

# 시각 율동을 이용한 샷 경계 검증

## (Shot Boundary Verification using Visual Rhythm)

김혁만<sup>†</sup> 이진호<sup>\*\*</sup>

(Heyeokman Kim) (Jinho Lee)

**요약** 샷 경계 검증 알고리즘은 영상 제작시 사용된 컷, 와이프, 디졸브 등의 편집 효과로 인해 완벽한 결과를 기대하기 어렵다. 따라서 정확한 샷 경계를 얻기 위해서는 수작업에 의한 검증이 필요하다. 본고에서는 시각 율동을 이용한 샷 경계 검증 방법을 제시한다. 시각 율동은 영상의 내용 변화를 요약한 장의 이미지이다. 편집 효과는 수직선, 사선, 곡선, 점진적 색상의 변화 등 시각적으로 인지 가능한 형태로 시각 율동에 표현된다. 따라서 영상을 재생시키지 않고도 시각 율동의 변화만을 파악하여 샷 경계로 의심되는 부분들을 쉽고 빠르게 찾아낼 수 있다. 또한 이 특성을 이용한 샷 검증기를 구현하여 시각 율동의 유용성을 보인다.

**Abstract** Recent works regarding video shot change detection algorithms show that abrupt shot changes are detected fairly well. However, gradual shot changes including wipes and dissolves are often missed or falsely detected. A robust shot change detection system, therefore, must include a shot verification step to further enhance the overall system performance. In this paper, we introduce the concept of the visual rhythm which is a single image, a subsampled version of a full video. On the visual rhythm, the different video edit effects such as cuts, wipes and dissolves manifest themselves as different patterns. Using this characteristic, it becomes possible, without sequentially playing the entire video, to find false positive shots as well as undetected shots. Thus, inclusion of the visual rhythm in the shot boundary verification process will aid the operator to exclude falsely detected shots as well as to find undetected shots fast and efficiently. For this purpose we have developed a new tool, a shot verifier incorporating the visual rhythm. The usefulness of the visual rhythm during the shot verification process will be presented.

### 1. 개요

최근 방송용 디지털 영상 편집, 디지털 라이브러리, 영상 데이터베이스, 멀티미디어 영상 콘텐츠 제작, 인터넷에서의 영상 스트리밍 등과 같이 디지털 영상을 이용한 응용 분야가 확대됨에 따라 영상 색인(video indexing)에 관한 연구가 매우 활발히 이루어지고 있다. 영상은 프레임의 집합이라는 균일한 요소로 이루어져 있기 때문에 재생, 빨리 감기, 되감기 등의 순차적인

VCR 제어 기능을 통해서만 전체적인 내용을 파악할 수 있으며, 특정 부분으로의 임의 접근이 매우 힘든 구조를 갖고 있다. 반면에 일반적인 텍스트 관련 응용에서는 특정 단어나 문장, 문단 혹은 차례에서 제공하는 장이나 절 단위로 특정 부분을 편하게 임의 접근할 수 있다. 영상 색인이란 영상에 관련된 응용에서도 영상의 특정 부분으로 임의 접근이 가능한 메카니즘을 의미한다.

영상 색인에서 가장 중요한 것은 샷(shot)과 장면(scene)의 경계 검증이다. 영상은 제작 단계에서 물리적인 단위인 샷과 논리적인 단위인 장면으로 구성된다. 샷이란 촬영 시에 카메라가 멈춤 없이 한 번에 기록한 연속적인 프레임울 의미하고 [2,3], 장면은 내용이 비슷한 인접한 샷의 집합이다 [1]. 일반적으로 샷은 1-10초 정도이며, 장면은 광고나 영화 등 영상의 내용에 따라 그 편차가 상당히 크다.

· 본 연구는 정보통신부의 대학기초연구지원사업으로 이루어진 것임.

† 정 회 원 : 국민대학교 컴퓨터학부 교수  
hmkim@cs.kookmin.ac.kr

\*\* 비 회 원 : 한국통신 연구원  
jino@kt.co.kr

논문접수 : 1998년 9월 28일

심사완료 : 1999년 6월 10일

영상 색인을 위해 샷 경계를 검출(shot change detection 혹은 segmentation)하고, 이들을 내용에 따라 장면 단위로 묶는(scene change detection 혹은 clustering) 작업은 수작업으로 할 경우 시간과 노력이 매우 많이 드는 고비용의 작업이다. 특히 다뤄야할 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 있어왔고 지금도 활발히 연구가 이뤄지고 있다 [3-6,8-11].

일반적으로 샷 경계를 검출하면 각 샷에서 대표화면(key frame)을 추출하고, 추출된 대표화면들을 분석하여 비슷한 것들을 묶음으로써 장면을 정의한다. 따라서 샷 경계의 정확한 검출이 장면 경계 검출의 정확도를 결정한다. 기존 대부분의 연구들은 샷 경계 검출 및 대표화면 추출 단계와 장면 경계 검출 단계를 연계하여 실행하고 있다. 따라서 샷 경계 검출이 부정확하면, 장면 경계 검출 또한 부정확해질 수밖에 없다. 그러나 모든 샷을 100% 완벽하게 검출하기란 거의 불가능한 일이다. (이에 대해서는 2장에서 자세히 서술한다.) 일반적으로 샷 경계 검출 알고리즘들은 샷 경계가 아닌 경우를 샷 경계로 오판(false positive shots)하거나, 존재하는 샷 경계를 검출 못하는 경우(missing shots)가 매우 많기 때문이다. 따라서 오검출과 미검출이 없는 완벽한 결과를 얻기 위해서는 수작업에 의한 확인 및 오검출을 제거하고 미검출을 찾아내 교정하는 샷 경계 검증(shot boundary verification) 혹은 샷 검증 과정이 반드시 필요하다. 본 고에서는 샷 경계 검출 결과를 효율적으로 검증하는 방법으로서 시각 율동(visual rhythm)을 이용하는 방법을 제안하고, 이를 소프트웨어 도구화한 샷 검증기를 제안한다.

본고의 구성은 다음과 같다. 2장에서는 기본적인 용어 정의와 샷 검증의 필요성을 살펴본다. 3장에서는 시각 율동의 정의 및 생성 방법, 그리고 시각 율동을 이용한 샷 경계 검증을 설명하고, 4장에서는 구현한 시각 율동을 이용한 샷 검증기의 기능을 설명한다. 마지막으로 5장에서는 결론 및 앞으로의 연구 방향을 서술한다.

## 2. 영상 편집 효과와 샷 검증

영상은 촬영한 원시 테이프에 있는 샷들을 선택하여 연결하는 편집 과정을 통해 제작된다. 편집시 샷을 연결하는 기법으로는 크게 컷(cut), 디졸브(dissolve), 와이프(wipe) 등의 편집 효과를 사용한다 [3]. 컷이란 인접한 샷을 단순히 잇는 것을 말하며, 이를 급진적 전환

(abrupt change)이라고 한다. 반면에 디졸브와 와이프는 인접한 샷을 몇 개의 프레임에 걸쳐 동시에 보여줌으로써 연결을 보다 자연스럽게 만드는 방법으로, 점진적 전환(gradual change)이라 한다. 디졸브는 사라지는 샷의 끝 부분은 페이드-아웃(fade-out)하고, 새로 나타나는 샷의 시작 부분은 페이드-인(fade-in)하면서 두 샷을 오버랩(overlap)시키는 방법을 말한다. 와이프는 새로 나타나는 샷이 사라지는 샷을 화면에서 밀어내는 방식을 말한다. 최근에는 디지털 편집 장비가 발달함에 따라 설명한 2차원 편집 효과 외에 rolling, flying, doppler 효과와 같은 점진적 3차원 편집 효과도 보편화되었다.

샷 경계 검출이란 이미 제작된 영상에서 역으로 편집시에 연결해 놓은 샷 경계를 찾아내는 것이다. 컷과 같은 급진적 전환은 일반적으로 전혀 다른 내용의 프레임이 인접하여 있으므로 샷 경계가 비교적 분명하다. 따라서 인접 프레임간의 픽셀 단위 신호 특성, 프레임간의 평균 신호 특성, 히스토그램, 모션 등의 차이를 이용하는 방법만으로도 어느 정도의 경계 검출이 가능하다. 그러나 이 경우에도 panning, zooming과 같은 카메라 동작이 많이 사용된 경우, 아주 밝거나 어두운 배경, 빠른 물체의 움직임이 있는 경우 등에서는 검출 정확도가 많이 떨어진다. 디졸브, 와이프, 3차원 편집 효과와 같은 점진적 전환은 여러 프레임에 걸쳐 연속적으로 일어나기 때문에 샷 경계가 불분명하여 경계 검출이 매우 어렵다. 기존 연구들도 거의 대부분 급진적 경계 검출 분야에서 이뤄졌으며 [5,6,8-11], 점진적 경계 검출 기법에 대한 연구는 아직 미미한 실정이다 [3,4].

표 1 급진적/점진적 전환의 빈도

영상 종류	프레임수(시간)	급진적 전환수	점진적 전환수
드라마	133204 (74분)	831	42
뉴스	81565 (45분)	293	99
영화	142507 (79분)	564	95
광고	51733 (28분)	755	254
기타	10706 (6분)	64	16
합계	419745 (233분)	2507	506

Boreczky 등[1]은 표 1의 영상을 분석하여 다음과 같은 결과를 얻었다. “전체 샷 경계 중 점진적 전환은 평균 16.8%이다. 그러나 장면 경계가 아닌 샷 경계에서 점진적 전환이 사용된 비율은 10%가 약간 넘는 정도인

반면에, 장면 경계에 해당하는 샷 경계에서 점진적 전환이 사용된 비율은 30%가 넘는다.” 따라서 만일 점진적 샷 경계를 검출하지 못한 채 샷 경계 검출의 다음 단계인 장면 경계를 검출하려 한다면, 장면 경계 검출 알고리즘이 아무리 정확하여도 30% 이상의 장면은 찾을 수 없게된다.

지금까지의 샷 경계 검출 알고리즘들은 영상에 디졸브와 와이프와 같은 점진적 전환이 많지 않다면 각각 약 10-20% 내외의 미검출 및 오검출율을 보인다 [1]. 또 점진적 전환을 검출하는 알고리즘도 전체적인 결과는 위의 데이터와 비슷하나, 점진적 전환 특히 와이프만을 대상으로 한다면 미검출율이 40%, 오검출율이 60%에 이른다 [3]. 현재의 연구로는 샷 경계의 완전한 자동 검출이 이뤄지지 않고 있으므로, 정확한 샷 경계 검출을 원할 경우 수작업에 의한 검증이 필요하다.

샷 검증을 위해서는 영상을 담은 테이프 혹은 파일을 영상 재생기(video player)의 VCR 기능을 이용해 자동 검출된 결과와 일일이 확인해야 한다. 그러나 일반적으로 매우 짧은 시간에 샷 경계가 변하기 때문에 확인을 위해서는 같은 구간을 계속 반복적으로 재생하는 경우가 많이 발생한다. 따라서 처음부터 수작업으로 하는 것보다 더 많은 시간과 노력이 필요할 수도 있다.

예를 들어, 표1의 경우 샷의 평균 길이는 4.6초이므로 1시간 짜리 영상에는 평균 782개의 샷이 있다. 만일 샷 경계 검출시 발생하는 오검출율이 실제 샷 경계의 20%, 미검출율이 10%라고 한다면, 이 영상에 대해 샷 경계 검출 알고리즘이 리포트하는 샷 경계의 수는  $782 + 156 - 78 = 860$ 개가 된다. 이 많은 샷 경계를 일일이 재생하면서 오검출을 확인하는 일은 여간 힘든 일이 아니다. 또한 미검출된 샷 경계를 찾아내기 위해서는 전체 영상을 거의 다 재생시켜 보아야 한다.

대부분의 Motion JPEG 영상 재생기는 프레임 단위(frame-accurate) 제어 기능이 있기 때문에 이런 방법이 가능하다. 그러나 MPEG 영상의 경우 프레임 단위 제어가 가능한 재생기가 거의 없기 때문에, 알고리즘이 리포트하는 정확한 프레임 위치 근처를 재생시키는 것조차 쉬운 일이 아니다. 또한 영상 재생기를 이용한 수동 확인은 단순 반복 작업이므로 많은 시간과 노력이 필요할 뿐만 아니라, 작업자의 집중도에 따라 검증 작업의 정확도가 떨어질 수 있다. 따라서 검증을 위한 소프트웨어 도구의 도움이 절실하다. 본고에서는 검증을 쉽고 빠르게 할 수 있는 메카니즘으로서 시각울동의 개념, 이를 이용한 샷 경계 검증, 그리고 이를 소프트웨어 도구화한 샷 검증기를 제안한다.

### 3. 시각 울동

#### 3.1 시각 울동과 화소 샘플링

샷 검증을 효율적으로 하려면 미검출 및 오검출된 샷 경계라고 의심되는 부분을 영상을 재생하지 않고도 눈으로 빠르고 정확하게 찾을 수 있는 방법이 필요하다. 본고에서는 이를 위해 시각 울동(visual rhythm)을 제안한다.

시각 울동의 정의에 앞서, 임의의 영상  $V$ 의 각 화소(pixel)값을  $f_V(x, y, t)$ 라 하면, 영상  $V$ 는 다음과 같이 정의할 수 있다.

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

영상  $V$ 의 각 프레임에 가로와 세로 모두  $r$ 배로 줄인 축소영상  $V_{Thumbnail}$ 의 화소값을  $f_{Thumbnail}(x, y, t)$ 이라 하면, 즉,

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

영상  $V$ 와 축소영상  $V_{Thumbnail}$ 의 관계는 이들 화소값을 통하여 다음과 같이 연결된다.

$$f_{Thumbnail}(x, y, t) = f_V(rx + k_x, ry + k_y, t), \quad x, y, t \in \{0, 1, 2, 3, \dots\}, \\ k_x, k_y \in \{0, 1, \dots, r-1\}$$

여기서  $k_x$ 와  $k_y$ 는 샘플링 오프셋(offset)을 의미한다.

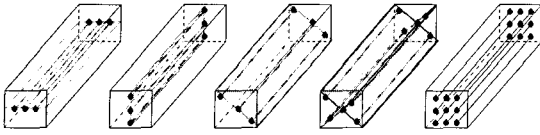
정의한 축소영상  $f_{Thumbnail}(x, y, t)$ 을 사용하여 임의의 영상  $V$ 에 대한 시각 울동  $VR$ 을 다음과 같이 정의한다.

$$VR = \{f_{VR}(z, t)\} = \{f_{Thumbnail}(x(z), y(z), t)\}$$

즉, 시각 울동  $VR$ 은 삼차원 축소영상의 화소값  $f_{Thumbnail}(x, y, t)$ 의 일부분을 샘플링한 이차원 이미지임을 알 수 있다. 다시 말하면, 시각 울동이란 2차원 프레임의 화소들을 부분적으로 샘플링하여 1차원에 배열함으로써, 3차원 공간의 영상 내용을 2차원 공간의 이미지로 요약한 것이다.

화소를 샘플링하는 방법에 따라 다양한 시각 울동을 만들 수 있다. 예를 들면, 임의의 평면  $y=y_0$ 에 대해  $(x(z), y(z)) = (z, y_0)$ 를 사용하여  $VR$ 을 만들면 이는  $y=y_0$ 에서의 수평 샘플링을 의미하고,  $(x(z), y(z)) = (x_0, z)$ 는 수직 샘플링,  $(x(z), y(z)) = (z, az)$ 는 대각선 샘플링을 의미한다. 그림 1은 가능한 몇 가지 예를 나타낸 것이다.

화소 샘플링 방법의 선택 기준은 영상 내용 변화의 표현 정도에 따라 결정한다. 그림 2는 그림 1의 방법중 대각선, 교차 그리고 지역 샘플링 방법을 이용해 동일



(a) 수평 (b) 수직 (c) 대각선 (d)교차 (e) 지역  
 그림 1 시각 울동 생성을 위한 화소 샘플링 방법

영상에서 만든 시각 울동의 예를 나타낸 것이다. 그림 2.(c)의 지역 샘플링은 신호 처리 측면에서 프레임의 전체 내용을 대표할 수는 있으나, 수평으로 나타나는 줄 때문에 눈으로 보기에는 불편하다. 교차 샘플링을 사용한 그림 2.(b)도 가운데 수평줄이 판독에 지장을 주며 한쪽 대각선의 내용이 너무 짧아 눈으로의 판독이 어려운 경향이 있다. 반면에 그림 2.(a)의 대각선 샘플링은 프레임 전체 내용을 대표하지는 못하나, 3.3 절에서 설명할 눈으로 검증하는데 편한 여러 특성을 갖고 있다. (그림에서 수직으로 나타난 선에 대해서도 3.3절에서 설명한다.) 본고에서는 대각선 샘플링을 이용한 시각 울동을 선택하였다. 이후의 시각 울동은 대각선 샘플링을 사용한 것을 의미한다.

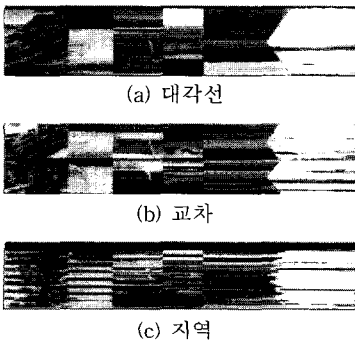


그림 2 화소 샘플링에 따른 시각 울동의 예

**3.2 시각 울동의 생성**

영상이 단순히 디지털화되어 파일로 저장되어 있다면 화소 샘플링은 매우 간단한 문제이다. 그러나 압축 저장되어 있다면 복원 과정이 필요하므로 시각 울동을 생성하는 알고리즘이 필요하다.

영상 압축 포맷중 M-JPEG(Motion JPEG) 혹은 MPEG의 I 프레임과 같이 DCT 방식으로 압축된 영상 일 경우에는 화소 샘플링을 효과적으로 할 수 있다. DCT에서는 입력 프레임을 8x8 블록으로 작게 나누어, 각 블록에 DCT를 적용하여 1개의 DC 계수(co-

efficient)와 63개의 AC 계수를 생성한다. DC 계수는 블록에 속한 64개 화소에 대응하는 신호의 평균(정확히는 평균의 8배)이다. 따라서 각 프레임의 DC 계수만으로 이미지를 만들면 1/64로 축소된 DC 이미지를 얻을 수 있다 [9]. 만일 DC 이미지를 축소화면으로 사용할 경우, 앞에서 정의한 축소영상  $V_{Thumbnail}$  과 영상  $V$ 의 화소 값은 다음과 같은 특성을 갖는다.

$$f_{Thumbnail}(x, y, t) = \frac{1}{8} \sum_{k_x=0}^7 \sum_{k_y=0}^7 f_V(8x+k_x, 8y+k_y, t),$$

$$x, y, t \in \{0, 1, 2, 3, \dots\}$$

DC 이미지를 축소화면으로 사용하면, 가장 시간이 많이 걸리는 IDCT(inverse DCT) 복원 과정 없이 압축된 영상에서 DC 계수만 읽어내어 시각 울동을 생성할 수 있다. 또 모션 벡터를 사용하는 MPEG1과 MPEG2의 B와 P 프레임에서도 최소한의 복원만으로 빠르게 DC 이미지를 생성하는 알고리즘이 이미 개발되어 있으므로 [7,9], 이를 이용하면 MPEG에서의 시각 울동도 효과적으로 생성할 수 있다.

**3.3 시각 울동의 특성과 샷 검증**

시각 울동은 전체 영상의 내용 변화를 표현하는 한 장의 이미지이다. 시각 울동에서 수직으로의 한 줄에 속한 화소들은 원시 프레임에서 추출된 축소화면의 대각선 화소이다. 특정 샷에 속한 프레임들에서 추출된 대각선 화소들은 거의 비슷한 시각적 특성을 지닌다. 따라서 시각 울동의 샷 경계 부근에서는 두드러진 시각적 변화가 나타난다. 이 변화들은 쉽게 눈으로 감지되므로, 이를 이용해 수작업에 의한 오검출 및 미검출된 샷들의 교정이 가능하다.

그림 3은 편집 효과가 시각 울동에서 어떻게 나타나는지를 보여주고 있다. 컷으로 연결된 샷 경계의 경우 그림 3.(a)와 같이 시각 울동에서는 수직선으로 나타난다. 만일 시각 울동에서 인접한 두 개의 줄이 서로 다른 샷에 속한다면 이들은 시각적으로 두드러진 차이를 보일 것이다. 이 차이가 시각 울동에서는 수직선으로 나타난다.

새로운 샷이 그전의 샷을 화면의 왼쪽에서 오른쪽으로 밀어내는 수평 와이프(horizontal, left-to-right wipe)의 경우 그림 3.(b)와 같이 시각 울동에서는 사선으로 나타난다. 화면의 위에서 아래로 새로운 샷이 나타나는 수직 와이프(vertical, top-to-bottom wipe)도 같은 형태의 사선으로 나타난다. 새로운 샷이 오른쪽에서 왼쪽으로 나타나는 수평 와이프나, 아래에서 위로 나타나는 수직 와이프의 경우에는 사선의 방향만 180도 바뀐 형태의 시각 울동이 생성된다.

새로운 샷이 화면의 중앙에서 나타나 그전의 샷을 주변으로 밀어내는 확장 와이프(expanding, center-to-outskirts wipe)의 경우에는 그림 3.(c)와 같이 시각 운동에서는 위에서 아래로 연결된 곡선으로 나타난다. (만일 새로운 샷이 시간축에 비해 일정 크기로 증가하면 곡선이 아니라, 180도 방향이 다른 두 개의 사선이 만나 각을 이루는 형태의 선으로 나타난다.) 새로운 샷이 화면의 주변에서 나타나 이전 샷을 중앙으로 밀어내는 흡수 와이프(absorbing, outskirts-to-center wipe)의 경우에는 곡선의 방향만 180도 바뀐 형태의 시각 운동이 생성된다.

디졸브에서는 새로운 샷과 이전 샷의 대응하는 두 프레임이 오버랩되어 새로운 프레임이 만들어지므로, 오버랩된 구간에 속한 프레임의 각 화소들은 두 샷의 신호 특성을 모두 갖는다. 따라서 컷이나 와이프와 같이 시각 운동에서 시각적으로 인식 가능한 선의 형태를 나타내지는 못한다. 그러나 디졸브는 그림 3.(d)와 같이 시각 운동에서 색상의 변화가 점진적으로 꾸준히 나타나므로, 주의 깊게 살펴보면 디졸브 여부를 판단할 수 있다.

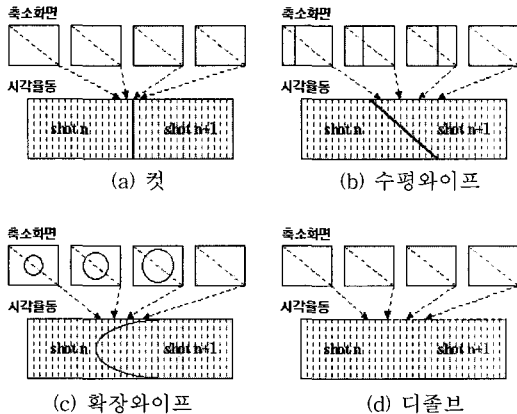


그림 3 시각 운동에서의 편집 효과

그림 4는 실제 영상으로부터 생성한 시각 운동이다. 앞서 설명한 대로 영상을 재생하지 않고도 시각 운동만으로 다양한 형태의 편집 효과를 쉽게 판별할 수 있다. 뿐만 아니라 그림 4.(c)에서 보듯이 오검출의 주요 원인인 줌-인/줌-아웃과 같은 카메라 동작, 카메라 플래시, 특정 객체의 움직임 등 다양한 내용 변화를 한 장의 이미지만 시각 운동을 통해 볼 수 있다. 우리는 설명한 시각 운동에서 나타나는 편집 효과의 특성을 이용한 샷

경계 검출 알고리즘을 개발하고 있다 [4,8]. 이 알고리즘은 기존 알고리즘과 달리 컷 뿐만 아니라 대부분의 와이프와 일부 디졸브의 검출도 가능하다. 그러나 이 알고리즘도 완벽한 샷 검출이 이뤄지는 단계는 아니므로, 오검출 및 미검출된 샷들을 찾아내어 교정하는 검증을 거쳐야만 정확한 샷 경계 검출 결과를 얻을 수 있다.

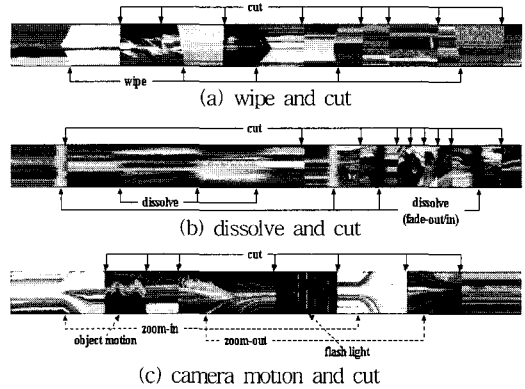


그림 4 시각 운동의 예

본고에서 교차와 지역 샘플링 대신 대각선 샘플링을 사용하는 이유는 3.1절에서 설명하였다. 수평이나 수직 샘플링 대신 대각선 샘플링을 사용하는 이유는 와이프 때문이다. 디졸브와 달리 와이프의 종류는 수백 가지에 이르며, 영상 편집자는 자기만의 독창적인 와이프를 만들 수도 있다. 와이프는 수평, 수직, 확장, 흡수, 기타 와이프와 같이 크게 5 종류로 나눌 수 있으며, 주로 사용되는 것은 앞의 네 종류의 변형들이다. 그림 5는 5 종류 와이프의 대표적인 형태를 나타낸 것으로, 기타의 경우는 부분 흡수 와이프를 나타내었다.

그림 5의 와이프들에 대해 각각 수평, 수직, 대각선 샘플링을 적용할 때 시각 운동에서 나타나는 선의 형태는 표 2와 같다. 수평 샘플링의 경우 수직 및 흡수 와이프, 수직 샘플링의 경우 수평 및 흡수 와이프가 시각 운동에서는 컷으로 나타남을 알 수 있다. 반면에 대각선 샘플링에서는 사용 빈도가 매우 높은 수평, 수직, 확장, 흡수 와이프는 모두 와이프로 나타나고, 사용 빈도가 낮은 부분 흡수 와이프만 컷으로 나타난다. 따라서 시각 운동에서의 표현 능력이 가장 좋은 대각선 샘플링을 사용하는 것이다. 한편 샷 검출의 측면에서 볼 때 표 2와 같이 와이프를 컷으로 오인하는 것은 미검출하는 것보다는 바람직한 현상이다. 왜냐하면 미검출된 샷을 눈으로 검출하는 작업은 오인된 샷을 교정하는 것 보다 몇 배 힘든 작업이기 때문이다. (엄밀히 말하면 오검출은

샷이 아닌 경우를 샷으로 판정하는 것이다. 따라서 와이프를 컷으로 오인하는 것은 오검출의 범주에 속하지 않는다.)



그림 5 와이프의 종류

표 2 수평, 수직, 대각선 샘플링과 시각 운동

	수평 와이프	수직 와이프	확장 와이프	축소 와이프	부분축소 와이프
수평 샘플링	사선	수직선 (컷)	곡선	수직선 (컷)	사선
수직 샘플링	수직선 (컷)	사선	곡선	수직선 (컷)	사선
대각선 샘플링	사선	사선	곡선	곡선	수직선 (컷)

#### 4. 샷 검지기

##### 4.1 샷 검증기의 기능

샷 검증기는 샷 경계 검출 알고리즘의 결과에 포함된 미검출 및 오검출된 샷을 눈으로 찾아내고, 이를 수작업으로 교정하는 검증 과정을 위한 소프트웨어 도구이다. 샷 검증기는 샷 경계 검출 알고리즘을 도구화한 샷 검출기와 함께 사용할 수도 있으며, 샷 검출기 없이 단독으로 사용할 경우에는 수작업만으로 샷 경계를 검출할 수도 있다.

단순히 영상 재생기의 VCR 기능만으로 영상을 반복적으로 재생하면서 검증하는 기존 방식 보다 샷 검증기를 사용하는 것이 훨씬 빠르고 편해야 한다. 이를 위해서 샷 검증기에는 다음과 같은 기능들이 필요하다.

- 전체 영상의 내용을 요약하는 기능
- 편집 효과를 시각적으로 보여주는 기능
- 특정 구간의 영상을 프레임 단위로 보여주는 기능
- VCR 기능
- 샷 경계 수동 설정 및 해제

본고에서는 720x480 M-JPEG 영상의 샷 경계 검출 결과를 시각 운동을 이용하여 검증하는 샷 검증기를 구현하였다. 설명한 것처럼 대각선 샘플링을 이용한 시각 운동은 컷, 와이프, 디졸브 등의 편집 효과를 수직선, 사선, 곡선 혹은 색상의 점진적 변화 등 눈으로 쉽게 인식

가능한 형태로 보여준다. 또한 시각 운동은 한 장의 이미지가므로, 디스플레이할 때 이 이미지를 시간축으로 적절히 확대/축소함으로써 전체 혹은 부분적 내용을 쉽게 파악할 수 있다. 그림 6은 8분 분량 영상으로 만든 시각 운동을 시간축을 변화시키며 보여주고 있다.

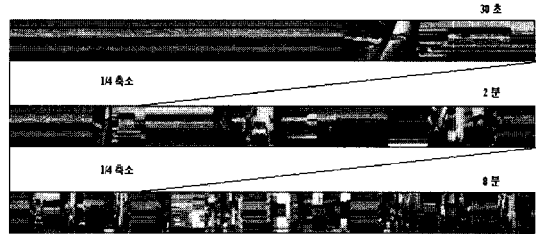


그림 6 시각 운동의 확대 및 축소

시각 운동에서 특정 구간의 편집 효과 사용 여부가 의심스러우면 그 부분을 프레임 단위로 살펴보거나, 재생할 수 있다. 구현한 샷 검증기에서는 각 프레임을 1/64로 줄인 축소화면을 이용해 특정 구간의 연속한 프레임을 보여주며, 사용자의 요구에 따라 원래 크기의 프레임도 디스플레이 가능하다. 또한 일반적인 VCR 기능을 이용한 영상 재생과 함께 샷 단위의 재생도 가능하다. 이 기능들을 이용해 의심스러운 구간들을 살펴봄으로써 미검출 및 오검출된 샷 경계들을 교정할 수 있도록 수동으로 샷 경계를 설정 혹은 해제하는 작업이 가능하다.

##### 4.2 샷 검증기의 사용자 인터페이스

그림 7은 구현한 샷 검증기의 사용자 인터페이스이다. 이 인터페이스는 크게 시각 운동부, 축소화면부, 제어부로 나눌 수 있다. 인터페이스 상단에 있는 시각 운동부는 생성한 시각 운동과 샷 검출기에서 찾아낸 샷 경계를 함께 디스플레이한다. 그림에서 알 수 있듯이 검출한 샷 경계는 시각 운동의 해당 위치 상단에 조그만 삼각형으로 표시한다. 시각 운동의 우측 상단에 있는 버튼은 시각 운동을 시간축으로 확대/축소하는 기능을 제공한다. 시각 운동 위의 커서(cursor)는 축소화면부의 중앙에서 보고있는 프레임의 위치를 나타낸다.

인터페이스 중앙 부분은 커서가 위치하고 있는 프레임, 그리고 전후로 인접한 각각의 열 개 프레임을 축소화면으로 나타낸다. 축소화면부의 중앙 좌측에 홀로 있는 축소화면이 현재 커서가 위치한 프레임이고, 그 위와 아래의 열 개 프레임이 커서가 위치한 프레임에 전후로 인접한 프레임들이다.

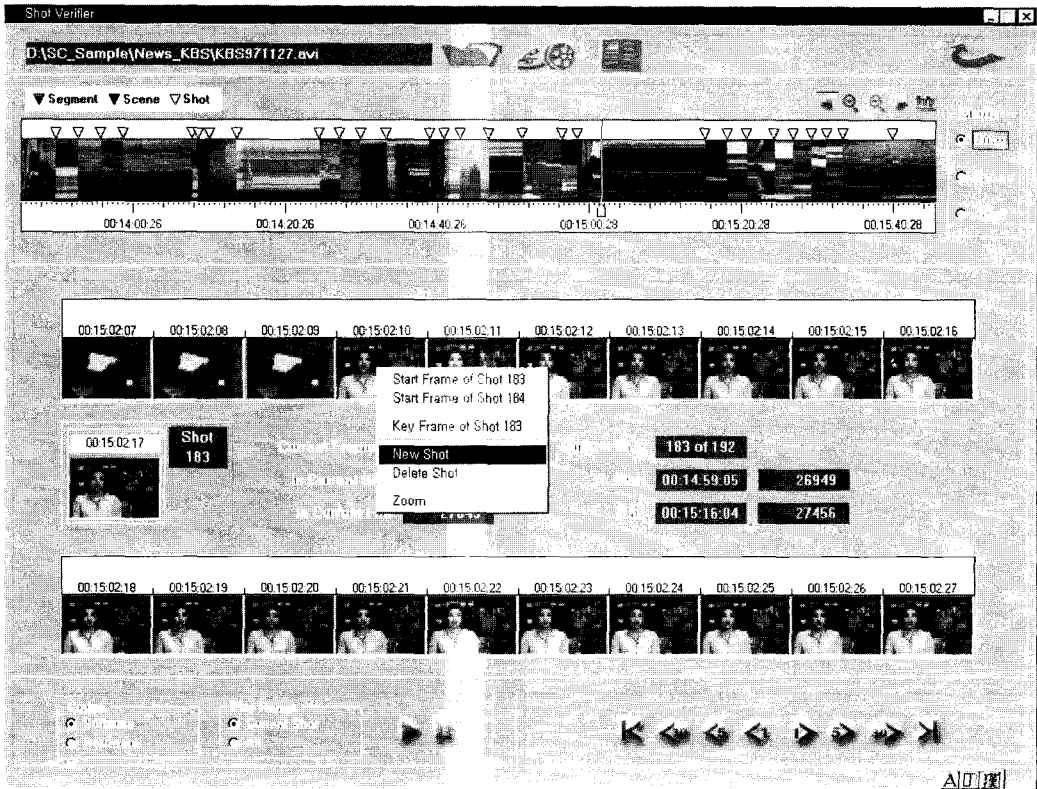


그림 7 샷 검증기의 사용자 인터페이스

인터페이스 하단에 있는 제어부는 왼쪽의 영상 제어부와 오른쪽의 프레임 제어부로 나눌 수 있다. 프레임 제어부는 커서의 위치를 프레임 단위로 제어하는 기능을 제공한다. 즉 현재의 커서 위치에서 전후로 1, 5, 30 프레임 단위로 세밀하게 이동하는 기능을 제공한다. 만일 임의의 위치로 이동하려면 시각 율동부의 커서를 직접 drag & drop할 수도 있다. 커서의 위치가 바뀌면 이와 대응하여 축소화면부에도 새로운 축소화면이 디스플레이된다. 영상 제어부는 샷 단위의 영상 재생 기능을 제공한다. 재생 버튼을 누르면 새로운 창이 생성되어 영상이 재생된다. 커서의 위치 이동은 시각 율동부에서 가능하므로 빨리감기나 되감기 기능은 제공하지 않는다.

샷 검증기를 이용한 검증 과정은 다음과 같다. 시각 율동부의 시각 율동을 살펴보다 의심스러운 부분이 있으면 커서를 그 부분으로 이동시킨다. 그림 7에서는 시각 율동부에서 컷으로 의심되는 직선 위에 컷 표시가 없으므로 커서를 해당 위치로 옮겼다. 그러면 축소화면

부에 해당 프레임과 전후의 20개 프레임이 나타난다. 이들을 살펴봄으로써 샷 경계 여부를 거의 대부분 판별할 수 있다. 그림 7의 경우 축소화면부의 네 번째 프레임부터 새로운 샷이 시작됨을 알 수 있다. 만일 그 이상의 프레임들을 보거나 영상을 재생하려면 제어부의 기능을 이용한다. 미검출 및 오검출된 샷 경계를 찾아내면 축소화면부의 해당 프레임 위에서 마우스를 클릭한다. 그림 7의 경우 네 번째 프레임 위에서 마우스를 클릭한다. 그러면 그림 7과 같은 작은 메뉴 창이 나타난다. 이 창에서는 미검출된 샷을 새로운 샷으로 정의하고, 오검출된 샷을 제거하는 기능이 제공된다. 미검출된 샷을 정의하면 시각 율동의 해당 위치에 작은 삼각형 표시가 나타나고, 축소화면부의 해당 프레임 위에 샷 시작 표시가 나타난다. 오검출된 샷을 제거하면 시각 율동의 해당 위치에 있는 삼각형 표시가 사라지고, 축소화면부의 해당 프레임 위에 있는 샷 시작 표시가 사라진다.

시각 율동을 사용하면 영상을 재생시키지 않고도 대

부분의 미검출 및 오검출된 샷 경계를 교정할 수 있다. 실제로 이 도구를 사용한 결과, 예를 들어 한 시간의 영상 내용을 불과 몇 분만에 검증하는 것이 가능하였다.

## 5. 결론

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. 본고에서는 샷 경계 검출 알고리즘이 자동 검출한 결과 중 미검출 및 오검출된 샷 경계를 쉽고 빠르게 수작업으로 찾아낼 수 있는 시각 운동의 개념을 소개하고, 이를 사용한 샷 검증기를 구현하였다.

시각 운동은 영상의 내용 변화를 요약한 한 장의 이미지이다. 특히 대각선 샘플링을 사용한 시각 운동은 컷, 와이프, 디졸브와 같은 편집 효과가 적용된 부분을 수직선, 사선, 곡선, 색상의 점진적 변화 등 눈으로 쉽게 인지 가능한 형태로 나타내는 특성을 갖는다. 또한 시각 운동은 시간축으로 확대/축소가 가능하며, 한 화면에 많은 분량의 영상 내용 변화를 디스플레이할 수 있다. 이와 같은 시각 운동의 특성을 이용한 샷 검증기는 시각 운동을 이용해 샷 경계 여부가 의심스러운 부분을 쉽게 판별하고, 커서를 이용한 임의 접근 기능으로 그 부분으로 용이하게 이동하며, 대부분의 경우 재생하지 않고도 축소화면만으로 샷 경계 여부를 판별할 수 있으므로 검증에 소요되는 시간과 노력을 크게 단축할 수 있다.

시각 운동은 영상 편집 효과 이외에도 카메라 동작, 플래시 라이트, 객체 동작 등을 시각적으로 표현할 수 있다. 또한 시간축으로의 확대/축소 및 특정 프레임으로의 임의 접근이 가능하다. 따라서 이들 특성을 이용하면 샷 검증기 외에도 디지털 영상 편집기, 영상 브라우저, 샷 검출기, 영상 색인기, 영상 모델링 등 디지털 영상에 관련된 각종 소프트웨어에 응용이 가능하다.

구현한 샷 검증기는 M-JPEG 영상만 다루고 있다. 앞으로는 현재 사용이 확산되고 있는 MPEG1 및 MPEG2 영상을 다룰 수 있도록 기능을 추가할 예정이다. 샷 경계 검출은 영상 색인의 기초로서 그 응용 분야가 매우 넓다. 앞으로는 미검출 및 오검출율을 최대한

줄일 수 있는 정교한 샷 경계 검출 알고리즘 개발에 주력할 예정이다.

## 참고 문헌

- [1] J. S. Boreczky, L. A. Rowe, "Comparison of video boundary detection techniques," Proc. of *SPIE Storage and Retrieval for Image and Video Database IV*, SPIE Vol.2670, pp.170-179, 1996.
- [2] G. Davenport, T. A. Smith, N. Pinciver, "Cinematic primitives for multimedia," *IEEE Computer Graphics and Applications*, pp.67-74, July 1991.
- [3] A. Hampapur, R. Jain, T. Weymouth, "Digital video segmentation," Proc. Of *ACM Multimedia*, pp.357-564, 1994.
- [4] H. Kim, S.-J. Park, J. Lee, W. M. Kim, S. M. Song, "Processing of partial video data for detection of wipes," Proc. of *SPIE Storage and Retrieval for Image and Video Databases VII*, SPIE Vol.3656, pp.280-289, 1999.
- [5] 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.
- [6] Y. Nakajima, K. Ujihara, A. Yoneyama, "Universal scene change detection on MPEG-coded data domain," Proc. of *SPIE Visual Communication and Image Processing*, SPIE Vol.3024, pp.992-1003, 1997.
- [7] J. Song, B.-L. Yeo, "Spatially reduced image extraction from MPEG-2 video: fast algorithms and applications," Proc. of *SPIE Storage and Retrieval for Image and Video Database VI*, SPIE Vol.3312, pp.93-107, 1998.
- [8] S. M. Song, T.-H. Kwon, W. M. Kim, H. Kim, B.-D. Rhee, "On detection of gradual scene changes for parsing of video data," Proc. of *SPIE Storage and Retrieval for Image and Video Database VI*, SPIE Vol.3312, pp.404-413, 1998.
- [9] B.-L. Yeo, B. Liu, "Rapid scene analysis on compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.5, No.6, pp.533-544, Dec. 1995.
- [10] M. M. Yeung, B.-L. Yeo, W. Wolf, B. Liu, "Video browsing using clustering and scene transitions on compressed sequences," Proc. of *SPIE Multimedia Computing and Networking*, SPIE Vol.2417, pp.399-413, 1995.
- [11] H. Zhang, K. Kankanhalli, S. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Systems*, Vol. 1, No. 1, pp.10-28, 1993.





김 혁 만

1985년 서울대학교 컴퓨터공학과 졸업 (학사). 1987년 서울대학교 대학원 컴퓨터공학과 졸업 (석사). 1996년 서울대학교 대학원 컴퓨터공학과 졸업 (공학박사). 1987년 ~ 1999년 한국통신 멀티미디어연구소 연구원. 1999년 ~ 현재 국민대학교 컴퓨터학부 교수. 관심분야는 디지털 라이브러리, 비디오 데이터베이스, 웹 방송 시스템 등임.



이 진 호

1991년 숭실대학교 전자계산학과 졸업 (학사). 1993년 숭실대학교 대학원 전자계산학과 (석사). 1993년 ~ 현재 한국통신 근무. 관심분야는 컴퓨터 비전, 멀티미디어, 웹 방송 시스템 등임.