

□ 기술개설 □

비디오 콘텐츠 제작 및 서비스 기술

한국통신 이병도·김혁만*

1. 개 요

최근 멀티미디어 기술이 급부상하여 정보 통신 분야의 핵심 이슈로 부상하고 있다. 이에 따라 정보의 표현이 문서 형태에서 듣는 형태와 보는 형태 쪽으로 점차 비중이 옮겨가고 있다. 디지털 비디오 라이브러리(이하 비디오 라이브러리)는 초고속 통신망을 이용한 비디오 서비스 전반을 포함하는 개념으로서 비디오 매체가 가진 정보 전달 능력, 정보 또는 지식 습득의 효율성, 그리고 그것이 갖는 상업성으로 인하여 최근 주목받고 있는 분야이다. 디지털 라이브러리¹⁾의 한 형태인 비디오 라이브러리에서 모든 비디오는 디지털 데이터 포맷을 가지며 내용의 직접 접근을 허용하고 또한 네트워크를 통하여 전송받아 재생시킬 수 있다.

비디오 라이브러리 시스템은 일반 아날로그 비디오로서 라이브러리에 수록될 비디오 클립을 생성하는 제작 부분, 많은 비디오 클립들을 데이터베이스에 관리하는 저장 부분, 컴퓨터 통신망을 통해 원격지에서 데이터베이스의 특정 비디오 클립을 찾아내는 검색 부분, 그리고 검색된 비디오 클립을 사용자에게 빠르게 전달하는 전송 부분 등으로 구성된다.

제작 부분은 방송 녹화 테잎이나 일반 비디오 테잎을 디지털화하고 장면의 경계를 찾아내어 클립을 생성한 후 각 장면 단위로 색인 정

보를 만드는 등 일련의 처리 과정을 거친다. 저장 부분은 대용량 비디오 클립과 관련 색인 정보를 관리하며, 다수의 사용자에게 동시에 전송할 수 있는 고성능 입출력 능력을 가져야 한다. 일반적으로 비디오 서버는 디스크 어레이와 같은 대용량의 고성능 저장시스템이나 CD-ROM 또는 테잎의 유크박스를 포함한 계층 기억장치를 사용하기도 한다. 검색 부분은 대개 내용 기반 검색을 의미하는 것으로서 제작 단계에서 추출한 이미지, 음성 또는 움직임 특성이나, 대본 등을 이용하여 원하는 비디오 클립이나 장면을 검색한다. 비디오의 경우 전송시에 타 매체보다 훨씬 큰 대역폭이 요구된다. 현재의 통신 환경으로는 비디오 서비스 사용자들을 전화망을 이용하는 저대역폭 사용자와 LAN을 이용하는 고대역폭 사용자로 나눌 수 있다. 따라서 비디오 전송은 저대역폭 및 고대역폭 비디오를 위한 압축 포맷, 전송 및 재생 기술 등이 필요하다.

본고에서는 비디오 라이브러리의 제작과 전송 분야의 기술을 주로 살펴본다. 본고의 구성은 다음과 같다. 2장에서는 비디오 클립을 만드는 과정에 소요되는 기술을 살펴보고, 3장에서는 한국통신에서 개발하고 있는 클립 생성 시스템에 대하여 설명한다. 제4장에서는 비디오 스트리밍 기술과 통신 프로토콜, 그리고 제5장에서는 한국통신 비디오 라이브러리 시스템 구성을 설명한다.

*정회원

1) 디지털 라이브러리는 그것이 취급하는 미디어의 형태에 따라 문서 라이브러리, 이미지 라이브러리, 오디오 라이브러리, 비디오 라이브러리 그리고 지도 라이브러리 등으로 분류된다.

2. 비디오 클립 제작 기술

아날로그 형태(NTSC 혹은 PAL)의 원시 비

디오로부터 검색 가능한 비디오 클립을 생성하려면 디지털 신호로의 변화 및 압축, 샷 경계 검출 및 장면 분할, 대표 화면 추출, 브라우징 및 색인 정보 구성 등의 과정을 거친다.

2.1 비디오 샷 경계 검출 및 장면 분할

영상은 제작 단계에서 물리적인 단위인 샷(shot)과 논리적인 단위인 장면(scene)으로 구성된다. 샷이란 촬영 시에 카메라가 멈춤 없이 한 번에 기록한 연속적인 동작을 의미하고, 장면은 논리적으로 서로 밀접한 관련이 있는 연속적인 샷의 묶음(cluster)이다[13]. 일반적으로 샷은 2-10초 정도이며, 장면은 영상의 내용에 따라 그 편차가 상당히 크다. 영상 제작자는 편집시에 이미 샷과 장면을 정의하여 영상을 제작하므로 이 샷과 장면 단위로 영상을 인덱싱하여 검색하는 것이 제작자 및 사용자의 입장에서 가장 자연스러울 것이다.

비디오 인덱싱을 위해 특정 비디오에서 샷 경계를 검출(shot change detection 혹은 segmentation)하고, 이들을 내용에 따라 하나로 묶어 장면 단위로 분할(scene change detection 혹은 clustering)하는 작업은 수작업으로 할 경우 시간과 노력이 매우 많이 드는 고비용의 작업이다. 특히 입력할 비디오가 많아지면 주어진 시간에 모든 샷과 장면을 찾아내어 검색에 필요한 인덱스를 구축하기 어려워진다. 따라서 대용량의 비디오 라이브러리 구축시에는 이를 자동으로 할 수 있는 소프트웨어 도구의 도움이 필수적이다.

샷 경계 검출은 비디오에서 샷이 바뀌는 경계 부분을 검출하는 기술을 의미하는 것으로서 클립 생성을 위한 비디오 장면 분할 작업의 시작 부분에 해당한다. 샷 경계 검출은 또한 내용 기반 검색 및 대표화면 추출에 없어서는 안 될 요소 기술이기도 하다. 비디오의 샷들은 단순히 샷을 이어 붙일(cut) 뿐만 아니라 dissolve, fade, wipe 등의 특수 편집 효과를 사용해 편집 시에 자연스럽게 이어지도록 제작된다. Cut은 전혀 다른 내용의 프레임이 인접하여 있으므로 샷의 경계가 분명하기 때문에 인접한 프레임 간의 픽셀 농도 차이를 이용하거나 컬러/명도 히스토그램의 차이를 이용하는

등의 통계적인 방법만으로도 경계 검출이 가능하다. 그러나 특수 효과가 사용된 경우 샷의 전환이 여러 프레임에 걸쳐 연속적으로 진행하기 때문에 샷 경계가 cut에 비하여 불분명하여 경계 검출이 어렵다. 특히 패닝(panning), 주밍(zooming)과 같은 카메라 동작이 많이 사용된 경우에는 이들을 샷의 경계로 오인하는 경우가 많다. 따라서 정확히 샷 경계를 검출하려면 카메라 동작의 분석이 동시에 이루어져야 한다.

샷 경계 검출 알고리즘은 사용하는 입력 비디오의 형태에 따라 비압축 비디오에 대한 검출 알고리즘과 압축 비디오에 대한 검출 알고리즘으로 구분할 수 있다. 비압축 비디오에서의 검출 알고리즘은 압축되지 않은 디지털 비디오 포맷을 입력으로 사용하는 것으로서 인접한 프레임간의 여러 종류의 변화를 감지하여 샷 경계를 검색한다. Zhang 등[16]은 인접한 두 프레임에서 동일 좌표에 위치한 화소간 또는 블럭간 intensity 값의 차이가 일정 크기보다 큰 화소들의 개수를 구하고, 그 수가 임계값 이상인 프레임을 샷 경계로 결정하였다. Hampapur 등[4]은 비디오의 제작 과정에 기반하여 샷의 경계를 만드는 과정을 수학적 모델로 만들어 샷 경계를 결정하였고, Zabih [14]는 프레임의 에지(edge) 영상으로부터 entering pixel과 exiting pixel의 개수를 계산하고 그들의 공간적 분포를 분석하여 샷 경계 검출을 시도하였다.

비압축 비디오를 사용하는 알고리즘의 가장 큰 문제는 그것의 방대한 볼륨으로 인한 저장 공간의 확보와 처리속도의 지연이다. 따라서 비디오의 효율적인 분석을 위하여는 압축된 데이터를 사용한다. 일반적으로 우리가 다루는 대개의 비디오 데이터는 압축된 형태이므로 압축된 상태에서의 영상 처리는 복원-처리-재압축의 단계를 생략할 수 있다. 압축 비디오에서의 탐색 알고리즘은 MPEG, M-JPEG 등으로 압축된 데이터를 대상으로 하는 것으로서 영상의 공간적 상관관계를 나타내는 DCT 계수나 이동 벡터(motion vector)를 이용한다. Arman 등[2]은 MPEG의 I 프레임에서 연속된 두 프레임 사이의 DCT 계수를 비교하여 그 차이가

임계값 이상인 프레임을 샷 경계로 정하였다. Yeو 등[13]의 알고리즘 역시 MPEG 데이터에서 DCT 계수를 이용하여 급속한 샷 전환과 점진적인 샷 전환을 탐색한다. 점진적인 샷 전환의 경우엔 인접한 프레임 대신 k만큼 떨어진 프레임을 이용한다. 이들은 알고리즘의 실행 시간 단축을 위하여 축소된 영상을 이용하기도 한다. 이동 벡터를 사용한 방법은 샷 경계에 해당하는 프레임에서의 대부분 이동 벡터의 값이 0이 된다는 사실에 근거하고 있다. Meng 등[8]은, I 프레임 직전에 있는 B 프레임이 디스플레이 상으로 앞의 I/P 프레임보다 뒤에 위치한 I 프레임과 유사한 정도를 비교하여 샷 경계를 결정한다. 이들 샷 경계 탐색 알고리즘은 카메라의 빠른 움직임, 빨리 움직이는 물체가 있는 샷, 밝기의 급격한 변화, 그리고 편집 특수 효과 등에 의하여 정확도가 떨어지는 문제점이 있으며 이를 해결해 나가는 방향으로 연구가 진행되고 있다.

비디오 장면 분할은 검출된 샷들을 분석하여 비슷한 내용의 샷들을 묶는 작업이다. 장면은 다분히 주관적 판단에 따라 결정되므로 이 작업을 자동화하기에는 상당한 어려움이 따른다. 예를 들어 뉴스 클립의 경우 일반적으로 앵커 샷과 그 뉴스를 설명하는 여러 배경 샷들로 구성된다. 이때 앵커 샷과 배경 샷들을 각각 장면으로 정의 할 수도 있고, 이들을 합하여 하나의 장면으로 정의할 수도 있다. 또 만일 배경 샷들의 내용들이 아주 다른 것들이라면 배경 샷들을 두 개 이상의 장면으로 분할할 수도 있다. 따라서 자동 장면 분할은 샷들의 신호 특성이 아주 유사한 것들만을 묶을 수 있으며, 이 이상은 보조적인 정보 혹은 사람의 주관에 개입될 수 밖에 없다. 가장 많이 사용하는 보조 수단은 영상 분석뿐만 아니라 음성의 변화를 분석하여 샷의 경계에서 음성이 이어질 때는 인접한 두 샷이 같은 장면에 속하고, 샷의 경계에서 음성이 단절되면 장면을 분할하는 것이다.

이 방법은 동시 녹음으로 제작된 비디오에서는 매우 유용하지만, 촬영과 녹음이 분리되어 제작되는 비디오에서는 정확도가 떨어지게 된다.

2.2 비디오 브라우저

비디오 응용에서는 비디오 시퀀스의 특정 위치를 지정하는 기능과 많은 비디오 클립 중에 일부를 선택하는 기능이 매우 중요하다. 이들 기초 기능을 구현하여 비디오 데이터를 탐색할 수 있도록 한 도구를(내용 기반) 비디오 브라우저라 한다. 비디오 브라우저는 사용자 인터페이스로서 비디오 파싱, 편집, 검색 등의 비디오 응용에 공통적으로 사용된다. 하나의 비디오 장면은 대표화면으로 표시되어 브라우저에 나타나며, 해당되는 비디오 시퀀스가 가지는 정보를 사용자에게 전달한다. 예를 들어 뉴스 기사를 담은 비디오 장면의 대표화면은 뉴스의 개략적 내용을 설명할 수 있으며, 비디오 클립 아이콘의 두께는 재생 시간을 나타낼 수 있고, 또한 아이콘의 측면을 이용하여 물체의 체적을 표현할 수 있다. 가장 기본적인 브라우저 방법은 대표화면을 나열하여 전체 비디오의 내용을 이해할 수 있게 한다.

Tonomura 등[12]은 비디오 장면의 내용을 볼 수 있는 몇 가지 방법을 제시하였다. 가변 속도 브라우저는 VCR의 조그/서플 기능을 흉내낸 기능으로서 브라우저상에 비디오의 조그/서플 버튼과 스피드 커서를 두었다. Sampling flash 브라우저는 장면 경계에 위치한 첫 화면을 단순히 디스플레이하는 것으로서, 첫 화면을 대표화면으로 지정한다. Rush 브라우저는 장면 분할 없는 상태에서 전체 비디오를 일정한 시간 간격으로 나누어 그 위치에 해당하는 화면을 나열한다. Time-space 브라우저에서는 비디오 시퀀스를 시간적으로 분할하여 여러개의 아이콘을 차례로 나열하였다. 이밖에도 이들이 제시한 비디오 브라우저는 비디오 클립들 간의 관계를 볼 수 있다.

Zhang[15], Mills[9] 등도 내용 기반 비디오 브라우저를 개발하였는데, 이들도 비디오의 장면을 기본 단위로 하여 탐색을 하며 그 기능 역시 Tonomura의 브라우저와 유사하다. Arman 등 [1]이 제시한 브라우저에는 비디오 클립의 테두리에 motion tracking region을 두어 화면 테두리 부분 물체의 움직임을 추적할 수 있게 하였다. 이들이 제안한 motion tracking 기법

은 장면 경계를 판단하는데 매우 유용한 수단이기도 하다.

3. 클립 생성 시스템

원시 비디오로부터 비디오 클립 생성시 가장 중요하고 어려운 기술은 샷 경계 검출 및 장면 분할을 자동으로 할 수 있는 기능이다. 대용량 입력 영상을 분석하여 샷과 장면을 자동적으로 인식하려면, 먼저 입력 영상을 하나의 화일에 압축 저장해야 한다. 비디오 압축에는 M-JPEG 또는 MPEG을 이용해 압축이 가능하다. M-JPEG은 압축률은 떨어지지만(논리적으로는 MPEG의 I 프레임만으로 구성됨), 압축된 영상을 프레임 단위로 조작(cut and paste)할 수 있다. MPEG은 높은 압축률을 제공하나, 일반적으로 12개 혹은 15개 프레임의 GOP(group of pictures) 단위로만 영상 조작이 가능하다. 따라서 일반적으로 비디오 편집 및 내용 분석 용으로는 M-JPEG을 사용하며, 최종적으로 만든 비디오 클립의 영구 저장 및 전송시에는 MPEG을 사용한다.

그림 1은 현재 우리가 개발 중인 비디오 클립 제작 시스템의 S/W 구조를 샷 경계 탐색 및 장면 분할 기능을 중심으로 보여준다. 비디오 클립 제작 시스템은 입력 비디오를 분석하여 몇개의 장면으로 이루어진 비디오 클립들로 나누어주는 시스템이다. 이 시스템에 의해 생성되는 비디오 클립들은 인덱스 정보와 함께 비디오 서버에 저장된다. 이 비디오 클립 제작 시스템에서는 입력 영상을 분석하여 샷과 장면

을 자동적으로 인식할 때는 M-JPEG을 사용해 정확한 프레임 위치를 인식하고 그 위치에서 영상을 분할하며, 실제로 분할된 영상을 압축하여 저장할 클립을 생성할 때는 MPEG을 사용하여 높은 압축률을 달성하는 방법을 택하였다. 이 클립 생성 시스템의 핵심 부분은 MH 서브시스템(M-JPEG Handling Subsystem)과 VA서브시스템(Video/audio Analysis Subsystem)으로 M-JPEG 형태의 비디오 데이터를 가지고 장면 분할, 대표화면 추출 등의 기능을 수행한다.

VA 서브시스템은 M-JPEG으로 압축된 비디오를 전달받아 샷의 경계를 검출하고 샷의 대표화면, 즉 Kframe(Key frame)을 추출한다. VA 서브시스템에서 샷의 경계를 찾아내는 비디오 파싱 알고리즘은 갑자기 변하는 샷 경계뿐만 아니라 디스솔브(dissolve), 페이드(fade), 와이프(wipe) 등과 같은 완만한 변화 역시 감지하며, 플리시 라이트와 같은 불빛의 급한 변화뿐만 아니라 줌(zoom), 팬(pan) 등과 같은 카메라 동작도 인식한다[5, 6, 11].

VA 서브시스템에서 샷의 경계를 찾아내는 비디오 파싱 알고리즘은 여러 형태의 장면 변화를 분석하는 부속 알고리즘들로 구성된다. 모든 샷 경계를 100% 정확하게 검출하는 시스템 또는 알고리즘은 아직 없다. 현재까지 발표된 샷 경계 검출 기법은 대개 제한된 샷 변화 패턴에서만 제 성능을 발휘하고 있다. 우리의 VA 서브시스템에서는 각 샷 변화 패턴에 효율적인 알고리즘을 조합하여 다양한 샷 경계 변화를 검출할 수 있게 하였다. 샷의 경계 검출과 아울러 VA 서브시스템에서는 각 샷의 대표화면인 Kframe, 그리고 샷의 카메라 동작과 히스토그램, Kframe에 대한 이미지 분석 결과 등과 같은 샷의 특성을 추출한다. 이 정보들은 샷에 대한 메타 정보로서 비디오 서버에 저장 관리되어 내용 기반 검색을 위해 사용된다.

이 서브시스템에서는 논리적으로 연결되어 있는 연속적인 샷들을 묶어 자동적으로 장면을 만들어낸다. 각 장면에 대해서도 샷의 경우와 마찬가지로 장면의 대표화면, 즉 Rframe(Representative frame)을 정의한다. Rframe은 해당 장면에 속한 샷들의 Kframe들 중 전체 장

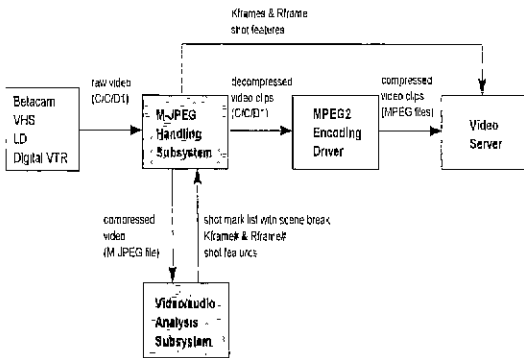


그림 1 클립 생성 시스템 S/W 구조

면을 가장 잘 나타내는 Kframe으로 선택한다. 장면을 자동적으로 생성하기 위해서는 영상과 음성을 동시에 이용하는 것이 가장 효율적이다. 현재 이를 자동화하는 알고리즘을 개발 중에 있다.

MH 서브시스템은 원시 비디오를 M-JPEG 압축 화일로 변환하여 VA 서브시스템에 전달한다. VA 서브시스템에서는 샷 경계를 검출하고 장면 분할 작업을 수행하여 그 결과를 MH 서브시스템에 전달한다. 샷과 장면 그리고 그들의 특성을 전달받은 MH 서브시스템은 이를 이용하여 샷과 장면이 올바르게 검출되었는지를 확인한다. 이 확인 작업에서는 알고리즘이 검출하지 못한 샷과 잘 못 검출한 샷들을 눈으로 식별하면서 찾아낸다. 이를 위해 비디오 브라우저가 사용되는데, 본 시스템의 브라우저는 일반 VCR의 기능뿐만 아니라 스트리보드, 앞서 언급한 샘플링 플래쉬(sampling flash), 그리고 모션 트래킹(motion tracking) 기능을 가진다. 특히 motion tracking 기능은 이 확인 작업을 빠르고 쉽게 할 수 있게 한다. 확인 작업이 완료된 후, 샷 및 장면 경계의 프레임 번호와 대표 화면 그리고 관련된 특성 데이터 들은 메타 정보로서 내용 기반 검색에 이용되고, M-JPEG 형태의 디지털 비디오는 비디오 서버 저장 형태인 MPEG으로 변환하기 위하여 이날로그로 복원한 후 다시 MPEG으로 압축한다. 이때 MPEG으로 압축되는 단위는 검출한 장면들 중 관련된 장면들을 몇개씩 묶은 클립 형태이다.

4. 비디오 스트리밍

비디오 라이브러리 응용의 대부분은 네트워크를 통하여 비디오를 전송받아 재생시키는 형태임은 의심할 바 없다. 이미 주문형 비디오나 대화형 TV 등은 CATV나 ADSL을 이용하여 상용화 된 바 있다. 또한 인터넷을 통해 뉴스나 영화, 광고 등과 같은 디지털 비디오 서비스도 점차 생겨나고 있다. 이와 같이 전화망이나 컴퓨터 망에서 비디오를 전송하는 것은 점차 보편화 되어가고 있으며 기술적으로도 큰 문제가 되지 않는다. 단지 비디오 전송의 문제

점은, 그것이 고대역폭을 요구하고 동시에 계속 고정된 속도를 유지해야 하는 제약을 가지는데 반하여 현재의 통신 환경이 저속의 전화망과 인터넷이므로 자연스런 비디오 재생이 어렵다는데 있다. 따라서 비디오 서비스, 특히, 인터넷 상에서의 비디오 서비스는 가용 대역폭에 따라서 고대역폭 비디오 스트리밍과 저대역폭 비디오 스트리밍으로 구분한다.

해상도가 낮고 프레임율이 15fps 보다 작은 저대역폭 비디오를 만들기 위해서는 MPEG과 같은 표준 압축 방식을 사용하지 않고, 화질은 떨어지나 압축율이 매우 좋은 특정 업체 고유의 압축 방식을 사용한다. 저대역폭 비디오 서비스의 응용 분야가 확장되면서 최근에는 저대역폭 비디오를 위한 표준 압축 방식인 H.263도 많이 사용되고 있다. 해상도와 음질이 좋고 프레임율이 15~30fps인 고대역폭 비디오에서는 MPEG1, MPEG2와 같은 표준 압축 방식을 사용한다.

4.1 인터넷 비디오 스트리밍 기술

IP에 기반한 네트워크인 인터넷이나 인트라넷에서 비디오를 전송하려면 다운로드하거나, 스트리밍(streaming)하는 방법을 사용한다. 스트리밍 방식은 압축 비디오를 전송 받으면서 동시에 재생이 이루어지는 것으로서 재생이 완료되면 전송받은 비디오 데이터도 사라지게 된다.

MPEG과 같은 고대역폭 압축 비디오의 경우, 모델을 사용하는 사용자는 수 분짜리 비디오 다운로드 받기 위해 몇 시간을 기다려야 하는 불편함이 발생한다. 또한 네트워크에서 데이터 패킷을 전송할 때 발생하는 노이즈에 의해 일부 프레임이 손실되면, 그 영향이 전체 영상에 미칠 수 있다. 반면에 스트리밍 방식은 전송과 재생이 거의 동시에 이루어지므로 사용자에게 대화형 느낌을 줄 수 있다.

인터넷 및 인트라넷에서의 스트리밍 방식은 다운로드 방식에 비해 다음과 같은 몇가지 기술적 문제를 야기한다. 첫째, HTTP 프로토콜을 사용하는 웹 서버는 클라이언트와 데이터 전송 중에 상호 작용을 할 수 없다[3]. 일단 웹 서버가 데이터 전송을 시작하면 클라이언트

는 그 데이터를 모두 받을 때까지 기다리거나, 데이터 전송을 일시 중단시킬 수 밖에 없다. 스트리밍 방식에서는 데이터 전송 중에 서버와 클라이언트가 상호 작용을 하며 네트워크의 상황에 따라 스트림 전송 속도를 조절할 수 있어야 한다. 이 기능의 작용 여부가 클라이언트에서의 자연스런 재생에 큰 영향을 미친다. 또한 클라이언트의 사용자는 재생중인 스트림에 대해 중단, 일시 중지 이외에 빨리 감기, 되감기, 스트림의 특정 부분으로의 이동 등과 같은 명령을 통해 마치 스트림을 다운로드 받은 상태에서 사용하는 것과 같은 편리성을 요구한다. 즉, 클라이언트는 전송받고 있는 스트림에 대한 완전한 제어권을 요구한다. 이를 위해서는 스트림 전송만을 위한 별도의 스트리밍 서버 소프트웨어가 필요하다.

둘째, 인터넷에서 IP 프로토콜의 상위 계층 프로토콜로 현재 사용하고 있는 TCP나 UDP 프로토콜은 클라이언트에서의 자연스런 스트림 재생을 보장할 수 없다. 인터넷에서 스트림을 데이터 패킷으로 전송할 때 노이즈에 의해 일부 패킷이 손상되거나 혹은 네트워크 과부하로 인한 전송 지연으로 일부 패킷이 사라지게 되면, 그와 관련된 프레임 또한 손상되거나 잃어버리게 된다. 따라서 이 경우 원하는 수준의 화질을 유지할 수 없을 뿐만 아니라 자연스런 재생도 힘들게 된다. 스트리밍 방식에서는 이렇게 손상되거나 사라지는 패킷 혹은 프레임에 대처함으로써, 화질이 원하는 수준 보다 약간 떨어지더라도 전송받고 있는 비디오를 자연스럽게 재생할 수 있어야 한다. 이를 위해서는 클라이언트에서의 자연스런 재생을 보장하는 새로운 스트림 전송 프로토콜이 개발되어야 한다.

셋째, 클라이언트는 전송되는 압축 스트림을 대역폭에 따라 즉시 처리할 수 있어야 한다. 이를 위해서는 전송되는 압축 스트림을 버퍼링하고 복원시키는 클라이언트 소프트웨어가 있어야 한다. 이 클라이언트 소프트웨어는 기능적으로는 디코더에 실시간 버퍼링, 그리고 스트리밍 서버와 상호 작용하면서 전송중인 스트림을 제어하는 기능이 추가된 형태이다. 일반적으로 클라이언트 소프트웨어는 독립적으로

동작하는 플레이어(player) 혹은 웹 브라우저의 plug-in 형태로 제공된다.

비디오 스트림 전송 모드는 비디오 데이터가 서버에서 사용자로 전송되는 형태에 따라 다음과 같이 구분된다. 유니캐스트(unicast)는 스트림과 사용자가 일대일 대응되어 특정 스트림을 특정 사용자에게 전송하는 방식이고, 브로드캐스트(broadcast)는 스트림과 사용자간에 일대다 대응되어 특정 스트림이 다수의 사용자에게 전송되는 방식이며, 멀티캐스트(multicast)는 브로드캐스트와 같이 스트림과 사용자간에 일대다 대응되어 특정 스트림이 다수의 사용자에게 전송되나 사용자는 동적으로 스트림을 전송받을지를 결정할 수 있는 방식이다. 따라서 스트리밍 서버는 멀티캐스트 서비스를 요청한 사용자의 리스트를 관리하여 원하는 사용자들에게만 선택적으로 스트림을 전송할 수 있게 해준다. VOD 와 같은 저장된 데이터를 이용한 주문형 서비스는 유니캐스트 모델을 따르며, 생방송이나 NVOD (Near VOD) 서비스와 같이 동시에 다수의 사용자에게 제공되는 서비스는 멀티캐스트 모델을 따른다[10].

4.2 비디오 전송 프로토콜

일반적으로 IP에 기반한 네트워크에서는 상위 계층 프로토콜로 TCP를 사용한다. TCP는 데이터 패킷의 전송을 보장하는 프로토콜이나 이 프로토콜의 패킷마다 응답 신호를 전송하는 메커니즘은 현재와 같은 과부하 상태의 인터넷이나 인트라넷에서는 빈번히 패킷을 재전송하게 되어 영상이나 음성 스트림을 전송할 경우 심각한 전송 지연을 유발한다. 클라이언트의 응답 신호를 요구하지 않는 UDP는 전송 과정에서 실종된 패킷은 무시하고, 전송 완료된 패킷만으로 데이터 스트림을 구성한다. 영상 스트림 전송의 경우 UDP를 많이 사용하는데, 그 이유는 UDP가 TCP 보다 지속적이고 빠른 데이터 전송을 보장할 수 있기 때문이다.

VDP(Video Datagram Protocol)은 잡음이 많고 전송 지연이 심한 IP 네트워크에서 비디오와 오디오 스트림을 빠르게 전송하기 위해 개발된 프로토콜로 비디오 브로드캐스팅을 목표로 하였다. VDP에서는 서버에서 가변 흐름

제어 기법을 사용하여 클라이언트의 부하에 따라 패킷 전송율을 서버가 조절한다. 이들 프로토콜은 대역폭과 등시성을 보장하지 못하는 문제점을 안고 있는데 이를 해결하기 위한 프로토콜로는 RSVP, RTP/RTCP, RTSP, IP Multicast 등이 있다[17].

5. KT 비디오 라이브러리

현재 KT가 개발중인 비디오 라이브러리는 웹 브라우저를 통하여 서비스되도록 설계하였다. 통신망 환경을 고려할 때, 당분간 인터넷에서의 고대역폭 비디오의 전송이 현실적으로 어려우므로 저화질 비디오와 고화질 비디오 서비스를 병행하도록 하였다. LAN을 이용한 인트라넷 사용자나 초고속 망에 연결된 인터넷 사용자는 고대역폭 MPEG 데이터를 스트리밍하여 볼 수 있으며, 전화망과 같은 저속망에 연결된 인터넷 사용자는 비디오 데이터를 다운로드 받거나 저대역폭 비디오를 스트리밍하여 볼 수 있도록 설계하였다.

그림 2는 현재 개발 중인 KT 비디오 라이브러리의 구조로서, 비디오 서버를 중심으로 다양한 도구들이 결합된 형태이다. 3장에서 이미 설명한 비디오 클립 생성기는 원시 입력 비디오를 장면 단위로 분할하여 압축 비디오 클립을 만들고, 이들에 대한 대표화면 및 카메라 동작과 같은 비디오 분석 정보들을 생성하는 도구이다. Video Annotator는 비디오 클립 생성기에서 검출한 각 샷과 장면들의 제목, 주제어, 내용 등 샷과 장면에 대한 메타 정보를 편

하게 입력시킬 수 있도록 하는 도구로, 이들 정보는 비디오 서버에 저장되어 비디오 검색에 사용된다. 이때 비디오의 모든 대사가 텍스트 형태로 주어진다면 그 내용으로도 검색이 가능하다. 비디오 클립 생성기와 Video Annotator는 비디오 서버의 입력 도구로서 비디오 서버 관리자가 사용하는 도구이다.

인터넷을 통해 비디오 서버를 액세스하는 일반 사용자는 다음의 도구들을 사용할 수 있다. Video Browser는 비디오 서버를 액세스하는 기본 도구로서, 각 클립의 장면들과 각 장면을 구성하는 샷들을 그들의 대표화면인 Kframe과 Rframe을 통해 임의 접근하여 재생할 수 있도록 한다. Story Board Viewer는 비디오 샷과 장면들의 내용 연결 관계를 그들의 대표화면을 이용한 그래프를 통해 보여줌으로써 전체 비디오의 내용을 한눈에 이해할 수 있게 하는 도구이다. Query Tool은 내용 기반 검색을 위한 도구이다. 일반 사용자는 Query Tool에서 텍스트 형태의 메타 정보뿐만 아니라 대표화면의 컬러, texture, shape 등을 입력으로 하여 원하는 비디오 장면을 찾을 때까지 반복적으로 검색을 할 수 있다. 비선형 비디오 편집 시스템인 MPEG editor는 MPEG 형태의 비디오 시퀀스에서 cut&paste 작업과 특수효과 편집 기능 등을 가진다. 이 편집기는 소프트웨어 도구로서 비전문가도 손쉽게 비디오 편집을 할 수 있도록 한다. 이 편집기를 통해 사용자들은 비디오 검색 뿐만 아니라 검색한 비디오 클립을 조합하여 자신만의 비디오를 제작할 수도 있다.

비디오 클립 생성기를 제외한 모든 도구들은 소프트웨어 도구들이다. 이들 도구들은 일반 사용자들의 다양한 하드웨어 및 소프트웨어 플랫폼에서 사용될 것이다. 이를 지원해 주기 위해 비디오 클립 생성기를 제외한 모든 도구들은 JMF(Java Media Framework)를 이용해 구현하고 있다.

6. 맺음말

본고에서는 비디오 라이브러리 요소 기술 전반에 걸쳐 살펴보고 아울러 우리가 개발 중인

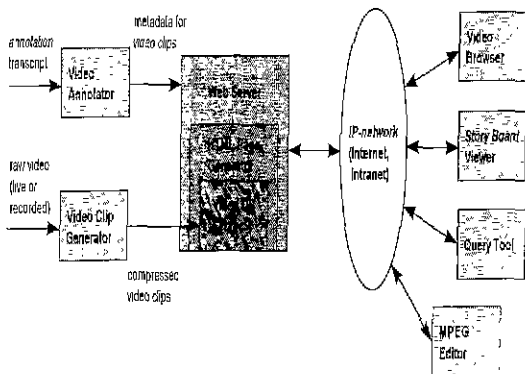


그림 2 KT 비디오 라이브러리

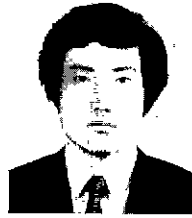
클립 생성 시스템과 KT 비디오 라이브러리에 관하여 설명하였다. 일반 비디오가 컴퓨터가 처리할 수 있는 정보로 만들어져 통신망을 통하여 원격 서비스되기 위하여는 샷 경계 검출 및 장면 분할, annotation, 비디오 스트리밍, 비디오 브라우징, 내용 기반 검색 등에 관련된 여러 기술들이 필요하다. 이들 기술들의 일부는 아직 실용화 단계에 이르지 못하여 계속 연구가 진행 중이다. 한국통신에서는 클립 제작, 내용 기반 검색, MPEG 편집 분야의 기술을 개발함과 아울러 관련 기술을 결합한 비디오 라이브러리 시스템을 개발하고 있다. 이 시스템은 인터넷 영상 서비스를 빠르고 쉽게 개발할 수 있는 기초 기술과 관련 도구를 제공하는 프레임워크로서, 한국통신의 인터넷 비디오 서비스 개발 부분에 사용될 것이다. 한국통신은 비디오 라이브러리를 이용하여 인트라넷을 통한 다양한 형태의 사내 비디오 서비스를 계획 중이며, 인터넷을 통해서도 다양한 비디오 서비스를 제공하게 될 것이다.

참고문헌

- [1] F. Arman, R. Depommier, A. Hsu, M.-Y. Chiu, "Content-based Browsing of Video Sequences", Proc. of *ACM Multimedia*, pp. 97-103, 1994.
- [2] F. Arman, A. Hsu, M. Chiu, "Feature Management for Large Video Databases", Proc. of *Storage and Retrieval for Image and Video Databases*, SPIE Vol.1908, pp. 2-12, 1993.
- [3] A. Covell, "Streaming audio and video on the internet", *Network Computing*, No. 712, Aug. 1996.
- [4] A. Hampapur, R. Jain, T. Weymouth, "Digital Video Segmentation", Proc. Of *ACM Multimedia*, pp. 357-564, 1994.
- [5] H. Kim, T. Kim, W. Kim, B.-D. Rhee, S. Song, "Fast algorithm for detection of camera motion", Submitted to *Real-time Imaging III*, SPIE Electronic Imaging 98, 1998.
- [6] W. Kim, S. Song, H. Kim, B.-D. Rhee, "Morphological scene change detection algorithms from the perspective of unified function theoretic approach", Submitted to *Storage and Retrieval for Image and Video Database VI*, SPIE Electronic Imaging 98, 1998.
- [7] A. C. Lewis, "QoS : Creating inequality in an equal world", *Network Computing*, No. 809, May 1997.
- [8] J. Meng, Y. Juan, S. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence", Proc. of *Digital Video Compression : Algorithms and Technologies*, SPIE Vol. 2419, pp. 14-25, 1995.
- [9] M. Mills, J. Cohen, Y. Wong, "Magnifier Tool for Video Data", Proc. of *ACM CHI*, 1992.
- [10] H. Retting, "Ms. Multicast : Precept CEO Judy Estrin spotlights VAR opportunity in network video", *VAR Business*, No.1302, Feb. 1997.
- [11] S. Song, W. Kim, H. Kim, B. Rhee, "On Detection of Gradual Scene Changes for Parsing of Video Data", Submitted to *Storage and Retrieval for Image and Video Database VI*, SPIE Electronic Imaging 98, 1998.
- [12] T. Y. Tonomura, S. Abe, "Content-oriented visual interface using video icons for visual database systems". *Journal of Visual Languages and Computing*, Vol.1, pp. 183-198, 1990.
- [13] B. Yeo, B. Liu, "Rapid Scene Analysis on Compressed Videos", *IEEE TCSVT-II*, Vol.5, No.6, pp. 533-544, 1995.
- [14] R. Zabih, "Feature-Based Algorithm for Detecting and Classifying Scene Breaks", Proc. of *ACM Multimedia*, pp. 189-200, 1995.
- [15] H. Zhang, K. Kankanhalli, S. Smoliar, "Automatic Partitioning of Full-mo-

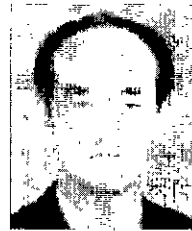
tion Video”, *Multimedia Systems*, Vol. 1, No. 1, pp. 10-28, 1993.

- [16] H. Zhang, C. Low, Y. Gong, S. Smoliar, “Video Parsing Using Compressed Data”, *Proc. of Image and Video Parsing II*, SPIE Vol. 2182, pp. 142-149, 1994.
- [17] 김혁만, 이병도, “스트리밍 기술과 인터넷 영상 서비스”, Technical Report, 한국통신, 1997.



이 병 도

1982 서울대학교 컴퓨터공학과 학사
 1984 서울대학교 컴퓨터공학과 석사
 1984~현재 한국통신 멀티미디어연구소 연구원
 1996 서울대학교 컴퓨터공학과 박사
 연구분야: 디지털 비디오 응용 분야, 디지털 라이브러리, 실시간 시스템



김 혁 만

1985 서울대학교 컴퓨터공학과 학사
 1987 서울대학교 컴퓨터공학과 석사
 1987~현재 한국통신 멀티미디어연구소 연구원
 1996 서울대학교 컴퓨터공학과 박사
 연구분야: 디지털 비디오 응용 분야, 멀티미디어 통신, 데이터베이스

● '97 하계 컴퓨터시스템 학술발표회 ●

- 일 자 : 1997년 9월 26일(금)
- 장 소 : 한국과학기술원(대전)
- 주 최 : 컴퓨터시스템연구회
- 문 의 처 : 중앙대학교 컴퓨터공학과 김성조 교수
 Tel. 02-820-5307