내용기반 검색을 위한 뉴스 비디오 인덱싱의 자동화

[0]이동섭* · 이지연* · 신성윤* · 전근환** · 배석찬* · 이양원*

* 군산대학교 컴퓨터과학과

** 군장대학 경영전산학부

Automation of News Video Indexing for Content-Based Retrieval

[0]Dong-seop Lee* · Ji-yon Lee* · Seong-yoon Shin* · Keun-hwan Jeon**

Seok-chan Bae* · Yang-weon Rhee*

* Dept. of Computer Science, Kunsan National Univ.

** School of Management & Computer Science, Kunjang College.

E-mail : syshin, scbae, ywrhee@cs.kunsan.ac.kr

요 약

다양한 분야에서 중요하게 사용될 수 있는 뉴스 비디오 데이터베이스를 구축하기 위해서는 비디오 색인의 자동화에 관한 연구가 필연적이다.

본 논문에서는 뉴스 비디오 색인을 자동화하는 방법으로, 이전에 제안한 컷 추출 방법을 이용하였다. 컷에 의해 추출된 키 프레임에서 앵커 인식 알고리즘으로 앵커 프레임을 자동으로 추출하여 색인을 부여하는 방법으로 비디오 스트림에 대한 색인을 자동화하였다. 구성되는 색인 구조의 형태는 앵커 프레임들이 시간에 따른 사건의 연결이 되고, 앵커 프레임 내에서는 종속되는 키 프레임들을 중심으로 원형을 형성한다. 이들 각각을 논리적으로 통합하면 사용자의 관점에 따라 여러 가지 방법으로 브라우징 되며, 사용자가 원하는 뉴스 비디오 씬들을 쉽게 선택하여 볼 수 있는 특징을 제공하는 장점을 부여한다. 또한, 색인화 된 비디오 스트림을 이용하면 자동적으로 비디오 편집을 수행 할 수 있는 비디오 저작도구의 기반을 제공할 수 있다.

Ⅰ. 서 론

멀티미디어 기술의 급속한 발전과 보급은 방대한 양의 멀티미디어 데이터를 효율적으로 저장, 관리, 검색 할 수 있는 멀티미디어 데이터베이스 구축을 필요로 하고 있다[1,2]. 다양한 멀티미디어 데이터에서 비디오 데이터는 인간에게 가장 친숙하며 자연스러운 형태로 많은 정보를 제공해 주는 동영상 데이터로서, 현재 드라마, 쇼, 뮤직 비디오, 뉴스, 다큐멘터리, 그리고 기타 분야에 대한 멀티미디어 데이터베이스의 연구가 활발히 진행되고 있다.

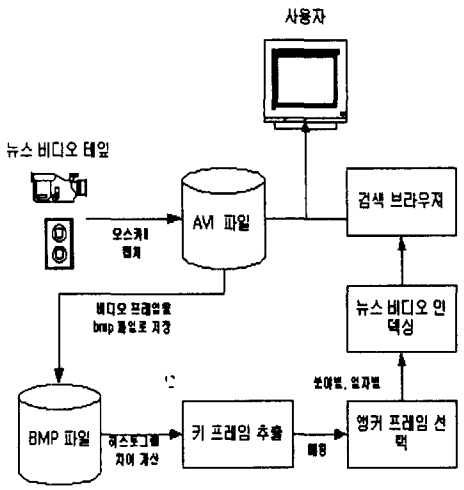
이러한 동영상 비디오 데이터 중에서 뉴스 비디오는 한 시대의 사회 전반에 대한 중요한 정보를 짧은 시간대에 함축적으로 표현한 것으로, 한 국가의 정치, 경제, 사회, 문화, 기타 부분의 전체적인 흐름과 상황을 분석할 수 있으며[1,2], 향후 역사적인 가치로서 뿐만 아니라 미래의 상황을 예측할 수 있는 귀중한 자료로서 중요한 연구 대상이다.

비디오 데이터를 처리하는데 사용되는 용어는 다양한 의미를 가지고 있는 경우가 있지만 본 논문에서는 다음과 같이 정의한다.

- 프레임(frame) : 비디오 자료를 구성하는 최소 단위로서 필름 한 장에 해당되는 하나의 정지영상 이다.
- 샷(shot) : 비디오를 분할할 때 기본 단위가 되는 일련의 프레임들의 모임이다. 보통 하나의 카메라로 촬영되며, 필름이 끊기지 않고 연결된 프레임들로 구성된다.
- 씬(scene) : 하나의 주 대상을 촬영한 연속된 샷들의 집합으로, 필름의 내용이 연결되지 않는다 하더라도 촬영 대상이 변하지 않거나 의미가 변하지 않는 범위 내에서의 샷들을 말한다.
- 컷(cut) : 샷과 샷 또는 씬과 씬 사이의 경계에 해당하는 전환점을 말한다.
- 키 프레임(key frame) : 첫 번째 프레임을 포함하여 컷 다음에 위치한 프레임을 말한다. 샷의 키 프레임과 씬의 키 프레임이 있다.

본 논문에서는 컬러 히스토그램과 χ^2 히스토그램을 조합한 방법[1,2] 을 사용하여 샷들의 키 프레임을 추출한 다음 앵커가 나타나는 프레임을 씬의 키프레임으로 설정한다. 정확한 앵커 프레임의 추출을 위하여 앵커 후보 프레임들은 앵커 인식 알고리즘을 사용하여 식별된다. 이렇게 앵커 프레임을 이용하여 색인화된 뉴스 비디오는 사용자 관점에 따라 다양한 방법으로 브라우징 되며 사용자가 쉽게 선택하여 볼 수 있는 편의성을 제공한다. 전체적인 시스템 구성은 (그림 1)과 같다.

본 논문의 구성은, Ⅱ장에서 앵커 프레임 추출 방법에 대해 설명하고, Ⅲ장에서는 추출된 앵커 프레임을 이용하여 색인화 하는 과정을 설명한다. Ⅳ장에서는 제안된 방법을 적용하여 실제 뉴스 비디오 데이터를 사용하여 실험을 실시하고, Ⅴ장에서는 결론을 맺도록 한다.



(그림 1) 시스템 구성도

Ⅱ. 앵커 프레임 추출

기존의 씬 변화에 따른 키 프레임 추출 방법은 히스토그램을 이용하는 방법, 객체의 이동을 계산하여 사용하는 방법, 에지를 추출하여 이용하는 방법, 그리고 전체적인 웨이블릿(wavelet)을 측정하여 이용하는 방법 등 많은 방법이 사용되었으며, 뉴스 도메인에는 주로 히스토그램 차이를 계산하여 이용하는 방법이 사용되었다 [3,4,5,6,7,8,9].

본 논문에서는 [1,2]에서 제안한 컬러 히스토그램과 χ^2 히스토그램의 장점을 조합한 (식 1)의 방법을 적용하여 프레임간의 차이 값을 계산하는 방법을 사용한다.

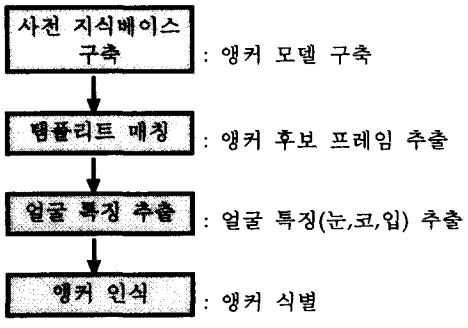
$$d(I_i, I_j) = \sum_{k=1}^{n} \left(\frac{(H_i^r(k) - H_j^r(k))^2}{H_i^r(k)} \times 0.299 + \frac{(H_i^g(k) - H_j^g(k))^2}{H_i^g(k)} \times 0.587 + \frac{(H_i^b(k) - H_j^b(k))^2}{H_i^b(k)} \times 0.114 \right) / 3$$

(식 1)

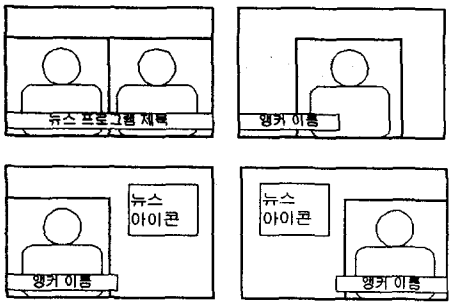
이렇게 하여 얻어진 히스토그램의 차이값에 따라 차이값 변화가 큰 부분의 전환점을 컷으로 설정하여 이 컷의 다음 프레임을 키 프레임으로 설정한다. 여기에서 키 프레임은 뉴스 비디오의 첫 번째 프레임과 각 샷들의 첫 번째 프레임이 된다.

뉴스 비디오에서 추출된 키 프레임들 중에는 앵커가 존재하는 프레임이 있는데 정확한 앵커 프레임의 추출이 필요하다. 이는 뉴스 비디오의 구조적인 특성상 비디오를 색인화 할 때 앵커 프레임이 하나의 사건 씬의 시작을 나타내는 씬의 키 프레임이 되기 때문이다.

앵커 프레임을 추출하는 방법은 (그림 2)와 같다. 먼저, (그림 3)과 같이 앵커 프레임의 공간적 구성을 바탕으로 앵커에 대한 사전 지식베이스를 구축한 후 입력된 프레임과의 템플리트 매칭을 통하여 앵커 후보 프레임들을 추출한다. 이렇게 추출된 앵커 후보 프레임들은 앵커의 얼굴에 대한 특징을 추출하여 인식하는 앵커 인식 알고리즘을 통하여 정확한 앵커로 식별된다. 앵커 인식은 얼굴의 특징 정보인 눈, 코, 그리고 입의 거리관계를 이용하여 수행되는데 본 논문에서는 색인화에 중점을 두어 구체적인 인식 과정은 생략하도록 한다.



(그림 2) 앵커 프레임 추출



(그림 3) 앵커 프레임의 공간적 구성

III. 색인화

추출된 앵커 프레임은 하나의 사건에 대한 씬의 키 프레임이 되며 이를 시작 프레임으로 하여 다음 앵커 프레임이 나타나기 전까지의 모든 샷들은 이 프레임에 종속된다. 즉, 하나의 뉴스 씬은 앵커 프레임이 나온 후에 다음 앵커 프레임이 나올 때까지이며, 이때까지의 이웃하는 각 샷들의 프레임들을 묶어 하나의 씬으로 구성한다.

키 프레임들은 추출된 순서에 의해 앵커 프레임을 중심으로 자동으로 일련의 번호에 의해 물리적으로 색인화 된다. 이렇게 물리적으로 색인화된 비디오 데이터에서 각각의 앵커 프레임들에 대해 분야별 또는 사건별로 의미를 부여하여 논리적으로 색인화 할 수 있다. 논리적인 색인은 사용자가 원하는 다양한 형태로 부여할 수 있는데, 물리적으로 색인화된 키 프레임들의 번호를 연결하여 뉴스 비디오의 어떠한 씬이라도 직접 접근이 가능하도록 한다.

IV. 실 험

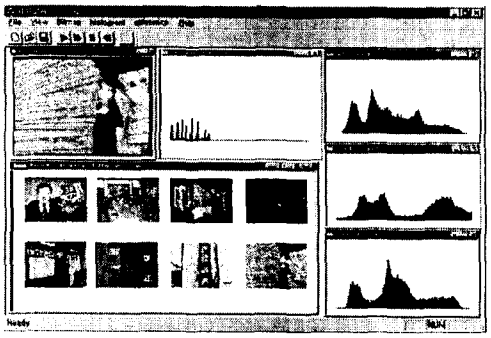
실험에는 펜티엄 166Mhz에 주기억 용량 32M의 PC와 비디오 플레이어 시스템에서 윈도우 98 환경의 MS Visual C++ 5.0 객체 언어를 사용하였다.

실험 데이터는 한국방송공사에서 판매한 이틀분(1997년 11월 22일과 12월 11일)의 KBS 저녁 9시 뉴스를 사용하였는데, 비디오 플레이어를 사용하여 비디오 테이프를 구동시킨 후, 오스카II 비디오 캡쳐 보드를 이용하여 AVI 파일 형태로 데이터를 획득하였다. 이 AVI 데이터에서 초당 5 프레임을 실험 데이터로 캡쳐한 다음 프레임 크기를 400 x 300으로 정규화 하여 실험에 사용하였다.

키 프레임의 추출은 (식 1)을 적용하여 각 프레임들 사이의 변화가 크게 나타난 프레임을 키 프레임으로 추출하는데 (그림 4)와 같다. (식 1)을 적용하여 얻어진 (그림 4)의 상단 중앙 히스토그램에서 보여주는 것처럼 프레임간의 변화가 심한 곳에서는 결과값이 높게 나타나며, 변화가 심하지 않는 곳에서는 결과값의 차가 거의 없다. 이렇게 적용된 히스토그램을 이용하여 적당한 임계치를 부여하여 키 프레임을 추출한다.

(그림 5)는 제안된 히스토그램의 차이값에 임계치를 적용하여 추출된 키 프레임에서 앵커 인식 알고리즘을 적용하여 앵커 프레임을 추출하고 이를 중심으로 순서에 의해 일련의 번호에 따라 색인화한 것이며, 의미를 갖는 논리적인 색인화는 물리적인 키 프레임 번호의 연결에 의하여 구성된다.

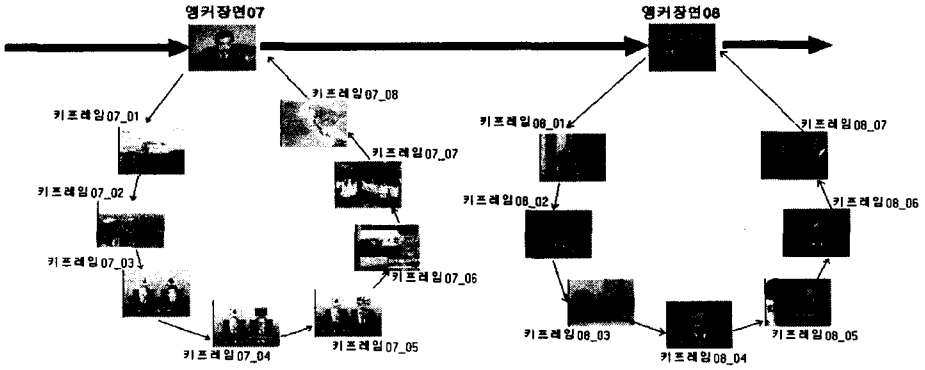
(그림 6)은 논리적으로 색인화된 뉴스 비디오 데이터를 검색할 수 있는 브라우저를 보여준다. 그림에서 앵커 프레임은 씬의 키 프레임을 나타내며 키 프레임은 하나의 씬을 형성하는 샷들의 키 프레임들이다. 색인화는 뉴스 분야와 세부분야 및 앵커 프레임으로 구성되어 있고 키 프레임은 하나의 앵커 프레임에 해당되는 순서화된 프레임들을 말한다. 따라서 사용자는 원하는 분야와 세부분야 및 앵커프레임을 선택하여 원하는 씬들을 쉽게 찾아 볼 수 있다.



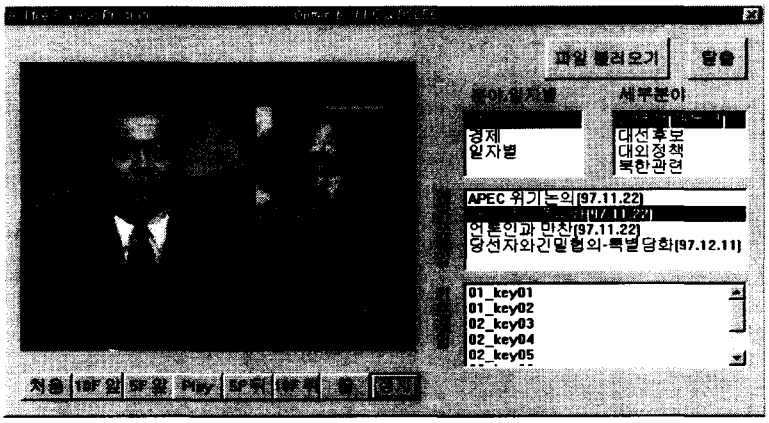
(그림 4) 키 프레임 추출 화면

V. 결 론

본 논문은 뉴스 비디오 데이터베이스를 구축하기 위해 이전에 제안된 새로운 씬 분할 기법을 사용하여 키 프레임을 추출하였고 앵커 인식을 통한 앵커 프레임의 자동 추출 및 색인 부여로 전체적인 뉴스 비디오 데이터의 색인을 자동화 하였다. 앵커 프레임을 사용자의 관점과 사건에 따라 논리적으로 색인화 하여 다양한 브라우징 방법과 편의성을 제공하고, 자동화된 비디오 편집을 위한 비디오 저작 도구의 기반을 제시하였다.



(그림 5) 뉴스 씬의 색인화 구조



(그림 6) 뉴스 비디오 검색 브라우저

참고문헌

[1] 이동섭, 김재원, 배석찬, 이양원, "비디오 장면변환 검출 알고리즘", 한국 퍼지 및 지능 시스템 학회 '98 춘계 학술대회 논문집, 제8권, 제1호, 1998.6.

[2] 이동섭, 이양원, "뉴스 비디오의 장면분할을 위한 키 프레임 추출 기법", 한국정보과학회 춘계 학술발표 논문집, Vol. 25, No. 1, pp. 613-615, 1998.

[3] E. Ardizzone and M. L. Cascia, "Automatic Video Database Indexing and Retrieval", Multimedia Tools and Applications, Vol. 4, No. 1, pp. 29-56, January 1997.

[4] J. C. Lee, Q. Li and W. Xiong, "VIMS : A Video Information Management System", Multimedia Tools and Applications, Vol. 4, No. 1, pp. 7-28, January 1997.

[5] B. Furht, S. W. Smoliar and H. J. Zhang, "Video and Image Processing in Multimedia System", Kluwer Academic Publishers, pp. 335-356, 1995.

[6] Minerva Yeung, Boon-Lock Yeo and Bede Liu, "Segmentation of Video by Clustering and Graph Analysis" Computer Vision and Image Understanding, Vol. 71, No. 1, pp. 94-109, July 1998.

[7] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-video Search for Object Apperance", IFIP:Visual Database Systems II, pp. 113-127, 1992.

[8] R. Kasturi and R. Jain, "Dynamic Vision", in Computer Vision : Principles, IEEE Comput. Soc., pp. 469-480, 1990.

[9] S. J. Dennis, R. Kasturi, U. Gargi and S. Antani, "An Evaluation of Color Histogram Based Methods in Video Indexing", Research Progress Report CSE-96-053 for the Contract MDA 904-95-C 2263, 1995.