

MPEG-7을 기반으로 한 뉴스 동영상 스키마 및 샷 종류별 키프레임을 이용한 요약 생성 방법

(A Scheme for News Videos based on MPEG-7 and Its
Summarization Mechanism by using the Key-Frames of Selected
Shot Types)

정진국[†] 심진선[†] 낭종호^{**}

(Jin-Guk Jeong) (Jin-Sun Shim) (Jong-Ho Nang)

김경수^{***} 하명환^{***} 정병희^{***}

(Gyung-Su Kim) (Myung-Hwan Ha) (Byung-Hee Jung)

요약 최근 정형화된 구조를 갖는 뉴스 동영상 비디오에 대한 아카이브 시스템을 구축하기 위한 많은 연구가 진행되어 왔다. 그러나 기존의 시스템에서는 기사에 대한 메타 데이터를 저장하는 방법이 서로 다르기 때문에 이런 데이터 사이의 호환성이 없다는 문제점을 가지고 있다. 본 논문에서는 이런 문제점을 해결하기 위하여 멀티미디어 내용 정보를 표현하기 위한 표준인 MPEG-7 MDS에 바탕을 둔 뉴스 동영상 스키마를 제안하였으며, 또한 기사를 이루는 각 샷들의 특징을 반영한 기사 요약 방법을 설계하고 구현하였다. 본 논문에서 제안한 뉴스 동영상을 위한 스키마는 뉴스 구조의 특성을 반영하여 MPEG-7 MDS의 VideoSegment, TextAnnotation 등과 같은 스키마를 이용하여 설계하였고, 요약 방법에서는 요약 비디오 데이터의 크기를 줄이기 위하여 샷들의 키프레임들만을 해당 오디오와 함께 상영하는 슬라이스 쇼 방법을 사용하였다.

키워드 : MPEG-7, 브라우징, 멀티미디어, 요약

Abstract Recently, there have been a lot of researches to develop an archive system for news videos that usually has a fixed structure. However, since the meta-data representation and storing schemes for news video are different from each other in the previously proposed archive systems, it was very hard to exchange these meta-data. This paper proposes a scheme for news video based on MPEG-7 MDS that is an international standard to represent the contents of multimedia, and a summarization mechanism reflecting the characteristics of shots in the news videos. The proposed scheme for news video uses the MPEG-7 MDS schemes such as VideoSegment and TextAnnotation to keep the original structure of news video, and the proposed summarization mechanism uses a slide-show style presentation of key frames with associated audio to reduce the data size of the summary video.

Key words : MPEG-7, Browsing, Multimedia, Abstraction

· 이 연구는 2001년도 서강대학교 교내 연구비 지원에 의하여 이루어졌음.

[†] 학생회원 : 서강대학교 컴퓨터학과
jjguk@mlneptune.sogang.ac.kr
icestar@unitel.co.kr

^{**} 종신회원 : 서강대학교 컴퓨터학과 교수
jhnang@ccs.sogang.ac.kr

^{***} 비회원 : KBS 기술연구소 연구원
odyssey@kbs.co.kr
mhha@kbs.co.kr

bhjung@kbs.co.kr
논문접수 : 2002년 2월 18일
심사완료 : 2002년 6월 26일

1. 서론

인터넷 및 컴퓨터 기반 기술이 발달함에 따라 디지털 TV 방송, 디지털 라이브러리 시스템, 전자 프로그램 가이드 등 사용자에게 다양한 서비스를 제공하는 멀티미디어 시스템들이 나오게 되었다. 특히 아카이브 시스템이나 요약 시스템과 같은 시스템들은 동영상에 대한 검색, 요약과 같은 서비스를 제공함으로써 해서 사용자가 전체 동영상을 보지 않고도 동영상의 내용을 확인할 수 있다.

록 하기 때문에 많은 연구가 진행이 되어 왔다. 이러한 시스템들은 다른 도메인에 비해 형식이 잘 갖추어져 있고 자료 보관 가치가 높은 뉴스 도메인에 한정되어서 구축이 되는 경우가 많다. 하지만 이러한 시스템에서 이용하는 자료들은 자료 구조나 데이터베이스 엔진이 서로 다르기 때문에 호환성을 갖지 못한다. 따라서 이러한 자료들을 교환하고 공유하기 위한 자료 구조와 저장에 대한 표준이 필요하다.

ISO/MPEG 그룹은 오디오/비디오 데이터에 대한 구조와 메타데이터를 기술하기 위한 표준인 MPEG-7[1]을 만들었다. MPEG-7은 멀티미디어에 관련된 각종 정보들을 기술하기 위한 표준이다. 하지만 이러한 정보들을 어떻게 얻어내고, 어떻게 메타데이터를 생성하며, 이런 기술 정보를 어떤 목적으로 사용할 것인가에 대한 내용은 포함되지 않는다. 즉, MPEG-7을 사용하는 시스템에서는 각 시스템 목적에 맞게 Description을 이용하면 되는 것이다. MPEG-7 MDS(Multimedia Description Scheme)[2]는 MPEG-7의 한 부분으로서 멀티미디어 생성에 관한 정보(제목, 장르, 만든 사람들에 대한 정보), 멀티미디어 내용의 고급 수준 정보(육하원칙에 따른 사건 설명), 저급 수준 정보(색상, 움직임, 카메라 기법 등) 등을 포함하고 있다.

본 논문에서는 MPEG-7 MDS를 기반으로 하여 뉴스 동영상을 위한 스키마를 설계하고 이러한 스키마에 맞는 데이터를 생성, 요약할 수 있는 도구를 구현하였다. 요구 사항을 분석한 결과 좀 더 다양한 질의를 서비스할 수 있도록 특정 단위로 인덱싱 되고 적절히 저장되어야 하며 분류가 되어야 함을 알 수 있었고, 또한 서비스 제공자가 뉴스 데이터를 효율적으로 관리하기 위해서는 동영상 파일의 위치나 제작에 참여한 사람들에 관한 정보를 유지해야 함을 알 수 있었다. 이런 조건들을 만족하기 위해 본 논문에서는 MPEG-7 MDS의 Video Segment, TextAnnotation, KeywordAnnotation, MediaInformation, Creation Information, Classification 등과 같은 스키마를 이용한다. 특히 본 논문에서 제안하는 요약 방법을 적용하기 위하여 Sequential Summary DS와 HierarchicalSummary DS를 이용하였다. 본 논문에서 제안한 MPEG-7 기반의 뉴스 동영상 스키마와 요약 생성 방법은 서로 다른 시스템간의 호환성을 보장하면서 효율적인 요약 기법을 제공할 것이다.

다음 장에서는 기존의 뉴스 아카이브 시스템에 대한 연구와 MPEG-7 MDS에 대해서 간략히 살펴본다. 3장에서는 MPEG-7 MDS를 이용하여 뉴스 동영상 스키마를 설계하고 요약 생성 방법에 대해 연구하고 4장에서

는 3장에서 제안한 스키마와 요약 방법을 이용하여 시스템을 구현하고 MPEG-7이 어떻게 유용하게 쓰일 수 있는지 분석한다. 마지막으로 5장에서는 MPEG-7과 뉴스 아카이브 시스템에 관련한 향후 연구를 제안하고 결론을 맺는다.

2. MPEG-7 MDS의 개요

MPEG-7[3]은 풍부한 오디오/비디오 정보의 기술자들을 제공함으로써 텍스트 기반의 검색 엔진보다 더 효율적이고 효과적으로 내용을 검색하고 브라우징 할 수 있도록 한다. 또한 MPEG-7은 다양하고 많은 환경에서 여러 가지 응용 시스템을 만들 수 있도록 유동적이고 확장 가능한 구조를 제공한다. MPEG-7은 7개의 파트로 구성이 되어 있다. 본 논문에서는 뉴스 동영상의 전체적인 구조를 다루고 있으므로 일반적인 특성(Feature)과 멀티미디어 기술 방법을 다루고 있는 MPEG-7 MDS를 이용한다. 특히 이번 절에서는 여러 가지 DS(Description Scheme)들 중에서 인덱싱 된 정보를 기술하는 VideoSegment DS와 요약 생성에 기반이 되는 Summarization DS에 대해서 살펴본다.

Segment DS는 멀티미디어 내용의 구조를 기술하기 위한 도구로서 기술하고자 하는 멀티미디어의 종류에 따라 StillRegion DS, ImageText DS, Mosaic DS, VideoSegment DS, AudioSegment DS, InkSegment DS 등으로 세분화된다. 이 중에서 VideoSegment DS는 하나의 프레임뿐만 아니라 전체 비디오 시퀀스와 같이 임의의 시퀀스에 대응될 수 있는 비디오 데이터 세그먼트를 기술한다. VideoSegment DS가 포함하고 있는 요소들 중에서 SpatialDecomposition은 비디오 세그먼트를 하나 이상의 세그먼트를 공간적으로 나누는 것을 의미하고 TemporalDecomposition은 시간적으로 나누는 것을 의미한다. MediaTime은 비디오 데이터의 시작 시간과 지속되는 시간을 규정하며 전체 비디오를 나타내는 비디오 세그먼트인 경우에는 생략할 수 있다. 대표 프레임들은 Mosaic이나 keyFor 관계로서 표현할 수 있다.

Summarization DS는 요약된 오디오/비디오 내용을 기술할 때 사용이 되며 오브젝트, 사건 등 다양한 관점에서 기술된다. Summarization DS에서는 오디오/비디오 내용에 대한 링크를 포함함으로써 많은 오디오/비디오 데이터를 저장하지 않고도 여러 버전의 요약을 생성하고 저장할 수 있도록 하며 이러한 요약에 대한 기술은 원본의 데이터와 분리되어서 저장될 수도 있다. Summarization DS는 크게 SequentialSummary와 HierarchicalSummary로 분류할 수 있다. SequentialSummary DS는 비선형

적이나 순차적인 탐색을 위해 오디오/비디오 요약에 대해 규정하는 DS로 오디오, 비디오, 관련된 텍스트 정보로 구성된다. 다양한 레벨에서 오디오/비디오를 요약할 수 있는 구성을 제공하는 HierarchicalSummary DS는 HighlightSegment DS로 기술되고 오디오/비디오의 시간으로 정의된 세그먼트들로 구성된다. 각각의 Highlight Segment는 관련된 키-비디오클립, 키-오디오클립, 키프레임, 키사운드에 대한 위치를 포함한다. 또한 각 Highlight Segment는 주제를 기술하기 위한 텍스트 정보를 포함할 수 있다

3. 뉴스 스키마 설계와 요약 생성 방법

본 장에서는 뉴스 동영상상을 일반 사용자에게 서비스하고 그에 대한 정보를 효율적으로 관리하기 위해 필요한 뉴스 동영상상에 대한 데이터 구조와 메타데이터를 분석하여 이를 토대로 MPEG-7 MDS를 이용한 뉴스 동영상 스키마를 설계한다. 또한 전체 뉴스를 보지 않고도 빠르고 효율적으로 뉴스를 이해할 수 있는 요약과 브라우징에 대한 방법을 제안한다.

3.1 뉴스 스키마 설계

뉴스 프로그램은 계층적으로 구성이 되어 있고 각 단계별로 메타데이터에 대한 정보를 유지한다. 이와 같은 계층적인 구조와 메타데이터를 효율적으로 관리하기 위해 본 논문에서는 MPEG-7 MDS를 사용한다. 이번 절에서는 MPEG-7 MDS를 이용하여 설계한 뉴스 동영상 스키마에 대해 자세히 설명한다.

3.1.1 전체 스키마 구조

일반적으로 동영상은 신호 샷으로 이루어져 있다. 그러나 뉴스에서는 하나의 샷이 하나의 신을 구성하는 경우가 많고 샷들 사이의 의미적인 연결성이 부족하다. 그리고 신의 의미보다는 기사의 의미를 더 자주 이용하게 된다. 따라서 본 논문에서는 신의 단계를 이용하지 않고 뉴스를 낱짜별 뉴스, 기사, 샷 이렇게 3단계의 계층적인 구조로 구분한다. 또한 동영상의 기본 단위인 샷은 보도의 유형에 따라서 앵커, 그래픽, 기사, 인터뷰, 회견/연설, 일반적으로 분류[4] 한다.

세그먼트, 분류, 데이터 관리, 요약에 대한 전체적인 관계와 구조는 <그림 1>과 같다. 사각형은 어떤 내용 정보를 기술하는지에 대해서 설명하며 괄호 안은 MPEG-7 MDS 내의 어떤 스키마를 이용하여 기술하는지를 나타낸다. 그림을 보면 알 수 있듯이 앞에서 설명한 뉴스의 구조, 즉 방영 낱짜별 뉴스, 기사, 샷이 반영이 되어 있다. 정적으로 생성되는 요약은 기사 단위로 이루어지게 되고 이를 연결하면 전체 뉴스를 요약하는

결과를 얻을 수 있다.

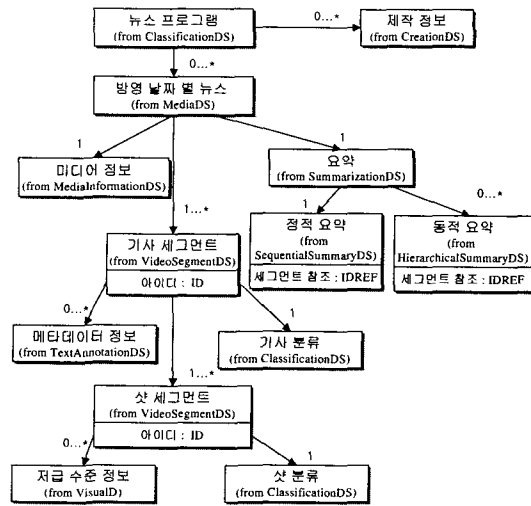


그림 1 뉴스 동영상 스키마 다이어그램

3.1.2 기본적인 데이터 관련 스키마

MPEG-7 MDS에서 내용의 구조를 표현하기 위한 스키마는 SegmentDS이다. 따라서 본 논문에서는 뉴스의 계층적인 구조를 표현하기 위해 Segment DS를 이용한다. 세그먼트[5]란 멀티미디어 데이터의 공간적 혹은 시간적인 단위를 의미하는 것으로서 예를 들어 비디오 세그먼트에 대해서는 시간상으로 특정 범위에 해당하는 프레임들이나 샷 등을 가르킨다. 뉴스는 앞 절에서 보는 바와 같이 기사, 샷으로 구성이 되므로 본 논문에서는 세그먼트 단위를 기사와 샷으로 한정한다. MPEG-7에서 세그먼트를 구분하는 기준은 공간적 구분(Spatial Decomposition)과 시간적 구분(Temporal Decomposition)으로 구별할 수 있는데 세그먼트 단위인 기사와 샷은 시간상으로 구분하는 것이므로 세그먼트에 대한 구분은 시간을 기준으로 서로 겹치지(Overlap) 않게 한다.

뉴스의 기사와 샷을 대표하는 프레임을 기술하기 위해서 본 논문에서는 새로운 타입을 추가한다. 물론 대표 프레임은 MPEG-7 MDS에서도 모자이크(mosaic)이나 관계 KeyFor를 이용하여 기술할 수 있다. 그러나 본 논문에서 이용되는 대표 프레임은 각각의 세그먼트당 하나이며 세그먼트를 보조하는 수단에만 사용되므로 위의 두 방법은 너무나 복잡하다. 따라서 대표 프레임을 기술하기 위해 <그림 2>에서 보는 바와 같이 Hierarchical Summary의 KeyFrame 타입을 추가하였다.

```
<complexType name="VideoSegmentType">
  <sequence>
    <element name="MediaInformation" type="mpeg7:MediaInformationType"
      minOccurs="0"/>
    <element name="MediaTime" type="mpeg7:MediaTimeType"/>
    <element name="KeyFrame" type="mpeg7:ImageLocatorType"/>
    <element name="TemporalDecomposition"
      type="news:VideoSegmentTemporalDecompositionType"
      maxOccurs="unbounded"/>
    ... <!-- More Types -->
  </sequence>
  <attribute name="id" type="ID" use="optional"/>
</complexType>
```

그림 2 비디오 세그먼트 스키마 예

기사의 제목과 스크립트, 키워드는 MPEG-7 MDS의 세그먼트에서 이용한 방법과 같이 TextAnnotationDS를 사용한다. TextAnnotationDS는 네 가지 데이터 타입인 FreeTextAnnotation, StructuredAnnotation, DependencyStructure, KeywordAnnotation 을 제공한다. 제목은 문장 길이가 짧고 구조적이거나 서로 종속되어 있지 않기 때문에 FreeTextAnnotation의 데이터 타입을 사용한다. 스크립트는 그 양이 많기 때문에 FreeTextAnnotation 타입이외의 다른 타입으로는 기술할 수 없다. 그러나 스크립트는 일반적인 텍스트와는 다른 특별한 의미를 가진다고 볼 수 있으므로 'Script' 타입을 항목을 추가하였다. 이러한 스키마를 사용하여 하나의 기사를 <그림 3>과 같이 기술한다.

```
<VideoSegment id="Article1">
  <KeyFrame>
    <MediaUri>pkbs_20000309_1302.jpg</MediaUri>
  </KeyFrame>
  <TextAnnotation>
    <Script>
      ©김종진 앵커:정부의 대북정책기조가 달라집니다. ...
    </Script>
  </TextAnnotation>
  <TextAnnotation type="NewsClassificationCS:News:Section">
    <FreeTextAnnotation>정치</FreeTextAnnotation>
  </TextAnnotation>
  <!--중략-->
</VideoSegment>
```

그림 3 기사 기술 예

기사와 샷의 타입을 직접 기술할 수 있는 것이 MPEG-7 MDS에는 없다. 그러나 MPEG-7 MDS의 세그먼트에서와 같이 TextAnnotation에 타입 속성을 지정함으로써 현재 기술하는 것이 어느 타입에 속하는

지 표현할 수 있다. 이러한 타입을 <그림 4>와 같이 정의하고 <그림 3>의 '정치'를 기술한 것과 같이 표현한다. 그러나 이것만으로는 기사와 샷의 타입 분류가 정확히 어떠한 것인지 나타낼 수 없다. <그림 3>의 'type' 속성이 참조한 것이 어떻게 용어 정의를 했는지 알아야 한다. 이를 위해서 <그림 5>와 같이 기사와 샷 타입의 분류를 정의한다.

```
<element name="TextAnnotation" minOccurs="0" maxOccurs="unbounded">
  <complexType>
    <complexContent>
      <extension base="mpeg7:TextAnnotationType">
        <attribute name="type" use="optional">
          <simpleType>
            <union memberTypes="mpeg7:termReferenceType string"/>
          </simpleType>
        </attribute>
      </extension>
    </complexContent>
  </complexType>
</element>
```

그림 4 샷과 기사 타입을 표현하기 위한 타입 정의

```
<ClassificationScheme>
  <Term termId="News">
    <Name>뉴스</Name>
    <Definition>일반적인 뉴스 프로그램</Definition>
  <Term termId="Section">
    <Name>섹션</Name>
    <Definition>기사의 종류</Definition>
    <Term termId="정치"><Name>정치</Name></Term>
    <Term termId="경제"><Name>경제</Name></Term>
    <!--중략-->
    <Term termId="스포츠"><Name>스포츠</Name></Term>
  </Term>
  <!--샷의 분류 중략-->
</ClassificationScheme>
```

그림 5 기사 분류 스키마 문서

3.1.3 요약 스키마

MPEG-7 MDS는 요약을 위한 스키마로서 Hierarchical SummaryDS와 SequentialSummary DS를 제공한다. HierarchicalSummaryDS는 다양한 단계의 계층적인 요약을 할 수 있게 하며 SequentialSummaryDS는 비선형적이거나 순차적인 요약을 할 수 있게 한다. 3.1.2에서 설명한 특정한 주제에 의한 요약은 Hierarchical SummaryDS에서만 주제에 따라 요약할 수 있도록 하므로 이 스키마를 이용한다. 이 요약은 사용자의 요구에 따라 곧바로 생성이 되는 것이므로 본 논문에서는 동적으로

생성되는 요약이라고 한다.

미리 생성되는 요약은 전체를 잘 표현할 수 있어야 한다. 계층적으로 요약을 하는 것은 너무나 다양한 결과를 생성할 수 있는 반면에 미리 생성되는 요약은 그에 비해 적은 양의 요약을 생성하게 되므로 보다 객관적인 형태의 요약이 되어야 한다. 본 논문에서는 Sequential SummaryDS를 이용하여 기술한다. 또한 본 논문에서는 이러한 요약을 미리 생성해 놓은 요약이라는 의미에서 정적으로 생성된 요약이라 부른다. 요약을 어떻게 할 것인지에 대한 것은 다음절에서 자세히 설명한다.

3.2 요약 생성 방법

어떤 사용자들은 동영상의 처음부터 상영하여 보지 않고 특정 부분부터 보기를 원한다. 하지만 단지 시간이나 텍스트 정보만으로는 사용자가 원하는 부분을 잘 선택할 수 없다. 따라서 비주얼한 정보가 제공되어야 하는데 그러한 인터페이스 중에서 가장 많이 사용하는 방법이 대표 프레임의 기반으로 한 사용자 인터페이스[6]이다. 대표 프레임을 선택하는 방법[7]은 응용 시스템의 요구 사항에 따라 다르게 되는데 본 논문에서는 뉴스 동영상의 정형화된 특징을 이용하여 대표 프레임을 선택하였다. 이번 절에서는 본 논문에서 제안하는 대표 프레임을 선택하는 방법과 정적으로 생성되는 요약 방법, 동적으로 생성되는 요약 방법에 대해서 설명한다.

3.2.1 샷의 대표 프레임 선택

기존의 대부분의 연구에서는 대표 프레임을 선택할 때, 색상 정보나 움직임 정보 등 저급 수준의 정보의 비교를 통해서 선택을 하거나 일률적으로 시작 프레임이나 중간 프레임들을 선택한다. 그러나 이러한 대표 프레임 선택은 중요한 정보를 나타내는 자막을 놓치는 경우가 많다. 본 논문에서는 샷의 타입을 이용하여 대표 프레임을 선택하는 방법을 제안한다. 샷의 타입에 따라 샷의 프레임들의 변화를 크게 처음 부분, 중간 부분, 끝 부분으로 나누어서 살펴보고 대표 프레임 선택 방법을 살펴보면 다음과 같다.

가. 앵커 샷

앵커 샷은 뉴스의 샷 중에서 가장 변화가 없는 정지 영상에 가까운 샷이다. <그림 6>에서 보는 바와 같이 처음 부분이나 중간 부분, 끝 부분의 이미지가 거의 같다. 따라서 어떤 프레임이 다른 프레임들보다 더 중요하다라고 할 수 없으며 대표 프레임은 어떤 프레임을 선택하든지 상관없다. 다만 점진적인 샷 변화(Gradual Change)에 의해서 시작 프레임과 마지막 프레임, 그 주변의 프레임들은 두 개 이상의 이미지가 겹쳐 보여 대표 프레임으로서 적합하지 않으므로 중간 프레임을 선택한다.



그림 6 앵커 샷의 프레임 변화

나. 그래픽 샷

그래픽 샷은 뉴스의 여러 샷들 중에서 가장 뚜렷한 특징을 보인다. 그래픽 샷은 도표나 그래프 등을 이용하여 정보들을 단계적으로 보여주므로 <그림 7>에서 보는 바와 같이 끝 부분의 프레임에 표현하고자 하는 모든 정보를 포함한다. 다만 날씨 예보와 같이 중부 지방의 날씨, 남부 지방의 날씨, 전국의 기온 등 여러 정보가 병렬 관계를 이루는 경우에는 끝 부분의 프레임 하나만으로는 전체 정보를 표현할 수 없지만 이러한 샷은 극히 일부이므로 무시할 수 있다. 따라서 그래픽 샷에서는 마지막 프레임을 대표 프레임으로 선택을 한다.



그림 7 그래픽 샷의 프레임 변화

다. 기자, 인터뷰, 회견/연설

기자, 인터뷰, 회견/연설 샷은 한 사람이 말을 하는 장면을 표현하며 거의 대부분 한 화면에는 한 사람의 모습만 나타내게 된다. 그리고 사람의 이름이나 소속 단체, 직업 등은 자막(Closed Caption) 처리가 되어 화면의 상단이나 하단에 보여지게 된다. 그러나 자막이 모든 프레임에 나타나는 것은 아니다. <그림 8>에서 보는 바와 같이 기자 샷의 처음 부분에는 자막이 없다가 10프레임 이상이 지난 후에 자막을 보여주고 끝 부분에는 자막이 사라지게 된다. 인터뷰, 회견/연설 샷도 같은 모양으로 나타난다. 자막에는 그 사람에 대한 부가적인 정보가 있으므로 자막이 보이는 프레임이 보이지 않는 프레임보다 더 중요하다고 볼 수 있다. 따라서 이러한 샷에서는 가장 가운데 프레임을 대표 프레임으로 선택을 한다.



그림 8 기자 샷의 프레임 변화

라. 일반 샷

일반 샷은 단지 부가적인 비주얼한 정보를 표현하기 위한 것이 대부분이다. 가끔 자막 처리로 샷의 부가적인 정보를 표현하긴 하지만 자막이 언제 나타나고 어느 위치에 나타나는지 정형화된 형태는 없다. 그러나 자막이 나오는 시간은 일정 시간 이상 되므로 자막이 나오는 시간(처음에 자막이 있다가 중간에 사라지거나 중간에 자막이 나오는 경우)에 상관없이 중간 부분에서는 항상 자막을 보여주게 된다. 따라서 자막이 없는 샷에서는 어떤 프레임이 더 중요하다고 볼 수 없으나 자막이 나오는 일부의 샷을 위해서 중간 부분의 프레임을 선택한다.

3.2.2 정적으로 생성되는 요약

멀티미디어에서 요약은 원래의 동영상이 가지는 주요 내용은 유지하는 상태에서 전체 동영상의 크기를 줄이는 것을 말한다. 예를 들어서 중요한 오디오 클립(Key-AudioClip), 비디오 클립(Key-VideoClip), 대표 프레임(Key-Frame)[1] 등이 될 수 있다. 이러한 요약 요소 중에서 어떤 내용을 선택해서 어떤 방식으로 요약 정보를 보여줄 지는 응용에 따라 다르다. 본 절에서는 슬라이드 기법을 이용하여 뉴스를 요약하고 MPEG-7 MDS의 SequentialSummary DS를 이용하여 기술하는 방법에 대해서 제안한다.

뉴스는 크게 앵커 샷과 나머지 샷으로 구분할 수 있다. 뉴스는 앵커 샷으로 시작을 하며 앵커 샷에서 앵커가 기사의 개요를 설명하고 나머지 샷들에서 자세히 설명하게 된다. 따라서 앵커의 말이 그 기사를 요약하는 것으로 볼 수 있으므로 뉴스에서 앵커 샷만을 추출해서 하나의 새로운 동영상을 만드는 것이 가장 쉬운 방법이다. 그러나 앵커 샷은 거의 정지 영상에 가깝고 기사마다 어깨걸이와 앵커만을 제외하고는 거의 같은 이미지처럼 보인다. 즉 하나의 기사에서 하나의 대표 프레임으로 기사의 전체 프레임들을 표현하는 것이 된다. 기사의 내용을 이해하는데 음성 정보가 결정적인 역할을 하므로 그러한 의미에서는 위의 동영상 요약이 좋은 요약이라고 할 수 있지만 비주얼한 정보를 전달하는데는 부족하다.

본 논문에서는 비주얼한 정보와 오디오 정보를 모두 표현하기 위해서 슬라이드 방법을 이용한다. 먼저 동영상을 비디오와 오디오로 분리한 후 비디오에서는 대표 프레임들을 추출하고 오디오에서는 앵커 샷의 오디오를 추출한다(<그림 9>). 기사의 대표 프레임들은 각 샷의 대표 프레임들로 선택하며 Sequential Summary의 FrameProperty D를 이용하여 프레임 정보들을 기술한다. 대표 프레임들의 개수는 앵커 샷의 길이(초단위)를 1.5초로 나눈 수 만큼 선택을 하는데 일반 샷이 아닌 샷

들을 모두 선택한 후 부족한 개수만큼 일반 샷에서 선택한다. 또한 기사의 제목은 그 기사의 내용을 함축적으로 표현한 것이라고 할 수 있으므로 TextProperty D를 이용하여 기술한 후 요약에 포함시킨다. 이러한 요약은 각각의 기사에 대해서 개별적으로 만들어서 기사 단위로 요약을 볼 수 있도록 한다. 이렇게 생성되는 요약을 MPEG-7 MDS에 맞게 기술한 예는 <그림 10>에서 보는 바와 같다.

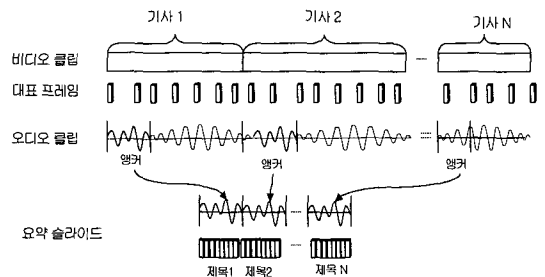


그림 9 기사의 제목, 앵커의 오디오, 대표 프레임들에 의한 정적 요약

```
<Summary id="기사1" xsi:type="SequentialSummaryType">
  <AudioSummaryLocator>
    <MediaUri>kbs_20000309.mp3</MediaUri>
    <BytePosition offset="191" length="698"/>
  </AudioSummaryLocator>
  <FrameProperty>
    <ImageLocator>
      <MediaUri>kbs_20000309_90.jpg</MediaUri>
    </ImageLocator>
  </FrameProperty>
  <- 대표 프레임들 종략 ->
  <TextProperty>
    <FreeTextAnnotation>배틀린</FreeTextAnnotation>
  </TextProperty>
</Summary>
<Summary id="기사2" xsi:type="SequentialSummaryType">
  <- 종략 ->
</Summary>
```

그림 10 정적으로 생성되는 요약 기술 예

3.2.3 동적으로 생성되는 요약

앞에서 설명한 바와 같이 동적으로 생성되는 요약을 위해서 본 논문에서는 HierarchicalSummary DS를 이용한다. MPEG-7 MDS의 HierarchicalSummary DS는 다양한 요약 시간과 조건에 맞는 계층적인 요약물 기술하기 위한 스키마다. HierarchicalSummary를 구성하는 요소[1]들은 keyAVClips, keyVideoClips, keyAudio Clips, keyFrames, keySounds, keyThemes이다. 본 논문에서는 오디오에 대한 인덱싱을 따로 하지 않으므로 keyAudioClips과 keySounds를 제외한 4개의 구성 요소를 사용한다. 또한 요약을 위한 정보들은 기사와 샷

세그먼트에 대한 기술을 바탕으로 한다. 뉴스의 기사 타입과 샷 타입은 뉴스를 구분하는 중요한 요소이므로 keyThemes를 위한 주제 리스트는 3.1.3절에서 정의한 분류 스키마의 항목들을 사용한다. 사용자는 Hierarchical Summary DS에서와 같이 이러한 주제 리스트에서 복수개의 항목을 선택할 수 있도록 하며 요약 시간을 함께 정할 수 있도록 한다.

요약의 세그먼트 단위는 사용자의 주제 선택에 의해서 자동으로 설정된다. 샷 타입에 관련된 항목을 선택하지 않고 기사의 섹션에 관련된 항목만 선택한 경우에는 요약 세그먼트 단위를 기사 단위로 하고 그 외의 경우에는 샷 단위로 한다. 사용자가 요약 시간은 정하지 않고 주제만 선택했을 경우에는 선택한 항목들에 해당하는 세그먼트 모두를 요약 세그먼트로 선택한다. <그림 11>은 사용자가 그래픽 샷을 선택한 후 인터페이스를 통해서 단계를 수정한 후의 요약용 기술한 예이다. 주제와 함께 요약 시간을 정했을 경우에는 주제에 해당하는 세그먼트 중에서 시간에 맞게 선택을 해야 한다. 일반적으로 뉴스[8]는 하이라이트를 제일 먼저 방송하고 중요한 사건을 다룬 후 스포츠와 기상보도로 마무리를 한다. 그러므로 같은 주제를 갖는 기사에서는 시간상으로 먼저 보도되는 기사가 더 중요하다고 볼 수 있다. 따라서 보도 순서대로 요약 시간을 넘지 않을 때까지 선택된 세그먼트 중에서 요약에 포함시킨다.

주제에 상관 없이 시간에 의해서만 요약을 할 경우에는 요약 시스템이 어떤 세그먼트를 포함할 것인가 선택을 해야 한다. 시간에 의존한 요약은 특정 기사를 보기

위한 요약이라기보다는 전체 뉴스를 개괄적으로 보기 위한 것이다. 따라서 그러한 요약을 위해서 기사는 너무 큰 단위가 되므로 요약 세그먼트 단위는 기사가 아닌 샷 단위로 한다. 뉴스에서 앵커의 말은 전체 기사를 요약하고 있으므로 가장 중요하며 그래픽 자료는 보통 기사의 내용을 일목요연하게 정리하여 표현할 때 많이 사용되므로 기사의 내용 전달력이 앵커 샷 다음으로 높다. 일반 샷은 샷 중에서 가장 중요도가 낮으므로 우선 순위를 가장 작게 한다. 따라서 앵커 샷, 그래픽 샷, 기사 샷/인터뷰 샷/회견 샷, 일반 샷 순서로 후보 세그먼트로 한 후 위의 과정을 반복한다.

4. 시스템 구현 및 분석

뉴스 동영상을 일반 사용자에게 서비스 하기 위해서 뉴스 동영상은 적절히 가공되어야 하고 가공된 데이터를 이용하기 위한 사용자 인터페이스가 필요하다. 본 장에서는 3장에서 제안한 뉴스 동영상 스키마에 맞는 응용 시스템을 설계하고 구현하며 이러한 스키마와 시스템이 사용자의 요구를 얼마나 반영할 수 있는지 분석한다.

4.1 시스템 구현

<그림 12>는 전체 시스템 구조를 나타내는 그림이다. 이 시스템에서는 방송 날짜별 동영상에 대한 메타데이터에 관한 기술 문서가 생성된다. 기술 문서에는 저급 수준 내용 정보 및 샷, 기사의 고급 수준 내용 정보가 저장된다. 본 시스템에서는 뉴스 동영상을 서비스하기 위해서 데이터를 생성하는 서비스 제공자와 그 서비스를 제공하는 일반 사용자로 구분하고 있다.

```

<ContentDescription xsi:type="SummaryDescriptionType">
  <Summarization>
    <Summary xsi:type="HierarchicalSummaryType" components="keyAVClips
    keyThemes" hierarchy="independent">
      <!-- 그래픽을 선택한 경우의 요약 -->
      <HighlightSummary id="Summary01" themelds="그래픽">
        <HighlightSegment>
          <keyAVClip>
            <MediaTime>
              <MediaTimePoint>126.292959</MediaTimePoint>
              <MediaDuration>7.640974</MediaDuration>
            </MediaTime>
            <keyAVClip>
              <keyFrame>
                <ImageLocator>
                  <MediaUri>kbs_20000309_4004.jpg</MediaUri>
                </ImageLocator>
              </keyFrame>
            </HighlightSegment>
            <HighlightChild level="0">
              <HighlightSegment>
                <!-- 세부 내용 요약 -->
                <HighlightSegment>
                  <!-- More HighlightSegment -->
                </HighlightSegment>
              </HighlightChild>
              <!-- More HighlightSegment and HighlightChild -->
            </HighlightSummary>
          </Summary>
        </Summarization>
      </ContentDescription>
    
```

그림 11 동적으로 생성되는 요약 기술 예

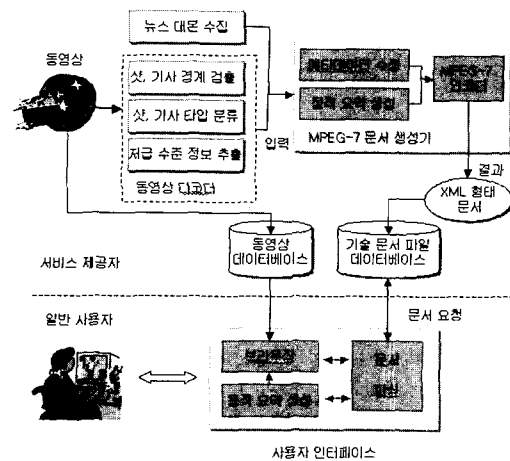


그림 12 시스템 구조

먼저 동영상에 대해 뉴스의 기본 데이터 구조인 기사와 샷의 경계를 자동으로 검출하고 샷의 보도 타입을 결정하게 된다. 그러나 자동 검출 도구는 100% 정확하지는 않기 때문에 인텍싱 도구를 통해서 이를 수정한다. 뉴스의 스크립트는 자동 검출을 하는 데 시간이 많이 소요되고 정확도가 떨어지기 때문에 웹 상에서 구현 후 시스템에 맞도록 수정하고 있다. 기사의 섹션 타입과 키워드 등의 고급 수준 정보들은 위에서 설명한 기사, 샷 경계 등의 저급 수준 정보와 스크립트를 바탕으로 MPEG-7 문서 생성기를 통해서 저작하게 된다. 또한 앞 절에서 설명한 정적 요약의 방법을 이용하여 기사 단위로 정적 요약을 생성한다. 이렇게 생성된 뉴스 동영상에 대한 메타데이터와 정적 요약 정보는 MPEG-7을 기반으로 한 뉴스 동영상 스키마에 맞게 XML 형태의 문서로 생성이 되고 방송날짜 인텍스와 함께 저장이 된다.

일반 사용자는 사용자 인터페이스를 통해서 특정 뉴스 프로그램의 방송날짜에 대한 문서를 요청할 수 있다. 요청된 문서는 인터페이스 내의 XML 파서(Parser)를 이용해서 분석되고 파싱된 데이터는 브라우저에 전달된다. 사용자는 관련 정보를 비주얼하게 볼 수 있고 특정 부분만을 재생해 볼 수 있다. 또한 사용자가 원하는 기사의 섹션 타입이나 샷의 타입에 맞는 세그먼트를 요약 시간에 맞추어 동적으로 요약해서 세그먼트 계층을 수정하고 재생하며 저장을 할 수 있다. 동적 요약 정보는 세그먼트 정보와 별도로 문서를 만들어서 저장할 수 있게 함으로써 사용자가 나중에 동적 요약 문서만 열어서 볼 수 있도록 하였다.

본 논문의 뉴스 아카이브 시스템은 C 언어를 이용하여 구현하였다. 사용한 샷 경계 검출 방법[4]은 MPEG

비디오 스트림의 칼라 히스토그램(Color Histogram)과 매크로블록(Macroblock)을 이용한 방법이고 기사의 경계와 샷의 타입은 샷의 길이와 칼라 정보를 이용하여[8] 검출하였다. 본 논문에서 제안한 스키마에 맞는 MPEG-7 문서 생성기는 MSXML 4.0 파서를 사용하여 구현하였고 스키마 문서는 XML Spy v4.0을 사용하여 편집하였다. 이러한 모든 정보들은 <그림 13>에서 보는 바와 같은 인터페이스를 통해서 비주얼하게 사용자들에게 보여진다.

4.2 관련 연구와의 비교

기존의 연구에서는 대부분 뉴스의 시간적인 구조[9]를 기반으로 샷, 샷의 그룹, 뉴스 아이템[10]으로 인텍싱 하였다. 더블린 시립 대학의 Fischlar 시스템[11] 또한 내용의 구조화를 위해서 비주얼 인덱스를 생성하였으며 이 인덱스들은 앞에서 설명한 바와 같이 계층적으로 구성이 된다. 이에 반해 본 논문에서는 뉴스 동영상을 나누는 기본적인 단위로 기사와 샷 세그먼트만을 이용하였고, 샷 타입[4]을 이용하여 세그먼트를 기술하는데 초점을 맞추었다. 본 논문에서 제안한 스키마를 이용하면 기존의 연구에 비해 연속되지 않은 세그먼트 사이의 관계나 하나의 기사안에서 각 세그먼트의 역할을 표현하는 데는 부족하였지만 본 논문에서 제안한 스키마에 의해 기사와 샷의 여러 타입들을 기술함으로써 세그먼트에 대한 정보는 자세히 기술할 수 있었다. 또한 샷 타입[4]을 이용하는 기존의 연구에서는 단순히 동영상의 대표 프레임을 만드는 데 타입 정보를 이용하였지만 본 논문에서는 이러한 정보를 통해 뉴스 동영상을 요약하는 데 있어서 사용자의 요구와 내용상의 중요도를 반영함으로써 해서 스트리밍 서비스가 가능하도록 하였다.

뉴스 동영상을 요약하는 가장 기본적인 방법은 앵커 샷을 나열하는 방법이다. 하지만 이러한 방법은 앞에서 설명한 바와 같이 비디오의 내용 전달이 부족하다는 단점을 가지고 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 오디오, 텍스트, 대표 프레임 모두를 이용하여 슬라이드로 보여주는 방법 - 정적으로 생성되는 요약 - 을 제안하였다. 이 방법에서는 앵커 샷의 비디오 데이터를 이용하는 것이 아니라 그 뒤에 나오는 기사 샷들의 키프레임들을 이용하였기 때문에 단순히 앵커 프레임만 나열되는 것보다는 비디오 내용 전달에 더 효율적이다. 사용자들을 대상으로 실험을 해 본 결과 이와 같은 특징은 기사 내용을 이해하는 데 많은 도움이 되는 것을 알 수 있었다. 또한 본 논문의 방법은 기존의 요약 방법보다 데이터 양이 훨씬 적게 이용이 된다. 예를 들어 60초짜리 동영상 요약 파일을 만드는 경우 기



그림 13 사용자 인터페이스

존의 방법과 본 논문에서 제안한 방법(60 프레임을 이용)을 비교하면 기존의 방법이 약 4.5배 정도 더 많은 데이터 양이 필요하다는 것을 알 수 있다. 그러므로 본 논문에서 제안한 방법은 웹이나 모바일 환경에서도 유용하게 이용될 수 있음을 알 수 있다.

동적으로 생성되는 요약은 기사의 섹션 타입과 샷의 보도 타입, 그리고 요약 시간에 의해서 생성이 되었다. 따라서 뉴스의 형식적인 측면에서는 사용자의 요구를 충분히 반영한다고 볼 수 있다. 하지만 내용적인 측면에서는 충분히 반영할 수 없었는데 즉, “앵커 샷에 의한 요약”과 같은 타입 기반으로 이루어지는 질의는 처리할 수 있으나 키워드 기반의 요약이나 스크립트를 이용한 요약은 할 수 없기 때문에 “김대중 대통령에 관한 요약”과 같은 내용 기반 질의는 처리할 수 없는 것이다.

기존의 뉴스 아카이브 시스템에서는 시스템에서 정의하는 형태로 데이터를 기술하기 때문에 다른 시스템과의 호환성이 없다는 큰 문제점을 가지고 있다. 본 논문의 시스템에서는 이와 같은 문제점을 해결하기 위하여 MPEG-7 MDS를 기반으로 데이터를 기술하였다. MPEG-7은 여러 가지 기술자들을 표현하기 위해 XML을 이용하는데 XML은 두 개 이상의 스키마를 사용하여 생성될 수 있다. 그러므로 본 논문의 시스템에서는 MPEG-7 MDS 스키마와 본 논문에서 제안한 스키마를 통해서 기존의 연구들의 가장 큰 문제점인 시스템간의 호환성을 보장할 수 있었다.

5. 결론 및 향후 연구 방향

뉴스에 관한 자료는 다른 어떤 종류의 동영상보다 자료 보관과 교환의 가치가 높다. 그러나 뉴스에 관련된 기존의 연구에서는 뉴스에서 앵커샷을 찾고 기사 경계를 찾는 인덱싱과 스크립트 분석이 주류를 이루었다. 본 논문에서는 보관과 호환성을 위해 뉴스 동영상을 기술하는 관점에서 MPEG-7을 기반으로 하여 뉴스 동영상 스키마를 제안하였다. 또한 본 논문에서 제안한 스키마에 맞는 문서를 생성하고 생성된 문서를 브라우징할 수 있는 도구를 구현하였고 MPEG-7이 뉴스 동영상에 관한 내용을 풍부하게 기술할 수 있음을 살펴보았다.

본 논문에서 제안한 스키마는 가장 기본적인 요소로서 세그먼트를 사용하였으며 세그먼트 단위 기준은 시간적인 기준으로 하였다. 그리고 기사, 샷 단위까지만 나누도록 하였으며 VideoSegment DS를 이용하여 기술하였다. 또한 대표 프레임과 스크립트를 기술하기 위해서 Keyframe 타입과 Script 타입을 추가하고 세그먼트의 미디어 정보는 항상 포함하도록 수정하였다. 기사와

샷의 타입을 기술하기 위해서 본 논문에서는 기사의 섹션 타입과 샷의 보도 타입을 정의하는 분류 스키마 문서를 생성하고 이러한 정의에 맞추어 기사와 샷의 타입을 기술하였다.

한편 본 논문에서 전체 뉴스를 빠르게 이해할 수 있는 요약과 브라우징을 위해서 샷 타입을 이용하여 대표 프레임을 선택하는 방법을 제안하였다. 또한 서비스 제공자가 미리 전체 뉴스를 요약하는 방법을 제안하였다. 이는 Sequential-SummaryDS를 기반으로 하고 기술하였으며 슬라이드를 통해서 사용자가 볼 수 있도록 하였다. 사용자의 질의를 받아서 동적으로 생성되는 요약은 HierarchicalSummaryDS를 기반으로 하였으며 트리 인터페이스를 통해서 요약 정보를 수정할 수 있도록 하였다.

앞으로의 연구에서는 사용자 프로파일 등의 정보와 보안 정보를 이용하여 개개인의 특성에 맞게 동영상 서비스를 제공할 수 있는 시스템을 제안할 것이다. 또한 본 논문에서는 다루지 않았던 방대한 분량의 데이터베이스에서 원하는 정보를 효율적으로 찾고 관리할 수 있는 방법도 연구를 할 것이다. 물론 이러한 부분은 상당 부분 XML에 관한 기반 기술에 의존하고 있어서 XML이 발전을 하게 되면 자연히 해결될 수 있을 것이다. 그러나 저급 수준의 정보까지 모두 포함하고 웹에서 빈번하게 이동을 한다고 가정할 때 텍스트 형태가 아닌 이진형태로 저장하고 효율적으로 관리하고 원하는 정보를 찾을 수 있는 방법이 따로 연구가 되어야 한다. 이런 연구들은 향후에 서로 다른 시스템 간에서도 자료를 교환할 수 있고 내용적으로도 풍부한 기술 문서를 생성할 것이며 사용자의 요구를 더 충족할 수 있는 시스템 개발에 밑바탕이 될 것이다.

참 고 문 헌

- [1] P. Beek, B. Ana, *Text of 15938-5 FCD Information Technology Multimedia Content Description Interface Part 5 Multimedia Description Schemes*, ISO/IEC JTC1/SC29/WG11/ N3966, March 2001.
- [2] M.Martinez, *Overview of the MPEG-7 Standard (version 5.0)*, ISO/IEC JTC1/ SC29/WG11/ N4031, March 2001.
- [3] N. Day, M. Martinez, *Introduction to MPEG-7 (v3.0)*, ISO/IEC JTC1/SC29/ WG11/N4032, March 2001.
- [4] 한근주, *샷의 타입을 이용한 뉴스 아카이브 시스템의 설계 및 구현*, 서강대학교 석사 학위 논문, 1999.
- [5] N. Day, *MPEG-7 Projects and Demos*, ISO/IEC JTC1/SC29/WG11/N4034, March 2001.
- [6] A. Girgensohn, J. Boreczky, L. Wilcox, "Keyframe

-Based User Interfaces for Digital Video," *IEEE Computer Magazine*, Vol. 34, pp. 61-67, 2001.

[7] L. Hyowon, F. Smeaton, et al, "Implementation and Analysis of Several Keyframe-Based Browsing Interfaces to Digital Video," *Proc. of the Fourth European Conference on Digital Libraries*, Vol.1823, pp. 206-218, 2000.

[8] B. Merialdo, K. Lee, et al, "Automatic Construction of Personalized TV News Programs," *Proc. of International Conference on Multimedia*, pp. 323-341, 1999.

[9] H. Zhang, Y. Gong, S. Smoliar, "Automatic Parsing of News Video," *Proc. of Multimedia Computing and Systems*, pp. 45-54, 1994.

[10] W. Qi, L. Gu, et al, "Integrating Visual, Audio and Text Analysis for News Video," *Proc. of International Conference on Image Processing*, Vol.3, pp.520-523, 2000.

[11] N. OConnor, C. Czirjek, et al, "News Story Segmentation in the Fischler Video Indexing System," *Proc. of International Conference on Image Processing*, Vol.3, pp.418-421, 2001.



김 경 수
 1983년 2월 서울대학교 공과대학 제어계측공학과 졸업(학사). 1985년 2월 서울대학교 학원 제어계측공학과 졸업(석사). 1985년 3월 ~ 현재 KBS 기술연구소 차장. 관심분야는 멀티미디어 방송 제작 시스템, 스트리밍 미디어 및 서비스, 콘텐츠 보호

및 관리 기술



하 명 환
 1993년 2월 경북대학교 공과대학 전자공학과 졸업(학사). 1995년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사). 1995년 2월 ~ 현재 KBS 기술연구소 연구원. 관심분야는 멀티미디어 방송 제작 시스템, 비디오 인덱싱, 영상 처리



정 병 회
 1994년 2월 이화여자대학교 전자계산학과 졸업(학사). 1996년 2월 한국과학기술원 전산학과 졸업(석사). 1996년 1월 ~ 현재 KBS 기술연구소 연구원. 2000년 9월 ~ 현재 한국과학기술원 전산학과 박사과정 재학중. 관심분야는 멀티미디어 방송 시스템, 미디어 아카이브, 초고속 네트워크 시스템



정 진 국
 1998년 서강대 전자계산학과 졸업. 2000년 서강대 컴퓨터학과 석사. 2000년 ~ 현재 서강대 컴퓨터학과 박사 과정



심 진 선
 1995년 2월 서강대학교 컴퓨터학과 입학. 1999년 2월 서강대학교 컴퓨터학과 졸업. 2000년 2월 서강대학교 컴퓨터학과 대학원 입학. 2002년 2월 서강대학교 컴퓨터학과 대학원 졸업. 2002년 3월 ~ LG전자/정보통신 CDMA 단말 사업부 신기술

개발팀



남 종 호
 1986년 서강대 전자계산학과 졸업. 1988년 한국과학기술원 석사. 1992년 한국과학기술원 박사. 1992년 ~ 1993년 Fujitsu연구소 연구원. 1993년 ~ 현재 서강대학교 컴퓨터학과 교수. 현재 서강대학교 컴퓨터학과 부교수