

## 대용량 비디오 데이터베이스 구축을 위한 비디오 개요 추출

신성운\*, 표성배\*\*, 이양원\*

# Video Abstracting Construction of Efficient Video Database

Shin Seong Yoon \*, Pyo Seong Bae \*\*, Rhee Yang Won \*

### 요약

대부분의 비디오는 대용량의 장시간 데이터로서 비디오 시청자들이 전반적인 내용을 이해하기에는 충분하지 못하다. 본 논문에서는 이러한 문제점을 해결하기 위하여 효율적인 장면 전환 검출 방법과 새로운 샷 클러스터링을 이용한 비디오 개요 추출 방법을 제시한다. 장면전환 검출 방법은 컬러 히스토그램과  $\chi^2$  히스토그램을 합성한 방법을 이용하여 추출하도록 한다. 클러스터링은 지역적 히스토그램의 차이값을 이용한 유사성 측정과 새로운 샷 병합 알고리즘을 통해 수행하도록 한다. 또한 실제 TV 방송 프로그램을 대상으로 비디오 개요 추출 실험 결과를 제시한다.

### Abstract

Video viewers can not understand enough entire video contents because most video is long length data of large capacity. This paper propose efficient scene change detection and video abstracting using new shot clustering to solve this problem. Scene change detection is extracted by method that was merged color histogram with  $\chi^2$  histogram. Clustering is performed by similarity measure using difference of local histogram and new shot merge algorithm. Furthermore, experimental result is represented by using Real TV broadcast program.

▶ Keyword : 장면전환검출(Scene Change Detection),  $\chi^2$  히스토그램( $\chi^2$  Histogram), 샷 클러스터링(Shot Clustering), 비디오 개요 추출(Video Abstracting),

• 제1저자 : 신성운

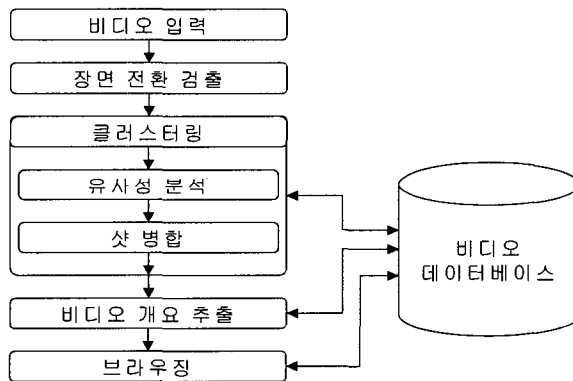
\* 군산대학교 컴퓨터정보학과

\*\* 인덕대학 소프트웨어개발과

## I. 서론

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 중요한 내용만을 함축적으로 표현할 수 있는 비디오의 개요를 추출하는 비디오 개요 추출 시스템을 구축하여 사용자에게 함축적인 내용을 보여주도록 한다. 전체적인 시스템 구조는 <그림 1>과 같다.



[그림 1] 시스템 구조  
Fig 1. System Architecture

## II. 관련연구

비디오 스키밍(video skimming)[3]은 다큐멘터리나 뉴스 방송을 요약하기 위하여 제시된 방법이다. 이 방법에서 사용된 비디오와 비디오 사본은 사용되는 단어의 순서에 따라 할당되며 언어의 분석을 통하여 사본에서 중요한 단어를 식별한다. 그리하여 이들 단어의 우선순위에 따라 비디오 클립들이 선택되어지는 방법으로서 많은 시각적 특징이 유실된다. 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 설정하여 이들 키 프레임을 포함하는 비디오 요약 시퀀스를 생성하기

위한 클러스터 유효성 분석(cluster validity analysis)을 사용하였다. Uchihashi 등(6)은 코믹 만화 형태를 닮은 그림 형태의 비디오 요약 생성 방법을 제시하였는데, 세그먼트의 회소성과 지속성을 기초로 한 중요도 측정을 사용하여 비디오를 요약하였다. 이들의 방법들 또한 많은 의미 정보들이 유실되며 다소 효율적이지 못하다. 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. 이들을 중요한 객체, 사람, 액션, 대화, 타이틀 텍스트 그리고 타이틀 음악의 클립을 선택하기 위하여 디지털 비디오의 기본적인 물리적 파라미터들에 대한 경험적 특징을 사용하였는데, 많은 특징을 이용하는 반면 매우 복잡하고 시간이 많이 소요된다.

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. 또한 비디오 분할을 위해서 자동 임계치 결정 알고리즘 [11]과 자막정보를 비디오 하이라이트 생성[12]방법 또한 새롭게 제시되었으나, 비디오가 갖는 중요한 정보들을 효율적으로 표현 할 수 있는 새로운 비디오의 요약 및 개요 추출 방법이 요구된다. 요약 비디오 및 비디오 개요는 원래 비디오가 갖는 중요한 정보를 함축적으로 표현할 수 있어야 하며, 중복되거나 무의미한 부분은 삭제되어야 한다.

### III. 장면전환검출

본 논문에서 제시하는 장면 전환 검출 방법은 컬러 히스토그램과  $\chi^2$  히스토그램의 장점을 합성한 아래 식 1의 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.

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

위 식의 컬러 히스토그램과  $\chi^2$  히스토그램을 합성한 방법은 컬러 히스토그램을 R·G·B 각각에 대하여 산출함으로써 이미지의 컬러를 구성하는 요소들을 신축성 있게 사용할 수 있으며,  $\chi^2$  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.

#### IV. 클러스터링을 이용한 비디오 개요 추출

##### 4.1 유사성 측정

본 논문에서의 유사성 측정을 위한 영역은 프레임 전체 영역을 대상으로 하지 않고 일반적으로 프레임 내의 중요 정보가 중심영역을 기준으로 상·하·좌·우에 위치한다는 특징을 바탕으로 <그림 2>와 같이 비디오 프레임의 9개의 영역으로 분할하여 수행한다.

	2	
5	1	3
	4	

【그림 2】 기준 영역  
Fig 2. Base Area

5개의 영역에 대한 히스토그램 평균값을 구하는 방법은 식 2와 같다. 식 2에서  $H_a$ 와  $H_b$ 는 이웃하는 두 샷의 키 프레임 a와 b에 대한 히스토그램 평균값을 나타내고  $H_{a_i}(x, y)$ 와  $H_{b_i}(x, y)$ 는  $(x, y)$  영역에 대한 컬러 히스토그램 값이다.

$$H_a = \frac{\sum_{i=1}^5 H_{a_i}(x, y)}{5} \tag{2}$$

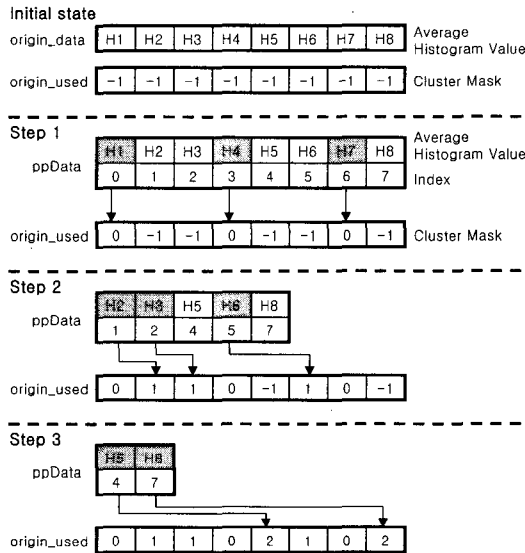
$$H_b = \frac{\sum_{i=1}^5 H_{b_i}(x, y)}{5}$$

식 3은 두 키 프레임 a와 b에 대한 히스토그램 평균값의 차이를 계산하는 방법으로서, 차이값이 임계치 내에 들면 유사한 프레임으로 간주하여 하나의 클러스터에 포함이 되고 임계치 이상이면 다른 클러스터에 해당되게 된다.

$$D(a, b) = |H_a - H_b| \tag{3}$$

### 4.2 클러스터링 알고리즘

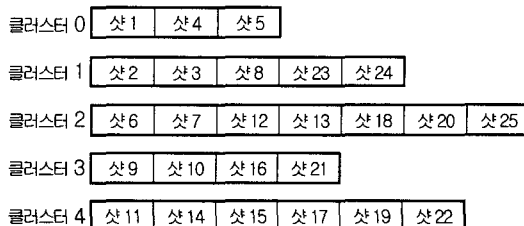
본 논문에서는 같이 유사성 측정을 위한 비교 회수와 기억공간 할당을 단계별로 줄여가는 새로운 샷 병합 알고리즘을 이용한 클러스터링 방법을 제시하는데, 전체적인 샷 병합 과정은 <그림 3>과 같다.



【그림 3】 새로운 샷 병합 과정  
 Fig 3. New Shot Merging Process

<그림 3>의 Initial state는 초기 단계로서, origin\_data는 각 프레임에 대한 히스토그램 평균값들을 갖고 origin\_used는 클러스터링을 위한 마스크로서 -1로 세트된다.

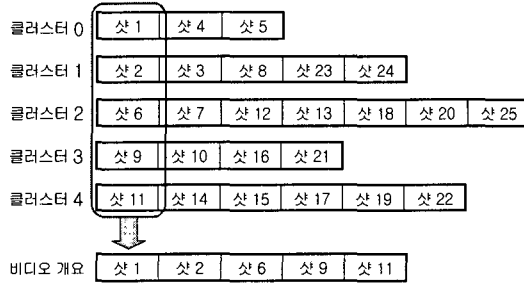
위와 같은 알고리즘을 사용하여 샷들을 클러스터링 하면 다음 <그림 4>와 같은 형태로 비디오가 클러스터 단위로 분할된다. <그림 4>에서는 총 25개의 샷들이 샷 병합 알고리즘을 통하여 5개의 클러스터로 구성된 형태이다.



【그림 5】 클러스터 구성 형태  
 Fig 5. Clustering Format

### 4.3 비디오 개요 추출

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 5>과 같다.



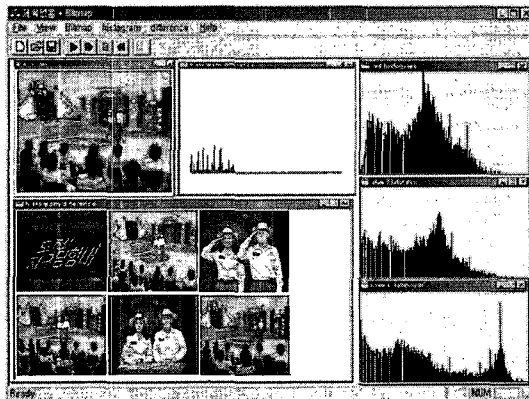
【그림 5】 첫 번째 샷을 이용한 비디오 개요의 예  
Fig 5. Example of Video Abstracting Using First Shot

위의 <그림 5>에서 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.

### V. 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 캡처 보드로 초당 5프레임을 실험 데이터로 캡처하여, 프레임 크기를 200X150으로 정규화 하여 사용하였다.

컬러 히스토그램과  $\chi^2$  히스토그램을 합성한 방법을 이용하여 장면 전환 검출에 의한 샷의 키 프레임 추출하는 화면은 <그림 6>과 같다.



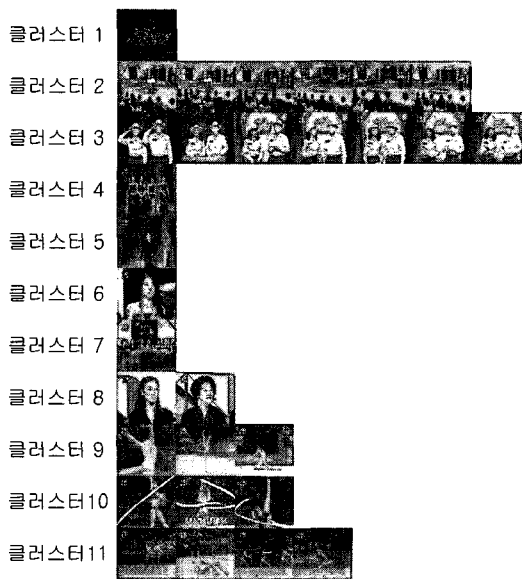
【그림 6】 장면 전환 검출에 의한 키 프레임 추출 예  
Fig 6. Example of Keyframe Extraction by Scene Change Detection

다음 <그림 7>은 본 논문에서 사용한 장면 전환 검출 방법을 이용하여 추출한 키 프레임들 중 30개의 키 프레임들의 예를 보여주고 있다.



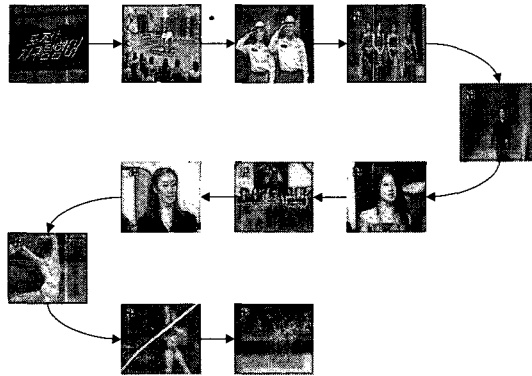
【그림 7】 추출된 키 프레임들의 예  
Fig 7. Example of Extracted Keyframes

비디오 프레임들을 <그림 2>에서 나타난 5개의 기준 영역을 바탕으로 히스토그램을 통한 유사성 측정을 통하여 얻은 클러스터의 예는 다음 <그림 8>와 같다.



【그림 8】 클러스터링 된 비디오의 예  
Fig 8. Example of Clustered Video

다음 <그림 9>는 <그림 8>의 클러스터를 대상으로 생성된 비디오 개요를 구성하는 샷들의 흐름을 나타낸다. 따라서 생성된 비디오 개요는 클러스터의 수와 같은 총 11개의 샷들로 구성된다.



【그림 9】 추출된 비디오 개요의 예  
Fig 9. Example of Extracted Video Abstract

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  
Table 1. Number of Extracted Keyframe

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

표 1의 결과는 컬러 히스토그램과  $\chi^2$  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 2】 비디오 개요 추출 결과  
Table 2. Result of Video Abstracting

구분	키 프레임 수	생성된 클러스터 수	비디오 개요 재생 시간
비디오 1	67	21	116초
비디오 2	72	23	121초
비디오 3	63	19	103초
비디오 4	66	23	112초



## VI. 결론 및 향후 연구 방향

본 논문에서는 장면 전환 검출과 클러스터링 알고리즘을 이용하여 새로운 비디오 개요 추출 방법을 제시하였다. 컬러 히스토그램과  $\chi^2$  히스토그램의 장점을 합성한 방법을 이용하여 효율적인 장면 전환 검출을 수행함으로써 비디오를 각각의 샷들로 분할하고 각 샷들의 키 프레임들을 설정하였고, 이렇게 추출된 각 샷들의 키 프레임들은 새로운 샷 병합 알고리즘을 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 Trans. Cir. & Sys. for Video Tech.*, Vol. 9, No. 8, pp. 1280-1289, Dec. 1999.
- [6] S. Uchihashi, J. Foote, A. Girgenson 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*.
- [11] 고경철, 이양원, "비디오 분할을 위한 자동 임계치 결정 알고리즘," *한국컴퓨터정보학회논문지*, 제10권, 제6호, pp.65-74, 2005.12.
- [12] 신성윤, 강일고, 이양원, "자막정보를 이용한 축구 비디오 하이라이트 생성," *한국컴퓨터정보학회논문지*, 제10권, 제5호, pp.65-76, 2005.11.

저자 소개



신성운

2003년 2월 군산대학교 컴퓨터학과 이학박사  
2006년~현재 군산대학교 컴퓨터정보학과 전임강사  
<관심분야> 비디오 인덱싱, 비디오 요약, 멀티미디어



표성배

2004년 2월 숭실대학교 컴퓨터학부 공학박사  
1992년~현재 인덕대학 소프트웨어개발과 교수  
<관심분야> 멀티미디어, 비디오 색인화, 비디오 요약



이양원

1994년 8월 숭실대학교 전자계산학과 공학박사  
1986년~현재 군산대학교 컴퓨터정보학과 교수  
<관심분야> 모바일 프로그래밍, 텔레메틱스, 가상현실