韓國 OA學會 論文誌
第7卷 第4號, 2002. 12.

2002-7-4-1-11

# 비디오 데이터베이스 구축을 위하여 장면전환 검출과 샷 클러스터링을 이용한 비디오 개요 추출

표 성 배*

# Video Abstracting Using Scene Change Detection and Sho Clustering for Construction of Efficient Video Database

Sung-bae Pyo*

## 요 약

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 x2 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 x2 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.

* 인덕대학 소프트웨어개발과

논문접수 : 2002. 10. 6
심사완료 : 2002. 12. 7

## Ⅰ. 서론

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 프레임(frame)들의 집합을 나타내며 비디오 정보를 구성하는 유효한 단위이다. 이에 따라서 최근 비디오 샷을 검출하고 샷의 내용을 특성화하기 위한 방법들에 대하여 관심이 집중되고 있다.

수많은 비디오에서 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 사용자들에게 함축적인 내용을 보여주도록 한다. 전체적인 시스템 구조는 그림 1과 같다.

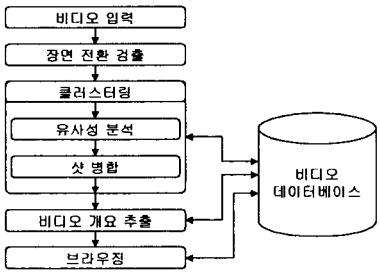


그림 2. 시스템 구조
Fig. 1 System Overview

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

## Ⅱ. 관련연구

비디오 요약 및 개요추출 에서 비디오 시퀀스 자체를 다양한 방법으로 요약하는 연구를 집중 수행하고 있으며 이를 살펴보면 다음과 같다.

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 희소성과 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. 이 방법은 중요한 객체, 사람, 액션, 대화, 타이틀 텍스트 그리고 타이틀 음악의 클립을 선택하기 위한 경험이 있는 물리적 파라미터들을 사용하였는데, 이는 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.

따라서 비디오가 갖는 중요한 정보들을 효율적으로 표현 할 수 있는 새로운 비디오의 요약 및 개요 추출 방법이 요구된다.

## Ⅲ. 장면전환 검출

본 논문에서 제시하는 장면 전환 검출 방법은 컬러 히스토그램과 x2 히스토그램의 장점을 합성한 아래 (식 1)의 방법을 이용하여 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.

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

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.

비디오에서 갑작스러운 장면 전환의 경우는 그 첫 번째 프레임을 키 프레임으로 설정하고 점진적인 장면 전환에서는 장면 전환의 시작 프레임과 끝 프레임 두 개를 키 프레임으로 설정한다.

본 논문에서 사용하는 방법은 갑작스러운 장면 전환과 점진적인 장면 전환을 동시에 검출 가능한 방법으로서, 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 첫 번째 프레임을 포함하여 추출된 각 샷들의 첫 번째 프레임이 된다.

## Ⅳ. 클러스터링 비디오 개요 추출

### 1. 유사성 측정

클러스터링은 장면 전환 검출에 의하여 추출된 샷들에 대하여 유사성 측정에 의한 연속적인 병합을 통하여 유사한 샷들끼리 그룹화를 병합을 수행하는 것이다. 유사성 측정은 샷들을 구성하는 전체 프레임들을 비교하지 않고 샷의 대표 프레임들을 비교하여 수행한다.

본 논문에서는 새로운 샷 병합 방법을 이용하여 클러스터링을 수행하게 되는데, 샷들을 클러스터링하기 위해서는 먼저 추출된 샷들의 키 프레임들에 대한 유사성 측정 방법이 설정되어야 한다. 따라서, 본 논문에서는 컬러 히스토그램 공간에 대한 평균 중심점으로서 기준이 되는 영역에 대한 평균 히스토그램 값을 비교하여 유사성을 측정하게 된다.

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 2와 같이 비디오 프레임을 9개의 영역으로 분할하여 수행한다. 즉, 중심 영역(1)을 기준으로 상(2)·하(4)·좌(5)·우(3)를 포함한 5개의 영역에 대한 히스토그램 평균값을 기준값으로 설정하여 유사성을 측정하기 위한 지역적 히스토그램 차이를 계산하여 비교를 수행하게 된다.

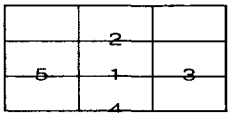


그림 2. 기준 영역
Fig. 2 Base Region

5개의 영역에 대한 히스토그램 평균값을 구하는 방법은 식 2와 같다. (식 2)에서 Ha와 Hb는 이웃하는 두 샷의 키 프레임 a와 b에 대한 히스토그램 평균값을 나타내고 Hai(x.y)와 Hbi(x,y)는 (x,y) 영역에 대한 컬러 히스토그램 값이다.

$$H_a = \frac{\sum_{i=1}^{5} H_{a_i}(x, y)}{5}$$
$$H_b = \frac{\sum_{i=1}^{5} H_{b_i}(x, y)}{5} \qquad \text{(식 2)}$$

(식 3)은 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.

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

결국, 각각의 샷들은 각각의 키 프레임들을 가지며 이 키 프레임들을 대상으로 유사성 측정을 통하여 클러스터링이 수행된다. 그리고, 샷의 내용을 표현하기 위해서 컬러 히스토그램 특징 공간에서 키 프레임들의 기준값을 계산하여 이 기준값을 바탕으로 샷들은 해당 클러스터로 병합되게 된다.

## 2. 클러스터링 방법

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

본 논문에서는 샷의 유사성 측정을 위한 새로운 샷 병합 알고리즘을 이용한 클러스터링 방법을 제시하는데, 전체적인 샷 병합 과정은 그림 3과 같다.

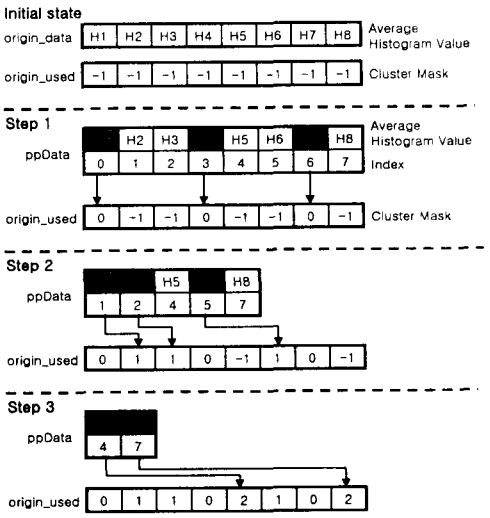


그림 3. 샷 병합 과정
Fig. 3 Shot Merge Process

그림 3의 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 3에서는 8개의 샷들을 대상으로 유사성 측정을 위한 히스토그램 평균값 차이를 계산하여 샷들을 병합하는 과정을 통하여 최종적으로 클러스터 0, 1, 2의 세 개의 클러스터가 생성된 것이다. 또한 비교 회수의 경우도 총 비교회수 12회로 단계별 평균 비교 회수는 4회가 된다.

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

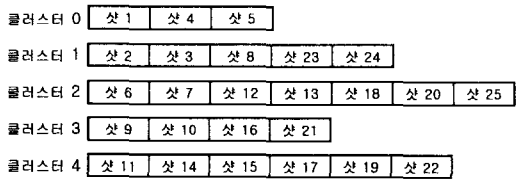


그림 4. 생성된 클러스터 구조
Fig. 4 Created Cluster Structure

### 3. 비디오 개요 추출

비디오 개요 추출은 샷들의 클러스터링을 통해서 생성된 클러스터를 바탕으로 수행된다. 각 클러스터는 특징들이 서로 유사한 샷들을 병합 하여 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 구성하게 된다. 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.

본 논문에서는 대표적인 방법인 클러스터를 구성하는 샷들 중 가장 첫 번째 샷을 대표 샷으로 설정하는 방법을 따르도록 하며 비디오 개요를 구성하는 형태는 다음 그림 5와 같다.

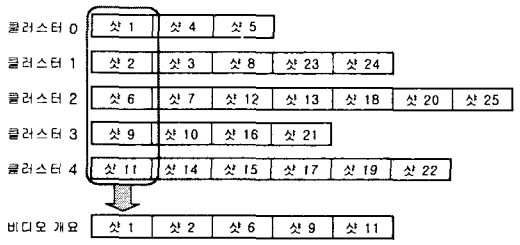


그림 5. 첫 번째 샷을 이용한 비디오 개요의 예
Fig. 5 Example of Video Abstract Using First Shot

위의 그림 5에서 나타내는 것처럼 각 클러스터를 구성하는 첫 번째 샷들로 요약된 비디오를 구성하여 요약된 비디오를 만들어 비디오의 전체적인 내용 파악을 쉽게하고 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 캡쳐 보드로 초당 5프레임을 실험 데이터로 캡쳐하여, 프레임 크기를 200x150으로 정규화 하여 사용하였다.

컬러 히스토그램과 x2 히스토그램을 합성한 방법을 이용하여 장면 전환 검출에 의한 샷의 키 프레임을 추출하는 화면은 그림 6과 같다.

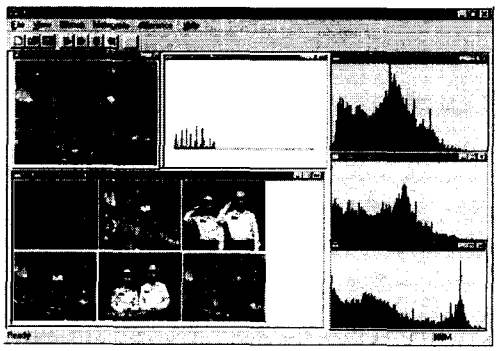

그림 6. 키 프레임 추출 화면
Fig. 6 Window of Key Frame Extraction

이렇게 추출된 키 프레임들은 각각의 샷을 구분하는 기준이 되며 결국, 샷의 키 프레임이 된다. 따라서, 하나의 샷은 이 키 프레임을 시작으로 다음 키 프레임이 나오기 이전까지의 프레임들로 구성되는 것이다.

다음 그림 7은 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 Key Frames

비디오 프레임들을 그림 2에서 나타낸 5개의 기준 영역을 바탕으로 히스토그램을 통한 유사성 측정을 통하여 얻은 클러스터의 예는 다음 그림 8과 같다.

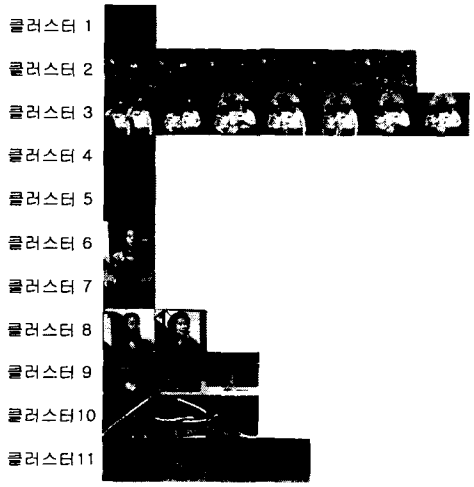


그림 8. 클러스터링 된 비디오의 예
Fig. 8 Example of Clustered Video

본 논문에서는 클러스터링 된 비디오 샷들에서 각 클러스터의 첫 번째 샷을 비디오 개요로 설정하기 때문에 그림 8의 예에서는 클러스터 1부터 클러스터 11까지의 첫 번째 샷들을 전체 비디오 개요로 설정한다.

다음 그림 9은 그림 8의 클러스터에서 추출된 샷 중에서 생성된 비디오 개요는 클러스터의 수와 같은 총 11개의 샷들로 구성된다.

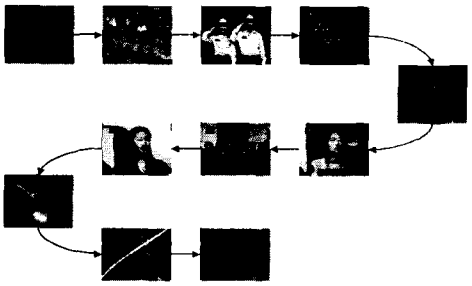

그림 9. 비디오 개요의 예
Fig. 9 Example of Video Abstract

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 1, 2와 같다.

표 1. 키 프레임 추출 결과
Table. 1 Result of Key Frame Extraction

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

표 1의 결과는 컬러 히스토그램과 x2 히스토그램을 합성한 방법을 이용한 장면 전환 검출에 의하여 추출된 키 프레임의 수를 나타낸다.

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 |

표 2에서 키 프레임 수는 장면 전환 검출에 의해 생성된 샷의 수를 나타내며 생성된 클러스터의 수는 비디오 개요를 구성하는 샷의 수를 나타낸다. 위 실험의 결과, 시간적으로 볼 때 5분 분량의 비디오를 3분의 1 정도로 요약되었다.

## Ⅵ. 결론 및 향후 연구 방향

본 논문에서는 장면 전환 검출과 클러스터링 알고리즘을 이용하여 새로운 비디오 개요 추출 방법을 제시하였다. 컬러 히스토그램과 x2 히스토그램의 장점을 합성한 방법을 이용하여 효율적인 장면 전환 검출을 수행함으로서 비디오를 각각의 샷들로 분할하고 각 샷들의 키 프레임을 설정하였고, 이렇게 추출된 각 샷들의 키 프레임들은 새로운 샷 병합 알고리즘을 이용하여 클러스터를 구성하게 되며, 이 클러스터를 바탕으로 비디오 개요를 생성하게 된다.

결론적으로, 본 논문에서는 대용량 비디오의 중요 내용을 한눈에 빠르고 쉽게 이해하여 시청하고자 하는 비디오 선택의 폭을 넓혀줌은 물론이며, 기술적인 측면에서 효율적인 장면 전환 검출 방법과 새로운 클러스터링 방법을 이용하여 비디오 개요를 추출함으로써 비디오 편집 분야에 적용할 수 있는 기반을 마련하였다.

향후 보다 의미적인 장면 전환 검출 방법과 클러스터링 방법을 개발하여 적용하고 다양한 장르의 비디오에 활용 가능한 방법들을 개발한다면 비디오 요약 및 개요 추출 기술을 보다 폭넓게 효율적으로 사용 할 수 있을 것으로 본다.

## 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.

## 저 자 소 개

표 성 배

1979년 숭실대학교 전산학과 졸

1984 - 1990 국방품질연구소
선임연구원

1992 - 현재 인덕대학
소프트웨어개발과 부교수

1997 - 현재 숭실대학교
전자계산전공 박사 수료

관심분야 :

Network security,

Web solution development,

Image processing,,

Multi-media 모델링