

비디오 샷 검증 시스템

정 지 문[†]

A Video Shot Verification System

Ji-Moon Chung[†]

ABSTRACT

Since video is composed of unstructured data with massive storage and linear forms, it is essential to conduct various research studies to provide the required contents for users who are accustomed to dealing with standardized data such as documents and images.

Previous studies have shown the occurrence of undetected and false detected shots. This thesis suggested shot verification and video retrieval system using visual rhythm to reduce these kinds of errors.

First, the system suggested in this paper is designed to detect the parts easily and quickly, which are assumed as shot boundaries, just by changing the visual rhythm without playing the image. Therefore, this enables to delete the false detected shot and to generate the unidentified shot and key frame.

The following are the summaries of the research results of this study.

Second, during the retrieving process, a thumbnail and keyword method of inquiry is possible and the user is able to put some more priorities on one part than the other between the color and shape. As a result, the corresponding shot or scene is displayed. However, in the case of not finding the preferred shot, the key picture frame of similar shot is supplied and can be used in the further inquiry of the next scene.

Key words : shot verification, video retrieval system, visual rhythm, shot boundary, movie image data, MPEG-7, XML, key picture frame.

1. 서론

색인을 위한 샷 경계를 구하는 작업은 완벽하게 구할

수 없는 것이 현실이다. 따라서 미검출 및 오검출을 해결 할 수 있는 방법이 필요하다. 이 논문에서는 비쥬얼 리듬을 이용하여 샷 경계의 오검출 및 미검출 문제를 해결하는 방법을 소개한다.

† 남서울대학교 컴퓨터학과 교수

* 본 논문은 2008년 남서울대학교의 교내연구비에 의하여 지원 되었음.

논문접수 : 2009년 6월 7일, 1차 수정을 거쳐, 심사완료 : 2009년 6월 15일

1.1 연구 배경 및 목적

비디오는 실세계의 다양한 내용들을 합축적으로 담

아낼 수 있는 유용한 도구로서 영상, 문자, 오디오와 같은 복합매체들로 구성이 된다. 특히 최근에는 컴퓨터와 통신, 그리고 데이터 압축 기술 및 표준기술의 발달로 비디오를 이용한 다양한 서비스가 가능하게 되었으며, 초고속 통신망의 보급과 멀티미디어 처리 기술로 인해 비디오는 다양한 분야에서 폭넓게 사용되고 있다. 이러한 멀티미디어 데이터의 양적 팽창과 급속한 확산은 기존의 텍스트 기반 검색으로는 멀티미디어 데이터를 효율적으로 검색하여 이용하는 것을 어렵게 만들고 있다.

이에 따라 최근 멀티미디어 데이터들이 포함된 시각적인 정보의 검색에 이용하는 내용 기반 검색 방법에 관한 연구들이 활발히 이루어지고 있다[2][3][5].

동영상은 프레임의 집합이라는 균일한 요소로 이루어져 있기 때문에 재생, 빨리 감기, 되감기 등의 순차적인 검색을 통하여 전체적인 내용을 파악할 수 있으며, 특정 부분으로의 임의 접근이 매우 힘든 구조를 갖고 있기 때문에 효율적인 영상 색인이 요구된다.

영상 색인을 위한 샷 경계 검출을 하고, 이들을 내용에 따라 장면 단위로 묶는 작업(clustering)은 수작업으로 할 경우 시간과 노력이 매우 많이 드는 고비용의 작업이다. 특히 다루어야 할 영상이 많으면 주어진 시간 내에 모든 샷과 장면을 검출하여 검색에 필요한 색인을 구축하기 어려워진다. 따라서 이를 자동화하는 소프트웨어의 도움이 필수적이다. 이 분야에 대해서는 최근 몇 년 동안 많은 연구가 이루어져오고 있다[1][3].

그러나 샷 경계 검출 알고리즘들은 샷 경계가 아닌 경우를 샷 경계로 오판하거나, 존재하는 샷 경계를 검출하지 못하는 경우가 매우 많다. 따라서 이러한 오검출과 미검출이 없는 완벽한 결과를 얻기 위해서 수작업에 의한 확인 및 오검출을 제거하고 미검출을 찾아내 교정하는 샷 경계 검증 과정이 반드시 필요하다.

이 논문에서는 수작업에 의한 샷 경계 검증시 활용할 수 있는 비쥬얼 리듬(Visual Rhythm)을 소개하고, 이를 이용하여 보다 쉽게 샷 경계 검출과 검증을 할 수 있는 방법을 제안한다. 또, MPEG-7 표준에 따라 동영상에 대한 메타 데이터를 생성하고 이를 이용하여 동영상에서 찾고자하는 장면(또는 샷)을 검색할 수 있는 시스템을 설계하고 구현한다.

2. 관련 연구

이 장에서는 이 논문과 관련된 선행 연구와 관련 연구를 기술한다.

2.1 텍스트 기반 검색

텍스트 기반 검색 방법은 인간이 인식할 수 있는 모든 내용을 정의하는 것이 가능하다는 장점을 가진다.

그러나 영상을 관찰하는 개개인마다 주관적인 평가에 의하여 영상을 정의하므로 동일한 영상일지라도 서로 다른 텍스트로 주석화를 수행하거나, 서로 다른 텍스트로 질의를 수행하는 상황이 발생한다. 또한 영상은 그 특성상 다양한 의미를 내포하는 경우가 빈번하고, 특정 단어의 나열로 그 의미를 정의하기 어려운 경우가 많을 뿐만 아니라 나날이 증가하고 발전하는 멀티미디어 데이터의 생산과 빠른 확산은 이러한 주석화 작업조차도 어렵게 만들고 있다. 따라서 멀티미디어 데이터를 효율적으로 색인하고 검색하여 이용하기 위해서는 내용 기반 검색 방법이 필요하다.

2.2 내용 기반 검색

내용 기반 검색 기법은 영상이 가지는 색상, 질감, 모양 등의 시각 정보를 이용하여 영상을 색인하고 검색에 이용하는 방법이다. 색상은 영상의 이동이나 회전 약간의 시점 변화에도 큰 영향을 받지 않으며 간편하고 빠른 처리 속도 때문에 여러 특정 정보들 중에서도 가장 널리 사용된다.

2.3 동영상 검색

이 절에서는 동영상의 특징을 알아보고, 동영상 검색 기법인 내용 기반 검색 기법을 설명하며, 현재 운영중이거나 연구되고 있는 동영상 검색 시스템의 특징을 비교한다.

2.3.1 카메라의 움직임과 동영상의 특징 정보

카메라 모션(motion)에는 카메라 위치가 고정되어 있는 모션으로 수평 선회하는 패닝(panning), 수직 선회하는 틸팅(tilting), 확대하는 줌 인(zoom in), 그리고 축소하는 줌 아웃(zoom out)이 있다. 카메라 위치가 움직이는 카메라 모션에는 좌우 이동하는 트랙킹(tracking), 상하 이동하는 봄잉(booming), 전후 이동하는 둘링(dollying)이 있다.

이러한 카메라 모션을 검출하기 위해 많이 사용되는 방법이 옵티컬 플로우(optical flow)를 이용하는 것이다. Ardizzone과 Cascia[4]는 카메라 모션을 계산하기 위해 각 프레임을 4등분하여 각각으로부터 옵티컬 플로우를

이용해 움직임의 세기 및 방향 히스토그램을 구하여 각 영역의 움직임 주 방향이 중심을 향해 발산하는 방향이면 줌 인이고, 수렴하는 방향이면 줌 아웃을 검출하고, 모든 영역의 주 방향이 일정한 방향이면 카메라 모션을 패닝 또는 틸트로 검출하는 방법을 제안하였다.

2.3.2 내용 기반 동영상 검색

일반적으로 동영상을 검색하기 위해서는 정지 영상의 검색에 사용되는 대부분의 방법들을 동영상을 구성하는 각각의 프레임에 적용시켜 사용할 수 있다.

MPEG-7에 등록되어 있는 GoF(Group of Frames) /GoP(Group of Picture) 컬러 기술자는 한 장의 영상이 가지는 색상 히스토그램에서 비디오 세그먼트가 가지는 여러 장의 영상 모음이 갖는 컬러 히스토그램까지 확장될 수 있다.

추가적인 2비트가 히스토그램 타입이 평균값(Average), 중앙값(Median), 교차값(Intersection)을 나타내기 위하여 사용된다.

3. 비디오 샷 검증 시스템

동영상에서 샷을 구분하기 위한 방법으로 영상의 변화를 요약한 한 장의 이미지를 나타내는 비쥬얼 리듬에 대하여 정의하고, 각 종 실험 영상을 통해서 비쥬얼 리듬을 구한다.

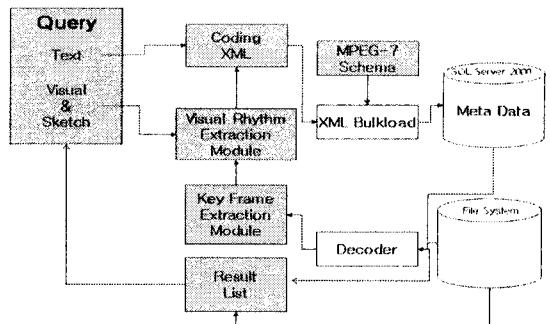
여러 실험 데이터를 통해서 샷 경계의 검출의 완전 자동화가 어렵다는 것을 밝히고, 이런 문제를 해결하고, 또한 다양한 질의로 검색할 수 있는 비디오 샷 검증 시스템을 제안한다.

3.1 시스템 구성도

비쥬얼 리듬을 이용한 비디오 샷 검증 시스템의 구조는 <그림 1>과 같다. 색칠한 부분이 제안하고 설계, 구현하고자 하는 모듈이다.

3.1.1 키 프레임 추출 모델

동영상 데이터를 다루는 부분으로 모든 동영상 데이터를 다루기 위하여 직접 동영상 데이터를 다루지 않고 디코더를 통하여 입력받는다. 이는 어떤 형태의 동영상 자료도 입력으로 사용할 수 있으므로 호환성에서 좋은 효과를 가져온다.



〈그림 1〉 시스템 구성도

3.1.2 비쥬얼 리듬 추출 모듈

추출된 키 프레임의 주요 색상 정보와 컬러 히스토그램, 모양 정보를 추출하여 Coding(XML) 모듈로 넘긴다. 윤곽선 정보를 추출하기 위하여 컬러를 흰도 성분으로 변환하고, 캐니 필터에 의한 윤곽선을 추출 후 팽창(dilation), 침식(erosion), 세선화 작업을 거친다.

3.1.3 XML Coding 모듈

비쥬얼 리듬 추출 모듈에서 받은 정보를 MPEG-7 표준 기술자로 기술하기 위한 작업을 수행하는 부분이다. 이렇게 기술된 XML 문서는 MPEG-7 스키마에 의해 유효성을 검증받아 데이터베이스에 저장된다.

3.1.4 질의 모듈

썸네일이나 디스플레이중인 영상을 캡쳐(capture)하여, 주요 색상(Dominant Color)이나 컬러 히스토그램, 모양 기술자(ContourShape)에 대한 정보를 비쥬얼 리듬 추출 모듈에 의해 추출하고, 입력된 키워드를 조합하여 Coding(XML) 모듈에 의하여 XML로 변환하여 데이터 베이스에 질의를 하게 된다. 질의에는 상황에 따라 가중치를 부여하는 기능이 내장되어 있다.

3.1.5 결과 리스트 관리 모듈

이 모듈에서는 질의의 결과로 생성된 정보를 관리한다. 비쥬얼 리듬 추출 모듈에서 요구하는 동영상 정보를 프레임 단위나 일부분 혹은 전체 동영상 정보를 비쥬얼 리듬 추출 모듈로 보낸다.

3.2 비쥬얼 리듬에 의한 샷 겹쳐과 겹증

비주얼 리듬은 영상의 내용 변화를 요약한 한 장의 이

미지이다. 특히 대각선 샘플링을 사용한 비쥬얼 리듬은 컷, 와이프, 디졸브와 같은 편집 효과가 적용된 부분을 수직선, 사선, 곡선, 색상의 점진적 변화 등으로 분명하게 나타나므로 이를 이용하여 샷 경계를 검출하고 검증하여야 한다.

3.2.1 비쥬얼 리듬의 정의

비쥬얼 리듬은 전체 영상의 내용 변화를 표현하는 한 장의 이미지이다. 비쥬얼 리듬에서 수직으로 한 줄에 속하는 화소들은 원시 프레임에서 추출된 축소 화면의 대각선 화소이다.

특정 샷에 속한 프레임들에서 추출된 대각선 화소들은 거의 비슷한 시각적 특성을 지닌다. 따라서 시각 유통의 샷 경계 부근에서는 두드러진 시각적 변화가 나타난다.

샷은 촬영시에 카메라가 멈춤 없이 한 번에 기록한 연속적인 프레임을 의미한다. 일반적으로 샷 경계를 검출하면 각 샷에서 키 프레임을 추출하고, 추출된 대표 화면들을 분석하여 비슷한 건들을 묶음으로서 장면을 정의한다. 따라서 샷 경계의 정확한 검출이 장면 경계 검출의 정확도를 결정한다.

영상 편집에서 사용되는 효과로는 컷, 디졸브, 와이프, 페이드 등이 있다.

<그림 2>는 대각선 교차, 그리고 지역 샘플링 방법을 이용해 동일 영상에서 만든 비쥬얼 리듬의 예를 나타낸 것이다.

임의의 영상 V 의 프레임 t 에서, 좌표 (x, y) 의 화소값을 $F_v(x, y, t)$ 라 하면, 영상 V 는 다음식과 같이 정의할 수 있다.

$$V = F_v(x, y, t), x, y, t \in \{0, 1, 2, \dots\}$$

영상 V 의 각 프레임을 가로, 세로 r 배로 줄인 축소 영상 $V_{\text{Thumbnail}}$ 의 화소값을 $F_{\text{Thumbnail}}(x, y, t)$ 라 하면 이는 다음식과 같이 표현할 수 있다.

$$V_{\text{Thumbnail}} = \{F_{\text{Thumbnail}}(x, y, t)\}, x, y, t \in \{0, 1, 2, \dots\}$$

영상 V 와 축소 영상 $V_{\text{Thumbnail}}$ 의 관계는 이들 화소값을 통하여 다음식과 같이 연결된다.

$$F_{\text{Thumbnail}}(x, y, t) =$$

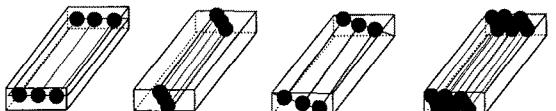
$$F_v(rx + k_x, ry + k_y, x, y, t \in \{0, 1, 2, \dots\}, \\ k_x, k_y \in \{0, 1, 2, \dots, r-1\})$$

여기서 k_x 와 k_y 는 샘플링 offset을 의미한다. 정의한 축소 영상 $F_{\text{Thumbnail}}(x, y, t)$ 를 사용하여 임의 영상 V 의 비쥬얼 리듬은 다음식과 같이 정의한다.

$$VS = \{F_{\text{vs}}(z, t)\} = \{F_{\text{Thumbnail}}(x(z), y(z), t)\}$$

비쥬얼 리듬 VS는 축소 영상들의 화소값

$F_{\text{Thumbnail}}(x, y, t)$ 의 일부분을 샘플링한 2차원 이미지이다.



<그림 2> 비쥬얼 리듬의 예

3.2.2 비쥬얼 리듬의 생성

비쥬얼 리듬을 만들기 위해서는 화소 샘플링에 의한 축소 영상을 먼저 생성하여야 한다. 원 영상이 MPEG의 I 프레임같이 DCT(discrete cosine transform)방식으로 압축된 경우 화소 샘플링은 다음과 같은 방법으로 효과적으로 할 수 있다.

DCT에서는 입력 프레임을 8×8 블록으로 나누어, 각 블록에 1개의 DC계수와 63개의 AC계수를 생성한다. 따라서, 각 DC계수만으로 $1/64$ 만큼 축소된 DC이미지를 얻을 수 있다.

축소된 이미지에서 비쥬얼 리듬에 반영될 화소의 선택은 수평, 수직, 대각선 화소 등 여러 가지가 있을 수 있으나, 대각선 화소가 이미지의 내용을 가장 잘 반영하는 것으로 알려지고 있다.

<그림 3>은 실험 영상에서 추출한 비쥬얼 리듬의 예이다. 여기서, 컷, 와이프, 디졸브 등이 수직선, 사선, 곡선, 색상의 점진적 변화 등으로 분명하게 나타나게 된다.



<그림 3> 실험 영상으로부터 추출한 비쥬얼 리듬

3.2.3 비쥬얼 리듬을 이용한 샷 경계 검출

필터링 작업과 세선화 작업과 같은 일련의 영상 처리 과정을 걸쳐 샷 경계를 검출한다.

비쥬얼 리듬에서 앞뒤 프레임의 대각선 정보들은 인접한 수직선에 위치한다. 이 논문에서는 각 수직선 화소값들을 비교하여 그 차이가 비쥬얼 리듬에 대한 표준 편차, 일정 지역 비쥬얼 리듬의 편차, 전체 평균치에 비하여 임계치를 넘으면 컷으로 판정한다.

비쥬얼 리듬 $I(z, t)$ 에 나타나는 명암의 불연속점이 수직선을 이루며 모여 있는 곳이 바로 컷(cut)이 발생한 장소이다. 컷을 찾기 위해 먼저 다음식과 같은 시간 방향 미

분을 수행한다.

$$d(z, t) = |I(z, t) - I(z, t-1)|$$

$d(z, t)$ 는 비쥬얼 리듬의 시간, 방향의 미분 이미지로서 화소값이 피크(peak)를 형성하는 곳이 명암의 불연속성이 나타나는 곳이다. 그것이 존재하는 피크를 효율적으로 검출하기 위해 $d(z, t)$ 를 다음식과 같이 일차원 함수로 변환한다.

$$d_{\mu} = \frac{1}{n} \sum_{z=0}^{n-1} d(z, t)$$

여기서 n 은 세로 방향 화소(z-axis)의 총 개수이다. 현재 변화가 이루어진 t 의 위치 주위에서 다음식과 같이 평균과 분산을 구한다.

$$\mu(t) = \frac{1}{N(B)} \sum_{k=B}^{B+n-1} d(t+k)$$

$$\sigma^2(t) = \frac{1}{N(B)-1} \sum_{k=B}^{B+n-1} (d_{\mu}(t+k) - \mu(t))^2$$

여기서 B 는 t 의 전후로 몇 개의 픽셀까지 영역을 포함할 것인가를 나타낸다.

$B = \{-16, -15, \dots, -1, 0, 1, \dots, 15, 16\}$ 이라면 $N(B)=32$ 가 된다. 따라서 다음식과 같이 검출할 수 있다.

$$b_{\mu}(t) = \begin{cases} 1, & d_{\mu}(t) > \mu(t) + m\delta(t) \\ 0, & d_{\mu}(t) \leq \mu(t) + m\delta(t) \end{cases}$$

$b_{\mu}(t)$ 의 값이 1이면 위치 t 에서 피크가 존재함을 의미하고, 값이 0이면 t 에서 피크가 없음을 의미한다. m 은 3에서 8사이의 값을 사용하면 무난하다.

3.2.4 시스템 인터페이스

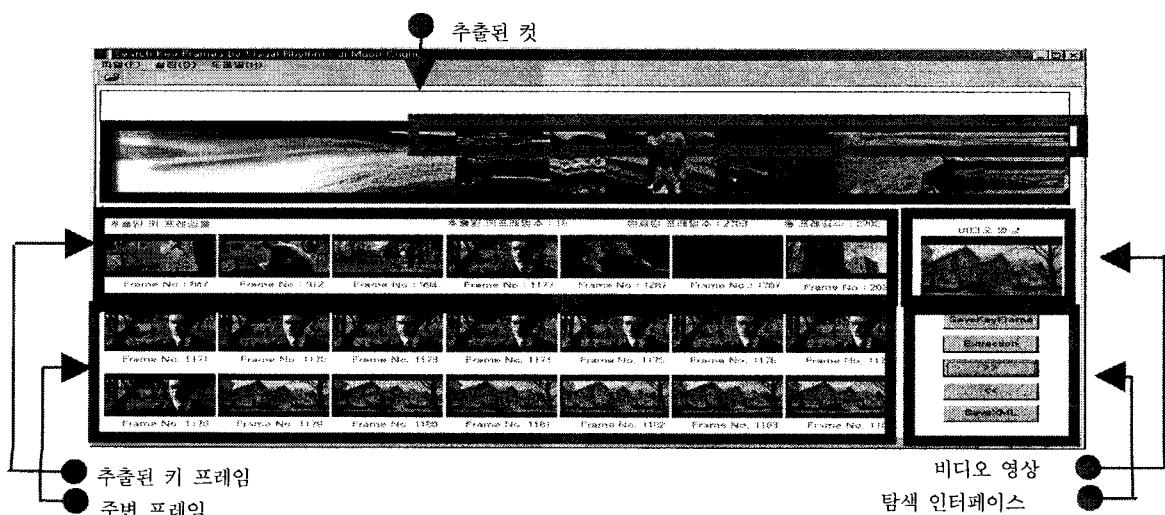
<그림 4>는 제안된 시스템의 사용자 인터페이스 부분이다. 것은 그림의 상단 부분에 비쥬얼 리듬에 대한 히스토그램과 함께 다이아몬드 모양으로 표시된다.

3.2.5 비쥬얼 리듬을 이용한 샷 검증

샷 경계 검증이란 샷 경계 검출 결과에 포함된 미검출 및 오검출된 샷 경계를 눈으로 찾아 교정하는 것이다. 샷 경계 검출 알고리즘이 미검출 및 오검출 없이 완벽하게 샷 경계를 검출한다면 샷 경계 검증은 불필요하다. 그러나 영상은 내용의 변화가 매우 다양하고, 와이프 및 디졸브와 같은 편집 효과를 사용하여 제작되므로 모든 경우를 완벽하게 찾아내는 샷 경계 검출 알고리즘을 개발하기는 거의 불가능하다. 특히 최근에는 디지털 영상 편집기를 사용하여 상상할 수도 없었던 2차원 및 3차원 효과로 영상을 편집하므로, 샷 경계 검출이 더욱 어렵게 되었다.

비쥬얼 리듬은 시간 축으로 확대/축소가 가능하여, 한 화면에 많은 분량의 영상 내용 변화를 표현할 수 있다. 이와 같은 비쥬얼 리듬의 특성을 이용하여 샷 경계 여부가 의심스러운 부분을 쉽게 확인할 수 있으므로 검증에 소요되는 시간과 노력을 크게 단축할 수 있다.

제안된 시스템은 사용자 인터페이스 상에 키 프레임화면과 그것 표시가 된 비쥬얼 리듬을 함께 배치하여 사용자가 직접 눈으로 확인하고 샷 경계 여부가 의심스러운 부분은 커서를 이용한 임의 접근 기능으로 그 부분으로



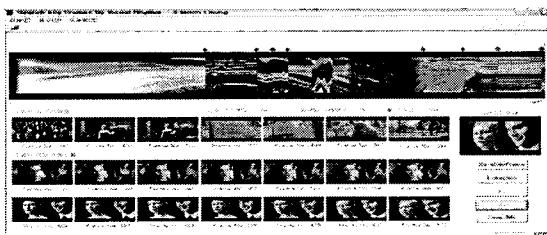
<그림 4> 시스템 인터페이스

이동하여, 샷 경계 여부를 하고 수정할 수 있도록 하였다.

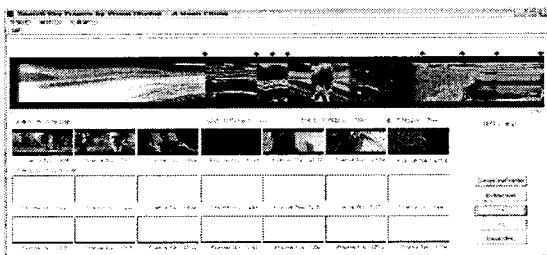
(1) 프레임 검색

인터페이스 상단에 출력된 비쥬얼 리듬은 1000장의 프레임의 요약 정보로써 사용자가 직관적으로 프레임의 전환 여부를 확인할 수 있도록 한다. 따라서 미심적은 프레임간의 경계에서 컷 검출 인식이 이루어지지 않은 경우 프레임 확인을 통한 면밀한 검토가 요구된다.

제안된 시스템은 추출된 비쥬얼 리듬의 일부분을 선택할 경우, 선택에 대한 해당 프레임들을 인터페이스 하단에 출력하며, 이를 통하여 사용자의 면밀한 검토와 정확한 피드백이 가능하도록 한다. <그림 5>는 비쥬얼 리듬의 선택을 통한 프레임 출력 인터페이스를 나타낸다.



(a) 프레임 탐색 전



(b) 프레임 탐색 결과

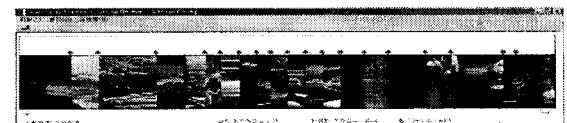
<그림 5> 비쥬얼 리듬의 선택을 통한 프레임 탐색

(2) 샷 경계의 추가 및 삭제

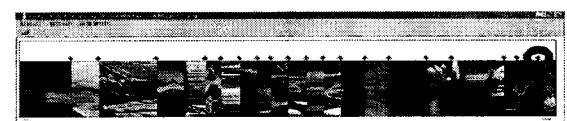
프레임 탐색을 통하여 구분되어야 할 샷 경계가 검출되지 않은 경우 사용자의 선택을 통하여 임의로 경계 설정을 추가하는 것이 가능하다. 뿐만 아니라 추출된 키 프레임 정보를 열람하여 삭제를 수행함으로써 잘 못 검출된 샷의 경계를 제거할 수 있다.

따라서 올바른 추출과 올바르지 못한 추출에 대한 완벽한 사용자 피드백을 사용자에게 제공하게 된다. <그림 6>과 <그림 7>은 검출되지 않은 샷의 경계를 사용자 피드백을 통하여 임의로 추가하는 경우를 보인다. <그림

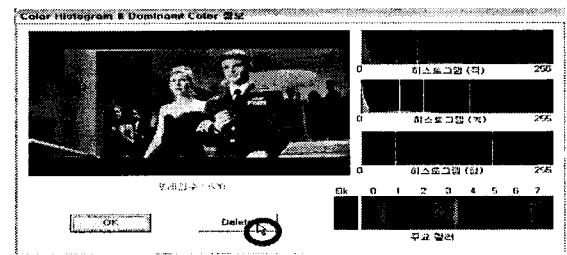
8>은 추출된 키 프레임의 정보를 열람하여 삭제를 수행하는 상황을 나타낸다.



<그림 6> 올바르게 검출되지 않은 샷이 존재하는 비쥬얼 리듬



<그림 7> 사용자의 피드백을 통한 임의의 샷 경계 추가



<그림 8> 사용자의 피드백을 통한 임의의 샷 경계 삭제

(3) XML 저장

사용자의 피드백을 통하여 올바른 검출과 올바르지 못한 검출에 대한 교정을 수행하고, 처리가 끝난 영상에 대한 프레임들의 정보를 XML로 저장하여 이용한다.

시스템 인터페이스 하단에 XML 저장 버튼을 통하여 완료된 영상의 각각의 샷 정보가 저장된다. <그림 9>는 XML 저장을 수행한 정보창과, 최종적으로 저장된 XML 데이터의 예이다. XML 데이터에는 추출 프레임의 프레임 번호, 영상에서의 프레임 위치, 영상의 경로, 저장된 프레임의 출력 경로, 모양 정보 및 색상 정보를 담고 있다.

```
<?xml version="1.0" encoding="utf-8" ?>
<Root>
  - <Objects>
    - <FrameNumber>365 </FrameNumber>
      <MediaTime>0:00:04.33 </MediaTime>
      <MediaTime>243333334 </MediaTime>
      <Offset>0 </Offset>
      <Offset>0 </Offset>
      <ImagePath>http://211.221.249.10/advert/spiderman3.wmv </FilePath>
      <ImagePath>http://211.221.249.10/advert/spiderman3.KeyFrame_0000365.bmp </ImagePath>
      <Moment1>-0.125724 </Moment1>
      <Moment2>-0.001726 </Moment2>
      <Moment3>-0.000019 </Moment3>
      <Moment4>-0.000005 </Moment4>
      <Moment5>-0.000001 </Moment5>
      <Moment6>-0.000001 </Moment6>
      <Moment7>-0.000001 </Moment7>
      <Hist01>+209 </Hist01>
      <Hist02>+137 </Hist02>
      <Hist03>+0 </Hist03>
      <Hist04>+109 </Hist04>
      <Hist05>+49 </Hist05>
      <Hist06>+44 </Hist06>
      <Hist07>+23 </Hist07>
      <Hist08>+138 </Hist08>
      <Hist09>+33 </Hist09>
  - <Objects>

```

<그림 9> 실험 영상에 대하여 저장된 XML 데이터

4. 실험 및 평가

실험은 팬티엄-IV 2.4GHz, Windows XP 서버 환경에서 Visual C++ 6.0 언어를 사용하여 구현한다. DBMS는 MS_SQL Server 2005를 이용하였고 XML 관련 처리는 MSXML 6.0 Parser, SQLXML 4.0, 화면 출력을 위하여 DirectX 9.0 SDK를 사용한다. 실험 영상은 영화 홍보 동영상과 광고 영상 등 150여의 다양한 영상으로 구성한다.

4.1 실험

실험을 통하여 동영상의 샷 검출과 실험을 통한 메타데이터 생성 단계, 비쥬얼 리듬을 이용한 샷 검증(수작업) 단계, 질의 단계로 나누어 진행한다.

4.1.1 비쥬얼 리듬 생성

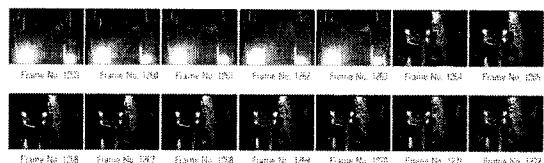
입력된 동영상의 모든 프레임에서 대각선을 이루는 화소들의 RGB 정보를 추출하여 비쥬얼 리듬을 생성한다. <그림 10>은 다양한 영상들에서 추출된 비쥬얼 리듬을 보인다.



<그림 10> 비쥬얼 리듬

4.1.2 샷 검증의 필요성

영상을 구성하는 샷의 검출은 다양한 요인들에 의하여 영향을 받는다. 일반적으로 영상들이 급격하게 변화하는 경우 영상 간의 시각 정보의 차가 위낙 뚜렷한 양상을 보이므로, 제안된 시스템뿐만 아니라 대부분의 관련 연구에서 안정적인 검출 성능을 보인다.



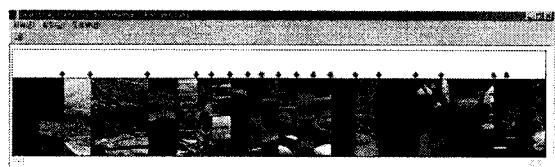
<그림 11> 급격한 장면 전환 예

<그림 11>는 급격한 화면의 전환으로 인한 명확한 검출이 가능한 실험 영상의 예를 보인다. 그러나 정확한 장면의 검출은 영상안에 포함된 조명의 밝기와 변화, 객체

들의 색상, 배경의 움직임 등의 영향을 많이 받는다. 뿐만 아니라 영상이 전환되는 속력이나 각각의 장면들을 연결하는 연결 부분의 편집 효과는 정확한 샷의 검출을 어렵게 만든다. 그밖에도 카메라의 이동이 효율적인 장면 검출을 방해하는 요인으로 작용한다.

4.1.3 샷 검증

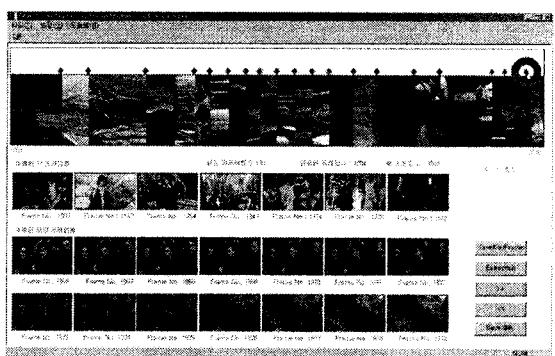
살펴본 바와 같이 제안된 시스템은 올바르게 검출되지 않은 샷을 사용자 피드백에 의하여 보완하고자 한다. <그림 12>은 실험 영상으로부터 추출된 정보를 나타내며 몇 개의 검색되지 않은 컷을 포함하고 있다.



<그림 12> 영상으로부터 추출된 비쥬얼 리듬과 샷 경계

<그림 13>에서는 사용자의 판단에 의하여 자동으로 검출되지 않은 샷 경계를 삽입하는 실험 화면이다.

이와 같이 자동적인 검출에 의하여 판단이 어려운 소수의 샷에 대한 피드백을 인덱싱 과정 이전에 수행함으로써 차후 인덱싱된 동영상 데이터를 검색하여 이용할 경우 보다 향상된 검색 성능과 정확도를 얻을 수 있다.



<그림 13> 인위적인 샷 경계의 추가

4.1.4 질의

썸네일 방식과 키워드 방식의 질의가 지원된다. 검색하고자 하는 장면의 내용에 따라서 색상 구성에 의한 주요 색상(dominant color)과 컬러 히스토그램 혹은 윤곽선으로도 질의할 수 있다.

정확한 결과를 찾지 못하였을 경우에도 검색 시 지정

한 유사도에 따라서 많은 이미지나 장면이 디스플레이된다. 원하는 샷(장면)을 찾지 못하였을 경우에는 디스플레이 중인 장면 중에서 검색하고자 하는 장면과 유사한 샷(장면)을 선택(capture)하여 다음 질의에 사용할 수 있도록 고려한다.

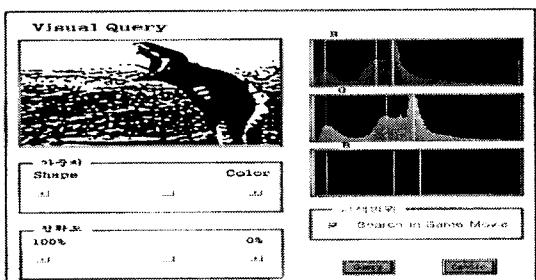
(1) 썬네일 방식 질의

썬네일의 히스토그램과 윤곽선 값을 추출하여 질의를 처리한다. <그림 14>와 같은 프레임을 갖는 동영상을 찾기 위하여 <그림 15>와 같이 히스토그램과 윤곽선 값을 추출하여 질의를 수행한다.

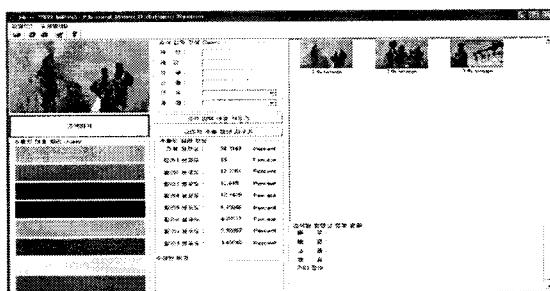
사용자는 검색하고자 하는 내용에 따라 장면 속에 있는 사람에 중심이면 모양에 파란 하늘에 푸른 들판과 같은 분위기를 찾고자 한다면 색상에 더 많은 가중치를 둘 수 있다.



<그림 14> 질의용 화면(썬네일)



<그림 15> 썬네일의 히스토그램과 윤곽선 값을 추출



<그림 16> 주요 칼라에 의한 질의와 검색 결과

<그림 16>은 주요 색상에 의한 질의와 검색 결과이다. 이는 색상에 의한 분위기 위주의 검색이며, 필요한 경우에는 키워드를 복합적으로 검색에 사용할 수 있다.

(2) 주석에 의한 질의

주석에 의한 질의는 영화명, 감독명, 주연 배우 등의 키워드 검색을 지원한다. 주요 색상과 함께 주석(키워드)을 사용하는 통합 검색도 가능하다. 키워드에 의한 검색의 경우에는 얼마나 많은 양의 주석 정보를 입력하는가에 따라 내용 검색이 가능하다. 이 실험에서는 극히 일부분의 주석만으로 제한하였으나, 사용 목적에 따라서, 사용자가 모두 입력하는 수동적 입력 부분과 음성 인식이나, 화면의 글씨를 인식하는 등의 자동화 부분을 추가하여 확장할 수 있다.

4.2 성능평가

<표 1>은 실험에 사용된 몇 가지 영상들에 대한 제안된 시스템의 샷 검출 결과를 정리한 결과이다.

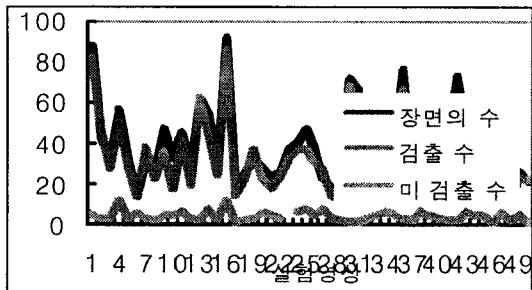
<표 77> 실험 동영상에 대한 샷 검출 결과

영상	총 프레임수	추출된 키 프레임수	정확도	미검출
디워 홍보영상	1850	32	98.2%	3
스파이더맨 홍보영상	2703	48	92.7%	4
해리포터 홍보영상	4083	73	94.1%	12
트랜스포터 홍보영상	2304	28	93.5%	6
자동차 광고 영상	882	13	96.0%	7
음료수 광고 영상	924	10	97.3%	2

여러 가지 실험 영상에 대한 제안된 시스템의 샷 검출 성능은 대부분 90%를 상회하는 높은 정확도를 보였다. 정확도는 추출된 샷들을 수작업으로 확인하여 검색된 샷에 대한 올바른 검출 비율을 계산한다. 같은 방법을 통하여 검출되지 않은 샷들이 어느 정도 존재함을 확인할 수 있었다. 전체 실험영상 150장 중 샷 검출에 대한 정확도는 평균 94%정도를 나타냄을 확인할 수 있었으며, 추출된 키 프레임 대비 추출되지 않은 미검출 샷은 7%~10%를 나타냄을 확인할 수 있었다.

<그림 17>은 실험영상들에 대한 검출 성능을 나타낸 그래프이다. 결과를 통하여 실험 영상에 존재하는 검출되어야 할 장면 전환을 거의 정확하게 검출하고 있음을 확인할 수 있으며, 약간의 미검출 장면이 존재함을 알 수 있다. 하지만 이러한 약간의 오류와 부족함은 사용자 편

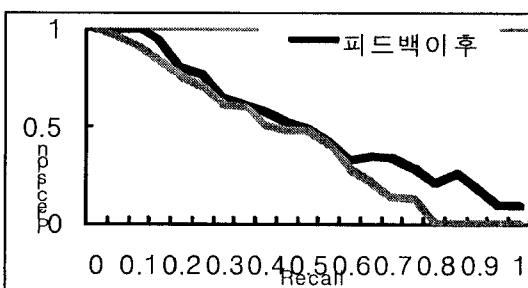
백에 의하여 정확하게 모두 제거하는 것이 가능하다.



〈그림 17〉 실험 영상들에 대한 검출 성능 결과

사용자의 피드백에 의하여 잘못된 샷의 검출을 제거하고, 검출되지 않은 샷을 추가함으로써 실험 영상에 대한 정확한 샷 경계를 정의할 수 있다. 잘못된 검출을 바로잡고, 검출되지 않은 샷을 추가한 뒤 실험 영상들을 대상으로 검색을 수행한 결과 <그림 18>과 같은 검색 성능의 향상 결과를 얻을 수 있었다.

사용자에 의한 적절한 피드백은 잘못 검출된 샷을 제거함으로써 검색 수행시 잘못된 결과가 검색될 가능성을 줄여 주며, 검출되지 않은 샷들을 추가해줌으로서 유사한 장면을 검색할 수 있는 가능성을 높여 줌을 알 수 있다.



〈그림 18〉 사용자 피드백에 따른 검색 성능의 향상

4.3 기존 동영상 검색 시스템과의 비교

<표 2>는 이 논문에서 제안하고 설계 구현한 시스템과 기존 동영상 검색 시스템과 비교한다. 상용화되어 있는 QBIC이나 Visual SEEK와 같은 시스템과 많은 부분들이 비슷한 수준의 성능을 보였다. 그러나 기존 시스템에서는 거의 지원하지 못하였던 메타 데이터의 호환성 부분은 본 시스템만의 장점이다. 또한, 시스템 구축 과정에서 비쥬얼 리듬에 의한 철저한 샷 경계 검증은 완벽한 메타 데이터의 생성을 보장하는 역할을 한다. VideoSTAR나 AVIS

같은 시스템도 메타 데이터의 호환이 일부 가능하지만, 이는 수작업에 의한 자료로 색상이나 모양과 같은 동영상의 내용에 관한 자료가 아니며, 내용 기반 검색에서 사용할 수 있는 메타 데이터가 아니다.

〈표 2〉 기존 동영상 검색 시스템과의 비교

시스템 이름	QBIC	Visual SEEK	Video STAR	AVIS	제안 시스템
메타 데이터 호환성	X	X	일부 가능	일부 가능	O
검색 가능한 특징 정보	주석, 색상, 질감, 모양	주석, 색상, 모양	주석	주석	주석, 색상, 모양
내용 기반 검색 지원	O	O	X	X	O
주석 기반 검색 지원	O	O	O	O	O
유사도 검색 지원	O	O	X	X	O
가중치 사용 지원	O	O	X	X	O
사용자 피드백 지원	X	X	X	X	O

5. 결론

비디오는 대용량이고 선형적인 형태의 비 구조화된 데이터로 구성되어 있어 문서나 이미지와 같은 정형화된 데이터를 다루는데 익숙해진 사용자들에게 필요한 내용을 제공해주기 위해서는 많은 연구가 필요하다. 기존 연구에서는 미검출과 오검출이 많이 발생한다. 따라서 이 논문에서는 비디오의 오검출과 미검출을 줄이기 위하여 비쥬얼 리듬을 이용한 샷 검증과 동영상 검색 시스템을 제안하였다.

제안된 시스템을 검증하기 위하여 샷의 검출된 정확도와 미검출 발생률을 실험하였다.

실험 데이터를 통하여 정확도는 평균 94%정도를 나타냄을 확인할 수 있었으며, 추출된 키 프레임 대비 추출되지 않은 미검출 샷은 7%~10%를 나타냄을 확인할 수 있었다. 미검출 및 오검출 된 샷은 사용자의 피드백에 의하여 잘못된 샷의 검출을 제거하고, 검출되지 않은 샷을 추가함으로써 실험 영상에 대한 정확한 샷 경계를 정의할 수 있었다.

제안된 시스템은 샷 검출 및 검증을 위하여 내용기반 검색의 기반 연구이며, 비디오의 장르에 영향을 받지 않

고 다양한 비디오에 대해서 좋은 결과를 보여주었다. 이 논문은 소리와 같은 매체를 이용하지 않았다. 따라서 다양한 매체의 특성을 복합적으로 이용한 샷 검출 방법에 대한 필요가 요구된다.

참 고 문 현

- [1] 원치선, “MPEG-7 시각서술자의 조사연구: 성능 비교 및 응용사례”(2006), *Telecommunications Review*, 제 16권, 3호.
- [2] A. Natsev, R. Rastogi, K. Shim(2004), “WALRUS: a similarity retrieval algorithm for image databases”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, pp.301-318.
- [3] Huamin Feng, Wei Fang(2005), Sen Liu, Yong Fang, “A new general framework for shot boundary detection and key-frame extraction”, *Multimedia Information Retrieval*, pp.121-126.
- [4] S. Adali(1996), “The Advanced Video Information System : data structuresand query processing,” *Multimedia Systems*.
- [5] Wang, J. Z., J. Li and G. Wiederhold, “SIMPLICity: Semantics-sensitive Integrated Matching for Picture. Libraries,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9):947-963, 2001.



정 지 문

- 1987 승실대학교
전자계산학과(공학사)
- 1990 연세대학교 공학대학원
전자계산전공(공학석사)
- 2007 충북대학교 대학원 전자계산
학과(이학박사)

1987.3~1991.2 한국방연구원(총괄팀장)
1991.3~1994.2 혜천대학 전자계산과 교수
1994~현재 남서울대학교 컴퓨터학과 교수,
학생처장, 전보전산원장, 정보전산교육원장,
컴퓨터학과장역임

관심분야 : 자료구조, 데이터베이스, 알고리즘,
멀티미디어컨텐츠.

E-Mail : jmchung@nsu.ac.kr