논문-00-5-1-09

효율적인 비디오 브라우징을 위한 동적 요약 및 요약 기술구조

김 재 곤*, 장 현 성*, 김 문 철*, 김 진 웅*, 김 형 명**

Dynamic Summarization and Summary Description Scheme for Efficient Video Browsing

Jae-Gon Kim*, Hyun Sung Chang*, Munchurl Kim*, Jinwoong Kim*, and Hyung-Myung Kim**

요 약

최근 디지털 비디오 데이터가 급격히 증가하고 대중화됨에 따라 이를 활용하기 위한 효율적인 접근 기법이 절실히 요구되고 있다. 비디오 요약(video summarization) 기법은 의미적으로 중요한 요점만으로 전체 비디오를 표현하는 것으로 비디오 내용에 대한 전반적인 개관(overview)을 제공할 뿐만 아니라 브라우징(browsing) 등의 유용한 접근 기능을 제공한다. 본 논문에서는 의미적으로 중요한 내용을 포함하는 비디오 주요구간(highlight segment) 검출을 통한 새로운 동적 요약(dynamic summarization) 기법과 생성된 요약 정보 표현을 위하여 MPEG-7에 제안한 요약 기술구조(DS: Description Scheme)에 대하여 기술한다. 본 논문의 기술구조는 다중 계층의 하이라이트(highlight), 계층적 브라우징, 사용자 주문형 요약 등의 기능을 통하여 비디오의 개관 및 효율적인 브라우징, 네비게이션(navigation)을 가능하게 한다. 또한, 제안하는 비디오 요약 기법 및 요약 기술구조의 실현 가능성 및 기능 구현을 확인하기 위하여 축구 비디오에 대한 적용 실례를 제시한다.

Abstract

Recently, the capability of efficient access to the desired video content is of growing importance because more digital video data are available at an increasing rate. A video summary abstracting the gist from the entirety enables the efficient browsing as well as the fast skimming of the video contents. In this paper, we discuss a novel dynamic summarization method based on the detection of highlights which represent semantically significant content and the description scheme (DS) proposed to MPEG-7 aiming to provide summary description. The summary DS proposed to MPEG-7 allows for efficient navigation and browsing to the contents of interest through the functionalities of multi-level highlights, hierarchical browsing and user-customized summarization. In this paper, we also show the validation and the usefulness of the methodology for dynamic summarization and the summary DS in real applications with soccer video sequences.

I. 서 론

최근 디지털 방송 및 인터넷을 중심으로 디지털 비디오의 활용이 급속히 확산되고 방대한 비디오 데이터가 대중화됨에 따라 이를 효과적으로 이용하기 위한 접근 및 검색 기법이 크게 요구되고 있다. 비디오 데이터의 압축, 전송 및 저장 기술에 비해 상대적으로 기술 개발이 미약했던 내용기반 색인(content-based indexing) 및 검색 기술과 관련하여 내용 분석(content analysis), 표현(representation), 접근(access) 기법을 중심으로 많은 연구가 이루어지고 있다. 방대한 비디오 데이터에 대한 효율적인 접근성은 표현 기법에 좌우되며, 이와 관련하여 현재 MPEG-7에서는 오디오-비쥬얼(audio-visual) 컨텐트(content)의 기

* 한국전자통신연구원 무선방송기술연구소 방송미디어연구부
Broadcasting Media Technology Dept., Radio & Broadcasting Technology Lab., ETRI

** 한국과학기술원 전자 전산학과 전기 및 전자공학 전공
Dept. of Electrical Engineering and Computer Science, KAIST

술(description)에 대한 표준화 작업을 진행하고 있다[14]. 비디오 내용 분석은 비디오 표현에 유용한 특징(feature) 정보의 추출에 관한 것이고, 접근 기법은 컨텐트 표현 데이터를 활용하여 사용자가 원하는 컨텐트에 접근하기 위한 검색, 브라우징 및 네비게이션 등의 기법에 관한 것이다.

비디오 요약은 원 비디오를 주요 핵심 부분으로 축약(compaction)하여 전반적인 내용을 빨리 파악할 수 있는 개관의 기능을 제공하는 오디오-비쥬얼 초록화(abstraction)로 검색 및 브라우징 과정에 있어서도 비선형 네비게이션(nonlinear navigation)으로 '빠른 재생'이나 '되감기' 등의 작업을 최소화함으로써 접근 효율성을 크게 향상시킬 수 있다. 이러한 측면에서 기존의 많은 연구에서 비디오 요약은 내용기반 비디오 표현의 기본적이고 강력한 기법으로 간주되었다[3][4][5].

비디오 요약은 그 표현 형식에 따라 정적 요약(static summary)[1][2][3][4]과 동적 요약(dynamic summary)[5][6][7]의 두 가지 형태로 분류할 수 있다. 현재까지의 대부분의 연구는 정적 요약에 관한 것으로, 대개 비디오를 가장 기본적인 단위인 샷(shot)으로 분할하고 각 샷 마다 소수의 키 프레임(key frame)을 선별한 후 이들을 적절히 배치한 스토리보드(story-board) 형태의 요약을 제공한다. 한편, 동적 요약은 원 비디오의 주요구간을 추출하여 짧은 길이로 축약한 또 하나의 비디오를 구성하는 기법이다.

MPEG-7에서는 오디오-비쥬얼 컨텐트를 구조화된 형태로 표현하여 간결한 기술, 부분 기술에 대한 접근의 용이성, 다양한 기술 정보의 효율적 관리 등을 위한 멀티미디어 기술구조(MDS: Multimedia Description Scheme)를 표준화하고 있다[15]. MDS의 세분화된 한 부분으로 비디오 요약을 효율적인 구조의 상호운용 가능한 메타 데이터(meta data)로 기술하기 위한 요약 기술구조(Summarization DS)를 설계하고 있다[15].

본 논문에서는 디지털 비디오의 효율적인 접근 기법을 제공하기 위한 방안으로 원 비디오로부터 동적 요약 비디오를 자동으로 생성하기 위한 방법론과 비디오 요약을 체계적으로 기술하기 위한 계층적 요약 기술구조(HierarchicalSummary DS)를 소개한다. 본 논문의 계층적 요약 기술구조는 MPEG-7에 제안하여 현재 위원회 작업안(WD: Working Draft)[15]에 채택된 것으로 동적 및 정적 요약 비디오를 적절히 함께 기술할 수 있는 통일된 기반구조로 동적 비디오를 통한 개관과 이를 바탕으로 정적 비디오의 접근 기능을 이용한 효율적인 비디오 네비게이션 및 브라우징 기능을 제공한다.

II장에서는 기존의 요약 기법을 검토하고 III장에서는 제안하는 동적 요약 기법을 IV장에서는 요약 기술구조를 각각 기술한다. V장에서는 비디오 요약과 기술에 대한 실험 내용 및 결과를 보이고 마지막으로VI장에서 결론을 맺는다.

II. 기존의 요약 기법

1. 정적 요약

비디오 요약과 관련한 초기의 연구들은 비디오 내용기반 색인을 위한 연구들로 비디오를 신태틱(syntactic)하게 구조화하는 기법이다. 즉, 스토리보드 형태로 비디오를 요약하고 색인하는 것으로 비교적 간단하며, 키 프레임 영상으로부터 해당 샷의 내용을 파악하고 그 키 프레임을 통하여 원하는 샷으로 직접 접근할 수 있는 기능을 제공한다. 하지만 정지영상만으로 샷을 표현하므로 비디오의 시간적인 정보 및 오디오를 함께 표현하지 못한다. 또한 많은 수의 키 프레임이 존재하게 되어 그들로부터 전체 비디오의 내용을 파악하기가 어렵고 원하는 샷을 선택하고 브라우징 하는데 많은 노력을 요하게 된다. 예를 들면, *Terminator - the Judgement Day*의 경우 300여장의 샷이 존재하며[2], 이는 곧 140여분의 전체 분량에서 각 샷 마다 하나씩만의 키 프레임을 추출하는 경우에 3천여장의 키 프레임이 존재함을 의미한다.

이러한 문제점을 개선한 기법으로 샷 군집화(clustering)에 대한 연구가 일부 진행되었다. 즉, 비디오를 샷들로부터 더 상위의 포괄적인 단위(일반적으로 씬(scene) 또는 에피소드(episode)라고 칭함)를 구성하고 이를 STG(Scene Transition Graph)[3]나 TOC(Table Of Contents)[2] 등의 다양한 형태로 구조화하여 사용자가 계층적으로 비디오 요약을 파악하고 브라우징 가능하도록 하는 방법이다. 그러나 이러한 기법은 씬이 결국 하나의 작은 이야기를 포함하는 의미적인 단위를 구성하는 것인데, 이러한 의미의 단위로의 분할에 한계를 갖고 있다. 모자이크 기법도 정적 요약의 한 예가 될 수 있는데 일반적으로 비교적 단순한 카메라 움직임을 갖는 샷 이내의 짧은 비디오 구간을 표현하는데 효과적이다[1].

2. 동적 요약

한편, 동적 요약 기법은 원 비디오의 주요구간을 추출

하여 짧은 길이로 축약한 또 하나의 비디오를 구성하고자 하는 것으로, 오디오 및 동영상의 움직임 정보를 포함하므로써 사용자 이해도의 측면에서는 전체 비디오를 보다 효과적으로 개관할 수 있다는 장점을 갖는다. 축약된 짧은 길이의 요약 비디오는 개인용 저장 장치를 갖는 디지털 방송 환경에서 EPG(Electronic Program Guide)나 브라우징 툴로 활용될 수 있다.

Informedia 연구[5]에서는 샷 경계, 카메라 움직임(camera motion), 영상 자막(embedded text) 및 얼굴 등의 비쥬얼 특징과 오디오 트랙의 음성인식 기법을 통한 주요단어 및 문장 정보를 함께 사용하여 요약에 포함될 비디오 구간을 실험적인 규칙에 의해서 검출하였다[5]. 현재까지 보고된 기존의 비디오 요약 기법은 비교적 안정적인 성능을 제공하는 것으로 알려져 있으나[5][6], 대부분 특징에 기반한 신택틱한 접근 방법을 사용하고 있으므로 의미적인(semantic) 측면에서의 중요한 부분을 포함하는 요약을 보장하지 못하는 단점이 있다[7].

또한 다양한 장르의 비디오에 적용할 수 있는 일반적인 요약 기법의 도출은 상당히 어려워 보이며 실제로 장르에 제한적인 요약 접근방법이 연구되고 있다. 동적 요약 측면에서 비디오 장르는 영화나 드라마와 같이 이야기의 전개가 중요시되는 이야기 중심(story-oriented)의 비디오와 비교적 중요한 사건이 분명히 정의되고 이들 사건이 중심이 되는 스포츠와 같은 사건 중심(event-oriented)의 비디오, 그리고 뉴스와 같이 특별한 구성 양식을 포함하고 있는 경우로 나눌 수 있다. 이야기 중심 비디오의 요약은 이야기의 흐름을 잘 전달하면서 중요한 내용을 포함하여야 하는데 이야기의 흐름을 파악하기가 상당히 어려우며 이야기 분석을 위한 특별한 접근 방법이 시도되고 있다[8]. 사건 중심 비디오는 중요한 사건을 정의하고 사건 검출의 지지기반이 되는 특징 정보의 분석을 통하여 중요 사건을 포함하는 비디오의 주요구간 즉, 하이라이트를 검출하는 기법을 사용할 수 있다. 뉴스의 경우는 앵커샷, 리포트 샷 등 형식화된 구조 분석에 기반한 요약 기법을 고려할 수 있다. 스포츠나 뉴스의 경우 안정적인 성능을 갖는 자동 요약 기법의 도출이 기대되며 비교적 많은 연구가 수행되었다[1][9].

III. 제안하는 동적 요약 기법

본 논문에서는 스포츠 등의 사건 중심의 비디오에 효과적으로 적용할 수 있는 것으로 의미적으로 중요한 내용을 포함하는 하이라이트들로 요약을 구성하기 위한 의미적(semantic) 접근 방법을 제시한다.

본 논문에서 제안하는 동적 요약 기법은 하이라이트를 검출하기 위하여 물리적인 현상(특징)과 내재된 의미 사이에 존재하는 상관성을 이용하고자 하는 것이다. 구체적으로, 비디오 장르 마다 중요한 내용에 해당하는 사건을 우선 정의한 후 관찰 및 실험 등으로 축적된 지식 기반의 요약 규칙을 적용하여 그 사건을 검출한다. 검출된 사건을 포함하는 비디오 구간을 중심으로 에피소드를 구성하여 최종적으로 요약 비디오에 포함될 요약 구간을 선택한다. 그림 1은 동적 요약 기법에 기반한 비디오 요약 시스템의 기능 구성도로 다음의 세 기능부로 구성된다.

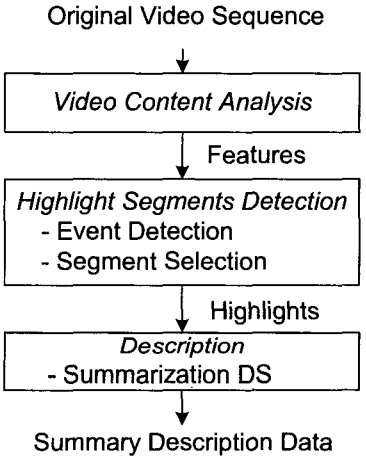

그림 1. 동적 요약 시스템의 기능 구성도
Fig. 1. Functional diagram for dynamic summarization

- 비디오 내용 분석 - 입력한 요약 대상 비디오의 내용 분석을 통하여 사건과 연관된 즉, 검출의 규칙에 활용되는 특징들을 추출한다.
- 하이라이트 검출 - 사건 검출과 요약구간 선택 기능으로 구성되며, 사건 검출은 요약 규칙을 적용하여 정의된 사건들을 검출하고 주요사건과 부대사건으로 구성된 에피소드를 검출한다. 요약구간 선택은 에피소드를 중심으로 비디오 편집 규칙 등을 고려하여 요약 비디오를 구성할 최종 요약구간들을 선택한다.
- 요약 기술 - 최종적으로 요약구간 등 요약에 포함되는 정보를 MPEG-7의 Summarization DS를 이용하여 기술한 요약 기술 데이터를 출력한다.

1. 비디오 내용 분석

요약을 위한 비디오 분석은 요약 기법과 비디오 장르에 따라서 유용한 특징 선택이 요구된다. 본 논문의 요약 기법

은 검출하고자 하는 의미적 사건이 전형적인 특징들을 동반한다는 관찰 결과를 전제로 한다. 예를 들면, 축구 비디오에서 중요한 의미적 사건으로 정의되는 골(goal)이 발생한 경우 슈팅(shooting)한 선수의 이름을 포함한 자막, 다른 각도에서 촬영된 골 장면의 느린 중복 재생(replay), 급격한 카메라 줌(zoom) 등을 동반하는 경우가 일반적이다.

본 논문에서는 축구 요약의 경우 비디오 분석을 통하여 샷 경계(컷(cut), 디졸브(dissolve), 와이프(wipe)), 영상자막, 카메라 움직임, 중복재생, 샷 분류(그라운드(ground), 벤치(bench), 관중석), 카메라 뷰(원경, 근경) 등의 특징들의 추출을 가정한다. 이 들 특징들의 추출 기법은 내용기반 색인과 관련하여 많은 연구결과가 보고된 바 있으며, 본 논문의 자동 요약 시스템 구현을 위해서 현재 개발 중이다. 본 논문에서는 동적 요약을 위한 비디오 내용 분석의 일환으로 개발한 MPEG 압축 영역에서의 카메라 움직임 분석 기법에 대해서 간략히 기술한다.

1.1 카메라 움직임 분석[11]

MPEG 비트열로부터 추출한 움직임 벡터(motion vector)를 전처리를 통하여 각 프레임에 해당하는 움직임 벡터장(MVF: motion vector field)을 구성한다. 각 프레임의 MVF를 affine 움직임 모델로 정합하는데, affine 움직임 모델에서 (x, y) 화소 위치의 매크로블록(macro-block)의 움직임 벡터(u, v)는

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_2 & a_3 \\ a_5 & a_6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_1 \\ a_4 \end{pmatrix} \tag{1}$$

으로 표현된다. 계수 벡터 $\Phi = (a_1, a_2, a_3, a_4, a_5, a_6)$는 MVF로부터 최소자승(least square) 방식으로 추정한다. 계수 벡터를 카메라 움직임의 정성적 특성을 더 잘 표현할 수 있는 식 (2)와 같은 형태로 변형한다. 여기서 변형된 계수 벡터 Φ'의 *pan*, *tilt*, *div* 그리고 *rot*는 각각 팬(pan), 틸트(tilt), 줌, 회전에 의해서 유도되는 MVF 성분을 나타낸다.

$$\begin{aligned} &\Phi' = (pan, tilt, div, rot, hyp_1, hyp_2), \; with: \\ &pan = a_1 \qquad tilt = a_4 \\ &div = \frac{1}{2}(a_2 + a_6) \qquad rot = \frac{1}{2}(a_5 - a_3) \\ &hyp_1 = \frac{1}{2}(a_2 - a_6) \quad hyp_2 = \frac{1}{2}(a_3 + a_5). \end{aligned} \tag{2}$$

변형된 계수들의 임계화를 통해서 줌, 회전, 팬, 틸트, 객체 움직임, 정지영역의 6가지 형태의 카메라 움직임 검출을 수행한다. 이 때 검출은 프레임 레벨(frame-level)과 구간 레벨(segment-level)의 두 단계에서 수행한다. 프레임 레벨의 임계값은 해당 계수가 지각할 수 있을 만큼의 카메라 움직임에 해당하는 MVF를 야기하는 값으로 비디오 종류와 상관 없이 안정적으로 설정할 수 있다. 구간 레벨의 임계값도 임의 패턴의 카메라 움직임이 최소 0.5 초 정도 이상 지속된다는 물리적인 카메라 움직임의 관찰 결과를 바탕으로 설정할 수 있다.

2. 하이라이트 검출[12]

대개 상위의 의미적인 요소를 반영하는 사건을 물리적인 현상에 의하여 직접 검출하기가 어렵다. 그러나, 사건의 발생 시점 부근에서 다른 사건이나 특징들을 동반하여 나타나는 경우가 많다는 사실을 이용하면 종합적인 문맥으로 추정할 수 있는 경우가 상당히 많다. 사건과 사건들 또는 특징들간의 인과성은 비록 그것이 결정적인 관계가 아니더라도 사건 검출에 유용할 수 있는데, 본 논문에서는 비디오 시퀀스에서 의미적인 사건을 검출하기 위하여 Bayes rule에 기반한 추론 모형을 제안한다. Bayes rule에 의하면, 사건 E와 관련된 특징 F 사이에 다음과 같은 확률식이 성립한다.

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)}. \tag{3}$$

식 (3)에서 $G_F(E) = \frac{P(F \mid E)}{P(F)}$ 라고 하면,

$$P(E \mid F) = G_F(E)P(E) \tag{4}$$

가 되는데, 이는 곧 $G_F(E)$가 "특징 F가 관찰될 때 사건 E의 발생확률에 대한 이득(gain)"으로 해석될 수 있음을 의미한다. 다시 말하면,

$$G_F(E) = \frac{P(F \mid E)}{P(F)} > 1 \tag{5}$$

을 만족하는 특징 F는 확률적인 의미에서 사건 E의 발생에 대한 암시를 제공할 수 있다. 이를 사건 E와 연관된 다수 개의 특징 $F = \{F_1, F_2, \cdots, F_N\}$에 대하여 확장하면,

$$P(E|F)=\prod_{k=1}^{N}\frac{P(F_k|F_1,\cdots,F_{k-1},E)}{P(F_k|F_1,\cdots,F_{k-1})}\cdot P(E) \qquad (6)$$

이 되고, 역시 식 (5)에서 확장된 다음의 조건식

$$P(F_i|J,E)>P(F_i|J),\ \forall J\subset F-\{F_i\} \qquad (7)$$

을 가정하면, 식 (6)의 조건부 이득항은 모두 1보다 크게 되어 각각의 특징들이 전반적인 이득에 일조하게 된다. 식 (7)를 만족하는 특징집합 F는 사건 E에 대한 지지집합을 구성하며, 이득이 충분히 큰 $(G_J(E)>\frac{1}{2P(E)})F$의 부분집합 J가 검출된 경우, 사건 E가 발생한 것으로 결정할 수 있다. 각 사건에 대하여 그 지지집합은 비디오의 장르에 따라 종속적이며 관찰이나 실험 등의 경험적인 방법에 의하여 얻을 수 있으며 축구에 대한 실험 예를 V장에서 기술한다.

하나의 사건은 그 자체로서 이야기 단위를 구성하기보다는 그 주위에서 보다 포괄적인 이야기 단위의 에피소드를 형성하는 경우가 많다[10]. 예를 들면 사건(골)이 발생한 장면 이후 연속하여 나타나는 슈팅을 한 선수에 대한 클로즈-업(close-up) 샷, 관중석 샷 및 슛 장면의 중복재생 샷 등이 하나의 에피소드를 구성한다. 이와 같이 에피소드를 구성하는 각 샷들은 서로 독립적인 것이 아니므로 비디오 요약구간을 추출하는 과정에서도 먼저 에피소드 경계를 추정한 후, 요약 비디오의 길이에 대한 요구사항이나 최종 구성될 요약 비디오가 시각적으로 거슬리지 않도록 편집 규칙을 고려하여 최종 요약구간을 선택한다[12][13].

IV. 요약 기술(Summary Description)[13][15]

본 장에서는 비디오 요약 정보의 기술을 통하여 원 비디오의 빠른 개관, 브라우징 및 네비게이션, 사건별 요약 등의 기능을 제공할 수 있는 계층적 요약 기술구조(HierarchicaSummary DS)에 대하여 기술한다. MPEG-7에서 정의한 요약 기술구조는 다 수의 요약을 포함하기 위하여 Summary DS의 집합으로 구성된 Summarization DS를 최상위 계층으로 한다. 각 Summary DS는 하나의 요약을 기술하는 단위로, 실제로 기술될 때에는 다중 레벨의 계층 구조로 요약을 표현하는 HierarchicalSummary DS 형태로 상속(inheritance)되어 구현된다.

기존의 HierarchicalSummary DS[16]는 요약 구간을 단순히 다중 레벨의 계층 구조로 연결하는 기본적인 형태의 기술구조였다. 본 논문의 요약 기술구조는 이러한 HierarchicalSummary DS를 확장 개선한 것으로서 1) 동적 요약과 정적 요약을 함께 기술할 수 있는 통일된 기술 구조를 통하여 개관과 브라우징을 유동적으로 결합한 효율적인 접근 기제(mechanism)를 제공하고, 2) 오디오 요약 정보를 기술할 수 있으며, 3) 사건 기반의 요약을 할 경우 기존의 기술 중복성과 탐색의 복잡성을 절감하는 효율적인 기술을 가능하게 한다[17][18][19].

1. HierarchicalSummary DS의 기본 구조

그림 2는 MPEG-7의 요약 기술구조로 제안된 몇 가지 요약 표현 기법들을 통합한 초기의 HierarchicalSummary DS[16]를 도시한 UML(Unified Modeling Language)도 이다. 이 초기 구조는 주요 비디오 구간들로 구성되는 동적 요약 비디오를 계층적으로 기술할 수 있는 것으로 원 비디오의 전반적인 내용을 대략적으로 파악할 수 있는 개관 기능을 제공한다.

HierarchicalSummary DS는 그림 2와 같이 HighlightLevel DS를 두어 연속적인 레벨 형태로 다중 길이의 요약을 표현할 수 있는데, 일반적으로 상위 계층은 보다 중요한 내용을 포함한 간략한 형태의 요약을 표현하고 하위 계층으로 내려갈수록 보다 자세한 내용을 전달할 수 있게 긴 길이를 갖는 요약으로 구성된다.

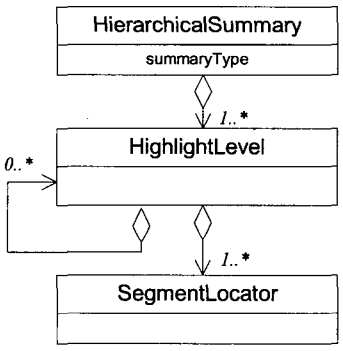

그림 2. HierarchicalSummary DS의 기본 구조
Fig. 2. Basic architecture of the HierarchicalSummary DS

2. HierarchicalSummary DS의 확장 및 개선.

기존의 HierarchicalSummary DS는 요약 비디오를 기술하는 기본적인 기능은 제공하지만 다음과 같은 제약을 갖는다. 요약 비디오를 통한 개관 후 요약 내용을 바탕으로

좀 더 자세한 내용 파악을 위하여 원 비디오를 접근하고자 하는 것이 일반적인 사용자의 비디오 접근 형태인데 반하여, 요약 비디오로부터 원 비디오의 접근을 제공하지 못한다. 또한 오디오의 요약 기술 기능을 충분히 제공하지 못하며, 사건 중심의 요약을 표현하고자 할 때 중복 기술과 탐색의 복잡성이 불가피해지는 단점을 갖고 있다[16].

이러한 문제점을 개선하기 위해서 기존의 Hierarchical-Summary DS를 확장 개선한 기술구조를 제안하였다[17][18][19]. 제안하는 기술구조는 그림 3과 같이 기존의 SegmentLocator DS 대신 VideoSegmentLocator DS, AudioSegmentLocator DS, ImageLocator DS 및 SoundLocator DS를 포괄하는 HighlightSegment DS를 가지며, 또한 SummaryThemeList DS 및 themeIds 속성이 추가되었다.

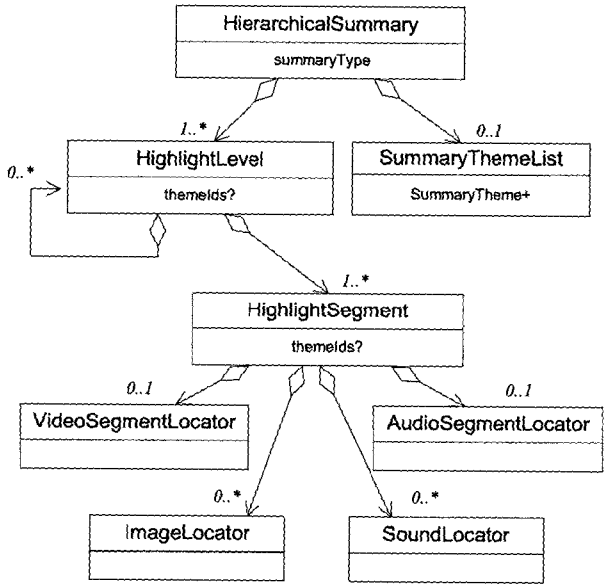

그림 3. 제안한 확장된 HierarchicalSummary DS의 구조
Fig. 3. Extension to the architecture of HierarchicalSummary DS

2.1 접근 기법의 개선

전술한 바와 같이 비디오의 충분한 내용 이해를 위해서는 동적 요약을 통한 개관과 함께 다음 단계로 브라우징 및 네비게이션 등의 접근 기능이 자연스럽게 연동되는 것이 필요하다. 본 논문에서는 효율적인 비디오 접근 기제을 제공하기 위하여 정적 요약의 접근 기능과 동적 요약의 빠른 개관 기능을 결합할 수 있는 통일된 기술 기반구조로 확장된 계층적 요약 기술구조를 제안하였다. 즉, 동적 요약을 구성하는 하이라이트 구간들의 키 프레임을 활용하여 동적 요약으로부터 원 비디오로 직접 접근할 수 있는 수단을 제공하는 것이다. 이는 그림 3와 같이 HighlightSegment DS를 도입하여 하이라이트 구간을 표현하기 위한 VideoSegmentLocator DS와 함께 그 하이라이트 구간에 해당하는 키 프레임을 표현 할 수 있는 ImageLocator DS를 함께 둠으로써 가능하게 된다.

즉, 동적 요약의 각 계층 사이 또는 동적 요약과 원 비디오 사이에서 계층적으로 구성된 키 프레임을 통하여 효율적인 접근 환경을 구성할 수 있다. 그림 4는 LoD (Level of Details)에 의하여 구성된 두 계층의 동적 요약과 원 비디오의 샷 정보를 이용한 계층적 접근 환경에서 가상 네비게이션 경로를 도시한 것으로, 화살표와 같이 동적 요약에서 원 비디오로의 다양한 접근 경로를 활용할 수 있다. 결국 제안된 요약 기술구조는 부가적인 기술(description) 없이 동적 요약과 정적 요약을 유동적으로 활용한 다양한 접

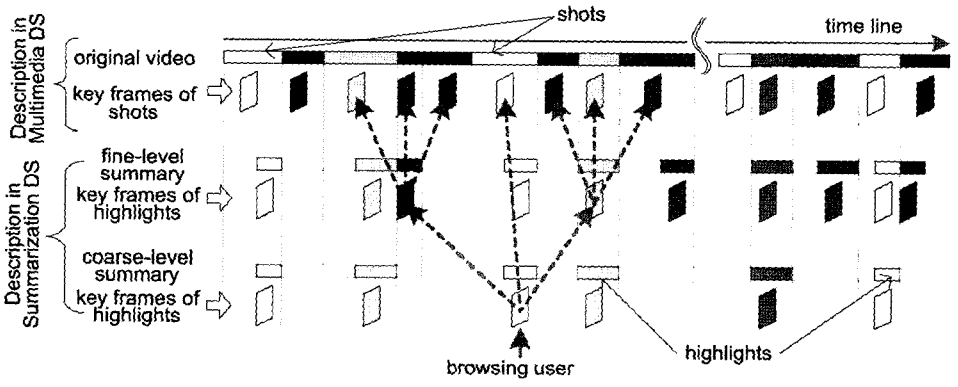

그림 4. 제안한 계층적 요약 기술구조의 다양한 접근 경로의 예
Fig. 4. An example of various navigation paths in the proposed HierarchicalSummary DS

근 경로를 통한 효율적인 브라우징 기능을 제공한다.

2.2 키 사운드의 도입

비디오 데이터의 오디오 트랙은 비쥬얼 정보와 연관된 오디오 정보를 포함하며, 일반적으로 의미적인 정보를 직접 표현하는 경우가 많으므로 전송대역이나 계산 능력이 제한되는 환경에서 특히 중요하게 다루어진다. 제한하는 HierarchicalSummary DS에서는 키 프레임에 대응되는 개념으로 요약구간의 음향 정보를 대표할 수 있는 키 사운드(key sound)를 기술할 수 있도록 SoundLocator DS를 도입하였다. 또한 AudioSegmentLocator DS를 추가하여 오디오 주요구간들로 구성된 오디오 요약을 기술할 수 있다. 축구 경기의 경우 골인, 슛 등 아나운서의 멘트나 관중의 함성 등 해당 구간의 의미를 잘 요약 대표할 수 있는 음성의 주요 단어나 특징적인 음향 효과 등을 키 사운드로 활용할 수 있다. 이들 키 사운드가 키 프레임과 함께 제공될 때 그 해당 구간의 내용을 훨씬 잘 이해할 수 있고 브라우징의 대상을 쉽게 줄여 나감으로써 더 효율적인 접근이 가능하다.

2.3 사건 기반 요약

사건 기반의 요약은 특정 사건 및 주제 별로 요약 비디오를 선택할 수 있는 사용자 주문형의 요약을 제공하는 HierarchicalSummary DS의 유용한 기능 중 하나이다. 한 예로 드라마를 인물, 배경 등의 관점에서 요약을 구성하여 특정 배우가 등장하는 장면 또는 특정한 배경을 갖는 장면 등을 기준으로 요약을 제공할 수 있다[8]. 그러나 초기의 HierarchicalSummary DS는 하나의 요약구간이 여러 사건을 포함하는 경우 그 구간이 불필요하게 해당 사건 수 만큼 중복하여 기술하는 문제가 발생한다. 또한 여러 사건의 조합으로 관심있는 구간을 브라우징할 경우 사건을 기술하는 모든 계층의 모든 요약 구간 정보를 파악해야 하는 탐색량의 증가와 조합 처리가 복잡해지는 문제가 있다[16].

이러한 문제점을 해결하기 위하여 확장된 기술구조[18]에서는 기술의 효율성과 탐색의 용이성을 고려하여 그림 3과 같이 HierarchicalSummary DS 아래에 SummaryThemeList DS가 추가되었으며, HighlightLevel DS와 HighlightSegment DS에 themeIds 속성이 추가되었다. SummaryThemeList DS는 사건 기반 요약의 기준이 되는 모든 사건이나 주제 등을 SummaryTheme 기술자(descriptor)를 이용하여 그 이름과 식별자 형태로 나열한다. 요약을 구성하는 각 요약구간을 기술하는 HighlightSegment DS에는 themeIds로 그 구간에 포함되는 사건의 종류를 표시한다. themeIds는 해당 구간에 포함되는 여러 가지 사건을 동시에 기술함으로써 한 요약구간이 여러 사건에 기술되는 기존의 중복성을 피할 수 있다. 한편, 기술의 효율성을 더욱 높이기 위하여 HighlightLevel DS에도 themeIds 속성을 두어 그 하위의 모든 HighlightSegment가 어느 공통의 사건을 포함할 경우 그 사건을 상위에서 대표적으로 기술할 수 있도록 하였다.

요약 정보를 이용하는 응용 시스템에서는 그림 8과 같이 SummaryThemeList DS에 기술된 요약기준을 목차의 형태로 제시함으로써 사용자가 요약에 포함된 사건들을 쉽게 파악하고 관심이 있는 사건 중심의 요약을 볼 수 있

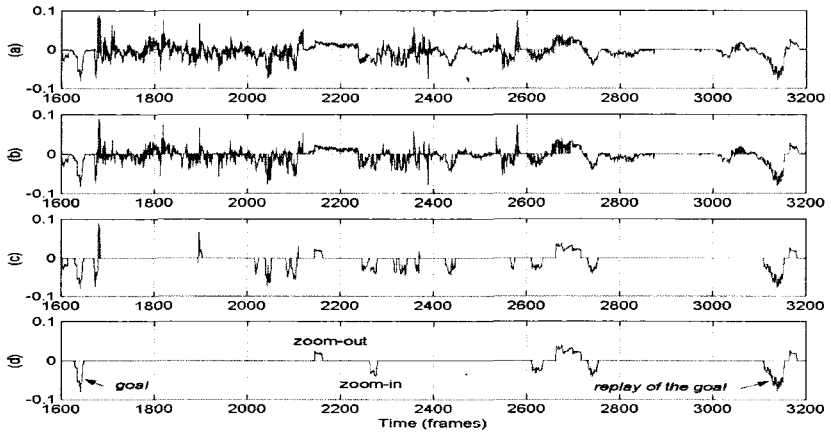

그림 5. 축구 비디오에 적용한 줌 검출 과정 (a) 각 프레임에 대한 MVF로부터 추정한 *div* 값 (b) a_2과 a_b의 부호 확인 결과 (c) 프레임 레벨의 임계화 결과 (임계값 = 0.015) (d) 구간 레벨의 임계화 결과 (임계값 = 15 프레임) 및 최종 검출된 줌-인/아웃 결과
Fig. 5. The zoom detection procedure applied to *Soccer1* sequence. (a) temporal evolution of the estimated *div* from the MVF sequence (b) the result of the sign validation (c) the result of the magnitude thresholding in the frame-level detection (d) the result of the temporal thresholding in the segment-level detection and the detected segments for zoom-in/out

도록 하는 사용자 주문형의 요약 기능을 제공한다.

V. 실험 결과

앞에서 기술한 동적 요약 기법 및 기술구조의 유용성과 타당성을 확인하기 위하여 축구 비디오를 대상으로 카메라 움직임 분석, 동적 요약 및 기술에 대한 실험 내용을 기술한다. 주 실험 데이터는 MPEG-1으로 압축된 5분 분량의 축구 시퀀스이다(*Soccer1*).

그림 5는 본 논문의 카메라 움직임 분석을 통한 줌 검출 과정을 도시한 것이다. 줌에 해당하는 *div* 계수를 프레임 레벨(그림 5(a)-(c))과 구간 레벨(그림 5(d))에서 임계화하는 과정을 나타낸다. 그림 5(b)는 *div*를 구성하는 과 가 같은 부호를 가져야 한다는 사실을 적용하여 서로 다른 부호를 갖는 프레임을 배제한 결과이다. 표 1은 이와 같은 검출 과정을 통한 5분 분량의 축구(*Soccer 1*)와 골프(*Golf*) 비디오에 대한 검출 성능이다.

사건 검출에 대한 실험 결과로 사건 검출을 위한 각 특징의 확률이득을 표 2에 나타내었다. 즉, 검출하고자 하는 골과 슛 사건이 대개 원경의 그라운드 샷에서 발생하며, 카메라 줌을 동반하고 자막 및 중복재생이 자주 검출되는 현상을 활용하여 이들 특징 및 사건 검출을 바탕으로 골 및 슛을 검출하고자 할 때 표 2과 같은 확률이득을 기대할 수 있다는 것이다.

그림 6은 사건 및 에피소드 검출을 바탕으로 최종 요약구간을 선택하는 요약 규칙의 적용 예를 보여준다. 그림 6에서 보듯이 (a)-(d)의 단계를 거쳐 (e)와 같은 최종적인 요약구간을 추출한다. 사건 샷 내의 요약구간 선택은 카메

표 1. 줌 검출 성능
Table 1. Performance of zoom detection

Sequences	# segments with zoom	Correctly detected	Missed	Falsely detected
Soccer1	19	17	2	2
Golf	13	13	0	0

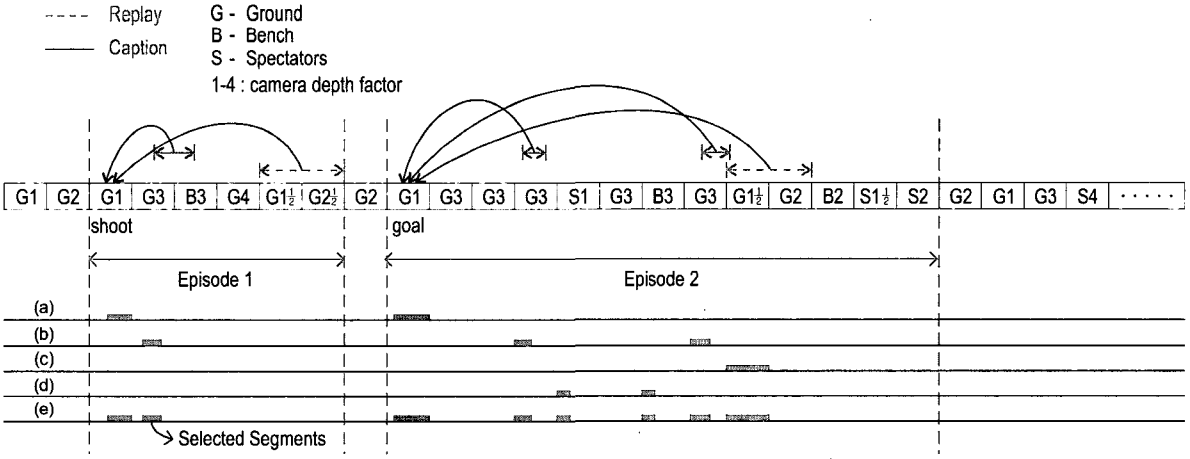

그림 6. 축구 비디오 *Soccer1*에서 사건 검출 및 에피소드 경계 검출의 실례(각 블록은 하나의 샷에 해당하며, 블록 내의 표기는 각각 샷 패턴 분류 정보(G, B, S)와 카메라 뷰 정도('원경(1)'에서 '근경(4)')를 나타낸다.)
(a) 사건 샷(골, 슛) (b) 자막영역 (c) 중복재생 샷 (d) 보조적인 내용(벤치, 관중석 샷)
Fig. 6. The event detection and episode boundary identification applied to *Soccer1*sequence (Each shot is represented by a block that carries the information of shot category and degree of camera views (from 'distant' (1) to 'close' (4)).)
(a) event shots (goal or shoot) (b) embedded caption (c) replayed shot (d) ausiliary contens (bench, audience shot)

표 2. 각 사건 검출을 위한 여러 특징에 대한 확률이득(실험값)
Table 2. Gain of each feature to the event occurrence probability (empirical)

Event	Goal/Shoot					Replay	
Feature	Zoom-in	Caption	Far view	Ground shot	Replay	Slow motion	Gradual transition
Gain	2.236	9.396	2.577	1.076	7.934	10.02	4.956

라 줌-인 구간을 선택하여 실제로 골, 슛이 발생하는 장면을 포함하도록 한다. 실제로 그림 5의 (d)에서 첫 번째와 마지막으로 검출된 줌-인은 실제 요약 과정을 보여주는 그림 6 (a)의 두 번째 골과 (c)의 골 장면의 중복재생에 해당한다.

요약구간들로 구성되는 요약 비디오가 시각적으로 거슬림 없이 자연스럽게 연결 재생되도록 요약구간들이 원 비디오의 샷 경계를 벗어나지 않도록 하고, 시청각 정보의 의미가 충분히 전달되도록 한 구간의 길이에 대한 하한값을 설정하는 편집규칙을 적용하였다.

표 3은 앞에서 기술한 요약 과정을 통한 몇 가지 축구 비디오에 적용한 최종 요약 결과 이다. 각각의 비디오 요약은 골과 슛을 기준으로 대략 15:1 내외로 요약되어 있으며, 사용자의 이해를 위하여 부대 내용(자막, 중복재생, 관중석 샷 등)도 어느 정도 포함하고 있다. 본 실험은 카메라 움직임이외의 특징들은 사전에 매뉴얼로 추출하였으며 요약 기법이 비교적 안정적이고 효과적임을 확인할 수 있었다.

제안한 계층적 요약 기술구조에 대한 검증 실험은 MPEG-7 CE(Core Experiment)[20][21]를 통하여 요약 기술 데이터 생성 및 파싱(parsing)과 소프트웨어 구현 등의 형태로 수행되었다. 본 논문에서는 축구 비디오 Soccer1에 대하여 그림 6과 같은 과정으로 생성된 동적 요약을 제안한 요약 기술구조에 따른 기술 예를 제시하고자 한다. 본 실험에서는 표 4와 같이 두 레벨(coarse와 fine)의 다중 길이의 동적 요약을 생성했으며 다 수의 사건을 정의하고 각 요약 구간에 해당하는 사건을 매뉴얼로 추출하였다. 그림 6은 fine-level의 요약을 구성하는 과정을 일부 도시한 것이다. 각 요약구간의 상세한 요약 정보는 표 5와 같다. 표 5에서 첫 번째 요약구간은 coarse 레벨에 속하고 192~280의 프레임 구간에 해당하며 사건 Goal/Shoot에 해당됨을 나타낸다. 표 5의 요약 정보가 요약 기술구조의 각 기술구조, 기술자, 속성에 어떻게 대응되어 기술되는지를 그림 7에서 나타내었다.

표 3. 축구 비디오 요약 결과
Table 3. Summarization results for *Soccer* sequences

Sequence	Length (min:sec)		Compaction Ratio
	Original	Summarized	
Soccer1	05:07	00:41	7.35:1
Soccer2	15:00	01:10	12.7:1
Soccer3	43:49	02:10	20.1:1

표 4. 축구 비디오의 요약 정보
Table 4. Summary information for *Soccer1* sequence

Test Sequence (Duration)	Summary Information		
	Level	Duration	Key Events
Soccer1(05:07)	coarse	0:27 (1/11.02)	Goal/Shoot; Close-up; Bench; Audience; Replay
	fine	0:41 (1/7.35)	

표 5. 축구 비디오의 상세 요약 정보
Table 5. Detailed summary information for *Soccer1* sequence

Segment ID	MediaTime		Level	Events				
	Start	End		Goal/Shoot	Close-up	Bench	Audience	Replay
1	192	280	coarse	o				
2	352	410	fine		o			
3	1,552	1,670	coarse	o				
4	1,869	1,942	coarse		o			
5	2,119	2,178	fine				o	
6	2,398	2,457	fine			o		
7	2,530	2,591	fine		o			
8	2,592	2,895	coarse					o
9	4,575	4,663	coarse	o				
10	4,732	4,790	fine		o			
11	5,697	5,785	coarse	o				
12	6,014	6,072	fine		o			
13	7,560	7,631	coarse	o				
14	7,714	7,772	fine		o			

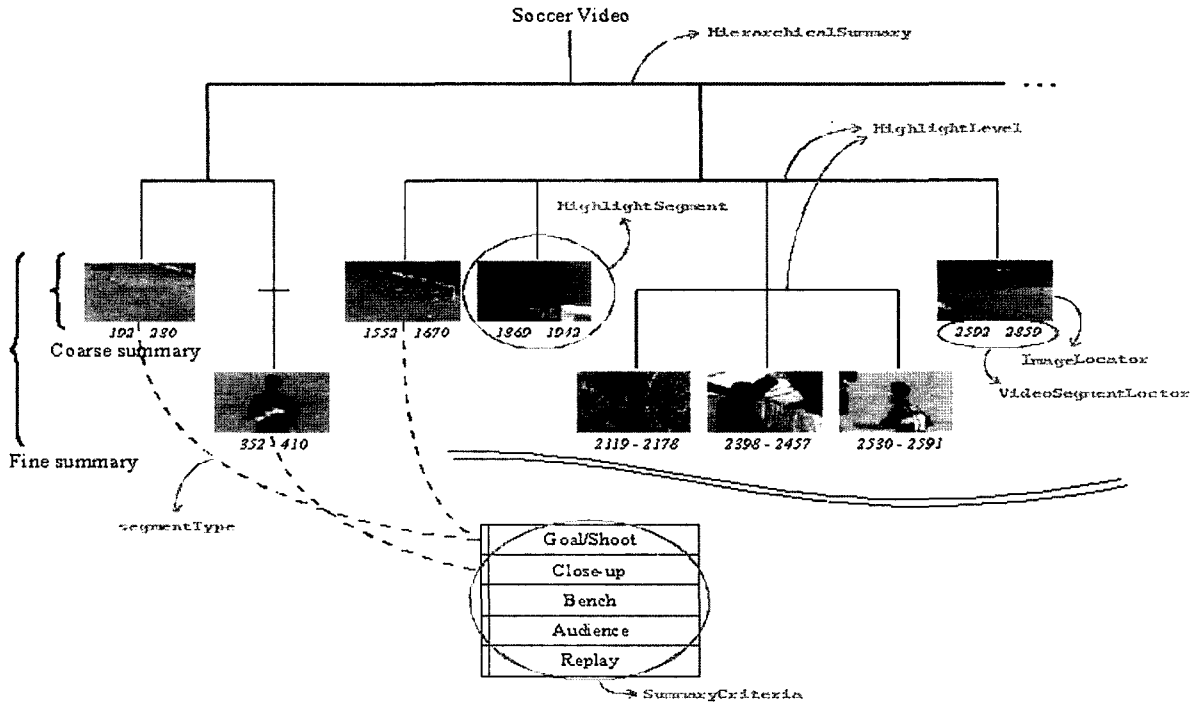

그림 7. 요약 기술구조에 따른 축구 요약 비디오의 기술 예
Fig. 7. Pictorial representation of the summary of *Soccer1* sequence

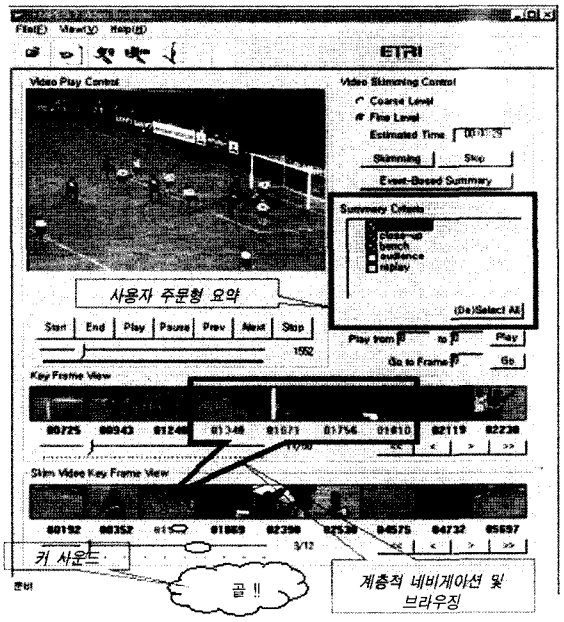

그림 8. 비디오 브라우저의 사용자 인터페이스
Fig. 8. The graphical user interface of the exemplary application system (A video browser)

실제로 본 논문에서 제안하는 계층적 요약 기술구조로 기술된 요약 비디오의 효율성과 실현 가능성의 확인을 위해 요약 기술 데이터를 입력하는 요약 기반의 비디오 브라우저(browser)를 구현하였다. 그 사용자 인터페이스는 그림 8과 같으며 브라우저의 실험을 통하여 1) 다중 길이의 동적 요약을 통한 개관, 2) 키 프레임 및 키 사운드를 이용한 계층적 네비게이션 및 브라우징, 3) 사건 정보를 활용한 사용자 주문형 요약 및 브라우징 등 주요 기능의 효율성과 유용성을 확인하였다.

VI. 결 론

본 논문에서는 방대한 디지털 비디오의 효율적 접근 기능을 제공하기 위하여 동적 요약 기법과 요약 비디오를 기술하기 위한 MPEG-7 기술구조를 제안하였다. 스포츠 등의 사건 중심 비디오에 적용할 수 있는 동적 요약의 새로운 접근방법으로 사건과 물리적 현상간의 상관성을 이용하는 Bayes rule에 기반한 기법을 제시하였고, 이의 자동 구현을 위한 비디오 내용 분석 중 카메라 움직임 분석 기법을 간략히 소개하였다. 축구 비디오의 적용 실험을 통

하여 제안한 요약 기법이 의미적인 측면에서 효과적인 요약 결과를 보이고 있음을 확인하였다. 또한 요약 비디오를 기술하기 위하여 MPEG-7에 제안한 Hierarchical-Summary DS의 제안 배경 및 이를 통하여 개선된 접근 기능을 살펴 보았다. 또한 브라우저 구현을 통하여 제안하는 기술구조가 효율적인 비디오 브라우징과 네비게이션 기능을 제공함을 확인하였다. 향 후 비디오 분석 자동화 등을 통한 동적 요약 및 기술(description)의 완전 자동화와 그 응용에 대한 연구가 계속 진행될 전망이다.

참 고 문 헌

[1] D. Yow, B.-L. Yeo, M. M. Yeung, and B. Liu, "Analysis and Presentation of Soccer Highlights from Digital Video," *in Proc. 2nd Asian Conf. Computer Vision*, vol. 2, Dec. 1995, pp. 499-503.

[2] Y. Rui, T. S. Huang, and S. Mehrotra, "Exploring Video Structure Beyond the Shots," *in Proc. IEEE ICMCS'98*, Jun. 1998, pp. 237-240.

[3] M. Yeung and B-L Yeo, "Segmentation of Video by Clustering and Graph Analysis," *Computer Vision and Image Understanding*, vol. 71, no. 1, pp. 94-109, Jul. 1998.

[4] H. S. Chang, S. Sull, and S. U. Lee, "Efficient Video Indexing Scheme for Content-Based Retrieval," *IEEE Trans. Circuits Sys. for Video Technol.*, vol. 9, no. 8, pp. 1269-1279, Dec. 1999.

[5] H. Wactlar, T. Kanade, M. Smith, and S. Stevens, "Intelligent Access to Digital Video: The Informedia Project," *IEEE Com.*, vol. 29, no. 5, pp. 46-52, May. 1996.

[6] S. Pfeiffer, R. Lienhart, S. Fischer, and W. Effelsberg, "Abstracting Digital Movies Automatically," *J. Vis. Commun. Image Represent.*, vol. 7, no. 4, pp. 345-353, Dec. 1996.

[7] R. Qian, N. Haering, and I. Sezan, "A Computational Approach to Semantic Event Detection," *in Proc. IEEE CVPR'99*, vol. 1, Jun. 1999, pp. 200-206.

[8] J. Saarela and B. Merialdo, "Using Content Models to Build Audio-Video Summaries," *in Proc. SPIE Storage and Retrieval for Image and Video Databases VII*, vol. SPIE-3656, Jan. 1999, pp. 338-347.

[9] A. Hanjalic, R. L. Lagendijk, and J. Biemond, "Semi-Automatic News Analysis, Indexing and Classification System based on Topic Preselection," *in Proc. Storage and Retrieval for Image and Video Databases VII*, Vol. SPIE-3656, Jan. 1999, pp. 86~97.

[10] A. Hanjalic, R. L. Lagendijk, and J. Biemond, "Automated High-Level Movie Segmentation for Advanced Video-Retrieval," *IEEE Trans. Circuits Sys. for Video Technol.*, vol. 9, no. 4, pp. 580-588, Jun. 1999.

[11] J.-G. Kim, H. S. Chang, J. Kim, and H.-M. Kim, "Efficient Camera Motion Characterization for MPEG Video Indexing," *to appear in Proc. IEEE ICME2000*, Aug. 2000.

[12] 장현성, 김재곤, 김진웅, "비디오 요약을 위한 주요구간 검출 및 표현 방법", *제12회 영상처리 및 이해에 관한 워크샵 논문집*, pp. 201-205, 2000년 1월.

[13] J.-G. Kim, H. S. Chang, M. Kim, J. Kim, and H.-M. Kim, "Summary Description Schemes for Efficient Video Navigation and Browsing," *in Proc. VCIP2000*, vol. SPIE-4067, Jun. 2000, pp. 1397-1408.

[14] "MPEG-7 Requirements Document," *ISO/ IEC JTC1/SC29/WG11*, (N2727) Mar. 1999.

[15] "MPEG-7 Multimedia Description Schemes WD (Ver. 3.0)," *ISO/IEC JTC1/SC29/WG11*(N3247), May. 2000.

[16] "MPEG-7 Description Schemes(V0.5)," *ISO/IEC JTC1/SC29/WG11 (N2844)*, Jul. 1999.

[17] M. Kim, J.-G. Kim, H. S. Chang, S.-Y. Kim, and J. Kim, "An Extended Summary DS for Navigation and Browsing," *ISO/IEC JTC1/SC29/WG11 (M5022)*, Oct. 1999.

[18] M. Kim, J.-G. Kim, H. S. Chang, and J. Kim, "An Efficient Hierarchical Summary DS for Event-Based Summarization," *ISO/IEC JTC1/SC29/WG11 (M5493)*, Dec. 1999.

[19] H. S. Chang, J.-G. Kim, M. Kim, and J. Kim, "Improved Structure of HierarchicalSummary DS," *ISO/IEC JTC1/SC29/WG11 (M6057)*, Jun. 2000.

[20] M. Kim, J.-G. Kim, H. S. Chang, and J. Kim, "Results of the Validation Experiments on Summary DS," *ISO/IEC JTC1/SC29/WG11 (M5492)*, Dec. 1999.

[21] H. S. Chang, J.-G. Kim, M. Kim, and J. Kim, "Results of Core Experiments on Summary DS," *ISO/IEC JTC1/SC29/WG11 (M5852)*, Mar. 2000.

저 자 소 개

김 재 곤

1990년 2월 : 경북대학교 전자공학과 공학사
1992년 2월 : 한국과학기술원 전기 및 전자공학과 공학석사
1992년 3월 ~ 현재 : 한국전자통신연구원 선임연구원
주관심분야 : 비디오 신호처리, 비디오 색인 및 검색

장 현 성

1997년 2월 : 서울대학교 전기공학부 공학사
1999년 2월 : 서울대학교 전기공학부 공학석사
1999년 3월 ~ 현재 한국전자통신연구원 연구원
주관심분야 : 영상 신호처리, 영상 내용 분석 및 이해

김 문 철

1989년 2월 : 경북대학교 전자공학과 공학사
1992년 12월 : University of Florida, Electrical and Computer Engineering, 공학석사
1996년 8월 : University of Florida, Electrical and Computer Engineering 공학박사
1997년 1월 ~ 현재 : 한국전자통신연구원 선임연구원
2000년 3월 ~ 현재 : 한국전자통신연구원 방송미디어연구부 실감영상연구팀장
주관심분야 : 멀티미디어 정보처리, 멀티미디어 통신, 디지털 신호/영상 처

김 진 웅

1981년 : 서울대학교 공과대학 전자공학과 학사,
1983년 : 서울대학교 대학원 전자공학 석사,
1993년 : Texas A&M Univ. Electrical Engineering 박사,
1983년 3월 ~ 현재 : 한국전자통신연구원 재직중, 책임연구원.
주관심 분야 : 디지털 신호 처리, 영상 통신, 디지털 라이브러리, VLSI 신호처리

김 형 명

1974년 2월 : 서울대학교 공학사
1982년 4월 : 미국 Pittsburgh 대학 전기공학과 석사
1985년 12월 : 미국 Pittsburgh 대학 전기공학과 공학박사
1986년 9월 ~ 현재 : 한국과학기술원 전기 및 전자공학과 교수
주관심분야 : 이동 통신 기술, 디지털 신호와 영상 처리, 다차원 시스템 이론, 다중사용자 검파기, 호 수락 제어