

동영상 콘텐츠의 메타데이터에 기반한 효율적인 브라우징 기법

(Efficient Browsing Method based on Metadata of Video Contents)

천 수 덕[†] 신 정 훈^{**} 이 상 준^{***}
(Sooduck Chun) (Junghoon Shin) (Sangjun Lee)

요 약 정보기술은 통신 및 멀티미디어 기술의 발전에 힘입어 빠르게 발전하고 있으며, 이에 따른 디지털 콘텐츠에 대한 수요가 증가하였다. 디지털 콘텐츠 중 비디오 콘텐츠는 VOD, NOD, 디지털 도서관, IPTV, UCC 등 다양한 응용 분야에서 점점 확산되고 있다. 비디오 데이터는 순차적인 특성이 있으며, 시간과 공간 정보가 결합한 3차원 데이터로서 계산 비용이 많이 드는 작업이기 때문에 검색이나 브라우징이 어렵다.

본 논문에서는 비디오 데이터를 구조화하는 동영상 내용 편집 도구(ATVC: Authoring Tool for Video Contents)를 제안한다. 동영상 내용 편집 도구는 비주얼 리듬을 이용하여 비디오 데이터에서 대표 프레임을 추출한 다음 XML을 이용한 태그 및 키워드 정보를 대표 프레임에 삽입하는 도구(Tool)이다. 비주얼 리듬은 3차원의 시공간적인 정보를 2차원으로 매핑한 정보로 IDCT과정 없이 픽셀 정보를 얻을 수 있어 처리속도가 빠르며 컷, 와이프, 디졸브 등의 편집 효과를 효과적으로 구분할 수 있다. 그리고 XML 데이터에는 태그 및 키워드 정보와 함께 대표 프레임의 정보까지 저장되므로 효율적인 브라우징을 제공할 수 있다.

키워드 : 메타데이터, 비디오, 대표프레임, 브라우징

Abstract The advancement of information technology along with the proliferation of communication and multimedia has increased the demand of digital contents. Video data of digital contents such as VOD, NOD, Digital Library, IPTV, and UCC are getting more permeated in various application fields. Video data have sequential characteristic besides providing the spatial and temporal information in its 3D format, making searching or browsing ineffective due to long turnaround time.

In this paper, we suggest ATVC(Authoring Tool for Video Contents) for solving this issue. ATVC is a video editing tool that detects key frames using visual rhythm and insert metadata such as keywords into key frames via XML tagging. Visual rhythm is applied to map 3D spatial and temporal information to 2D information. Its processing speed is fast because it can get pixel information without IDCT, and it can classify edit-effects such as cut, wipe, and dissolve. Since XML data save key frame information via XML tag and keyword information, it can furnish efficient browsing.

Key words : metadata, video, key frame, browsing

· 이 논문은 2008년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(KRF-2008-331-D00485)

† 비 회 원 : 알티베이스 연구원

caddie04@ssu.ac.kr

** 학생회원 : 숭실대학교 컴퓨터학부

coolhoony@ssu.ac.kr

*** 정 회 원 : 숭실대학교 컴퓨터학부 교수

sangjun@ssu.ac.kr

논문접수 : 2009년 12월 8일

심사완료 : 2010년 3월 2일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다. 정보과학회논문지: 컴퓨팅의 실제 및 테러 제16권 제5호(2010.5)

1. 서 론

정보기술의 발전은 통신 및 멀티미디어 기술의 발전에 힘입어 빠르게 발전하고 있다. 이에 따른 데이터베이스의 기술이 공간데이터, 텍스트, 비디오, 음성, XML 등과 같은 다양한 멀티미디어 데이터분야에 적용되고 있으며, 이에 따른 콘텐츠에 대한 수요가 증가하였다. 문화 콘텐츠는 '문화적 요소'가 체화되어 경제적 가치를 창출하는 문화상품(Cultural Commodity)을 의미하며 그 내용에 따라 '음악', '공연', '방송 영상', '게임', '애니메이션', '만화' 등이 될 수 있다. 이러한 문화 콘텐츠를 디지털화시킨 것이 디지털 콘텐츠이다. 특히 디지털 콘

텐츠 중 비디오 콘텐츠는 IPTV, UCC, 디지털도서관 등의 다양한 응용 분야에서 점점 확산하고 있다.

비디오 콘텐츠(동영상)의 크기는 수 메가바이트에서 수 기가바이트에 이르는 대용량이기 때문에 컴퓨터상에서의 많은 처리량과 시간을 필요로 한다. 그리고 시간과 공간 정보가 결합한 이동 객체(Moving Object)가 존재하며, 시간에 의해 표현되는 연속매체이다. 이러한 특징들 때문에 비디오 내의 원하는 부분을 찾는 내용 기반 검색이나 브라우징이 대단히 어렵다. 따라서 컴퓨터 처리량의 감소와 사용자가 원하는 정보를 쉽게 찾도록 도와줄 수 있는 효율적인 비디오 검색과 브라우징 기법이 요구되고 있다.

본 논문은 동영상 내용 편집 도구[1]의 연구 및 디지털 콘텐츠의 브라우징 기법[2]을 확장하여 이러한 구조화된 데이터를 이용하여 원하는 부분을 쉽게 볼 수 있을 뿐만 아니라, 검색을 위한 주석의 내용까지 보여줌으로써 사용자에게 효율적인 계층적 브라우징 기법을 제공하고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 비디오 콘텐츠에 대한 이전 연구들에 대해 살펴본다. 3장에서는 기존에 행하였던 동영상 내용 편집도구에서 효율적인 브라우징을 위해 Adobe Flex[3]를 이용하여 브라우징을 구현한 부분을 설명한다. 마지막으로 4장에서는 결론 및 앞으로의 연구 방향을 서술한다.

2. 관련 연구

2.1 비디오 데이터의 구조

비디오 데이터를 내용 기반 검색이나 브라우징을 위해서는 비디오 데이터를 샷으로 분할해야 한다. 그리고 각 샷에 대해 대표 프레임을 선택하는 것이 필요하며, 그림 1과 같은 구조로 되어 있다[4].

샷(Shot)은 한 카메라의 연속적인 동작에 의하여 생성된 프레임의 집합을 말한다. 일반적으로 샷의 경계 부근에서 커다란 변화를 보여, 자동검출이 가능한 부분이며 시간적 특징을 가진다. 샷의 검출 기법의 종류로는 화소 값의 편차를 기반을 둔 접근[5], DCT 계수를 이용한 접근[6], 모션 벡터(motion vectors)를 이용한 접근[7], 에지(Edge) 특징을 이용한 접근[8], 히스토그램의 차이를 이용한 접근[9] 등이 있다.

대표 프레임(Key Frame)은 각 샷에서 가장 잘 표현되는 하나 또는 여러 개의 프레임을 의미하며 공간적 특징을 가지고 있다. 대표 프레임의 검출 기법은 각 샷에서의 처음과 마지막 프레임을 추출하는 방법[10], 샷의 움직임 지표에 의한 방법[11], 시각적 내용 복잡 지표를 이용한 방법[12] 등이 있다.

신(Scene)은 비슷한 배경 또는 줄거리 상의 사건에 따라 하나로 묶일 수 있는 샷의 집합이다[13].

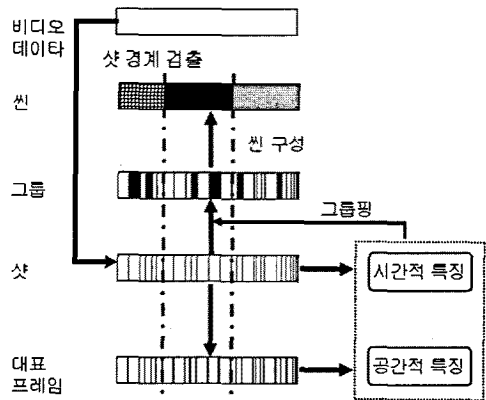


그림 1 비디오 데이터의 계층적 구조

2.2 비디오 브라우징

비디오 브라우징은 적은 수의 대표 프레임을 이용하여 보여줌으로써 비디오 전체 내용을 쉽게 파악할 수 있으며, 사용자가 원하는 비디오 콘텐츠를 찾을 수 있도록 하는 기능이다[14]. 브라우징 기법은 VCR 식 순차탐색법, 대표 프레임 정적표현법, 계층적 브라우징 기법, 신 기반 브라우징 기법이 있다.

VCR 식 순차탐색 기법은 아날로그 비디오 매체에서 주로 사용되던 방식을 디지털 비디오에 적용한 기법으로 사용자가 잘 아는 장점이 있지만 임의접근(Random Access)이 가능한 특성을 활용하지 못한 단점이 있다. 대표 프레임 정적표현법은 각 샷에서 하나의 대표 프레임을 추출한 브라우징 기법이다. 즉, 비디오를 구성하는 프레임을 보이는 방식이다. 전체 대표 프레임을 한눈에 살펴볼 수 있는 장점이 있지만, 각 샷을 가장 잘 표현하는 프레임을 찾아 대표 프레임으로 하는 것이 어렵다[15]. 그리고 영상의 시간이 증가함에 따라 너무 많은 대표 프레임이 존재한다는 단점이 있다. 계층적 브라우징 기법[14]은 대표 프레임 정적표현법을 발전시킨 기법이다. 전체 대표 프레임을 모두 보이지 않고, 그것 중 대표성을 가진 일부만을 표시하여, 현재 나타나 있는 대표 프레임으로부터 보이지 않는 대표 프레임으로 확장하는 브라우징 기법이다. 대표 프레임을 시간적인 순서대로 배열하여 보여줌으로써, 빨리 비디오의 대략적인 내용을 살펴보는데 용이하다. 그러나 샷들의 위치만을 중요시하고 의미를 무시하여, 비효율적인 브라우징을 가져올 수 있는 단점이 있다. 신 기반 브라우징 기법[13]은 비슷한 배경의 샷을 모아 신을 구성하고, 이를 기반으로 원하는 샷을 찾아가는 브라우징 기법으로 신을 구성하는 샷이 다양하게 존재하면 신간의 순서나 비디오 콘텐츠 내에서의 시간적인 위치를 파악하는 데 어려움이 있다.

3. 메타데이터를 이용한 효율적인 브라우징의 구현

3.1 동영상 내용 편집 도구

비디오 브라우징을 하기 위해서는 먼저 컷의 정확한 검출이 중요하며, 각 샷을 가장 잘 표현하는 대표 프레임 추출이 필요하다.

샷의 검출 기법의 종류는 2장에서 설명하였다. 그 중 가장 보편적인 검출 방법은 컬러 히스토그램 비교 방법 [8]과 화소 단위 비교 방법 [4]이다. 컬러 히스토그램 비교 방법은 같은 샷 내의 프레임들은 서로 유사한 색상 분포를 가진다는 특성을 이용한 방법으로, 인접한 프레임들의 히스토그램 차이를 계산하여 주어진 임계값과 비교함으로써 컷을 검출한다. 화소 단위 비교 방법은 같은 샷 내에서는 화소 값의 변화가 적다는 특성을 이용한 방법으로, 연속하는 한 쌍의 프레임에서 대응하는 화소 값을 비교함으로써 컷을 검출한다.

기존의 히스토그램을 사용한 방법은 국소적인 특징을 반영하지 않으며, 화소 차를 이용한 방법은 전체적인 특징을 반영하지 못한다. 또한, 비디오 데이터의 원시 데이터는 대용량이므로 MPEG 등으로 인코딩되어진다. 비디오 데이터를 디코딩한 다음 컷을 검출하는 부분은 상당한 처리시간이 필요하다.

이러한 문제점을 해결하기 위해 국부적이고 전체적인 특성을 같이 보여주는 비주얼 리듬 [16]을 이용하여 대표 프레임을 추출한 다음 XML을 이용하여 메타 데이터를 삽입하는 동영상 내용 편집 도구 (Authoring Tool for Video Contents)를 제안하였다 [1]. 비주얼 리듬은 전체 영상의 내용변화를 표현하는 한 장의 이미지이다. 비주얼 리듬에서 수직으로 한 줄에 속하는 화소들은 원시 프레임에서 추출된 축소화면의 대각선 화소이며, 그림 2와 같다. 특정 샷에 속한 프레임들에서 추출된 대각선 화소들은 거의 비슷한 시각적 특성을 지닌다. 따라서 시각 운동의 샷 경계 부근에서는 두드러진 시각적 변화가 나타난다. 샷은 한 카메라의 연속적인 동작에 의하여 생성된 영상을 의미한다. 일반적으로 샷 경계를 검출하면 각 샷에서 대표 프레임을 추출하고, 추출된 대표 프레임들을 분석하여 비슷한 것들을 묶음으로써 장면을 정의한다. 따라서 샷 경계의 정확한 검출이 장면 경계 검출의 정확도를 결정한다. 비주얼 리듬은 독특한 화소 샘플링에 의하여 3차원 영상을 2차원 이미지로 요약한 것으로서 다양한 영상 편집 효과들에 대하여 시각적 구분이 뚜렷한 패턴을 보여주는 특징을 가지고 있다.

그림 3 및 그림 4는 본 연구의 선행연구의 결과로 비주얼 리듬을 활용하여 대표화면을 검출하여 대표화면에 대한 메타데이터를 편집하는 도구의 실행을 보여주고 있다.

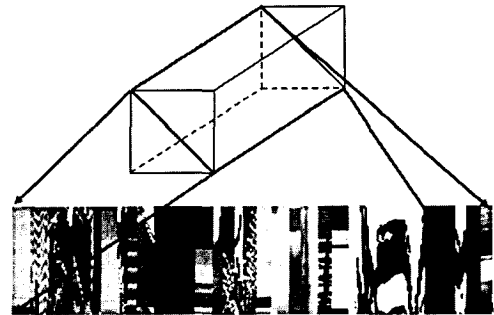


그림 2 비주얼 리듬의 샷 경계 추출



그림 3 동영상 대표화면 검출



그림 4 ATVC를 이용한 메타데이터 태깅

본 논문은 [1,2]의 연구를 확장하여, 이러한 구조화된 데이터를 이용하여 원하는 부분을 쉽게 볼 수 있을 뿐만 아니라, 검색을 위한 주석의 내용까지 보여줌으로써 사용자에게 효율적으로 제공할 수 있는 계층적 브라우징을 구현하였다.

3.2 유사한 대표 프레임의 병합

효율적인 브라우징은 적은 수의 대표 프레임을 이용하여 전체 내용을 쉽게 파악하는 것이 핵심이므로, 유사한 대표 프레임을 식별, 병합하는 것이 중요하다. 대표 프레임은 공간적 특징을 가지고 있으므로 유사한 대표 프레임은 색상, 질감, 텍스트, 공간상의 관계 등을 이용

하여 병합할 수 있다. 컬러를 기반으로 한 유사도 측정은 일반적으로 영상의 컬러 히스토그램에 대하여 유클리디언 거리(Eucliden distance)나 히스토그램 인터섹션(histogram intersection)방법[17]이 사용되고 있다. 본 논문에서는 추출된 대표 프레임의 히스토그램 인터섹션을 사용하여, 유사도 측정을 한 후, 유사도의 편차가 일정 범위 이내인 것을 대표 프레임으로 병합한다.

그림 5는 동영상 내용 편집 도구를 이용하여 대표 프레임을 추출한 부분이다. 대표 프레임 5개 중에 1~3번 대표 프레임은 히스토그램 상 유사한 프레임이다.



그림 5 ATVC에서 추출된 대표 프레임

이러한 유사한 대표 프레임은 그림 6과 같이 상위 레벨과 하위 레벨로 구분하여 1차 레벨의 병합을 한다. 대표 프레임을 순차적으로 비교하여 병합되는 첫 번째 대표 프레임을 상위 레벨의 대표 프레임으로, 나머지 대표 프레임은 하위 프레임으로 정의한다. 본 논문에서는 시간적으로 연속된 대표프레임에 대하여 1차 레벨의 병합을 사용하였다.

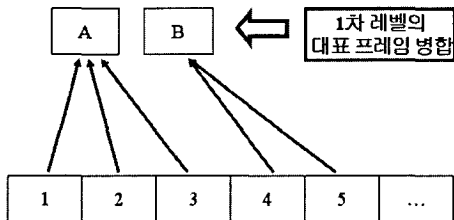


그림 6 1차 레벨의 대표 프레임 병합

3.3 플렉스를 이용한 계층적 브라우징 구현

본 논문에서는 메타데이터가 저장된 XML 파일을 플렉스를 이용하여 계층적인 브라우징을 구현하였다. XML 파일은 문서의 구조에 대한 정보만 저장할 뿐 외양에 관한 정보는 저장하지 않는다. 그래서 XML만으로는 웹 브라우저에서 실행시켜도 구조화된 노트 방식의 텍스트로만 표현된다. 결국, 추가로 웹 브라우저에 출력하는 역할을 하는 XSL과 XSLT기술이 필요하다.

플렉스[2]는 XML 태그로 된 스크립트(*.mxml)를 작성해서 컴파일러로 컴파일하면 플래시로 된 화면을 만들 수 있는 리치 인터넷 애플리케이션(RIA) 개발 솔루션이다. 즉, 웹 애플리케이션에 플래시를 도입시킨 것으로 동적인 사이트에서 한층 더 발전시킬 수 있다. 또한,

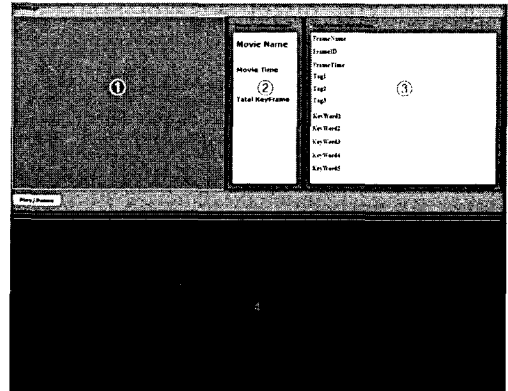


그림 7 계층적 브라우징을 구현한 폼

표 1 계층적인 브라우징 폼의 기능

기능	
①	영상 데이터가 재생되는 레이어이다. NetStream클래스를 이용하였으며, ④에서 임의의 대표 프레임을 클릭하면, 클릭된 대표 프레임부터 재생이 가능하다. 즉, 랜덤 액세스가 가능하다.
②	메타 데이터의 상위 부분의 데이터를 출력한다. 영상의 이름, 영상의 시간과 대표 프레임의 개수를 나타낸다.
③	메타 데이터에서 대표 프레임의 시간 정보와 제작자가 넣고 싶은 광고나 키워드를 저장한 부분을 출력해 준다.
④	병합된 대표 프레임을 보여준다. 시간적인 순서대로 배열하여 보여줌으로써, 빠른 시간 내에 비디오의 대략적인 내용을 살펴보는 데 용이하다.

플래시 플레이어(Flash Player)가 동작한다면 운영체제, 브라우저, 디바이스를 가리지 않고 실행된다. 즉, ActiveX 나 JVM 등에 관계없이 인터넷 익스플로러(IE), 파이어폭스(Firefox), PDA, 휴대전화기, 위성 단말기 등 어디에서든지 인터넷으로 같은 애플리케이션을 공유할 수 있는 환경을 제공할 수 있다. 물론 플래시 플레이어(Flash Player)의 설치를 필요로 하지만 현재 전 세계의 90% 이상의 컴퓨터에 이미 설치가 되어 있다는 Adobe 사의 공식 통계[18]를 고려해 볼 때 사용자 측에 플래시 플레이어에 대한 부담감은 없을 것이다.

그림 7은 플렉스를 이용하여 계층적 브라우징을 구현한 폼(Form)이며, 폼의 각 기능은 표 1에 기술되어 있으며, 그림 8은 유사한 대표 프레임을 병합한 계층적 브라우징의 실행 화면을 보여주고 있다.

4. 결론 및 향후 계획

본 논문에서 비디오 데이터를 구조화하는 동영상 내용 편집 도구(Authoring Tool for Video Contents)의 연구를 확장하여, 구조화된 데이터를 이용하여 원하는 부분을 쉽게 볼 수 있을 뿐만 아니라, 검색을 위한 주석

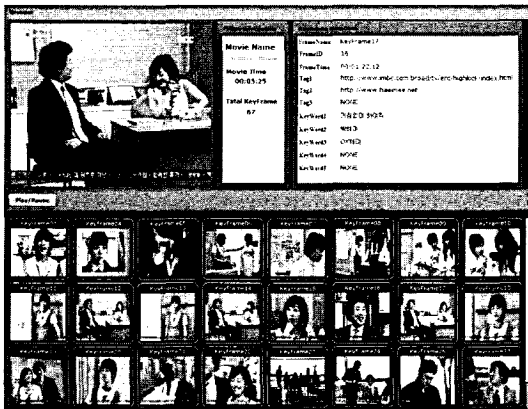


그림 8 유사한 대표 프레임을 병합한 계층적 브라우징의 실행 화면

의 내용까지 보여줌으로써 사용자에게 효율적으로 콘텐츠를 제공할 수 있는 브라우징을 제안하였다. 대표 프레임을 시간적인 순서대로 배열하여 보여줌으로써, 빨리 비디오의 대략적인 내용을 살펴보는 데 용이하다. 또한, 플렉스를 이용하여 브라우징을 구현되어 플래시 플레이어가 동작한다면 운영체제, 브라우저, 디바이스를 가리지 않고 독립적으로 실행되는 장점을 가지고 있다.

참 고 문 헌

[1] S. Chun, S. Joo, S. Lee, "Development of Digital Contents Authoring Tool using Metadata," *Proc. of the 34th KIISE Fall Conference*, vol.34, no.2(C), pp.50-54, 2007. (in Korean)

[2] S. Chun, J. Shin, S. Lee, "Implementation of an Efficient Browsing using Metadata of Digital Contents," *Proc. of the KIISE Korea Computer Congress 2008*, vol.35, no.1(C), pp.7-10, 2008. (in Korean)

[3] L. Sorokin, F. Montero, C. Martin, "Flex RIA Development and Usability Evaluation," *Proc. of WISE 2007 WorkShops*, LNCS 4832, pp.447-452, 2007.

[4] Y. Rui, T. S. Huang, S. Mehrotra, "Constructing table-of-content for videos," *ACM Multimedia Systems Journal, Special Issue Multimedia Systems on Video Libraries*, vol.7, no.5, pp.359-368, 1999.

[5] R. Kasturi and R. Jain, "Dynamic Vision," *Computer Vision*, IEEE CS Press, pp.469-480, 1991.

[6] F. Arman, A. Hsu, M. Y. Chiu, "Feature Management for Large Video Databases," *Proc. of SPIE Storage & Retrieval for Image and Video Databases*, vol.1908, pp.2-12, 1993.

[7] B. L. Yeo, "Efficient Processing of Compressed Images and Video," *Ph.D. dissertation*, Princeton University, 1996.

[8] R. Zabih, J. Miller, K. Mai, "A Feature-based Algorithm for Detecting and Classifying Scene Breaks," *Proc. of ACM Conf. on Multimedia*, pp.189-200, 1995.

[9] G. Ahanger, T. Little, "A Survey of Technologies for Parsing and Indexing Digital Video," *Journal of Visual Communication and Image Representation, Special Issue on Digital Libraries*, vol.7, no.1, pp.28-43, 1996.

[10] H. Zhang, C. Y. Low, S. W. Smoliar and D. Zhong, "Video Parsing, Retrieval and Browsing: An Integrated and Content-based Solution," *Proc. of ACM Conf. on Multimedia*, pp.15-24, 1995.

[11] W. Wolf, "Key Frame Selection by Motion Analysis," *Proc. of Int. Conf. on the Acoustics, Speech, and Signal Processing*, pp.1228-1231, 1996.

[12] P. O. Gresle, T. S. Huang, "Gisting of video documents: A Key Frames Selection Algorithm Using Relative Activity Measure," *Proc. of Int. Conf. on Visual Information System*, pp.279-286, 1997.

[13] M. M. Yeung, W. Wolf, B. Liu, "Video Browsing using Clustering and Scene Transitions on Compressed Sequences," *Proc. of IS&T/SPIE Conf. Multimedia Computing and Networking*, pp.399-413, 1995.

[14] D. Zhong, H. J. Zhang, S. F. chang, "Clustering Methods for Video Browsing and Annotation," *Proc. of SPIE Storage & Retrieval for Image and Video Databases*, vol.2670, pp.239-246, 1996.

[15] Y. Taniguchi, A. Akutsu, Y. Tomomura, "Panorama-Excerpts: Extracting and Packing Panoramas for Video Browsing," *Proc. of ACM Conf. on Multimedia*, pp.427-436, 1997.

[16] H. M. Kim, J. H. Lee, J. H. Yang, S. H. Sull, W. K. M. Kim and S. M. H. Song, "Visual Rhythm and Shot Verification," *Multimedia Tools and Applications*, vol.15, no.3, pp.227-245, 2001.

[17] M. J. Swain, D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol.7, no.1, pp.11-32, 1991.

[18] Adobe, Adobe Flash Player Version Penetration, "http://www.adobe.com/products/player_census/flashplayer/version_penetration.html"



천 수 덕

2009년 숭실대학교 대학원 컴퓨터학과 졸업(석사). 2009년~현재 알티베이스 연구원. 관심분야는 데이터베이스, 멀티미디어, P2P



신 정 훈

2010년 숭실대학교 대학원 컴퓨터학과 졸업(석사). 2010년~현재 숭실대학교 대학원 컴퓨터학과 박사과정. 관심분야는 데이터베이스, 멀티미디어, 분산처리



이 상 준

1996년 서울대학교 컴퓨터공학과 졸업(학사). 1998년 서울대학교 대학원 컴퓨터공학과 졸업(석사). 2004년 서울대학교 대학원 전기컴퓨터공학부 졸업(박사). 2005년~현재 숭실대학교 컴퓨터학부 조교수
관심분야는 데이터베이스, 멀티미디어,

P2P 시스템