

인터랙티브 멀티미디어 플레이어 위한 편집기의 설계 및 구현

고재필, 안준한, 최익원, 변혜란
연세대학교 컴퓨터과학과

Design and Implementation of Authoring Tool for Interactive Multimedia Player

Jaepil Ko, Junhan Ahn, Ickwon Choi, Hyeran Byun
Dept. of Computer Science, Yonsei University

요 약

최근 인터넷 방송 및 디지털 방송에 관한 관심이 고조되고 있다. 이들 방송이 기존의 방송과 크게 다른 점은 정보전달의 양방향성에 있다. 즉, 시청자는 단방향의 정보만을 받아들이는 것이 아니라, 방송 프로그램과의 인터랙션을 통하여 시청자가 원하는 정보를 얻을 수 있다는 점이다. 이를 위해서는 방송 프로그램은 기존의 동영상 뿐 아니라 부가정보에도 충실한 콘텐츠를 생성해야 한다. 이때 영상물의 각 장면과 동기화 된 데이터 정보는 사용자의 편리성을 증대할 것이다. 또한 장면내의 객체에 대한 정보를 부가정보에 포함한다면, 영상에 대한 직접적인 인터랙션이 가능해 질 것이다. 이를 위해 본 논문에서는 영상, 비디오 처리 기법을 적용하여, 영상전체에 대한 데이터 정보가 아닌 프레임단위 또는 장면단위로 동기화 된 데이터 정보를 용이하게 생성하고, 재생할 수 있는 편집기 및 미디어 플레이어를 설계하고 이에 대한 간단한 프로토타입을 구현하였다.

1. 서론

최근 ADSL과 같은 초고속 통신망 인프라가 확산됨에 따라 인터넷상의 동영상 콘텐츠가 급속도로 늘고 있다. 특히, 스트리밍 기술에 힘입어 동영상 콘텐츠를 전문적으로 제공하는 인터넷 방송국 역시 기하급수적으로 늘고 있다. 또한, 인터넷 방송과 더불어 기존의 아날로그 방송을 대체할 디지털 방송이 국내에서도 본 방송을 앞두고 엄청난 관심을 끌고 있다[1].

디지털 방송 및 인터넷 방송이 기존의 방송과 크게 다른 점은 정보의 흐름이 양방향성이라는 점이다. 기존의 공중파 방송은 방송편성에 따른 단방향 서비스로 시청자의 요구를 즉각적으로 반영하지 못하고 있다. 그러나, 인터넷 방송이나 디지털 방송은 인터넷 채널을 이용하여 시청자와의 상호작용이 가능하다. 예를 들어, 퀴즈쇼 프로그램에 간접적으로 참여하여 자신의 점수를 확인해 볼 수도 있고[그림1], 다양한 형태의 날씨 정보를 받아 볼 수도 있다. 보다 상업적으로는 드라마나 쇼프로그램에 나오는 연예인의 정보나, 프로그램에 연계된 상품 정보 등을 제공하고 이를 쇼핑물과 연계할 수도 있을 것이다.

이러한 방송에서 사용자와의 인터랙션이 가능한 이유는 영상뿐 아니라 데이터 정보를 포함하고 있기 때문이다. 데이터 정보를 통해 사용자가 원하는 특정 주식정보, 뉴스정보, 날씨 정보 등을 제공할 수 있으며, 또한 홈쇼핑, 여론조사, 전자 메일 등이 가능해진다[2]. 향후 방송에 있어서 데이터 정보에 대한 요구는 더욱 증가할 것이다.

그러나, 현재 시범 서비스되고 있는 방송은 동영상 전체 또는 3-5분 단위로 갱신되는 데이터 정보를 기반으로 하고 있어, 동영상내의 특정 장면과는 개연성이 떨어진다. 예를 들어, 쇼프로에서 현재 노래를 부르고 있는 가수의 정보를 알고 싶은 사용자는 쇼프로 전체에 대해 제공되고 있는 정보로부터 해당 가수를 검색해야 하는 불편함이 예상된다. 또 다른 예로서 드라마에 나오는 어떤 상품이나 카페와 같은 특정 장소에 대한 정보를 알고자 할 때 검색을 통해야 한다면 사용자의 호기심이나 구매력은 떨어질 것이다.

본 논문에서는 MS윈도우즈 스트리밍 기술에 기반하여 영상전체에 대한 데이터 정보가 아닌, 프레임 단위 또는 장면 단위로 동기화 된 데이터정보를 용이하게 생성하고, 재생할 수 있는 편집기 및 미디어플레이어를 설계하고 이에 대한 간단한 프로토타입을 구현하였다.



그림 1 인터랙티브 프로그램 예

2. MS윈도우즈 스트리밍 솔루션

MS윈도우즈에서 제공하는 DirectShow는 멀티미디어 스트리밍 기술로 다음과 같은 구조를 가진다(그림2).

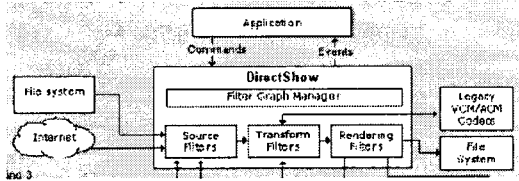


그림 2 DirectShow 구조
출처 : Microsoft. MSDN

DirectShow는 크게 4부분으로 소스필터, 전송필터, 렌더링필터, 그리고 필터그래프관리자로 구성된다. 소스필터는 각종 입력장치로부터 로우 데이터를 읽어온 후 전송필터로 데이터를 전달한다. 전송필터는 로우 또는 부분적으로 처리된 데이터를 입력으로 하여 다양한 처리를 한다. 과상, 압축해독, 포맷 변환 등의 기능이 이에 해당한다. 렌더링필터는 완전히 처리된 데이터를 입력으로 하여 다양한 출력을 담당한다. 응용프로그램 개발자는 각 필터를 목적에 맞게 설계한 후 필터그래프관리자를 통하여 추가, 재배열등을 통하여 응용 프로그램을 개발할 수 있다.

3. 인터랙티브 동영상 서비스 개념도

사용자는 그림3과 같이 동영상 및 부가정보를 웹으로부터 전송 받아 영상 및 이와 동기화 된 부가 정보를 재생할 수 있다. 또한 부가 정보에 포함된 물체의 위치정보를 통해 물체를 표현하고, 사용자의 클릭에 의해 관련 사이트로 이동하여 보다 상세한 정보를 얻을 수 있다.

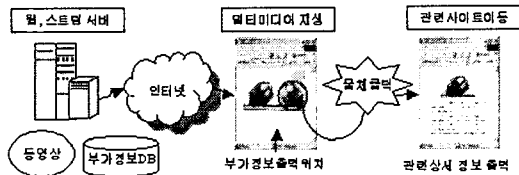


그림 3 인터랙티브 서비스 개념도

4. 멀티미디어 부가정보 생성 편집기 및 플레이어

동영상 스트림의 서비스 질을 높이기 위하여 기존의 동영상을 위한 부가 정보를 생성하여야 한다. 이를 위해서 다음과 같은 미디어 편집기를 설계하였다.

편집기는 크게 5부분으로 나뉜다. 첫째, 연속적인 스트림으로부터 처리 가능한 이산적인 프레임 정보를 추출하는 프레임 추출단계, 둘째, 장면단위의 부가 정보 편집 및 객체 추출을 위한 장면 추출 단계, 셋째, 장면단위의 객체 트래킹 단계, 마지막으로 사용자가 부가정보를 입력할 수 있는 편집 단계로 나뉜다. 사용자는 각 단계를 거치면서 스트림의 시간정보와 동기화 된 프레임정보, 장면정보, 객체정보, 문

자 정보 등의 부가정보를 생성한다.

그림4는 미디어정보 편집기의 구성을 보여준다.

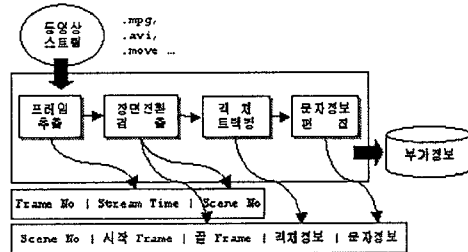


그림 4 미디어 정보 편집기 구성

4.1 프레임 추출

스트림은 연속적인 시간축으로 표현되는 데이터이므로 처리 가능한 이산적인 프레임으로 구분하여야 한다. 프레임 추출 단계에서는 프레임번호 및 이와 동기화 된 스트림 시간을 구한다. 이렇게 프레임을 추출해 냄으로서 기존의 영상처리 기법을 스트림 영상에 그대로 적용할 수 있다.

4.2 장면전환 검출

프레임단위의 객체정보 추출은 무의미하며 사용자의 작업량만 증가시킨다. 따라서, 장면단위로 구분하는 것이 필요하다. 비디오에서 장면단위는 하나의 카메라에 의해 기록된 연속적인 일련의 프레임을 말한다[4].

같은 장면에서는 객체의 움직임이 크지 않기 때문에 다음 단계의 객체 추출도 용이하게 이루어 질 수 있다.

장면전환 검출 기법은 화소기반, 통계기반, 변환기반, 특징기반, 히스토그램기반 등의 범주로 나눌 수 있다[5].

본 논문에서는 프레임 전체의 특징을 갖는 히스토그램 차이방법(HDM) 및 지역적 특성을 잘 반영하는 화소차이법(PDM)을 적용하여 장면을 검출한다.

프레임 크기가 $M \times N$ 인 k 번째 쌍의 $PD(f_k, f_{k+1})$, $HD(f_k, f_{k+1})$ 를 각각 화소 차이와 히스토그램 차이 수치로 표시하면 식(1)과 같다.

$$d_{i,j}(f_k, f_{k+1}) = \begin{cases} 1 & \text{if } |I_{i,j}(f_k) - I_{i,j}(f_{k+1})| > 0 \\ 0 & \text{Otherwise} \end{cases}$$

$$PD(f_k, f_{k+1}) = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N d_{i,j}(f_k, f_{k+1}) \quad (1)$$

PDM에서 식(1)과 같이 해당 화소가 다를 때만 1로 선택한 이유는 화소의 작은 변화에 민감하지 않게 하기 위함이며, 화소 차이의 잡음을 피하기 위해 각각의 채널은 32개의 빈(bin)으로 미리 균일하게 양자화 한다. HDM은 식(2)로 구한다.

$$HD_{i,j}(f_k, f_{k+1}) = \frac{1}{M \times N} \sum_{j=1}^{256} |H_i(j) - H_{i+1}(j)| \quad (2)$$

위 식을 RGB색상 모델에 대해 적용하고, 두 식의 결과 값

을 중첩시킨 결과를 적절한 임계값으로 장면전환 여부를 결정한다. 이때 오 검출을 줄이기 위해 식 (1),(2)의 결과에 미디안 필터를 적용한다. 그림5는 1136장의 프레임축에 대한 PD, HD값을 보여준다. 그림에서 (a)는 수식의 결과이고 (b)는 (a)의 미디안 필터링 결과이다. 그림에서 보듯이 더 나은 결과를 얻을 수 있다.

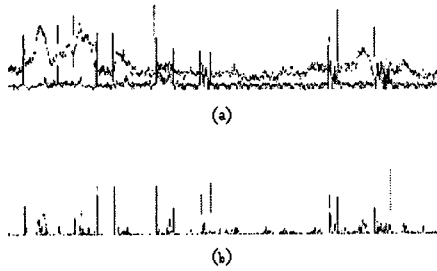


그림 5 (a)원PD,HD 신호, (b)미디안 필터링 결과신호

장면전환 검출을 통해 유사한 프레임들의 시작과 끝을 장면번호와 함께 부가정보 DB에 저장한다. 이는 트랙킹 단계에서, 추적하고자 하는 객체가 포함된 프레임의 범위를 지정하게 된다.

4.3 객체트랙킹[6]

사용자가 관심 있을 것으로 예상되는 프레임내의 객체에 대한 정보를 생성하여 덤으로서 사용자의 클릭이벤트를 적절히 처리할 수 있다. 즉 객체와 이와 관련된 링크정보를 통하여 사용자는 관련 홈페이지에 자동으로 접속하여 상세한 정보를 얻을 수 있다. 이를 위해서는 프레임 단위로 객체의 위치정보를 유지하여야 한다. 그러나, 매 프레임마다 편집자가 이를 입력해 줄 수는 없다. 이 문제를 해결하기 위해 객체 트랙킹 기법을 적용하였다.

장면의 대표 프레임을 장면의 첫 번째 프레임으로 가정하고, 이 프레임에 대해서만 편집자는 트랙킹 하고자 하는 객체의 위치를 지정해 준다. 그러면 객체 트랙킹을 통해서 장면안에서의 객체 위치를 자동으로 생성할 수 있다.

트랙킹을 위한 방법론은 크게 두 가지로 나누어진다. 첫 번째는 트랙킹 하고자 하는 대상을 미리 선정하여 대상에 대한 모델을 만들고 전체영상에서 특정모델의 위치를 계속 추적하는 모델기반 방식[7][8]과 특정 모델이 아닌 움직임을 추적하는 움직임 기반 방식이 그것이다[9]. 본 부가정보 미디어 플레이어 위한 방법은 트랙킹 대상이 고정적인 것이 아니라 비디오마다 대상이 변하기 때문에 움직임 기반 기법을 사용하는 것이 타당하다.

움직임기반 기법에 가장 많이 사용되는 블록정합 방법 중 3단계탐색 알고리즘을 사용하여 구현하였다. 3단계 탐색 알고리즘은 검색 점의 개수를 대수적으로 줄이는 방법이다. 첫 번째 단계에서 자신의 위치와 검색 영역크기의 절반 지점인 8개의 지점을 선택하여 비교한다. 두 번째 단계에서는

첫 번째 단계에서 비교한 9개의 지점 중 가장 유사한 지점에서 다시 검색 영역 크기의 절반 지점인 8개의 지점을 비교한다. 예를 들어 검색 영역의 크기가 7개의 화소이면 비교하여야 할 점의 개수는 25 (9+8+8)가 된다.[그림6]참조.

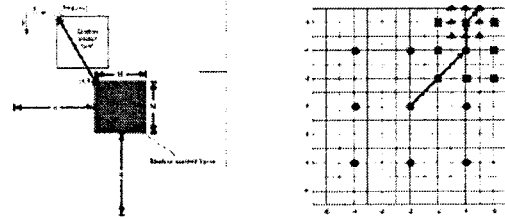


그림 6 블록매칭의 3단계 탐색 알고리즘

T프레임에서 편집자가 선택한 영역(안쪽 사각형)은 T+1 프레임에서 탐색 윈도우영역(바깥쪽 사각형)의 여러 점들과 비교하여 가장 유사한 영역이 선택된다[그림7].

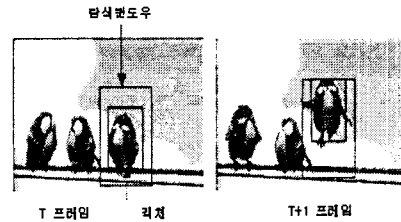


그림 7 객체 트랙킹 예

4.4 문자정보 편집

문자 편집기를 통하여 동영상 및 이와 연관된 문자정보를 저장할 수 있다. 문자정보는 프레임단위가 아닌 장면단위로 편집한다. 프레임단위로 편집하여 1/30초 단위의 문자정보 갱신은 사용자에게 의미가 없기 때문이다.

4.4 인터랙티브 미디어 플레이어

미디어 플레이어에서는 편집기에서 생성된 부가정보를 이용하여 동영상 및 관련 정보를 출력한다. 동영상 스트림이 재생되는 동안 프레임번호와 장면번호에 대한 정보를 가져온다. 이때 장면이 바뀔 때마다 이와 관련된 문자정보를 갱신한다. 문자정보는 HTML 브라우저에서 출력되어 링크를 포함할 수 있다. 따라서, 사용자는 장면단위의 문자정보로부터 관련된 사이트로 이동이 가능하다.

만약, 사용자가 영상위로 마우스를 옮겨가면 다음 프레임부터는 프레임 단위로 저장된 객체의 위치정보를 가지고 와서 화면에 객체의 위치를 출력하여 준다. 이때 사용자가 해당 객체를 클릭하면 관련 사이트로 이동할 수 있다.

5. 결론 및 향후연구계획

본 논문에서는 동영상 정보에 데이터 정보를 부가하여 사

용자의 인터랙션 서비스를 강화할 수 있는 멀티미디어 플레이어
이어를 위해 영상 및 비디오 처리기법을 사용한 미디어 편집
집기를 설계하고 간단한 프로토타입 시스템을 구현하였다.

본 시스템에서는 객체정보를 통해 영상내의 물체에 대한
사용자 클릭 이벤트를 처리하여 사용자의 편리성을 증진할
수 있다. 인터넷 방송이나 디지털 방송에서 교육용 또는 상
업용으로 그 활용가치가 매우 클 것으로 기대된다.

그러나, 상용화를 위해서는 앞으로 많은 보완이 이루어져
야 한다. 먼저, MS 윈도우즈에서 제공하는 스트리밍 기술
에 의존하지 않고 독자적인 스트리밍 서버 구현 및 클라이
언트 프로그램을 개발해야 한다. 이렇게 함으로서 동영상
및 부가 정보 모두 스트리밍 서비스가 가능해지기 때문이
다.

참고문헌

- [1] 대한전자공학회, 인터넷 방송 워크샵, 2000년 7월
- [2] 박종석, 김민수, "디지털TV", 정보과학회지, 제18권 제9
호, 통권 136호 2000년 9월, pp.13-20
- [3] 최익원, 변혜란, "비디오 표현과 추상화를 위한 클러스터
링 비디오 계층 구조 구축", 정보과학회논문지 제 27권
제5호, 2000년 5월
- [4] Yong Rui, Thomas S. Huang, and Sharad Mehrotra,
"Constructing Table-of-Content for Videos," to appear
in *ACM Multimedia Systems Journal*, Special Issue
Multimedia Systems on Video Libraries, Sept, 1999.
- [5] J.S. Boreczky and L.A. Rowe. "Comparison of Video
Shot Boundary Detection Techniques" Proc. of SPIE
Conference on Storage and Retrieval for Image and
Video Database IV Vol. SPIE 2670, pp.170-179, 1996
- [6] Tekalp, Digital Video Processing, Prentice Hall, 1995
- [7] J.R. Bergen, P.Anadan, K.J.Hanna, and R.
Hingorani.Hierachical model-based motion estimation.
In Proceedings of European Conference on Computer
Visin 1992, pages 239-252,1992
- [8] Sumit Basu, Irfan Essa, Alex Pentland. Motion
Regularizatin for Model-based Head Tracking. 13th
Int'l. Conference on Pattern Recognition(ICPR'96),
Vienna, Austria, August 25-30, 1996.
- [9] R. LI, B.Zeng, and M. L. Liou, "A new three-step
search algorithm for block motion estimatioin," IEEE
trans. Circ, Syst, and Video Technol., vol. 4, p.
438-442, Aug. 1994