

주석정보와 특징정보를 이용한 비디오데이터 검색 시스템

이근왕*

Video Data Retrieval System using Annotation and Feature Information

Keun-Wang Lee^{1*}

요약 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. 설계하고 구현한 시스템은 실험을 통한 성능 평가에서 비디오 데이터의 장면 검색에 대하여 기존 시스템 보다 재현율과 정확률의 증가를 보였다.

Abstract In this thesis, we propose a semantics-based video retrieval system which supports semantics-retrieval for various users of massive video data. Proposed system automatically processes the extraction of contents information which video data has and retrieval process using agent which integrate annotation-based retrieval and feature-based retrieval. From experiment, the designed and implemented system shows increase of recall rate and precision rate for video data scene retrieval in performance assessment.

Key words : 비디오 데이터 검색, 주석기반 검색, 특징기반 검색, 에이전트, Video data retrieval, Annotation-based retrieval, Feature-based retrieval, Agent

1. 서론

비디오 데이터의 효율적인 유지와 관리를 위해서는 사용자가 원하는 데이터를 빠르고 정확하게 검색할 수 있도록 비디오 데이터의 정보를 체계적으로 분류하고 통합하는 기술이 필요하다. 비디오 데이터를 체계적으로 분류하고 통합하기 위해서는 비디오 데이터가 처음 제작될 때 생성되는 제작 지식정보뿐만 아니라 비디오 데이터 자체를 다양한 방법으로 분석한 후 비디오 데이터가 가지고 있는 특징들을 추출하여 체계적으로 분류하는 것이 무엇보다도 중요하다.

현재 비디오 데이터에 대한 내용기반 검색에 대한 연구는 크게 두 가지로 분류할 수 있다. 첫 째, 키 프레임에서의 특징을 추출하여 그 특징의 유사성을 이용한 특징기반 검색(feature-based retrieval)이 있으며 둘째, 주석자가 키 프레임에 대한 내용을 요약한

주석을 입력하여 저장한 후, 사용자의 질의와 주석자의 주석을 비교 검색하는 주석기반 검색(annotation-based retrieval)이 있다[1].

본 논문에서는 이러한 1차 주석기반검색을 수행한 후에 2차적으로 특징기반 검색을 수행하여 검색의 효율성을 높였다. 또한 검색에 사용된 키워드는 에이전트에 의해 학습하여 검색의 결과를 이후의 검색에 활용하여 불필요한 데이터를 검색으로부터 줄여주었다.

2. 관련연구

2.1 주석기반 비디오 검색 시스템

주석기반 비디오 검색 시스템은 일본 고베 대학에서 개발된 OVID(Object-oriented Video Information Data System)[2], MIT 대학에서 개발된 Algebraic Video 시스템[3], 노르웨이 공대에서 개발된 멀티미디어 데이터베이스 시스템(VideoSTAR) [4], 미국 메릴랜드 대학에서 개발된 비디오 데이터베이스 시스템AVIS(Advanced Video Information System)[5], 서울대학교에서 개발된 주석기반 비디오

이 논문은 2006년 청운대학교 교내연구비의 지원에 의하여 연구되었음

¹청운대학교 멀티미디어학과

*교신저자: 이근왕(kwlee@chungwoon.ac.kr)

검색 시스템인 VIRON(Video Information Retrieval On Notation) 등이 있다.

2.2 특징기반 비디오 검색 시스템

특징기반 비디오 검색 시스템은 대만 Tsing Hua 대학에서 개발된 비디오 데이터베이스 시스템 Venus[6], CMU(Carnegie Mellon University)에서 개발한 Informedia[7], IBM Almaden 연구소에서 개발한 QBIC(Query by Image and Video Content)[8], 콜롬비아 대학에서 개발된 Visual SEEK[9] 등이 있다.

2.3 주석 및 특징기반 비디오 검색 시스템

주석 및 특징기반 비디오 검색 시스템은 미국 콜롬비아 대학에서 개발된 VideoQ[10], 서울대학교에서 개발한 TIVRON(The Integrated Video Retrieval System based on OODBMS), 숭실 대학교에서 개발한 의미기반 인덱싱 에이전트 시스템 등이 있으며, 사용자의 질의에 의해 자동으로 주석이 갱신되고 색상 히스토그램 기법으로 비교 영역을 학습하지만 DB의 전체 키 프레임에 계산 대상으로 하여 부정확하고 객관적이지 못하다.

3. 제안 시스템 설계

3.1 제안하는 비디오 검색 시스템의 구조

제안하는 비디오 검색 시스템은 비디오 데이터가 입력되면 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 전체 구조는 그림 1과 같다.

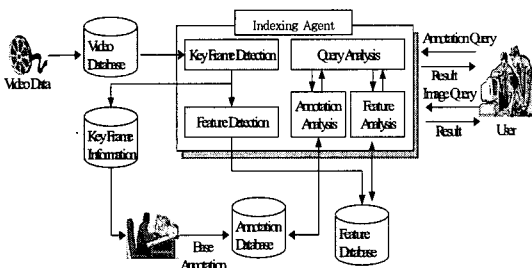


그림 1. 전체 시스템 구조

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 키워드가 입력되면 사용자 질의에 의해 주석데이터베이스에서 매칭이 되는 프레임들이 추출되고, 주석데이터베이스의 주석 정보와 사용자질의의 키워드와 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 보여주게 되고 키 프레임 리스트에서 사용자가 선택한 특정 키 프레임 이미지는 전체영역과 중심영역의 색상분포에 의한 특징을 왜도와 첨도에 의해 계산한 후 특징기반 검색을 실시하여 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 자동 주석처리 기법

비디오 키 프레임에 대한 주석의 입력은 주석입력자의 많은 수고와 시간을 요구 한다. 사용자의 입력 키워드와 선택된 키 프레임의 주석정보를 학습함으로 자동으로 주석을 관리할 수 있는 방법이 필요하다.

3.2.1 키워드 분류

비디오 데이터에서 찾고자 하는 장면을 검색하기 위해 시스템에 단어 한 개 이상으로 구성된 사용자 질의가 입력되면 이전 사용자들의 입력에 의한 주석 중 키워드가 추출되고, 추출된 사용자 키워드가 포함된 키 프레임들을 데이터베이스에서 검색한다.

자동 주석처리를 하기 위한 전처리인 키워드 분류는 그림 2와 같다. 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 상이한 키워드로 분류하며, 그림 2에서 각 키워드들의 정의는 표 1과 같다.

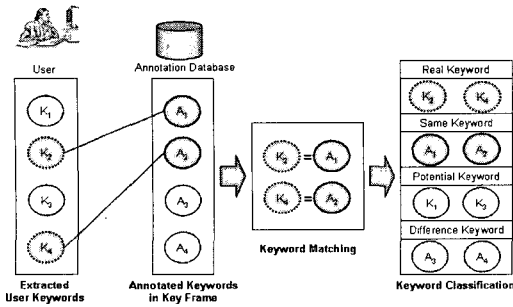


그림 2. 키워드 분류

표 1. 사용자 키워드 정의

키워드	키워드 분류	의미
사용자 키워드	실 키워드	사용자가 질의를 위해 질의 입력창에 입력하는 키워드
	잠재적 키워드	사용자 키워드 중 DB에 있는 주식과 매칭되지 않은 키워드
주식 키워드	동일 키워드	DB에 있는 주식 중 사용자 키워드와 매칭된 키워드
	상이한 키워드	DB에 있는 주식 중 사용자 키워드와 매칭되지 않은 키워드

3.2.2 의미가중치

에이전트는 동일 키워드를 포함하고 있는 키 프레임들을 추출하여 사용자에게 키 프레임 리스트를 디스플레이하고, 키 프레임 리스트 중 사용자가 특정 키 프레임을 선택하게 되면 선택된 키 프레임이 가지고 있는 각 키워드들에 대한 의미가중치 계산을 하게 된다.

선택된 키 프레임의 키워드에 대한 의미가중치를 계산하는 알고리즘은 다음과 같다. 키 프레임에서의 주식 키워드가 동일 키워드일 경우 새로운 의미가중치 $W_{Keyword_new}$ 는 식(1)과 같다.

$$W_{Keyword_new} = W_{Keyword_old} + \left(\frac{1}{N_{Kframe_SK}} * \frac{N_{Matching_KW}}{N_{Kword_User}} \right) \quad (1)$$

- $W_{Keyword_new}$: 주식 키워드에 대한 새로운 의미가중치
- $W_{Keyword_old}$: 주식 키워드에 대한 이전 의미가중치
- N_{Kframe_SK} : 동일한 키워드가 포함되어있는 키 프레임의 개수
- N_{Kword_User} : 입력된 사용자 키워드 개수
- $N_{Matching_KW}$: 선택된 이미지와 입력된 사용자 키워드의 매칭된 주식 수

키 프레임에서의 주식 키워드가 상이한 키워드일 경우 새로운 의미가중치 $W_{Keyword_new}$ 는 식(2)와 같다.

$$W_{Keyword_new} = W_{Keyword_old} - \left(\frac{1}{N_{Kframe_SK}} * \frac{N_{Matching_KW}}{N_{Kword_User}} \right) \quad (2)$$

- $W_{Keyword_new}$: 주식 키워드에 대한 새로운 의미가중치
- $W_{Keyword_old}$: 주식 키워드에 대한 이전 의미가중치
- N_{Kframe_SK} : 동일한 키워드가 포함되어있는 키 프레임의 개수
- N_{Kword_User} : 입력된 사용자 키워드 개수
- $N_{Matching_KW}$: 선택된 이미지와 입력된 사용자 키워드의 매칭된 주식 수

3.3 키 프레임 특징 추출 기법

입력된 비디오에서 추출된 키 프레임의 특징정보는 색상정보, 모양정보, 움직임정보 등 많은 특징을 추출할 수 있다. 그러나 비디오 검색시스템에서는 빠른시간에 특징정보를 추출하여 검색할 수 있어야 함으로 특징추출에 대한 연산량을 최소화 하면서 정확한 특징을 추출할 수 있어야 한다. 제안하는 특징정보추출에 대한 방법은 다음과 같다.

3.3.1 전체영역과 중심영역에 대한 색상분포 계산

사용자로부터 입력된 질의이미지와 비디오 데이터에서 추출된 키 프레임의 이미지에 대한 색상 유사도를 빠르고 정확하게 계산하기 위하여 질의이미지와 키 프레임의 이미지를 전체영역과 중심영역에 대해 계산한다.

3.3.2 이진영상 히스토그램에 의한 이미지 모양 정보

사용자의 이미지 질의에 대해 색상분포로 유사이미지 검색은 계산과정이 복잡하지 않고 빠르게 유사 이미지를 추출할 수 있다. 하지만 비슷한 분포를 가지고 있는 서로 다른 이미지도 유사이미지로 오인할 수 있으므로 색상분포에 의한 결과 프레임에 대하여 이진영상 변환과 가로/세로 히스토그램을 이용한 모양 정보를 분석하여 유사이미지 검출율을 높일 수 있다. 기존의 이진영상 변환은 0~255의 색상값의 절반인 128을 기준으로 검정색과 흰색을 변환하는데, 이러한 방법은 어두운 색상이나 밝은색이 많이 분포되어 있는 이미지에 대해서는 정확한 모양 정보를 추출할 수 없으므로 평균에 의한 이진영상을 구하여야 한다.

4. 실험 및 평가

제안하는 시스템의 전체적인 검색에 대한 정확도를 계산하기 위해 500회 이상의 사용자 질의를 실행하여 제안한 시스템의 검색에 대한 실험을 하였다. 제안하는 시스템의 검색 적합성[11]을 평가하기 위하여 재현율(Recall)과 정확률(Precision)에 의한 성능평가 척도를 이용하여 실험을 하였다. 재현율은 시스템이 적합한 정보를 검색하는 능력을 말하며 검색의 완전성을 측정하고, 정확률은 시스템이 부적합한 정보를 검색하지 않는 능력을 말하며 검색의 정확도를 측정한다 [13]. 그림 3은 시스템의 검색에 대한 재현율과 정확도를 나타낸 것이다.

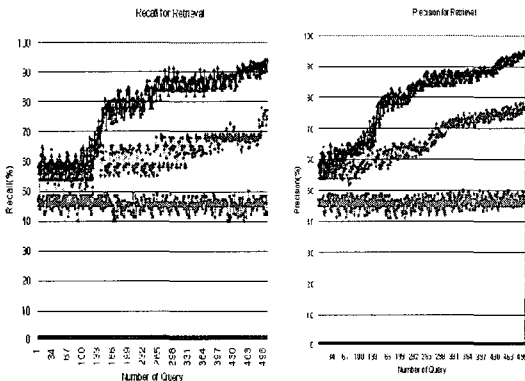


그림 3. 제안시스템의 재현율과 정확도

그림 3은 제안하는 시스템에서 주석기반 검색을 수행 하였을 때 그래프의 중간에 위치하는 데이터를 얻었고, 특징기반 검색을 수행하였을 때 재현율과 정확율에서 가장 좋지 않은 결과가 보여 진다. 주석기반검색과 특징기반이 혼합검색을 수행하였을 때 위쪽에 분포되는 결과를 나타내었다. 따라서, 제안 시스템의 재현율과 정확도가 한가지의 방법을 적용하였을 때 보다 높게 나타났으며, 질의어가 증가할수록 에이전트의 학습과 주석정보의 갱신으로 인해 검색에 대한 정확도가 계속 증가하는 것을 알 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. 제안

하는 의미가중치 계산을 이용하여 주석기반 검색시 사용자의 키워드 입력에 대한 오류를 최소화 시켜 더욱 의미가 정확하고 구체적인 키워드로 질의가 될 수 있도록 하였고 이진 영상 변환을 이용하여 히스토그램의 왜도와 첨도를 계산하여 이를 키 프레임 이미지의 특징으로 이용함으로써 특징기반 검색의 정확도를 더욱 높일 수 있었다.

향후에는 음성 인식 기술을 적용하여 비디오 검색시 키워드 검색뿐만 아니라 사용자의 음성으로 비디오 장면을 검색할 수 있도록 비디오 데이터에서의 음성을 추출하고 유사 비교 검색을 할 수 있는 기법에 대한 연구가 요구된다.

참고문헌

- [1] Chong-Wah Ngo, Ting-Chuen Pong, and Hong-Jiang Zhang, "Clustering and retrieval of video shots through temporal slices analysis," *IEEE Trans on Multimedia*, Vol.4 No.4 pp.446-458, 2002.
- [2] Herwig Rehatschek, and Heimo Muller, "A Generic Annotation Model for Video Database," *The 3rd International Conference on VISUAL '99*, Amsterdam, Netherlands, pp.383-390, 1999.
- [3] Ron Weiss, Andrzej Duda, and David K. Gifford, "Content-based access to algebraic video," In *Proceedings of the International Conference on Multimedia Computing and System*, pp.140-151, 1994.
- [4] Rune Hjelmsvold and Roger Midström, "Modelling and Querying Video Data," In *Proceedings of the 20th VLDB Conference*, Santiago, Chile, pp.686-694, Sep. 1994.
- [5] Eenjun Hwang and V.S. Subrahmanian, "Querying Video Libraries," *UMIACS-TR-95-66*, University of Maryland, June 1995.
- [6] Tony C.T. Kuo and Arbee L.P. Chen, "A Content-Based Query Language for Video Databases," In *Proceedings of IEEE MULTIMEDIA '96*, pp.209-214, 1996.
- [7] Alexander G. Hauptmann and Witbrock M.J., "Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval," In Maybury, M.T.(ed.), *Intelligent Multimedia Information Retrieval*, AAAI Press, 1997.
- [8] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika

- Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker, "Query by Image and Video Content : The QBIC system," IEEE Computer, Vol.28 No.9, pp.23-32, 1995.
- [9] John R. Smith and Shih-Fu Chang, "VisualSEEK : a fully automated content-based image query system," In Proceedings of ACM Multimedia Conference, Boston, pp.87-98, Nov. 1996.
- [10] Shih-Fu Chang, William Chen, Horace J. Meng, Hari Sundaram, and Di Zhong, "VideoQ: An Automated Content-based Video Search System using Visual Cues," In Proceedings of ACM Multimedia '97, pp.313-324, 1997.
- [11] Dennis Shasha and Tsong-Li Wang, "New Techniques for Best-match Retrieval," ACM Transactions on Information Systems, Vol. 8, No. 2, pp.140-158, 1990.
- [12] Gerard Salton and Michael J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1983.
- [13] William B. Frakes and Richard Baeza-Yates, Information Retrieval : Data Structure & Algorithms, Prentice-Hall, 1992.

이근왕(Keun-Wang Lee)

[중신회원]



- 1993년 2월 : 한밭대학교 전자계산학과 (공학사)
- 1996년 : 숭실대학교 컴퓨터학과 (공학석사)
- 2000년 : 숭실대학교 컴퓨터학과 (공학박사)
- 2001년 ~ 현재 : 청운대학교 멀티미디어학과 조교수

<관심분야>

멀티미디어 통신, 멀티미디어 응용, 모바일 통신