메타데이타를 삽입한 디지털 콘텐츠 생성 도구 개발

천수덕° 주상욱 이상준
숭실대학교 컴퓨터학부
caddie04@comp.ssu.ac.kr, jupiterpen@comp.ssu.ac.kr, sangjun@ssu.ac.kr

Development of Digital Contents Authoring Tool using Metadata

Sooduck Chun°, Sangwook Joo, Sangjun Lee
School of Computing, Soongsil University

요 약

정보기술은 통신 및 멀티미디어 기술의 발전에 힘입어 빠르게 발전되고 있으며, 이에 따른 데이타베이스의 기술이 공간데이타, XML, 비디오, 음성과 같은 다양한 멀티미디어 데이터 분야에 적용되고 있다. 비디오 데이타는 순차적인 특성을 가지며, 시간과 공간정보가 결합된 3차원 데이타로서 처리시간이 높은 작업이기 때문에 검색이나 브라우징이 대단히 비효율적이다.

본 논문에서는 비주얼리듬을 이용하여 비디오 데이타에서 대표 프레임(Key Frame)을 추출한 다음 XML을 이용한 태그 및 키워드 정보를 대표 프레임에 삽입하여 검색이나 브라우징을 할 수 있는 동영상 내용 편집 도구(Authoring Tool for Video Contents)를 제안한다. 비주얼리듬은 3차원의 시공간적인 정보를 2차원으로 매핑한 정보로 IDCT(Inverse Discrete Cosine Transform)과정 없이 픽셀 정보를 얻을 수 있어 처리속도가 빠르며 컷, 와이프, 디졸브 등의 편집효과를 효과적으로 구분할 수 있다. 그리고 XML 데이타에는 태그 및 키워드 정보와 함께 대표 프레임의 정보까지 저장되므로 유사 화면 검색이나 내용 기반 검색을 제공할 수 있다.

1. 서 론

정보기술의 발전은 통신 및 멀티미디어 기술의 발전에 힘입어 빠르게 발전되고 있으며 이에 따른 데이타베이스의 기술이 공간데이타, 텍스트, 비디오, 음성, XML등과 같은 다양한 멀티미디어 데이타분야에 적용되고 있다. 최근에 비디오는 VOD(Video On Demand), NOD(News On Demand), 디지털 도서관, IPTV(Internet Protocol Television), UCC(User Created Content)등 다양한 응용 분야에서 점점 확산되고 있는 추세이다.

비디오 데이타는 시간에 의해 표현되는 연속 매체이다. 크기 자체가 수 메가바이트에서 수 기가바이트에 이르는 방대한 양을 가지고 있으며 컴퓨터상에서의 처리량의 부담과 많은 시간을 필요로 한다. 그러므로 내용 기반 검색이 어렵고 브라우징이 대단히 곤란하다. 따라서 처리량의 감소나 사용자에게 보다 원하는 정보를 찾도록 도와주는 응용기술 연구는 크게 비디오 검색과 브라우징으로 나뉠 수 있다. 비디오 데이타의 내용을 기반으로 검색 및 브라우징을 위해서는 비디오 데이타를 구성하고 있는 기본 단위인 샷(Shot)으로 분할하고, 각 샷에 대해 대표 프레임(Key frame) 추출 기법이 필요하며, 대표 프레임을 이용하여 사용자의 다양한 검색 질의를 보다 효과적으로 처리하기 위한 비디오 색인 기술이 필요하다 [1].

본 논문은 비주얼리듬[2,3]을 이용하여 비디오 데이타에서 대표 프레임을 추출한 다음 XML을 이용한다. 태그 및 키워드 정보를 대표 프레임에 삽입하여 검색이나 브라우징을 할 수 있는 동영상 내용 편집 도구(Authoring Tool for Video Contents)를 제안한다.

본고의 구성은 다음과 같다. 2장에서는 비디오의 연구 분야에 대해 살펴본다. 3장에서는 비주얼 리듬을 이용하여 대표 프레임을 추출한다. 그리고 XML을 이용하여 태깅을 하는 동영상 내용 편집 도구를 설명한다. 마지막으로 4장에서는 결론 및 앞으로의 연구 방향을 서술한다.

2. 관련 연구

비디오 데이타는 방대한 양과 시간에 의해 표현되는 연속 매체이다. 그러므로 비디오 데이타를 브라우징 및 검색을 하기 위해서는 복잡한 기술이 요구된다. 이와 관련된 연구 분야는 [그림 1]과 같다.

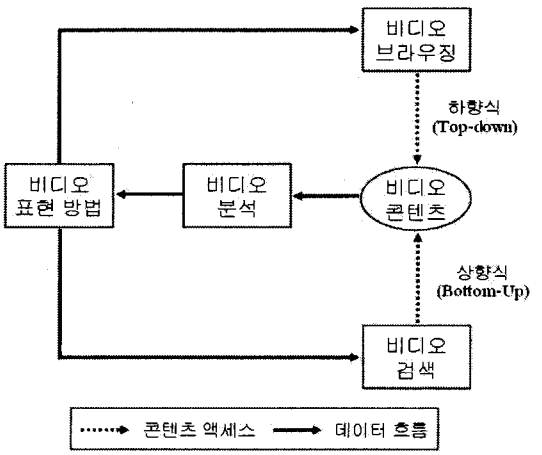


그림 1. 비디오의 연구 분야

* 본 연구는 서울시 산학연 협력사업(10581 cooperate Org 93112)의 지원에 의하여 수행되었음

비디오 분석은 경계의 검출, 대표 프레임의 추출, 키 오브젝트의 추출, 오디오 분석, 자막 분석 등이 있다.

샷 경계의 검출에서는 비디오 클립을 샷을 분해하고 샷 레벨에서 시그널 프로세싱을 한다. 자동적인 샷 경계 검출 기법의 종류로는 화소값에 기반한 접근[4], 화소값의 편차를 기반한 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 프레임의 추출에서는 샷 경계를 한 후, 각 샷에서 두드러진 프레임을 추출하는 작업으로 각 샷에서의 처음과 마지막 프레임을 추출하는 방법[10], 샷의 움직임 지표에 의한 방법[11], 시각적 내용 복잡 지표를 이용한 방법[12]등이 있다.

비디오 표현방법에서는 비디오 분석을 통해 공간적인 특징 벡터나 대표 프레임을 사용하여 트리형식의 계층적으로 구성하거나, 다양한 특징정보를 이용하여 상세한 검색을 제공할 수 있다.

비디오 브라우징은 하향식 접근으로 사용자가 자신도 원하는 부분을 접하기 전까지는 그 부분에 대해 묘사할 수 없는 경우로, 많은 분량의 비디오를 적은 수의 대표 프레임을 이용하여 보여줌으로써 사용자가 원하는 비디오 콘텐츠를 쉽게 찾을 수 있도록 하는 기능이다. 계층적 브라우징 기법[13]은 대표 프레임을 시간적인 순서대로 배열하여 보여줌으로써, 빠른 시간 내에 비디오의 개략을 살펴보는데 좋으며, 비디오 콘텐츠 전체를 메모리에 저장할 필요가 없는 장점이 있지만, 샷들의 위치만을 중요시하고 의미를 무시하여 비효율적인 브라우징을 초래할 수 있다. 씬 기반 브라우징 기법[14]은 비슷한 배경의 샷을 모아 씬을 구성하고, 이를 기반으로 원하는 샷을 찾아가는 브라우징 기법으로 씬을 구성하는 샷이 다양하게 존재하면 씬 간의 순서나 비디오 콘텐츠 내에서의 시간적인 위치를 파악하는데 어려움이 있다.

상향식 접근의 비디오 검색은 사용자가 자신이 찾고자 하는 부분에 대해 분명한 정보를 가지고 있어서 검색용 질의를 구성할 수 있는 경우로 비디오 색인은 두 가지로 나뉠 수 있다. 하나는 주석자에 의해 키워드나 텍스트에 만들어진 주석을 통하여 비디오 섹션의 특성을 부여하는 방법과 다른 하나는 비디오 신호 특성의 분석을 통해 자동적으로 세션을 상세화하여 검색하는 방법이다. 트리 구조를 이용한 계층적 색인화 기법[15], 대표 컬러를 이용한 색인화 기법[16], 미리 저장된 장면을 이용하여 검색하는 장면기반 검색 기법[1]등이 있다.

3. 동영상 내용 편집 도구

기존의 히스토그램을 사용한 방법은 국소적인 특징을 반영하지 않으며, 화소차를 이용한 방법은 전체적인 특징을 반영하지 못한다. 그리고 영상 제작시 사용된 컷, 와이프, 디졸브 등의 편집 효과로 인해 완벽한 샷 경계의 검출을 기대하기 어렵다[17].

본 논문에서는 국부적이고 전체적인 특성을 같이 보여주는 비주얼 리듬을 이용하였다. 비주얼 리듬은 동영상의 시공간적 정보를 모두 가지면서도 동영상의 일부만으로 동영상 전체를 충분히 표현할 수 있다.

비디오 데이타의 원시 데이타는 대용량이므로 MPEG 등으로 압축되어 진다. 디코딩이 상당한 처리시간이 요한다는 점을 감안하면 비주얼 리듬은 DC영상을 굳이 복호화 필요 없이 I-프레임에서 쉽게 얻을 수 있으므로 DCT를 기반으로 하는 영상 압축표준에서는 빠른 속도로 추출될 수 있다[18].

동영상 내용 편집 도구는 비주얼 리듬을 이용하여 비디오 콘텐츠에서 대표 프레임을 추출한 다음 XML (eXtensible Markup Language)을 이용한 주석의 태깅 및 정보를 입력하는 색인 방법을 구현하였다. XML을 이용하고 있어 그 데이타에 대한 처리, 접근 및 검색에 따른 XML 관련 프로그램이 필요하다. 그래서 본 논문에서는 마이크로소프트의 MSXML을 사용하여 XML 메타데이타를 처리하였고, 이 XML 파서에서 해석하는 표준으로 국제 표준화 기구에서 제정한 DOM(Document Object Model)을 사용하였다.

3.1 동영상 내용 편집 도구의 구조

동영상의 각 프레임에 원하는 메타데이타를 삽입하기 위해 태깅 툴을 이용해서 콘텐츠를 제작할 수 있다. 본 논문에서 제안하는 동영상 내용 편집 도구의 실행과정은 [그림 2]와 같다. 먼저 비주얼 리듬을 이용하여 대표 프레임을 추출하고, 그 대표 프레임의 아이템에 키워드를 삽입시켜 링크 시켜놓은 것을 XML 파일로 저장한다. 끝으로 동영상 파일에 결합하여 새로운 포맷인 IMF(Including Metadata File)을 생성한다.

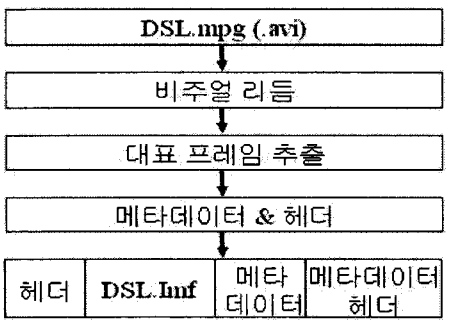


그림 2. 동영상 내용 편집 도구 실행과정

3.2 비주얼 리듬(Visual Rhythm)

비주얼 리듬은 전체 영상의 내용변화를 표현하는 한 장의 이미지이다. 비주얼 리듬에서 수직으로 한 줄에 속하는 화소들은 원시 프레임에서 추출된 축소화면의 대각선 화소이다. 특정 샷에 속한 프레임들에서 추출된 대각선 화소들은 거의 비슷한 시각적 특성을 지닌다. 따라서 시각 율동의 샷 경계 부근에서는 두드러진 시각적 변화가 나타난다.

샷은 촬영시에 카메라가 멈춤없이 한 번에 기록한 연속적인 프레임을 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.

그림 3. 샷 경계 추출

비주얼 리듬의 픽셀 샘플링 방법은 크게 수평, 수직, 대각선, 교차, 지역의 5가지 샘플링 방법으로 분류할 수 있다. 수평 샘플링은 수평 방향의 움직임을 검출할 수 없고, 수직 샘플링은 수직 방향의 움직임을 검출할 수 없다. 교차 대각선과 영역의 샘플링은 전체 내용을 대표할 수는 있으나 수평으로 나타나는 줄 때문에 판독에 지장을 주며 육안으로 보기 힘들다. 반면 대각선은 프레임 전체 내용을 대표하지는 못하나 컷, 와이프, 디졸브와 같은 편집 효과가 적용된 부분을 수직선, 사선, 곡선, 색상의 점진적 변화 등 육안으로 쉽게 인지 가능한 형태로 나타내는 특성을 갖는다. 이러한 이유로 비주얼 리듬은 대부분 대각선 샘플링을 사용한다[3].

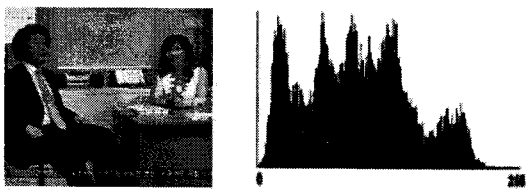
비주얼 리듬은 독특한 화소 샘플링에 의하여 3차원 영상을 2차원 이미지로 요약한 것으로서 다양한 영상 편집 효과들에 대하여 시각적 구분이 뚜렷한 패턴을 보여주는 특징을 가지고 있다. 이것을 이용하여 컷(cut)과 와이프(wipe)검출을 효과적으로 할 수 있다[2].

컷(cut)검출은 [그림 6]의 그래프에서 알 수 있듯이 비주얼 리듬의 시간 방향 미분이미지이다. 화소값이 피크(peak)를 형성하는 곳이 명암의 불연속성이 나타나는 곳이며, 명암의 불연속점이 수직선을 이루며 모여있는 곳이 컷(cut)이 발생한 장소이다.

3.3 이미지 프로세싱(Image Processing)

3.3.1 히스토그램(histogram)

영상 히스토그램은 영상의 명암값의 정보를 보여주기 위해 사용되는 데 매우 유용하다. 이 히스토그램을 사용하여 영상의 구성 즉, 명암 대비 및 명암값 분포에 대해 자세히 알 수 있다. 영상 히스토그램은 단지 화소가 가진 명암값들을 막대그래프로 표현한 것이다. 화소가 가질 수 있는 명암값은 x축상에 그려지며 각 명암값이 갖는 빈도수는 y축상에 그려진다. [그림 4]는 영상에 대한 샘플 히스토그램이다.



(1) 전체 영상 데이터 (2) 히스토그램

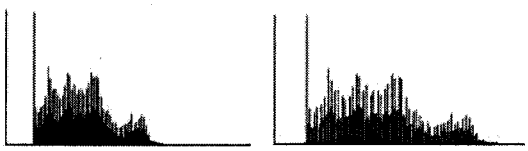
그림 4. 영상 데이타를 히스토그램으로 변환

3.3.2 히스토그램 평활화

영상의 명암값 분포가 빈약할 때 히스토그램 평활화 작업에 의해 향상될 수 있다. 히스토그램 평활화의 궁극적인 목적은 일정한 분포를 가진 히스토그램을 생성하는 것이다. 따라서 평활화를 수행한 히스토그램은 보다 균일한 분포를 가질 것이다. 히스토그램 평활화는 히스토그램을 평탄하게 하는 것이 아니라 명암값 분포를 재분배하는 것이다. 히스토그램 평활화는 포인트 처리이기 때문에 새로운 명암값이 영상에 추가 되지는 않는다. 즉, 기존의 명암값은 새로운 값으로 설정되지만 명암값의 실질적인 개수는 입력영상의 명암값의 개수와 동일하거나 적다.

히스토그램 평활화는 다음과 같은 3단계로 이루어진다.

(1) 히스토그램을 생성한다.
(2) 히스토그램의 정규화된 합을 계산한다.
(3) 입력영상을 변형하여 결과 영상을 생성한다.



(1) 히스토그램 (2) 히스토그램 평활화

그림 5. 히스토그램과 히스토그램 평활화 비교

3.4 대표 프레임(Key frame) 추출

비주얼 리듬을 이용하여 MPEG 동영상의 I-프레임을 추출하고 각각의 대표 프레임을 비트맵 객체에 저장한다. [그림 6]은 동영상 내용 편집 도구 인터페이스로 한 화면에 10개의 대표 프레임을 보여준다. 사용자는 이를 통해 대표 프레임을 선택하여 메타데이타를 태깅할 수 있고 이를 XML파일과 새로운 포맷인 IMF 파일로 저장할 수 있다.

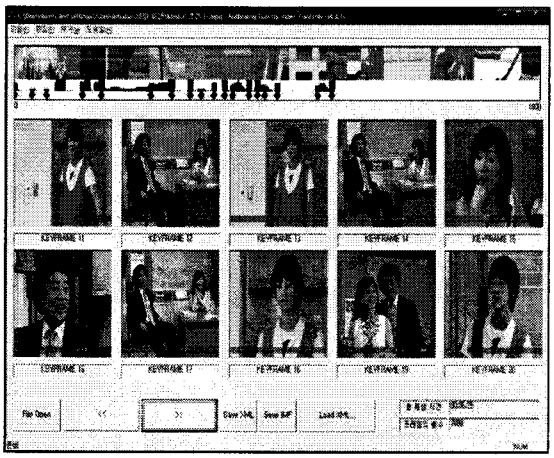


그림 6. 비주얼 리듬을 이용한 대표 프레임 추출

본 논문에서 대표 프레임 추출을 위해 구현한 컴포넌트의 주요 기능은 다음과 같다.

첫 번째로 히스토그램을 얻기 위해 히스토그램에 관련된 데이타들을 얻어주는 모듈을 구현하였다. 여기서 밴드 이미지의 화소 추출 기준을 설정하는데 색상(Hue) 및 RGB값을 균등하게 하고 Blue, Green, Red 위주로 설정한다. 밴드 이미지의 분석은 1000 프레임 단위로 이루어지므로 밴드의 가장 첫 번째 픽셀 스트라이프(pixel stripe)는 정상적인 시간축 방향 미분값을 구할 수 없으므로 이전 밴드의 가장 마지막 픽셀 스트라이프의 화소값을 유지해서 현재 밴드의 첫 번째 픽셀 스트라이프와 비교해야 한다. 그리하여 두 프레임 간의 화소차를 구하고 프레임의 색상 평균값을 계산하고 계산한 각 프레임의 색상 평균값을 모두 더한 후 1000개의 프레임에 대한 전체 색상 평균값을 구한다.

두 번째로 위의 모듈에서 생성된 시간축 방향 화소 미분치 배열에 대해서 정해진 범위 내에서 국소평균(Local Average)과 해당 국소평균을 기준으로 표준편차값을 구하는 모듈을 구현하였다. 대상 픽셀을 중심으로 좌우로 특정 너비(width)만큼의 영역에서 국소평균을 구하되 밴드의 끝 또는 시작 지점에 가깝게 있어서 특정 너비만큼의 영역을 가질 수 없는 경우에는 밴드의 경계까지 만을 대상영역으로 잡아서 계산한다. 각 국부영역(Local region)의 국소평균을 계산하고 구해진 국소평균을 기준으로 표준편차를 구하는 경우에도 같은 방식으로 대상영역을 제한한다.

세 번째로 시간축 방향 최소 미분치와 국소평균 및 표준편차에 기초하여 각각의 픽셀에서 피크(peak)가 존재하는 지 여부를 판정하고 피크가 존재한다면 해당 지점에 대한 누적 프레임값을 배열에 저장하는 모듈을 구현하였다. 해당 픽셀지점에서의 표준편차값의 가중치와 전체 평균의 가중치를 적용하였는데 그 이유는 피크를 단순한 절대적인 값으로 판정하지 않고 그 주변의 시간축 방향 화소 미분값의 평균변화량에 대해 상대적으로 판정하기 위해서이다. 이를 통해 장면 전환이 일어나는 지점들을 저장하여 측정할 수 있고 검색할 수 있게 된다.

3.5 메타데이타를 이용한 태깅 방법

추출된 대표 프레임의 프레임에서 영역을 잡고 그곳에 제작자가 넣고 싶은 광고하고 키워드를 저장하게 된다. 넣고 싶은 대표 프레임에 키워드를 모두 입력하고 저장을 하면 동영상에 XML파일이 합쳐지게 된다.

메타데이타를 이용한 태깅 방법에는 XML DOM parsing 을 이용하였으며 프레임 정보를 DOM을 이용하여 관련한 메타데이타를 XML 파일로 저장한 구조는 [그림 7, 8]과 같다.

XML 파일에는 동영상의 파일 이름과 플레이 타임 및 경로를 저장하고, 각 프레임별로 메타데이타가 태깅된 프레임의 시간을 따로 기록해 두었으며 각 프레임을 id로 구분하고 각 프레임내의 태깅 정보는 link id로 구분하였다. 이와 같은 방식으로 메타데이타를 이용하여 태깅을 하고 이를 동영상 파일에 삽입하여 새로운 포맷인 IMF(Including Metadata File)를 생성한다.

IMF는 기존의 동영상 파일과 함께 XML 데이타를 함께 저장하기 위한 구조로 XML 파일 크기 및 IMF의 크기를 나타내는 헤더 정보와 추후 IMF 포맷인지 아닌지 확인하기 위한 Signal을 나타내는 정보를 포함한다.

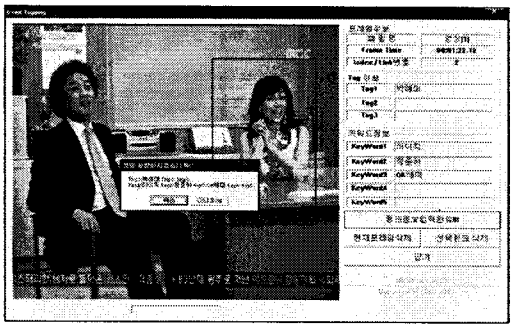


그림 7. 메타데이타를 이용한 태깅

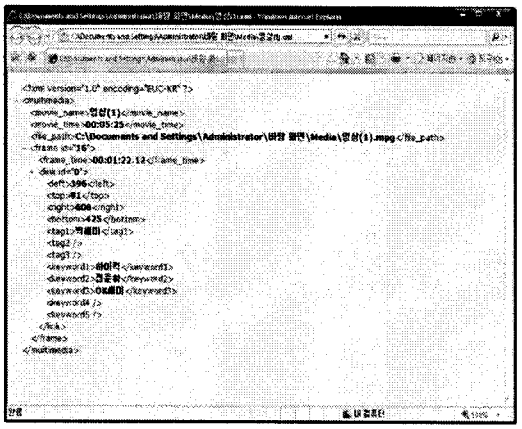


그림 8. 메타데이타가 저장된 XML 파일

4. 결론 및 향후 연구과제

본 논문에서 비주얼리듬을 이용하여 비디오 데이타에서 대표 프레임을 추출한다. 그리고 XML을 이용한 태그 및 키워드 정보를 대표 프레임에 삽입하여 검색이나 브라우징을 할 수 있는 동영상 내용 편집 도구(Authoring Tool for Video Contents)를 제안하였다. 비주얼리듬은 3차원의 시공간적인 정보를 2차원으로 매핑한 정보로 IDCT(Inverse Discrete Cosine Transform)과정 없이 픽셀 정보를 얻을 수 있어 처리속도가 빠르다. 또한 여러 편집효과를 효과적으로 구분할 수 있다.

또한, XML 데이타에는 태그 및 키워드 정보와 함께 대표 프레임의 정보까지 저장되므로 유사 화면 검색이나 내용 기반 검색을 제공할 수 있다.

향후, 본 논문에서는 대표 프레임을 추출하는데 I-프레임을 비교하였으나 특수효과 처리된 비디오 데이타의 경우는 B-프레임에서도 샷 경계 전환이 추출될 가능성이 있다. 이에 대한 연구가 필요하며 생성된 XML 태그 및

키워드 정보를 효율적으로 이용할 수 있는 브라우저 방법도 필요하다.

참고 문헌

[1] B. Furth, S. Smoliar and Zhang, "Video and Image Processing in Multimedia Systems," Kluwer Academic Publishers, 1995

[2] 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, Kluwer Academic Publishers, Vol.15, No.3, pp.227-245, 2001

[3] H. M. Kim, J. H. Lee, M. H. Huh, D. H. Choi, S. M. H. Song, "Method for Producing a Visual Rhythm Using a Pixel Sampling Technique," US Patent 6,549,245 B1, 2003

[4] H. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic Partitioning of Full-motion Video," ACM Multimedia Sys., Vol.1, No.1, pp.1-12, 1993

[5] R. Kasturi and R. Jain, "Dynamic Vision," Proc. of Computer Vision: Principles, IEEE Computer Society Press, 1991

[6] F. Arman, A. Hsu and M. Y. Chiu, "Feature Management for Large Video Databases," Proc. 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, and K. Mai, "A Feature-based Algorithm for Detecting and Classifying Scene Breaks," Proc. 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. ACM Conf. on Multimedia, pp.15-24, 1995

[11] W. Wolf, "Key Frame Selection by Motion Analysis," ICASSP, pp.1228-1231, 1996

[12] P. O. Gresle and T. S. Huang, "Gisting of video documents: A Key Frames Selection Algorithm Using Relative Activity Measure," The 2nd Int. Conf. on Visual Information System, pp.279-286, 1997

[13] D. Zhong and H. J. Zhang and S. F. Chang, "Clustering Methods for Video Browsing and Annotation," SPIE Vol.2670, pp.239-246, 1996

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

[15] U. Gargi, S. Oswald, D. Kosiba, S. Devadiga and R. Kasturi, "Evaluation of Video Sequence Indexing and Hierarchical Video Indexing," SPIE Storage & Retrieval for Image and Video Databases, pp144-151, 1995

[16] J. C. Lee, Q. Li, W. Xiong, "VIMS: A Video Information Management System," Multimedia Tools and Applications, Vol.4, No.1, pp.7-28, 1997

[17] A. Hampapur, R. Jain, T. Weymouth, "Digital Video Segmentation," Proc. ACM Multimedia, pp.357-364, 1994

[18] B. L. Yeo and B. Liu, "Rapid Scene Analysis on Compressed Video," IEEE Transactions on Circuit and Systems for Video Technology, Vol.5, pp.533-544, 1995