

# 개인화된 비디오 요약 서비스를 위한 비디오 스플리터

김원철<sup>o</sup> 황인준

아주대학교 정보통신전문대학원 정보통신공학과  
{wc323, ehwang}@ajou.ac.kr

## Video Splitter for Personalized Video Summary Services

Woncheol Kim<sup>o</sup> Eenjun Hwang

Graduate School of Information and Communication, Ajou University

### 요약

멀티미디어 관련 기술이 발전하고 인터넷 사용이 보편화되면서 모바일 단말기 상에서 비디오 데이터를 검색하려는 요구가 증가하고 있다. 그러나 모바일 단말기의 경우 낮은 CPU 처리율이나 대역폭, 배터리 용량 등의 제약으로 인해 비디오를 그대로 검색하기에는 어려움이 많다. 최근 들어 비디오 데이터의 요약물 통해 모바일 환경의 제약점을 극복하고 효율적으로 비디오를 검색하기 위한 연구가 활발히 진행되고 있다. 본 논문에서는 기존의 단편적인 비디오 데이터 요약 기술에서 벗어나 요약된 비디오 데이터에 특징이나 중요도를 MPEG-7을 이용해서 주석 처리하여 사용자에게 보다 효과적인 검색 환경을 제공하고자 한다. 이러한 요약 방법은 모바일 환경에서 사용자의 우선 순위나 요구하는 특징에 적합한 동영상상을 볼 수 있고 비디오의 전송시 모바일 장비의 성능에 따라 차별적으로 요약 정보를 제공함으로써 모바일 환경의 제약을 상당히 완화시킨다.

### 1. 서론

최근 정보기술과 이동통신 서비스 망의 발전으로 모바일 단말기를 이용하여 빠르고 편리하게 인터넷 사용이 가능하게 되었다. PDA, 휴대폰(cell phone), '스마트폰(smart phone) 등과 같은 모바일 단말기의 가격이 크게 하락하고 있으며 성능이 비약적으로 향상되고 있다. 이러한 추세에 발맞추어 고성능 모바일 단말기를 이용한 멀티미디어 데이터 서비스 요구가 증가하고 있다. 요구하는 멀티미디어 데이터는 대용량 동영상 데이터가 주를 이루고 있다.

그러나 모바일 단말기는 데스크톱 컴퓨터에 비해 상대적으로 낮은 CPU 처리율, 낮은 대역폭, 배터리 용량, 작은 화면 등의 제약을 가지고 있다[1, 2]. 특히 낮은 대역폭과 CPU 처리율은 원활한 멀티미디어 데이터를 제공하려는 노력에 큰 걸림돌이 되고 있다. 모바일 단말기 상에서 비디오 데이터를 효과적으로 재생하기 위해서는 1) 단말기의 CPU 성능 향상, 2) 네트워크 기술의 발달, 그리고 3) 전송되는 비디오 데이터의 요약 기술 등이 선행되어야 한다. 이러한 모바일 단말기 상에서의 제약을 극복하고 효과적으로 비디오를 요약하기 위한 연구들이 활발히 진행되고 있다[3, 4, 5]. 하지만 이러한 비디오 요약은 사용자의 요구를 반영하기보다는 비디오의 각 장면나 형태를 분석해서 비디오를 요약하는 방법이다. 이러한 요약 방법은 자칫 사용자가 보고 싶어하는 비디오의 내용을 간과하기 쉽고, 사용자가 원하지 않는 부분을 보여주게 된다.

본 연구에서는 사용자의 요구와 모바일 단말기의 특징을 고려해 보다 개인화 된 서비스를 제공하고자 한다. 이를 위해 요약된 비디오 데이터는 MPEG-7[6, 7]을 이용하여 그 특징을 주석 처리함으로써 사용자의 요구에 맞게 필터링 할 수 있다. 또한 모바일 단말기의 특징을 고려하여 요약된 동영상상을 검색할 때 요약된 내용을 모두 보여주는 것이 아니라 요

약된 각 샷의 키 프레임만을 보여주고 음성과 문자로 정보를 처리해 장비의 제약을 극복하였다.

본 논문의 구성은 다음과 같다. 2장에서는 그 동안 연구되었던 요약기법과 멀티미디어에 관련된 연구들을 살펴보고, 3장은 본 시스템의 비디오 요약에 관한 기법을 설명하고 4장은 본 논문에서 제공한 시스템에 대한 전체적인 구조를 보여주고 마지막 5장에서는 결론과 향후 계획에 대해 논의한다.

### 2. 관련 연구

비디오 데이터는 다른 데이터들에 비해 크기가 크고 복잡하기 때문에 대용량 기억 장소가 필요하다. 비디오 데이터를 효율적으로 다루기 위해 비정형의 데이터를 구조적으로 재구성을 해야 한다. 동영상을 계층적인 구조로 나누어보면 기본적인 정적 이미지인 프레임이 있다. 프레임들이 모여 샷을 이루고, 여러 개의 샷이 모여 의미적인 단위인 장면(Scene)을 이룬다. 장면들 모여 하나의 비디오 데이터를 구성하게 된다. 샷에 대한 경계 검출[8]을 위한 가장 기본적인 방법은 인접한 프레임간의 화소 값이 얼마나 변했는가를 상대적으로 계산하는 것이다. 즉 프레임간의 화소 값 차이를 계산해서 값이 임계값 이상이면 샷의 경계구역으로 판단하고 구분하는 것이다. 또한 MPEG 형식으로 된 비디오의 경우는 DCT 상수나 움직임 보상 벡터 등을 이용해서 샷 경계를 검출한다. 샷은 비디오 처리의 기본단위로 경계를 검출하기 쉬우나 이는 단지 물리적인 방법으로 경계를 나눈 것이다. 조금 더 의미있는 처리를 위해서는 샷들을 하나의 의미있는 단위인 장면으로 만드는 과정이 필요하다. 동영상의 요약을 위해서는 동영상의 각 샷이나 장면의 중요도를 판단하고 필요없는 부분을 제거해줄 방법이 있어야 한다. 이러한 방법 중의 한가지는

샷 내의 움직임 벡터를 구하거나, 샷간의 컬러 히스토그램[8]을 이용할 수 있다. 현재 여러 대학이나 연구기관들이 멀티미디어에 관한 연구를 하고 있다. 대표적인 시스템으로는 IBM의 QBIC, Carnegie Mellon University의 Informedia, MIT의 Photobook, Virage Logic Corporation의 Virage등이 있다.

### 3. Video Splitter 비디오 요약기술

비디오 서버는 비디오에 대한 요약 수행하고 요약된 비디오 내용에 대해 주석 처리한다. 본 시스템에 대한 요약의 단계는 그림 1과 같다.

첫 번째 단계에서 비디오를 분석하는 분석기는 비디오를 구조적이고 관리 가능하게 비디오의 프레임 유사도를 측정해 유사한 것을 모아 샷으로 구분하게 된다. 본 Video Splitter 시스템에서 샷은 요약 시스템에서 처리할 수 있는 가장 작은 단위로 정의한다. 두 번째 단계에서는 샷으로 구분된 비디오 데이터들이 비디오 요약 처리기로 보내지게 된다. 비디오 요약 처리기에서는 임계 값보다 작은 유사한 샷들을 모으고 이 샷들을 하나의 집합으로 구성한다. 세 번째 단계에서는 비슷한 샷들을 모은 후에 그 집합에서 가장 길이가 긴 샷을 그 집합을 대표할 수 있는 대표 샷으로 선택 하게 된다. 네 번째 단계에선 선택된 대표 샷들 중에 길이가 1.5초 이하인 것을 찾아 그 그룹을 삭제한다[5]. 대표 샷은 그 그룹 중 제일 긴 샷이기 때문에 대표 샷의 길이가 1.5초 이하라는 것은 모든 샷들이 1.5초 이하라는 것을 뜻한다. 다섯 번째 단계에선 대표 샷들을 시간 순으로 정렬을 한 후에 그 대표 샷 이외의 부분을 제거함으로써 1차적인 요약을 마무리한다

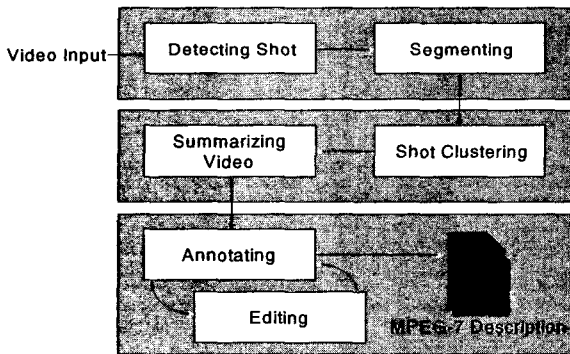


그림 1. 비디오 요약 서버

1차적인 요약 즉 물리적인 요약을 마무리한 후에 요약된 각 샷의 특징을 주석처리를 함으로써 보다 의미 있는 정보를 포함한다.

### 4. Video Splitter 시스템

그림2는 비디오 요약시스템에 대한 전체적인 구성을 보여준다. Video Splitter는 클라이언트 장비에 대해 유연성이 있고 사용자의 요구를 보다 효과적으로 처리할 수 있도록 설계하였다. MPEG-7을 사용하였고 비디오 프록시는 클라이언트와 서버간의 동기 및 비디오 데이터를 클라이언트에 스트리밍하며 자막 및 클라이언트 장치의 상태에 따라 필요한 사항을 지원해준다. 비디오 서버는 비디오 데이터를 요약하고 요약된 각 샷에 의미적인 정보를 주석 처리하여 의미적인 요약

한다.

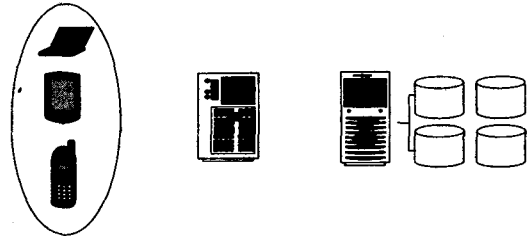


그림 2. Video Splitter 구성도

#### 4.1 요약 동영상을 위한 스키마

비디오 서버는 비디오를 분석하여 요약을 하며 요약된 비디오에 보다 의미적인 정보를 제공하기 위해 주석 처리한다. 이때 사용되는 XML 스키마는 그림 3과 같다.

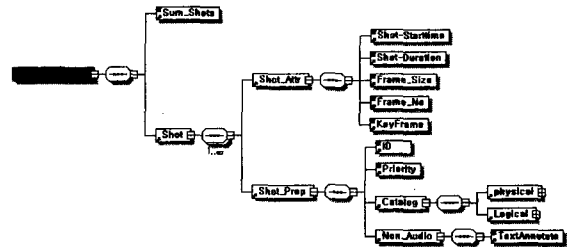


그림 3. 요약 동영상을 위한 스키마

스키마상의 Sum\_Shots은 요약된 비디오 샷에 대한 전체 개요이고 Shot은 각 요약된 샷들에 대해 주석 처리하는 부분이다. Shot\_Attr은 샷의 속성을 나타내는 부분이다. Frame\_No은 샷에 포함되는 프레임의 수를 나타낸다. 샷에 포함된 프레임 중에서 키 프레임을 표시하는 KeyFrame이 있어 키 프레임을 효율적으로 표현해준다. Shot\_Prop은 샷에 대한 특징을 나타내는 부분이다. Priority는 샷이 어느 정도의 우선순위를 가지고 처리할 수 있는가를 나타내고 각 샷에 대한 특징을 나타내는 Catalog가 있다. 이는 물리적인 특징과 논리적인 특징을 표시하기 위해 각각 Physical, Logical을 가지고 있다. Non\_Audio은 사용자의 장비가 오디오 기능을 가지고 있지 않을 경우 더 많은 문자 정보로 처리하기 위해 사용된다. 사용자마다 요구하는 요약의 정도가 다르고 보고자 하는 대상도 다르다. 모든 사용자에게 일방적으로 보여주는 것보다는 사용자가 요구하는 요약의 정도를 주석으로 표시해서 사용자의 성향에 따라 요약 정도를 맞추어 주게 된다. 요약된 샷에 주석을 달아 보다 쉽게 사용자에게 요약된 비디오를 제공해준다. 만일 사용자가 휴대폰으로 요약된 데이터를 요구할 경우에 요약된 샷들을 모두 보여주는 것은 단말기에 무리가 따르게 된다. 따라서 이 경우를 위해서 요약된 샷들에 키 프레임들만을 나타내는 KeyFrame에 나타난 프레임들만 휴대폰에 제공해 주게 된다.

#### 4.2 비디오 프록시

비디오 프록시는 클라이언트의 요청을 받아 사용자의 요구 사항에 맞게 처리해주는 부분으로 내부 모듈은 다음과 같다.

##### 1) Profile Analysis

사용자 및 장비의 프로파일을 분석하고 사용자의 요구가 이

상이 없는지 판단하고 본 시스템에서 사용하는 질의 언어로 변환한 후 클라이언트 요구 분석 결과와 클라이언트 장비에 대한 정보를 Device Manager에 전달한다.

2) Device Manager

Device Manager는 장비와 사용자의 요구사항에 대한 정보를 가지고 있으며 요약된 동영상을 필터링 할 때 사용자의 요구를 분석해서 Filter Manager로 보내주는 역할을 하고 요구분석결과를 Filter Manager와 동기화 작업에 대한 정보를 제공해 준다.

3) Filter Manager

Data Request 모듈에서 서버에 비디오 데이터를 요구한 후 서버로부터 전송되는 비디오의 각 샷에 대해 Device Manager에서 제공해준 사용자의 우선순위 및 요구사항 정보에 맞게 필터링 해준다. 이때 판단은 MPEG-7주석과 비교하여 판단하게 된다.

4) Process Manager

Process Manager는 필터링 된 비디오와 Device Manager를 통해 클라이언트의 장비를 파악한 후 클라이언트의 장비가 휴대용인 경우 비디오 샷을 전부 보내줄 필요없이 각 샷의 키 프레임만을 선택해서 클라이언트에 보내준다. 이때 Sync Manager가 자막과 음성에 대한 동기화를 제공해준다.

5) Stream Manager

Stream Manager는 멀티미디어 디바이스에 공통적으로 적용될 수 있는 제어 인터페이스를 가지고 있으며 디바이스간의 연결 설정시 송수신 디바이스간 미디어 형식을 지원해준다.

4.3 클라이언트

클라이언트는 자신의 선호사항이나, 요약의 우선순위, 장비의 정보가 포함되어 있는 클라이언트 프로파일을 가지고 비디오에 대한 요청을 한다. 클라이언트 프로파일에 대한 XML 스키마는 그림 4와 같다. 스키마상에서 user 는 사용자의 프로파일을 저장하기 위한 부분이다. caption 은 사용자가 요약된 비디오를 받을 때 자막 보기를 원하는가를 표시해준다.

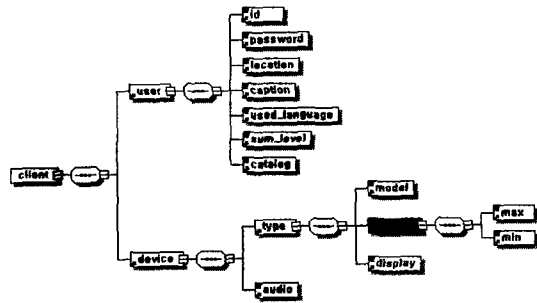


그림 4. 클라이언트 프로파일 스키마

스키마상의 sum\_level은 요약된 비디오를 받을 때 요약의 정도를 나타내는 것이다. catalog부분은 요약할 때 어떤 장면을 위주로 요약할 해야 하는지에 대한 정보를 나타내고 있다. device부분은 장비에 대한 프로파일로 장비의 모델을 나타내는 model이 있다. 사용할 수 있는 battery의 최대 용량과 최소 용량을 나타내는 max, min이 있고, 장비가 오디오 기능의 정보를 나타내는 audio가 있다. 그림 5는 비디오 요약시스템의 내부적인 구조를 나타낸다. 클라이언트의 프로파일을 비디오 프록시가 보고 처리후 사용자에게 맞게 개인화 된 서버

스를 제공해준다.

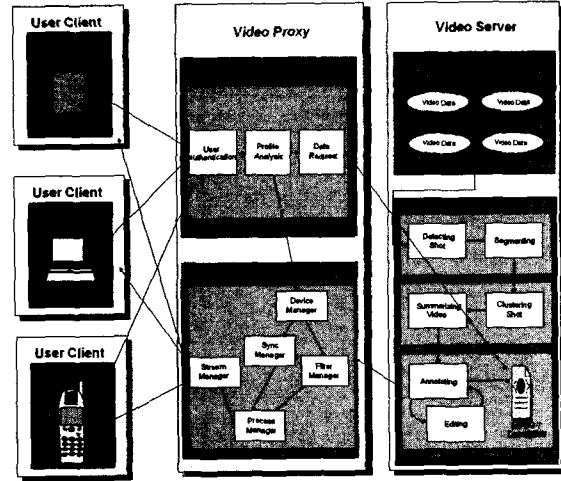


그림 5. Video Splitter 전체 모듈

5. 결론 및 향후 연구 과제

멀티미디어 관련 기술과 인터넷 속도의 향상으로 많은 비디오 데이터들에 대한 사용요구가 증가했다. 그러나 이러한 모든 정보를 검색하기란 불가능하다. 따라서 요약된 정보를 보고 전체적인 흐름을 파악하여 이 콘텐츠가 처음부터 순차적으로 볼 만한 것인가를 판단함으로써 많은 시간을 단축시킬 수 있고 클라이언트 장비에 따라 요약 비디오 데이터를 차별적으로 제공함으로써 각 장비에 대한 제약사항을 극복할 수 있다. 본 논문은 개인화 비디오 요약 시스템으로써 클라이언트 장비에 따라 차별적으로 요약된 정보를 제공해 줄 수 있다. 비디오 내용요약 한 후 요약된 각 샷에 대하여 주석을 줌으로써 사용자 요구를 보다 다양하게 받아들일 수 있다.

향후 계획은 보다 편리한 사용자 인터페이스를 개발하고 여러 종류 클라이언트에 효율적으로 처리할 수 있는 방법을 개발하는 것이다.

6. 참고문헌

- [1]T. Kamba, S. A. et al., "Using small screen space more efficiently", Proceedings of CHI'96, ACM Press, 1996
- [2]O.Buyukkokten, H. et al., "Power Browser: Efficient Web Browsing for PDAs" In Proceedings of CHI'2000, ACM Press
- [3]Uchihashi, S. and Foote, J., "Summarizing Video Using a Shot Importance Measure and a Frame-Packing Algorithm", Proc. ICASSP '99, Vol. 6, pp. 3041-3044, 1999
- [4]Jeho Nam, Ahmed H. Tewfik, "Dynamic Video Summarization and Visualization "Proceedings of the seventh ACM international conference on Multimedia October 1999
- [5]Yihong Gong and Xin Liu, "Summarizing video by minimizing visual content redundancies" IEEE 2001
- [6]Nack, F. Lindsay, A.T. "Everything you wanted to know about MPEG-7. 2" IEEE Multimedia, Vol. 6 No. 4, Oct-Dec. 1999, pp. 64 -73
- [7]http://www.mpeg.org/
- [8]Yihong Gong, and Xin Liu, "Video Shot Segmentation and Classification" Research Laboratories, NEC USA, Inc. 110 Rio