장면전환 검출과 클러스터링을 이용한 비디오 개요 추출

신성윤* · 강일고** · 이양원*

* 군산대학교 컴퓨터정보과학과, ** 전라북도청 공보정보화과

Video Abstracting Using Scene Change Detection and Clustering

Seong-Yoon Shin* · Il-Ko Kang** · Yang-Won Rhee*

* Dept. of Computer Information Science, Kunsan National University

** A Provincial Office, Jeollabuk-do

{syshin, ywrhee}@cs.kunsan.ac.kr

요 약

비디오를 시청하기 위하여 원하는 비디오를 선택하고자 할 때 사용자들은 비디오의 전반적인 내용을 알 수 있는 방법이 많지 않다. 따라서 비디오 시청을 원하는 사용자들에게 비디오의 전반적인 개요를 보여주어 선택할 수 있는 방법이 요구된다. 본 논문에서는 전환 검출 방법과 샷 클러스터링을 이용한 비디오 개요 추출 방법을 제시한다. 장면전환 검출 방법은 컬러 히스토그램과 χ2 히스토그램을 합성한 방법을 이용하여 추출하도록 한다. 클러스터링은 히스토그램의 차이값을 측정과 샷 병합 알고리즘을 통해 수행하도록 한다.

1. 서 론

비디오 데이터베이스(video database)의 경우 비디오 내용을 샷(shot)이나 클립 레벨(clip level)의 구조로 묘사하는 방법이 있다[1]. 비디오에서 샷은 일정한 시간이나 공간에서 연속적인 동작을 나타내는 하나 이상의 연속적인 프레임들의 집합을 나타 내며 비디오 정보를 구성하는 유효한 단위이다. 이에 따라서 최근 비디오 샷을 검출하고 샷의 내용을 특성화하기 위한 자동화된 방법과 반자동화된 방법들에 대하여 커다란 관심이 집중되고 있다.

매우 긴 비디오의 개요 추출(video abstracting)은 종종 비디오를 시청하고자 하는 사람들에게 비디오를 시청할 가치가 있는지를 결정하는데 있어서 유용한 정보를 제공한다. 비디오의 개요 추출은 크게 비디오 요약 시퀀스(video summary sequence)와 비디오 하이라이트(video highlight)의 두 가지 방법이 있다. 요약 시퀀스는 전체 비디오의 내용에 대하여 매우 적당한 전체적인 개요를 제공하기 때문에 다큐멘터리(documentary)에 적합한 반면, 하이라이트는 대부분 매우 흥미 있는 관심 분야의 비디오 세그먼트(video segment)들만을 포함하기 때문에 비디오 예고편 등에 적합하다[2].

본 논문에서는 대용량의 긴 비디오 데이터를 요약하여 중요 내용만을 표현할 수 있는 비디오 개요 추출 시스템을 구축하여 사용자들에게 함축적인 내용을 보여주도록 한다. 전체적인 시스템 구조는 그림 1과 같다.

본 논문의 2장에서는 관련연구를 살펴보고, 3장에서는 장면 전환 검출에 대해 설명하며, 4장에서는 클러스터링을 이용한 비디오 개요 추출을 설명한다. 5장에서는 실험 및 결과를 제시하고 6장에서 결론을 맺도록 한다.

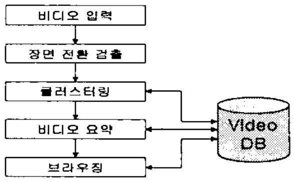

그림 1. 시스템 구조

2. 관련연구

비디오 요약 및 개요추출 분야 또한 많은 연구자들이 계속해서 관심을 갖고 지속적으로 연구해 온 분야이다. 많은 연구자들이 비디오 시퀀스 자체를 다양한 방법으로 요약하는데 중점적으로 연구를 수행하였다.

비디오 스키밍(video skimming)[3]은 다큐멘터리나 뉴스 방송을 요약하기 위하여 제시되었다. 여기서 비디오와 비디오 사본은 사용되는 단어의 순서에 따라

할당되며 언어의 분석을 통하여 사본에서 중요한 단어를 식별한다. 그리하여 이들 단어의 우선순위에 따라 비디오 클립들이 선택되어진다.

Yeung 등[4]은 장면 변화 그래프(scene transition graph)라는 스토리 흐름을 이용한 샷 기반 구조를 제시하였다. Hanjalic 등[5]은 키 프레임을 추출하고 비디오 샷을 설정하며, 이들 키 프레임을 포함하는 비디오 요약 시퀀스를 생성하기 위하여 클러스터 유효성 분석(cluster validity analysis)을 사용하였다. Uchihashi 등[6]은 코믹 만화 형태를 닮은 그림 형태의 비디오 요약 생성 방법을 제시하였다. 이들은 세그먼트의 희소성과 지속성을 기초로 한 중요도 측정을 사용하여 비디오를 요약하였다.

또 다른 형태의 비디오 요약은 하이라이트 추출이다. Lienhart 등[7,8]은 저수준 시각적/오디오 특징, 모션 정보, 그리고 컬러 정보를 탐색하여 무비 트레일러(movie trailer)의 자동 생성 방법을 제시하였다. 이들은 중요한 객체, 사람, 액션, 대화, 타이틀 텍스트 그리고 타이틀 뮤직의 클립을 선택하기 위하여 디지털 비디오의 기본적인 물리적 파라미터들에 대한 경험적 특징을 사용하였다.

Babaguchi[9]는 이벤트 기반 비디오 색인화 방법을 이용한 스포츠 비디오 요약 방법을 제시하였다. 이 방법은 비디오 요약에 유용하긴 하지만 의미를 표현하는 중요한 특징들이 많이 유실된다.

따라서 비디오가 갖는 중요한 정보들을 효율적으로 표현 할 수 있는 새로운 비디오의 요약 방법이 요구된다. 요약된 비디오는 원래 비디오가 갖는 중요한 정보를 함축적으로 표현할 수 있어야 하며, 중복되거나 무의미한 부분은 삭제되어야 한다.

3. 장면전환 검출

본 논문에서 제시하는 장면 전환 검출 방법은 컬러 히스토그램과 x2 히스토그램의 장점을 합성한 아래 식의 방법을 이용하는데, 이 방법은 각 프레임 사이의 차이 값을 계산하고 차이 값의 크기에 따라 각 샷들의 경계를 추출하는, 즉 프레임 단위로 연산을 수행하여 입력된 비디오 스트림을 샷 단위로 분할하는 방법이다.

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

위 식의 컬러 히스토그램과 x2 히스토그램을 합성한 방법은 컬러 히스토그램을 R·G·B 각각에 대하여 산출함으로서 이미지의 컬러를 구성하는 요소들을 신축성 있게 사용할 수 있으며, x2 히스토그램이 갖는 검출 성능이 우수하다는 특징을 적용하여 보다 효율적으로 장면 전환을 검출 할 수 있는 방법이다. 위 식에 곱한 세 개의 상수 값들은 R·G·B 컬러 공간을 완전한 YIQ 공간이 아닌 반-YIQ 공간으로 바꾸기 위한 값들이다.

일반적으로 갑작스러운 장면 전환의 경우 장면 전환이 발생한 첫 번째 프레임을 키 프레임으로 설정하고 점진적인 장면 전환에서는 장면 전환의 시작 프레임과 끝 프레임 두 개를 키 프레임으로 설정한다. 또한 비디오 데이터의 특성과 내용에 따라 키 프레임 설정 방법은 유연하게 바뀔 수 있다.

본 논문에서는 갑작스러운 장면 전환과 점진적인 장면 전환을 동시에 검출하기 위해서 장면 전환 검출 단계에서 얻어진 히스토그램의 차이 값에 따라 차이 값 변화가 큰 부분의 전환점이 되는 프레임을 찾아 이 프레임을 키 프레임으로 설정하도록 한다. 따라서, 본 논문에서는 어떤 형태의 비디오를 대상으로 하더라도 키 프레임은 추출된 각 샷들의 첫 번째 프레임이 된다.

4. 클러스터링을 이용한 비디오 개요 추출

4.1 클러스터링

장면 전환 검출에 의하여 추출된 샷들의 키 프레임들은 앞절에서 제시한 유사성 측정 방법에 의하여 유사한 샷들끼리 병합을 통하여 클러스터를 형성하게 된다. 따라서 하나의 클러스터는 유사한 장면들로 구성된 샷들로 구성된다.

본 논문에서는 같이 유사성 측정을 위한 비교 회수와 기억공간 할당을 단계별로 줄여가는 새로운 샷 병합 알고리즘을 이용한 클러스터링 방법을 제시하는데, 전체적인 병합 과정은 그림 2와 같다.

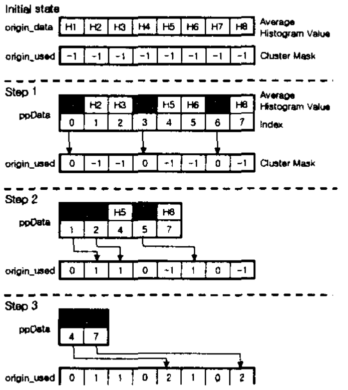

그림 2. 새로운 샷 병합 과정

그림 2의 Initial state는 초기 단계로서, origin_data는 각 프레임에 대한 히스토그램 평균값들을 갖으며 origin_used는 클러스터링을 위한 마스크로서 -1로 세트된다.

첫 번째 단계인 Step 1에서는 ppData를 생성하여 각 프레임별 히스토그램 평균값과 색인을 갖도록 하며, 유사성 측정을 위한 비교는 먼저 첫 번째 프레임의 히스토그램 평균값(H1)과 나머지 프레임들의 히스토그램 평균값(H2~H8)을 순차적으로 식 3의 공식에 의하여 차이값을 측정하게 된다. 이때 차이값이 임계치 이내이면 유사한 샷으로 간주하여 해당 색인에 맞는 origin_used를 초기 클러스터 0으로 마스크하고 메모리 할당을 해제한다. 그림 3에서는 H1과 H4, H7이 유사한 샷으로서 색인 0, 3, 7번이 클러스터 0으로 마스크 된 것이다.

두 번째 단계인 Step 2에서는 나머지 ppData를 대상으로 첫 번째 프레임의 평균 히스토그램값(H2)와 나머지 프레임의 히스토그램 평균값(H3, H5, H6, H8)을 비교하여 차이값을 측정하고 임계치를 적용하게 된다. 마찬가지로 유사한 샷들은 해당 색인에 맞는 origin_used를 다음 클러스터 1로 마스크하고 메모리 할당을 해제한다. 그림 3에서는 H2과 H3, H6이 유사한 샷으로서 색인 1, 2, 5번이 클러스터 1으로 마스크 된 것이다.

세 번째 단계인 Step 3에서는 나머지 ppData로서 H5와 H8을 비교하여 색인에 맞게 origin_used를 세 번째 클러스터 2로 마스크하고 메모리 할당을 해제하고, 더 이상 데이터가 없으므로 클러스터링이 종료된다.

따라서 그림 2에서는 8개의 샷들을 대상으로 유사성 측정을 이용하여 샷들을 병합하는 과정을 통하여 최종적으로 클러스터 0, 1, 2의 세 개의 클러스터가 생성된 것이다. 또한 비교 회수를 보면 Step 1에서는 7회, Step 2에서는 4회, Step 3에서는 1회로서, 단계별로 비교 회수는 크게 감소함을 알 수 있고 총 비교회수는 12회이며 단계별 평균 비교 회수는 4회가 된다.

4.2 비디오 개요 추출

비디오 개요 추출은 샷들의 클러스터링을 통해서 생성된 클러스터를 바탕으로 수행된다. 각 클러스터는 특징들이 서로 유사한 샷들을 병합하여 형성된 데이터 구조로서 대체로 비슷한 내용을 갖게된다. 따라서 샷들이 병합되어 각 클러스터의 전체 내용을 재생하게 되면 유사한 내용들이 계속해서 나타나게 되므로 클러스터를 구성하는 샷들 중에서 클러스터를 대표하는 샷을 선택해야 한다.

클러스터를 대표하는 대표 샷을 각 클러스터 당 1개씩 지정하여 비디오 개요 추출에서 각각의 클러스터를 대표하는 샷들을 병합하여 하나의 비디오 개요를 구성한다. 따라서, 각 클러스터를 대표하는 각각의 샷들을 결정하는 방법이 명확해야 한다. 결국 클러스터를 대표하는 샷들은 클러스터의 내용을 가장 잘 표현하여야 하며 클러스터 내의 샷들이 갖고있는 특징을 가장 잘 반영하여야 한다. 하지만 클러스터를 구성하는 전체 샷들에 대하여 샷의 키 프레임이 갖고있는 전체 내용을 분석하는 것은 많은 시간과 비용이 소요되는 작업이므로 샷들을 분석하는 방법은 지양하도록 한다.

클러스터 대표 샷 설정 방법 중 가장 편리하고 효율적인 방법은 클러스터를 구성하는 샷들 중 가장 첫 번째 샷을 대표 샷으로 설정하여 비디오 개요를 생성하는 방법이다. 본 논문에서는 클러스터를 구성하는 샷들 중 가장 첫 번째 샷을 대표 샷으로 설정하는 방법을 따르도록 하며 비디오 개요를 구성하는 형태는 다음 그림 3과 같다.

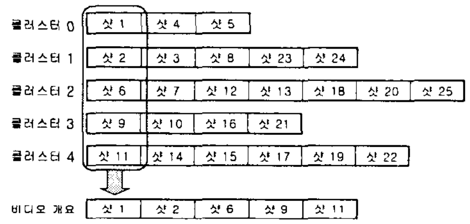

그림 3. 첫 번째 샷을 이용한 비디오 개요의 예

위의 그림 3에서 나타내는 것처럼 각 클러스터를 구성하는 첫 번째 샷들을 추출하여 빠르고 편리하게 요약된 비디오를 구성하여 사용자들은 이 요약된 비디오를 빠른 시간에 시청하여 비디오의 전체적인 내용을 쉽게 이해하고 원하는 비디오를 편리하게 선택할 수 있도록 한다.

5. 실험 및 결과

본 논문의 실험은 펜티엄 Ⅳ 1.3GHz, Windows 2000 Server 환경에서 Visual C++ 6.0 언어로 프로그래밍 하였으며, 비디오 자료는 KBS 도전 지구탐험대(5분 분량)를 대상으로 AVI 압축 형태의 비디오를 OSCAR Ⅱ 캡쳐 보드로 초당 5프레임을 실험 데이터로 캡쳐하여, 프레임 크기를 200X150으로 정규화 하여 사용하였다.

컬러 히스토그램과 χ^2 히스토그램을 합성한 방법을 이용하여 장면 전환 검출에 의한 샷의 키 프레임을 추출하는 화면은 그림 4와 같다.

이렇게 추출된 키 프레임들은 각각의 샷을 구분하는 기준이 되며 결국, 샷의 키 프레임이 된다. 따라서 하나의 샷은 이 키 프레임을 시작으로 다음 키 프레임

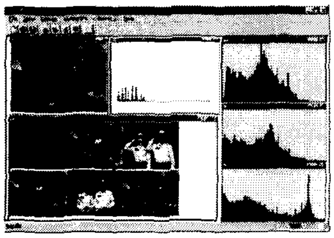

그림 4. 키 프레임 추출 예

그림 5 추출된 키 프레임들의 예

이 나오기 이전까지의 프레임들로 구성되는 것이다. 그림 5는 본 논문에서 사용한 장면 전환 검출 방법을 이용하여 추출한 키 프레임들 중 30개의 키 프레임들의 예를 보여주고 있다.

비디오 프레임들을 히스토그램을 통한 유사성 측정을 통하여 얻은 클러스터의 예는 다음 그림 6과 같다.

본 논문에서는 이와 같이 클러스터링 된 비디오 샷들에서 각 클러스터의 첫 번째 샷을 비디오 개요를 구성하는 샷들로 설정하기 때문에 그림 6의 예에서는 클러스터 1부터 클러스터 11까지의 첫 번째 샷들을 전체 비디오 개요를 구성하는 샷들로 설정한다. 그림 7은 그림 6의 클러스터를 대상으로 생성된 비디오 개요를 구성하는 샷들의 흐름을 나타낸다. 따라서 생성된 비디오 개요는 클러스터의 수와 같은 총 11개의 샷들로 구성된다. 이렇게 생성된 비디오 개요는 전체 비디오의 내용을 시청하지 않고도 전반적인 내용을 빠르고 편리하게 이해할 수 있도록 하며, 비디오를 시청하고자 하는 사람들에게 선택의 폭을 보다 넓게 부여하게 된다.

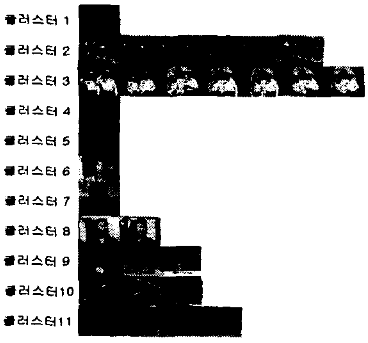

그림 6. 클러스터링 된 비디오의 예

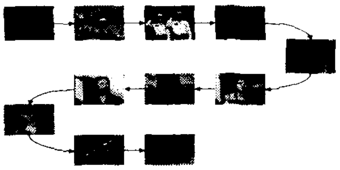

그림 7. 추출된 비디오 개요의 예

본 논문에서 실험에 이용한 5분 분량의 KBS 도전 지구탐험대 비디오 4편을 대상으로 장면 전환 검출과 새로운 클러스터링을 통하여 비디오 개요 추출을 수행한 결과는 다음 표 1과 표 2와 같다.

표 1. 검출된 키 프레임 수

구분	검출된 키 프레임 수	비고
비디오1	67	5분 분량, 임계치 60,
비디오2	72	
비디오3	63	
비디오4	66	

표 2. 비디오 개요 추출 결과

구분	키 프레임 수	클러스터수	개요 재생시간
비디오 1	67	21	116초
비디오 2	72	23	121초
비디오 3	63	19	103초
비디오 4	66	23	112초

표 1의 결과는 컬러 히스토그램과 $x2$ 히스토그램을 합성한 방법을 이용한 장면 전환 검출에 의하여 추출된 키 프레임의 수를 나타낸다. 장면 전환 검출에서는 주어진 임계치에 따라 키 프레임의 수가 다르게 검출될 수 있다.

표 2에서 키 프레임 수는 장면 전환 검출에 의해 생성된 샷의 수를 나타내며 생성된 클러스터의 수는 비디오 개요를 구성하는 샷의 수를 나타낸다. 클러스터링을 수행할 때 임계치는 20을 부여하였는데, 이 임계치 또한 사용자가 유동적으로 변경할 수 있다. 비디오 개요 추출 결과, 시간적으로 볼 때 5분 분량의 비디오를 3분의 1 정도로 요약함을 알 수 있다.

6. 결론

본 논문에서는 장면 전환 검출과 클러스터링 알고리즘을 이용한 새로운 비디오 개요 추출 방법을 제시하였다. 컬러 히스토그램과 $x2$ 히스토그램의 장점을 합성한 방법을 이용하여 효율적인 장면 전환 검출을

수행함으로서 비디오를 샷들로 분할하고 각 샷들의 키프레임을 설정하게 된다. 이렇게 추출된 각 샷들의 키프레임들은 새로운 샷 병합 방법을 이용하여 클러스터를 구성하게 되며, 이 클러스터를 바탕으로 비디오 개요를 생성하게 된다.

참고문헌

[1] G. Davenport, T. Smith, and N. Pincever, "Cinematic Primitives for Multimedia," Computers and Graphics, Vol. 15, pp. 67-74, 1991.

[2] K. Hang-Bong, "Generation of Video Highlights Using Video Context and Perception," Proc. of SPIE, Storage and Retrieval for Media Databases 2001, Vol. 4315, pp. 320-399, 2001.

[3] M. Christal, M. Smith, C. Taylor and D. Winkler, "Evolving Video Skims into Useful Multimedia Abstractions," Proc. CHI'98, pp. 171-178, 1998.

[4] M. Yeung, B. Yeo and B. Liu, "Segmentation of Video by Clustering and Graph Analysis," Computer Vision and Image Understanding, Vol. 71, No. 1, pp. 94-109, 1998.

[5] A. Hanjalic and H. Zhang, "An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis," IEEE Taans. Cir. & Sys. for Video Tech., Vol. 9, No. 8, pp. 1280-1289, Dec. 1999.

[6] S. Uchihashi, J. Foote, A. Girgenshon and J. Boreczky, "Video Manga: Generating Semantically Meaningful Video Summaries," Proc. ACM MM'99, 1999.

[7] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video Abstracting," Communications of the ACM, Vol. 40, No. 12, pp. 54-62, 1997.

[8] S. Peiffer, R. Lienhart, S. Fisher and Effelsberg, "Abstracting Digital Movies Automatically," Int. Jour. Visual Communication and Image Representation, Vol. 7, No. 4, pp. 345-353, 1996.

[9] N. Babaguchi, "Towards Abstracting Sports Video by Highlights," Proc. ICME'00, Aug. 2000.

[10] J. Platt, "Auto Album : Clustering Digital Photographs using Probabilistic Model Merging," IEEE Workshop on Content-Based Access to Image and Video Libraries 2000.