따라서 사용자의 질의는 모두 에이전트에 의해 학습 되고 학습된 사용자의 질의로 인해 주석 정보가 자동으로 갱신된다. 에이전트는 정확히 키워드로 매칭 된 주석 정보를 가지고 있는 키 프레임들을 추출하여 유사성이 높은 순으로 사용자에게 키 프레임의 이미지를 리스트 형태로 디스플레이 한다. 에이전트에 의한 자동 주석처리의 구조는 [그림 2]와 같다.



[그림 2] 에이전트의 자동 주석처리

3.3 특징기반 검색구조

사용자의 입력된 주석에 의해 주석기반 검색을 수행하여 사용자에게 제시된 키 프레임 리스트에서 사용자로부터 선택된 키 프레임은 특징기반 검색을 위한 질의이미지가 된다.



[그림 3] 키 프레임의 특징 검출

질의이미지는 제안하는 전체영역 색상분포 특징과 중심영역 색상분포 특징, 이진영상변환 가로/세로

히스토그램의 특징정보에 의해 특징데이터베이스에 있는 모든 비디오 데이터의 키 프레임들의 특징값과 비교하여 유사도를 계산한 후 유사성이 높은 키 프레임부터 낮은 키 프레임 순으로 사용자에게 키 프레임을 디스플레이한다. 제안하는 프레임의 특징기반 검색에 대한 구조는 [그림 3]과 같다.

4. 성능평가

본 논문에서 제안한 의미기반 인덱싱 에이전트 시스템을 다양한 질의를 통해 성능을 평가해 보았다. 제안하는 시스템의 전체 검색에 대한 재현율과 정확률이 [그림 5-11]과 [그림 5-12]에서와 같이 높게 나타났으며, 질의어가 증가할수록 질의어 학습에 의한 주석정보의 자동 갱신을 하는 학습으로 인해 검색에 대한 재현율과 정확률이 계속 증가하는 것을 알 수 있다. 1차 주석기반 검색과 2차 특징기반 검색을 모두 수행한 결과 재현율은 최대 95.7%까지 증가했으며, 정확률은 최대 94.2%까지 증가하였다. 재현율은 값의 분포가 다소 넓게 퍼질 정도로 데이터 및 질의에 따라 그 결과값이 일정치 않게 나타났다. 하지만, 정확도는 값의 분포가 다소 고르게 밀집 분포되어 데이터 및 질의에 상관없이 그 결과값이 일정하게 나타난 것을 확인하였다. 제안하는 시스템의 처리와 특징에 대해 기존 시스템과의 비교는 [표 1]과 같다.

[표 1] 기존 시스템과 제안 시스템의 처리기능 비교

시스템	항목	주석기반	특징기반	의미기반	인덱싱 자동화	비디오 처리	예외 처리
OVID		O	X	X	X	X	X
VideoSTAR		O	X	X	X	X	X
Algebraic Video		O	X	X	X	X	X
Venus		X	O	O	X	X	X
VideoQ		O	O	X	X	X	X
IVIS		O	O	X	X	X	X
TIVRON		O	O	X	X	X	X
제안 시스템 (SSU)		O	O	O	O	O	O

[표 1]과 같이 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 대용량 비디오 처리를

지원하며 사용자의 질의시 키워드의 오류 입력을 최소화하고 더욱 의미있고 구체적인 키워드로 질의를 할 수 있는 키워드 예외처리를 가능하게 하였다.본 논문에서 제안한 주석/특징기반 비디오검색 시스템의 응용 분야는 다음과 같다.

- 방대한 멀티미디어 데이터를 저장 및 관리할 수 있는 멀티미디어 데이터베이스 관리 시스템에 적용하여 운용자 및 관리자의 주석 처리 및 인덱싱 등 수많은 작업들을 에이전트가 대행해 주어 업무의 작업량을 현저히 줄일 수 있다.
- 영화를 제작하는 국내외 영화사나 제작된 영화를 보관하는 영화 박물관 및 각종 방송용 프로그램을 제작하는 방송국에서 수많은 비디오 데이터를 인덱싱하여 보관하고 검색하는데 용이하다.
- 인터넷이나 모바일로 영화 및 뮤직 비디오, CF 등 동영상 서비스를 하는 각종 업체들의 비디오 데이터를 효율적으로 관리하여 검색 및 서비스를 용이하게 할 수 있다.

5. 결론

본 논문에서는 대용량의 비디오 데이터에 대한 주석기반 검색과 특징기반 검색을 이용하여 사용자의 다양한 의미검색을 지원하는 에이전트 기반에서의 개선된 내용기반 비디오 검색 시스템을 제안하였다. 또한, 설계에 입각하여 시스템을 구현하였으며, 성능평가를 위해 시스템의 검색에 대한 재현율과 정확률을 측정하였다. 본 논문에서 제안하는 시스템은 비디오 장면 검색시 사용자의 질의가 입력되면 사용자의 질의에 대한 키워드를 분류한 후 의미가중치와 의존가중치를 계산하여 데이터베이스의 키 프레임에 대한 주석과 비교하여 최적의 키 프레임들을 검출한 후 이를 이미지화하여 사용자에게 디스플레이 한다. 따라서 인덱싱 에이전트가 사용자들의 질의를 지속적으로 분석하고 학습하여 키 프레임에 대한 주석을 지속적으로 갱신한다. 또한, 주석기반 검색에 의해 디스플레이된 키 프레임 이미지 중에 사용자에게 의해 선택된 특정 이미지는 중심영역 색상분포와 이진영상의 히스토그램(histogram) 비교법을 이용하여 선택된 키 프레임 이미지와 가장 유사한 키 프레임을 가지고 있는 비디오 데이터의 장면을 데이터베이스에서 검색하여 그 결과를 사용자에게 디스플레이 한다. 평가 결과와 같이 구현된 시스템으로 검색에 대한 실험을 하여 비디오 데이터의 의미기반 장면 검색의 기존 시스템에 비해 정확률을 더욱 높일 수 있

었다. 본 논문에서 제안한 개선된 의미기반 비디오 검색 시스템의 장점은 다음과 같다. 첫째, 제안하는 시스템은 기존의 비디오 검색에서 도메인으로 사용했던 광고, 테마 뉴스 등 저용량 비디오 파일이 아닌 대용량 영화 비디오 파일을 도메인으로 하여 시스템에 입력된 모든 영화에 대해 완전한 키 프레임에 대한 주석 정보와 특징 정보를 추출하여 인덱싱을 처리하였다. 둘째, 기존의 주석기반 비디오 검색 시스템에서는 주석자의 일관적이고 주관적인 주석 입력을 하여 사용자의 검색을 위한 장면의 주석이 정확히 일치하지 않으면 검색하지 못하였으나, 제안하는 시스템은 사용자의 질의를 이용하여 인덱싱 에이전트가 지속적으로 사용자의 질의에 대한 의미를 자동으로 주석을 처리 하여 키 프레임의 주석을 자동으로 갱신한다. 셋째, 개선된 의미가중치 계산을 이용하여 사용자의 질의와 1차적인 주석기반 검색 결과에서의 사용자 선택에 의해 각 주석의 의미가중치 값을 재계산하여 키 프레임의 주석에 대한 의미를 더욱 구체화시켰다. 넷째, 제안하는 의존가중치 계산을 이용하여 주석기반 검색시 사용자의 키워드 입력에 대한 오류를 최소화시켜 더욱 의미가 정확하고 구체적인 키워드로 질의가 될 수 있도록 하였다. 다섯째, 특징 기반 검색에 있어서 키 프레임 이미지의 주 객체 요소가 많이 분포되어 있는 전체영역과 중심영역에 대한 색상분포를 계산하여 색상 비교 검색에 있어서 유사도 계산량을 줄일 수 있었다.

[참고문헌]

- [1] Rafael C. Gonzalez, and Richard E. Woods, "Digital Image Processing(SE)" Addison-Wesley, 1995.
- [2] 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.4, pp.446-458, 2002.
- [3] 이동근, 색상 및 영역 특징기반 이미지 검색 시스템, *숭실대학교 석사학위 논문*, 1999.
- [4] 이종희, 멀티미디어 데이터베이스를 위한 의미기반 인덱싱 에이전트 시스템, *숭실대학교 박사학위 논문*, 2003.
- [5] 이태동, 김민구, "통합된 비디오 인덱싱 방법을 이용한 내용기반 비디오 데이터베이스의 설계 및 구현", *한국정보과학회 논문지(C)*, 제7권, 제6호, pp.661-683, 2001.