

# Visual Rhythm의 방향성을 이용한 장면변환 검출

준회원 윤상호\*, 정회원 유지상\*

## Scene change detection using visual rhythm by direction

Sangho Yoon\* Associate Member, Jisang Yoo\* Regular Member

요약

최근 디지털 콘텐츠 관리의 중요성이 점점 증대되는 가운데, 동영상 콘텐츠의 중복장면을 최대한 줄이고, 대량의 비디오 데이터를 효과적으로 요약하기 위하여 장면변환 검출과 관련된 연구가 계속 진행되고 있다. 기존의 히스토그램과 화소 차를 이용한 장면변환 검출 기법은 조명변화와 움직임에 민감한 결괏값을 나타내었다. 따라서, 최근에는 계산량이 적고, 장면특성을 비교적 잘 나타내는 visual rhythm을 사용한 검출 방법이 많이 연구되어지고 있다. 본 논문에서는 visual rhythm의 방향성을 이용하여 장면을 검출하는 기법을 제안하였다. 제안하는 기법에서는 계산량을 효과적으로 줄일 수 있으며, 방향성 특성으로 인해 움직임이 있는 장면에서도 지속적인 성능을 가질 수 있다. 실험 결과에서는 기존에 사용되던 히스토그램을 이용한 장면변환 검출과 비교하여 약 30%의 성능효과를 보였고, 움직임이 많은 광고와, 뮤직비디오에서 비슷한 성능이 유지됨을 확인 할 수 있었다.

Key Words visual rhythm, cut, dissolve, scene change

ABSTRACT

While the management of digital contents is getting more and more important, many researchers have studied about scene change detection algorithms to reduce similar scenes in the video contents and to efficiently summarize video data. The algorithms using histogram and pixel information are found out as being sensitive to light changes and motion. Therefore, visual rhythm gets used in recent work to solve this problem, which shows some characteristics of scenes and requires even less computational power. In this paper, a new scene detection algorithm using visual rhythm by direction is proposed. The proposed algorithm needs less computational power and is able to keep good performance even in the scenes with motion. Experimental results show the performance improvement of about 30% comparing with conventional methods with histogram. They also show that the proposed algorithm is able to keep the same performance even to music video contents with lots of motion.

### I 서론

오늘날 디지털 멀티미디어 분야의 급속한 발달에 따라, 디지털 콘텐츠 관리의 중요성이 점점 증가되고 있는 추세이다. 이에 따라, 각 콘텐츠의 중요 정보를 추출하는 기법이 요구되어지고 있으며, 특히 동영상에서는 각 구간을 대표하는 대표프레임을 추출을 통하여 동영상의 중요 정보만을 요약하고 있다.

이를 위해서, 연속적인 장면 중에서 장면의 변화가 있는 부분을 추출하여 표시할 수 있는 방법이 요구되어지고 있다.

이러한 장면 변화는 컷(cut)과 같은 급진적인 장면변화와 디졸브(dissolve), 페이드(fade) 등의 점진적인 장면변화로 구분할 수 있다. 가장 대표적인 급진적인 장면변화 추출에는 히스토그램간의 차이를 이용한 방법<sup>[1]</sup>과 화소차<sup>[2]</sup>를 이용한 방법이 있다. 그

\* 광운대학교 전자공학과 디지털 미디어 연구실(shyoon@image.gwu.ac.kr, jsyoo@daisy.kw.ac.kr),

논문번호 040136-0331, 접수일자 2004년 03월 31일

※ 본 연구는 대학 IT 연구센터 육성 지원 사업의 연구 결과로 수행되었음

러나 이러한 방법들은 조명변화와 움직임에 민감하다는 문제를 가지고 있다. 그 이유는 히스토그램을 사용한 방법은 국소적인 특징을 반영하지 않고, 화소차를 이용한 방법은 전체적인 특징을 반영하지 못했기 때문이다. 이러한 문제점을 보완하기 위해서 위의 두 가지 방법을 병합하는 방법<sup>[3]</sup>, 에지를 이용한 방법<sup>[4]</sup>들이 제안되었다. 점진적인 장면변화 추출에서는 일반적으로, 앞에서 언급한 방법들을 이용하여 두 개의 임계치를 사용하는 방법<sup>[5]</sup>과, 점진적인 장면 변화가 전 장면에 비해 점진적으로 차이가 발생한다는 특성을 이용한 방법<sup>[6][7]</sup>이 이용된다.

또 다른 장면검출 방법으로 계산량을 효과적으로 줄이면서, 국부적이고 전체적인 특성을 같이 보여주는 visual rhythm을 이용한 방법이 있다<sup>[8]</sup>. 이를 시각적으로 보면, 장면변화에 대한 특성이 정확히 보이나, 자동화를 위해서는 필터링하여 단순화시킨 후 장면추출을 하는 방법<sup>[10]</sup>과 통계적인 특성을 이용하여<sup>[9]</sup> 장면을 추출하는 방법 등 추가적인 방법이 필요하다.

그러나 visual rhythm을 이용하는 기존의 알고리즘은 필터링을 위해서 복잡한 연산을 요구하거나 전체적인 특징의 통계를 이용하기 때문에 움직임이 많은 동영상에서는 성능이 떨어지는 문제가 있다. 본 논문에서는 이러한 문제점 해결을 위하여 visual rhythm의 방향성을 정의하고, 이를 이용하여 장면변환을 효과적으로 추출하는 알고리즘을 제안한다.

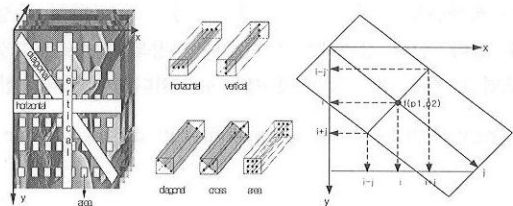
제한한 기법에서는, 움직이는 장면에서도 visual rhythm의 변화를 효과적으로 측정하기 위하여 visual rhythm의 방향성을 정의하였는데 기존의 visual rhythm간의 변화를 단순히 같은 위치의 차 값으로 구하는 것이 아니라, 방향성 정보 또는 움직임 정보를 이용하여 visual rhythm간의 변화를 구하기 위한 것이다.

실험을 통하여, 본 논문에서 제안된 알고리즘이 기존의 히스토그램을 이용한 장면검출 방법보다 약 30% 가량 성능이 향상되는 것을 확인할 수 있고, 움직임이 많은 동영상에서도 거의 비슷한 성능을 유지하는 것을 알 수 있다.

본 논문의 구성은 제 II장에서 visual rhythm을 소개하고, 제 III장에서는 새로이 제안하는 visual rhythm의 방향성을 이용한 장면 검출 기법을 설명하며, 제 IV장에서는 모의실험을 통한 제안된 기법의 성능을 비교한다. 그리고 마지막으로 V장에서 결론을 맺는다.

## II. Visual rhythm

Visual rhythm은 시간 축을 고려한 3차원적 동영상을 다시 2차원적으로 표시하여 장면추출의 효율성을 높이는 방법으로써, 그림 1(a)와 같이 주사 방식에 따라 수평형(horizontal), 수직형(vertical), 대각선형(diagonal), 십자형(cross), 영역형(area) 등 총 5가지의 형태로 구분되어 진다<sup>[8]</sup>. 본 논문에서는 특정 부분만을 표시하지 않고, 일정한 형식을 유지시키는 대각선 방향의 visual rhythm을 사용하였다. 그림 1(a)에서 x는 프레임의 폭, y는 프레임의 높이를 나타내며, t는 시간을 나타낸다. 원점을 좌측 상단으로 한 것은, 일반적으로 영상이 좌측 상단부터 그려지기 때문이다.



(a) Visual rhythm의 여러 가지 형태 (b) 대각선방향의 visual rhythm  
그림 1. Visual rhythm

그림 1(b)에서는 대각선 방향의 visual rhythm의 특성을 보였다. 식 (1)은 M×N의 크기를 갖는 프레임에서, 그림 1(b)와 같이 폭 j를 가지는 대각선 방향의 visual rhythm을 수식으로 표현한 것이다<sup>[10]</sup>. 여기서,  $p_1, p_2$ 는 프레임의 대각선 값만을 가지는 x축, y축 각각의 원소가 된다.

$$d_i = \sum_{p_1=i-j}^{i+j} \sum_{p_2=i+j}^{i-j} f(p_1, p_2) \quad (1)$$

그림 2는 대각선 방향의 visual rhythm을 실제 동영상에서 추출하여 표현한 그림이다. 급진적인 장면인 cut은 일정하게 지속된 특정한 패턴이 갑자기 끊기는 부분에서 나타나고, 점진적인 장면은 dissolve와 같이 일정하게 지속되던 특정한 패턴이 다른 패턴으로 서서히 바뀌어 가는 특징을 가진다는 것을 그림을 통해 알 수 있다.

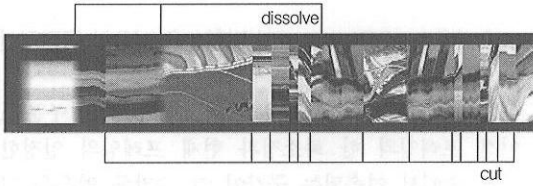


그림 2. 대각선 방향의 visual rhythm

### III. Visual Rhythm의 방향성을 이용한 장면추출

Visual rhythm을 사람이 직접 시각적으로 볼 때는 일정한 패턴의 연속 정도에 따라 장면전환의 판단이 가능하지만, 자동으로 장면 전환 부분을 찾기 위해서는 먼저 이것을 단순화하는 과정이 필요하다. 본 논문에서는 이와 관련하여, 일정한 패턴이 지속되는 성질을 visual rhythm의 방향성(visual rhythm by direction)을 사용하여 정의하고, 정의된 방향성의 특징을 조사하여 장면검출을 수행한다.

#### 1. Visual Rhythm의 방향성 추출

이미 언급한 바와 같이 visual rhythm은 일정한 패턴의 지속여부로서 장면 변환을 판단하는 방법이다. 만약 일정한 패턴들을 정량화할 수 있다면, visual rhythm을 단순화하는 것이 가능하다. 이를 위하여 현재 프레임의 visual rhythm과 다음 프레임의 visual rhythm의 비유사성을 계산하여 visual rhythm의 일정한 패턴을 방향성으로 표시하였다.

사용되어지는 visual rhythm의 집합,  $VR$ 의 하나의 원소  $VR(i_t)$ 는 식 (1)의 대각선 방향의 visual rhythm  $d_i$ 에 시간 정보를 포함시킨,  $d_{i_t}$ 를 의미한다. 식 (2)에 이것을 표현하였다.

$$VR(i_t) = d_{i_t} \quad (2)$$

여기서  $t$ 는 시간 또는 프레임을 의미하고,  $VR(i_t)$ 는 HSI 색상 모델에 근거하여, 식 (3)과 같이 3개의 성분으로 구성된다.

$$VR(i_t) = VR_H(i_t), VR_S(i_t), VR_I(i_t) \quad (3)$$

먼저, visual rhythm의 방향성을 측정하기 위해서는 두 화소점 간의 유사성 또는 비유사성 정도를 계산하기 위한 식이 필요하다. 본 연구에서는 두 화소간의 비유사성( $\epsilon$ )을 식 (4)와 같이 정의하였다.

이는 이전 프레임의 visual rhythm 중 하나의 화소( $i$ )와 현재 프레임 visual rhythm 중 대응하는 화소( $i_{t+1}$ )간의 차를 통하여 비유사성( $\epsilon$ )을 계산하는 것이다. 비유사성( $\epsilon$ )의 값이 작을수록 두 화소는 비슷할 가능성이 높다.

$$\begin{aligned} \epsilon(i_{t-1}, i_t) = & VR_H(i_{t-1}) - VR_H(i_t) \\ & + \alpha \times VR_S(i_{t-1}) - VR_S(i_t) \\ & + (2 - \alpha) \times VR_I(i_{t-1}) - VR_I(i_t) \\ & , 1 < \alpha < 2 \end{aligned} \quad (4)$$

여기서  $i$ 는 그림 1(b)에 표시한 것과 같이, 대각선 방향의 visual rhythm 집합  $VR$ 의 인덱스를 의미하고,  $t$ 는 시간 또는 프레임을 나타낸다. 그러므로  $VR_H(i_t)$ 은 visual rhythm중 현재 시간  $t$  또는  $t$  프레임에서,  $i$  인덱스의 색상(H)값을 나타낸다. 그리고,  $\alpha$  값은 색상(H)과 채도(S)의 가중치를 위한 값이다. RGB 모델에서는 각각의 성분이 장면 검출 시 같은 중요성을 지니지만, HSI 색상 모델에서는 장면 검출에서 각 성분마다 중요도가 틀리므로 가중치를 달리하여 장면검출에 효과를 줄 수 있다.

Visual rhythm의 방향성을 판단하기 위해서는 이전 프레임의 한 화소와 현재 프레임의 여러 화소점과의 비유사성을 그림 3과 같이 계산하여야 한다.

식 (5)에서 정의된 집합  $E_i$ 는 이전 visual rhythm의 한 화소와 현재 프레임 visual rhythm의 일정 범위의 화소들과의 비유사성을 계산하여 놓은 것이다.

$$E_i(i) = \{x | x = \epsilon(i_{t-1}, i_t + n), -k < n < k, k > 0\} \quad (5)$$

여기서,  $k$ 는 구간의 크기를 나타내며,  $(i_t + n)$ 은 현재 프레임의 화소  $i$ 와 비유사도를 검사할 다음 프레임 또는 다음 시간의 visual rhythm 화소 값이다. 집합  $E_i$ 의 원소 중 가장 최소 값이 되는 원소의  $n$ 값이  $n_{min}$ 이 되며, 이것은 그림 3의 우측에 표시되어진 값과 같이 이전 지점에서부터 현재 지점  $VR(i_t)$ 으로의 방향이 된다. 이를 통하여, visual rhythm의 방향성(VRD : visual rhythm by

direction)은 방향성을 나타내는  $n_{min}$  값과 방향성이 없는  $none$  값으로 식 (6)과 같이 표현된다.

$$VRD(i) = \begin{cases} n_{min}, & MIN E_i(i) \leq T_\epsilon \\ none, & MIN E_i(i) > T_\epsilon \end{cases} \quad (6)$$

방향성이 없는  $none$  원소는 집합  $E_i$ 의 원소 중에서, 최소 값을 갖는 원소값( $MIN E_i(i)$ )이 비유사성 임계값  $T_\epsilon$ 을 초과하였을 경우에 발생한다. 이 경우 현재의 화소와 비슷한 화소를 다음 프레임의 화소에서 찾을 수 없다는 것을 의미한다.

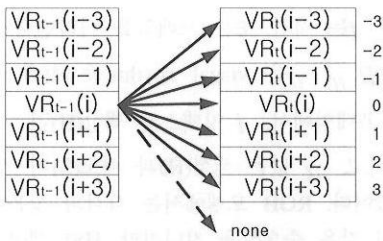
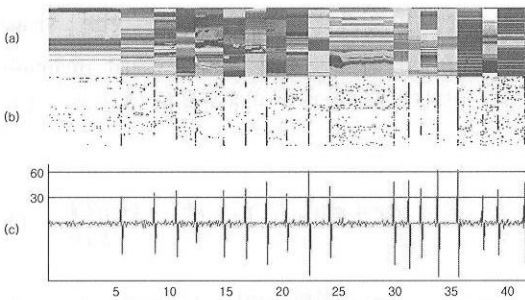


그림 3. Visual rhythm의 방향성

위의 과정을 통해, visual rhythm 방향성중, 방향성이 없는  $none$  원소만을 검은색으로 표시하면, 그림 4(b)같이 시각화가 가능하다.



(a) Visual rhythm (b)VRD (c)  $\Delta CVRD_{none}$

그림 4. VRD의 그래프화

그림 4(a)는 visual rhythm을 표시한 것이고, (b)는 visual rhythm의 방향성중  $none$ 값을 가지는 부분만을 표시한 것이며, (c)는 각 시간에서 방향성이 측정되지 않는  $none$ 의 총합과 그 이전 프레임 총합과의 차분을 그래프로 나타낸 것이다.

## 2. Visual Rhythm의 방향성을 이용한 급진적인 장면추출

그림 4의 (b)에서, 검은색 화소점이 많은 부분은, 이전 프레임의 한 화소점과 현재 프레임의 일정한 구간 중에서 연속되는 구간이 그 수만큼 없음을 의미한다. 다시 말해서 이는 방향성이 측정되지 않는 부분( $none$ )이며, 한 프레임에서 이 부분이 많으면 많을수록 장면전환의 가능성이 높은 곳이다. 따라서 장면전환 검출을 위해서는 방향성이 측정되지 않는 화소의 총 개수 ( $CVRD_{none}$  : count of visual rhythm by direction)의 계산이 필요하다. 이것을 식 (7)에서 정의하였다.

$$CVRD_{none}(t) = \sum_{i=0}^{height} c(i) \quad (7)$$

$$c(i) = \begin{cases} 0, & VRD(i) \neq none \\ 1, & VRD(i) = none \end{cases}$$

위 식은 한 프레임의 visual rhythm에 대하여, 방향성을 표시하는  $VRD(i)$ 값이  $none$  값을 가질 때만  $CVRD_{none}(t)$  값이 증가하는 것을 의미한다. 식 (7)에서  $height$ 는 프레임의 최대 높이 또는 visual rhythm의 최대 길이를 의미한다.

식 (7)을 이용하여, 현재 장면( $f(t)$ )에서의 급진적인 장면전환의 검출은 식 (8)과 같이 현재 visual rhythm에서 방향성이 없는 화소의 총 개수 ( $CVRD_{none}$ )가 임계값  $T_{none}$ 을 초과할 때 검출된다.

$$f(t) = \begin{cases} cut, & CVRD_{none}(t) > T_{none} \\ normal, & CVRD_{none}(t) \leq T_{none} \end{cases} \quad (8)$$

여기서,  $T_{none}$ 은 고정 임계값을 나타낸다.

그러나, 급속한 장면변화의 판단은 식 (8)에 정의한 것과 같이 단순히 방향성이 없는 화소 수 값의 합을 임계값을 사용하여 판단하는 것보다, 변화량을 이용하여 판단하는 것이 유리하다. 이는 장면변화가 주변 장면과의 변화를 보고 판단되기 때문이다. 식 (9)에  $CVRD_{none}(t)$ 의 변화량인  $\Delta CVRD_{none}(t)$ 를 정의하였다.

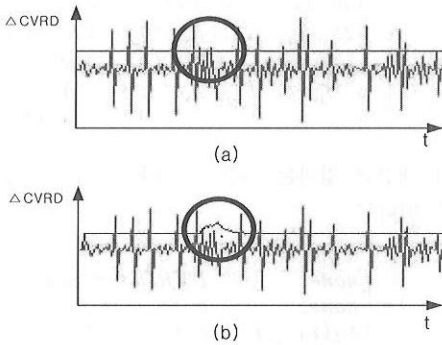
$$\Delta CVRD_{none}(t) = CVRD_{none}(t) - CVRD_{none}(t-1) \quad (9)$$

식 (9)를 식 (8)에 적용하면, 현재의 장면 ( $f(t)$ )은 식 (10)과 같이 판단되어 질 수 있다.

$$f(t) = \begin{cases} cut, & \Delta CVRD_{none}(t) > T'_{none} \\ normal, & \Delta CVRD_{none}(t) \leq T'_{none} \end{cases} \quad (10)$$

식 (10)은 방향성이 없는 화소의 총 개수의 변화량 ( $\Delta CVRD_{none}(t)$ )을 가지고, 장면변환의 판단을 수행하는 것이다. 따라서 새로운 임계값  $T'_{none}$ 을 사용하였다.

지금까지 언급한 장면 변환 검출은 고정임계값  $T_{none}$ ,  $T'_{none}$ 을 사용한 것이었다. 이러한 고정 임계값을 사용하면 조명변화 및 움직임이 많은 곳에서는, 그림 5(a)와 같이 중복된 장면이 추출될 가능성이 높다. 이와 같은 중복 검출을 방지하기 위해서, 본 논문에서는 각 변화값에 따라 임계값을 달리 적용하는 동적 임계값을 사용하였다. 동적 임계값을 적용한 그래프를 그림 5(b)에 보였다.



(a)정적임계값 (b)동적임계값

그림 5.  $\Delta CVRD_{none}(t)$ 와 임계값

본 논문에서는 식 (11)과 같이 주위의 평균값을 사용하여 임계값을 동적으로 정하는 방법을 사용하였다.

$$D_{none}(t) = \begin{cases} \frac{\beta}{t - t_{cut}} \sum_{n=t_{cut}}^t |\Delta CVRD_{none}(n)|, & D_{none}(t) > D_{min} \\ D_{min}, & D_{none}(t) \leq D_{min} \end{cases} \quad (11)$$

여기서  $t_{cut}$ 은 현재 샷의 제일 앞부분에 위치한 컷 지점을 표시한 것이고,  $\beta$ 값은 비례상수이다. 그리고 계산된 동적 임계값  $D_{none}(t)$ 가 최소임계값,  $D_{min}$ 보다 작으면,  $D_{none}(t)$ 값은  $D_{min}$ 으로 대

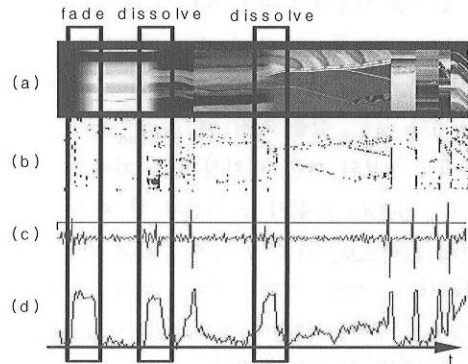
체된다. 이러한 사실을 이용하면, 동적 임계값은 그림 5(b)와 같이 방향성 측정이 잘 되지 않은 구간 즉, 움직임이 많거나 조명변화가 심한 구간에서는 높이고, 그렇지 않은 구간에서는 최소 임계값으로 고정시키는 것이 가능하다. 최종적으로, cut을 판단하기 위해서 식 (11)을 식 (10)에 적용하면 식 (12)를 얻을 수 있다.

$$f(t) = \begin{cases} cut, & \Delta CVRD_{none}(t) > D_{none}(t) \\ normal, & \Delta CVRD_{none}(t) \leq D_{none}(t) \end{cases} \quad (12)$$

따라서, 현재의 장면 ( $f(t)$ )에 대한 급진적인 장면 변환 여부는 식 (12)에 나타난 바와 같이, 인접한 visual rhythm간에 방향성이 측정이 되지 않는, 화소수의 변화량 ( $\Delta CVRD_{none}(t)$ )과 동적 임계값 ( $D_{none}(t)$ )의 비교로 판단되어지게 된다.

### 3. Visual Rhythm의 방향성을 이용한 점진적인 장면 추출

점진적인 장면은 장면의 변화에서 두 개의 장면이 서서히 변화하는 과정을 말한다. 이러한 장면을 검출하기 위해서는 구간에 대한 장면변화 특성을 계산하는 과정이 필요하다. Visual rhythm 방향성중  $\Delta CVRD_{none}(t)$ 값을 이용하면, 그림 6(c)에서 보는 것과 같이 점진적인 장면변화인 경우 장면변화의 특징이 나타나지 않는다. 이것은 점진적인 장면에서, 연속되어지는 장면의 비유사도가 항상 작기 때문이다.



(a) visual rhythm (b) 무 방향성 화소

(c)  $\Delta CVRD(t)$  (d)  $LCVRD(t)$

그림 6. 점진적인 장면변화에서의 visual rhythm

그림 6(a)는 visual rhythm을 나타내고, (b)는 방향성이 측정이 되지 않는 부분을 검은색으로 표시한 것이다. 그림 6(c)는  $\Delta CVRD(t)$  값으로써, 방향성이 측정되지 않는 화소의 총합 변화량을 표시한 것이며, (d)는 이후에 언급되어질  $LCVRD(t)$  값을 그래프로 표시한 것이다.  $LCVRD(t)$ 는 근접거리의 비유사성을 이용한 것이 아니라, 원거리 비유사성을 이용하여 구한 값이다. 그림 6(d)를 보면 점진적인 장면에서, 특이한 패턴의 그래프가 나타나는 것을 알 수 있다.

먼저, 점진적인 장면의 검출에서도 화소간의 비유사도( $\epsilon$ )를 구한다. 그러나, 점진적인 장면에는 구간의 개념이 포함되어 있으므로,  $l$  만큼의 거리가 떨어진 두 화소간의 원거리 비유사도,  $\epsilon_L$ 를 구하여야 한다. 식 (4)를 이용하면, 식 (13)과 같이  $\epsilon_L$ 를 정의할 수 있다.

$$\begin{aligned} \epsilon_L(i_{t-1}, i_t) = & VR_H(i_{t-1}) - VR_H(i_t) \\ & + \alpha \times VR_S(i_{t-1}) - VR_S(i_t) \\ & + (2-\alpha) \times VR_I(i_{t-1}) - VR_I(i_t) \end{aligned} \quad , 1 < \alpha < 2 \quad (13)$$

식 (4) 다른 점은  $i_{t-1}$  대신에  $i_{t-l}$ 을 사용한 것이다. 그러나, 만약 점진적인 장면 구간 내에서 움직임이 발생한다면, 이 식은 화소간의 비유사성을 정확하게 나타내지 못할 것이다. 그 이유는 식 (13)에서, 두 화소간의 비유사성을 계산할 때, 같은 위치의  $i$  값만을 사용하기 때문이다. 식 (6)에서 계산된  $VRD(i_t)$  값, 즉 visual rhythm 방향성을 이용하면 이러한 문제점을 해결할 수 있다. 이것은  $VRD(i_t)$ 가  $n_{min}$  값을 가질 때,  $n_{min}$  값이 화소간의 움직임 정보와 같은, 방향성 값을 가지고 있기 때문이다.  $VRD(i_t)$  값이  $t$  시간의 한 화소에 대한 방향성 값이므로, 이를  $l$  시간 구역에 적용시키기 위해서는 그림 7과 같이 누적 방향성,  $LVRD(t)$ 을 구하는 과정이 필요하다. 식 (6)을 이용하여 정의하면 식 (14)와 같다.

$$LVRD(t) = \sum_{m=t-l}^{t-1} VRD(m) \quad (14)$$

식(14)에서  $VRD(m)$  값이  $none$ 이면, 두 화소간

의 연속성이 이어지지 않으므로,  $LVRD(t)$  값은  $none$ 이 된다. 그림 7에서 보면,  $LVRD(t)$ 는 진한 음영색이 있는 두 부분의 방향성이 된다.

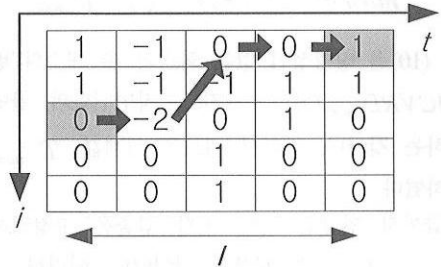


그림 7. visual rhythm 의 누적 방향성

$LVRD(t)$  값이 구해지면, 식 (13)과 (14)를 이용하여 원거리 비유사도,  $\epsilon_L$ 을 식 (15)와 같이 구할 수 있다.

$$\begin{aligned} \epsilon_L(i_{t-1}, i_t) = & VR_H(i_{t-l+LVRD(t)}) - VR_H(i_t) \\ & + \alpha \times VR_S(i_{t-l+LVRD(t)}) - VR_S(i_t) \\ & + (2-\alpha) \times VR_I(i_{t-l+LVRD(t)}) - VR_I(i_t) \end{aligned} \quad , 1 < \alpha < 2 \quad (15)$$

식 (15)가 계산된 결과는 다음과 같이 식 (16)으로 나타낼 수 있다.

$$\epsilon_L(i_{t-1}, i_t) = \begin{cases} none, & LVRD(t) = none \\ none, & \epsilon_L(i_{t-1}, i_t) > T_\epsilon \\ \epsilon_L(i_{t-1}, i_t), & \epsilon_L(i_{t-1}, i_t) \leq T_\epsilon \end{cases} \quad (16)$$

식 (16)에서 원거리 비유사도 값이  $none$ 이 되어질 경우는, 첫 번째 항목에 있는  $LVRD(t)$  값이  $none$ 이 되어질 경우나 원거리 비유사도 값이 임계값( $T_\epsilon$ )을 초과하였을 경우이다.

점진적인 장면변화 검출을 위해서는 급진적인 장면변환에서 방향성이 측정되지 않는  $none$  값의 총합을 사용했던 것처럼, 현재 프레임내의 원거리 비유사도( $\epsilon_L$ ) 값이  $none$ 인 화소수의 총합( $LCVRD_{none}(t)$ )을 사용한다. 이를 식으로 표현하면 식 (17)과 같다.

$$LCVRD_{none}(t) = \sum_{i=0}^{height} c(i) \quad (17)$$

$$c(i) = \begin{cases} 0, & \varepsilon_L(i_t - i_{t-d}) \neq none \\ 1, & \varepsilon_L(i_t - i_{t-d}) = none \end{cases}$$

식 (17)은 원거리 비유시도,  $\varepsilon_L$ 를 사용,  $l$  만큼의 거리가 떨어진 두 최소값의 방향성을 세심하여 병형성이 측정되지 않는, 즉 *none* 값을 가진 화소의 총 수를 계산하는 식이다 이 값이 크면 클수록 비교되는 프레임간의 인위성이 없음을 의미한다

그림 6(d)는 식 (17)의  $LCVRD_{none}(t)$  값을 그래프로 표시한 것이다 잠정적인 장면변화는 변환 구간에 따라 비유사성 크기가 일정하게 유지되어지는 특징을 보인다 그리고 구간 내에서 급진적으로 변화는 특징이 없으므로, 급진적인 장면검출을 위해 사용하였던 식 (9)의  $\Delta CVRD_{none}$  값은 적은 값으로 유지되어진다 그러므로,  $LCVRD_{none}(t)$  값이 일정하게 유지되어지는 구간이 임계구간을 초과하게 되면, 이러한 장면을 잠정적인 장면변화 구간으로 판단하게 된다 먼저  $LCVRD_{none}(t)$  값이 일정하게 유지되어지는 구간의 거리  $dis(t)$ 를 계산하기 위해서 식 (18)을 사용하였다

$$dis(t) = \begin{cases} dis(t-1)+1, & LCVRD(t)_{none} > T_{none} \\ 0, & LCVRD(t)_{none} \leq T_{none} \end{cases} \quad (18)$$

식 (18)에서,  $T_{none}$ 은  $LCVRD_{none}(t)$  값을 위한 고정 임계값이며,  $LCVRD_{none}(t)$  값이  $T_{none}$ 을 초과할 때  $dis(t)$  값이 증가한다 그리고,  $T_{none}$ 을 초과하지 못할 경우에는  $dis(t)$  값을 초기화한다 그러므로  $dis(t)$  값은  $LCVRD_{none}(t)$  값이 임계값  $T_{none}$ 을 초과하는 구간의 거리를 의미하게 된다

마지막으로, 잠정적인 구간의 장면변환 판관을 위해서 식 (18)의  $dis(t)$  값을 사용하면 식 (19)로 표현이 가능하다

$$f(t) = \begin{cases} dissolve, & dis(t) > T_l \text{ and } f(t_c) \neq cut, \\ & t - dis(t) \leq t_c < t \\ normal, & dis(t) \leq T_l \text{ and } f(t_c) \neq cut, \\ & t - dis(t) \leq t_c < t \end{cases} \quad (19)$$

식 (19)는  $LCVRD_{none}(t)$  값이 고정임계값,  $T_{none}$ 을 초과하는 구간의 길이,  $dis(t)$ 가 임계구간  $T_l$ 을 초과할 때,  $dis(t)$  구간을 잠정적인 장면변화의 구간이라 판단하는 식이다 그리고, 식(19)에서는 컷(cut)을 잠정적인 변환보다 우선으로 하여 잠정적인 장면변화 구간 내에서 컷이 발생 시 잠정적인 장면변화로 판단하지 않도록 하였다

#### IV 실험결과 및 분석

본 실험은 windows2000 운영체제에서, Microsoft에서 제공한 DirectShow 라이브러리를 사용하여 수행되었다 사용한 컴퓨터는 펜티엄4 1.8GHZ이며, 사용된 테스트 동영상은 표 1에 보인 것과 같다

표 1 실험에 사용한 테스트 동영상

장르 \ 정보	size	length(sec)	frames
광고	352×240	1.01	1833
뉴스	352×240	1.29	2670
뮤직비디오	352×240	4.00	7210

뉴스는 대체적으로 컷 구분이 뚜렷하고, 샷의 길이가 길며 움직임이 적은 반면, 광고와 뮤직비디오는 움직임이 많고, 샷의 길이가 짧은 양상을 보였니 뮤직비디오의 경우는 dissolve 장면과 fade 장면 같은 잠정적인 구간이 포함되어 있다

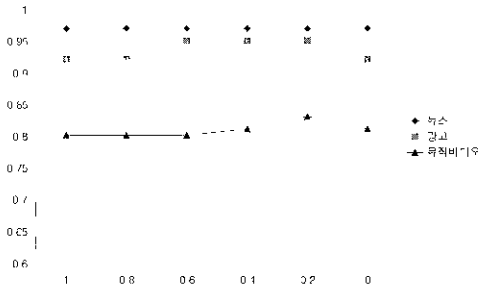
실험결과에 쓰여지는 recall과 precision 값은 일반적으로 컷의 정확성을 판단하는 평가 기준이 된다 따라서 이들을 식 (20)와 식 (21)에 나타내었다

$$Recall = \frac{N_c}{N_c + N_m} \quad (20)$$

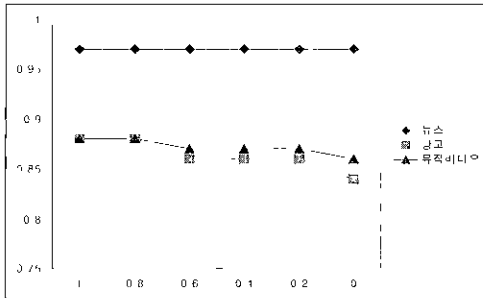
$$Precision = \frac{N_c}{N_c + N_f} \quad (21)$$

여기서  $N_c$ 는 정확히 검출된 컷 수를 나타내고,  $N_m$ 은 갖지 못한 컷 수,  $N_f$ 는 잘못 찾은 컷 수를 의미한다 따라서 recall은 수치가 높을수록 놓친 컷이 적다는 것을 뜻하고, precision은 수치가 높을수록 잘못 찾은 컷이 적다는 것을 의미한다 보통

recall 수치가 클수록 precision 수치가 낮아진다 이는 낮은 임계값이니, 적은 비례상수를 사용하여서 컷 검색을 지세히 하기 때문이다 Recall과 precision의 적절한 타점점을 찾기 위해 관련된 임계값과 비례상수를 조절하여, 그 수치를 확인하기 위해서 다음 실험을 행하였다



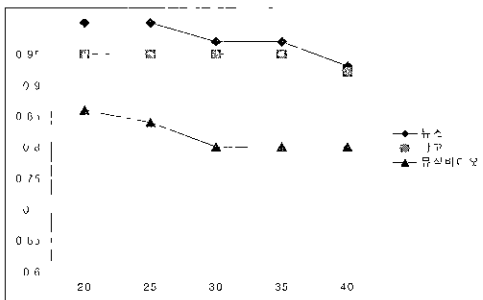
(a) α값에 따른 각 장르의 recall 값



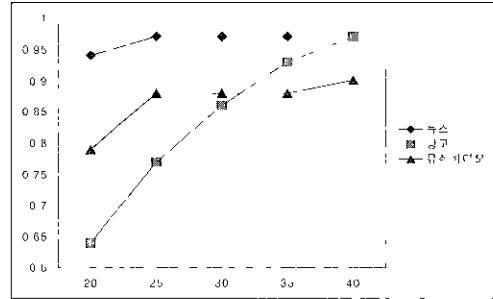
(b) α값에 따른 각 장르의 precision 값

그림 8 α값에 따른 각 장르별 실험치

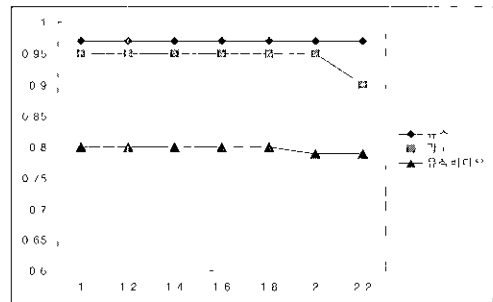
처음 실험은 식 (4)를 기준으로, α값을 변화시키며 각 장르별 recall과 precision을 조사하였고 그 결과들을 그림 8에 나타내었다 그림 8(a)는 recall 값을 (b)에는 precision 값을 보였다



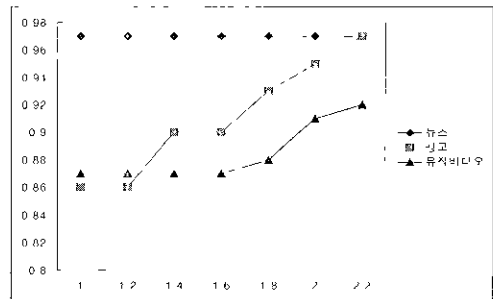
(a) 비유사도의 임계값에 따른 각 장르의 recall 값



(b) 비유사도의 임계값 따른 각 장르의 precision 값  
그림 9 비유사도 임계값 따른 각 장르별 실험치



(a) β값에 따른 각 장르의 recall 값



(b) β값에 따른 각 장르의 precision 값

그림 10 β값에 따른 각 장르별 실험치

그림 9의 (a)와 (b)는 식 (6)의 비유사도 임계값  $T_\epsilon$ 의 변화에 따른 결과로써,  $T_\epsilon$  수치가 높을수록 recall 값은 점점 작아지고 precision 값은 커지는 것을 알 수 있다 이는  $T_\epsilon$  수치가 높을수록 방향성이 없는 성분의 수가 적어지기 때문이다 그리고 그림 10(a)와 (b)는 식 (11)의 β값과 관련된 실험 결과이다 뉴스보다는 움직임이 비교적 많은 광고와 뮤직비디오에서 값이 크게 변화하는 것을 알 수 있다

각 상수 값의 선택은 recall 과 precision 중 어느



것에 중요성을 두느냐에 따라 결정되나 본 실험에 시는 recall 값에 중요성을 두고  $\alpha$ ,  $\beta$ ,  $T_c$  의 값을 각각 0.2, 1.8, 25로 설정하였다 본 논문에서 세로이 제안된 알고리즘을 기존의 히스토그램을 이용한 컷 검출 방법<sup>[3]</sup>과 visual rhythm의 통계적 특징을 이용한 방법<sup>[9]</sup>과 비교하여, 그 결과를 표 2에 보았다 표 2에서 보는 것과 같이 실험을 한 모든 테스트 동영상에 대하여 제안된 알고리즘을 사용하는 경우 recall과 precision이 히스토그램을 사용한 검출 방법보다 약 30% 정도 개선된다는 것을 확인할 수 있었다 뮤지비디오의 경우는 17개의 디폴트 화면이 존재한다

표 2 히스토그램을 사용한 컷 검출과 visual rhythm의 방향성을 이용한 검출의 비교

장르	뉴스		광고		뮤지비디오	
	recall	precision	recall	precision	recall	precision
히스토그램을 사용한 검출	1.0	0.78	0.72	0.76	0.72	0.59
visual rhythm의 통계적 특징을 이용한 검출	0.64	0.48	0.69	0.66	0.90	0.43
visual rhythm의 방향성을 이용한 검출	1.0	0.97	0.95	0.90	0.85 (0.88)	0.87 (0.75)

(0)는 디폴트 수치임

### V 결론

본 논문에서는 계산량이 적고 장면진환을 효과적으로 판단할 수 있는 visual rhythm을 이용하여 장면진환을 자동으로 검출할 수 있는 알고리즘을 제안하였다 제안된 알고리즘에서는 먼저 visual rhythm를 단순화하기 위해서 visual rhythm의 방향성을 정의하였고 움직임이 많고, 변화가 심한 동영상에서 장면 진환을 효과적으로 검출하기 위해 동적 임계값을 사용하여 중복검출 가능성을 줄일 수 있었다

제안된 알고리즘은 점진적인 장면이 많은 뮤지비디오를 제외하고는 recall과 precision 모두 90% 이상 검출할 수 있는 성능을 보였고, 원거리 비유사도의 개산을 통하여 점진적인 장면비치의 추출도 가능하도록 하였다

### 참고 문헌

- [1] Yoshinobu Tonotmura, "Video handling based on structured information for hyper media systems," ACM Proc of international conference on Multimedia Information Systems, 91, pp 333-344, 1991
- [2] Kiyotaka Otsuji, Yoshinobu Tonomuta, Yuji Ohba, "Video Browsing using brightness data," Proc of SPIE VCIP'91, pp 980-989, 1991
- [3] 김승철, 박성환, "중간값 필터의 신경망 회로를 사용한 자동 컷 검출 알고리즘", 제 12회 영상 지리 및 이해에 관한 워크샵 발표 논문집, pp 485-488 2000 1
- [4] W J Heng, K, N Ngam, "The implementation of object based Shot boundary detection using edge tracing and tracking," IEEE International Conference on Image Processing(ICIP '99), Japan, 1999
- [5] Thomas S Huang, Yong Rui, Trausti Kristjansson, Milind Naphande and Yueting Zhuang, "Video analysis and representation", Proposal of Univerisity of Illinois at urbana-chmpaign, 1998
- [6] 김우생, 엄상주, "동적 임계값과 컷 프레임 차를 이용한 점진적 선택 검출 기법", 한국정보처리 학 회, 정보처리학회논문지, 2002
- [7] 이광국, 김형준, 김희윤, "미디어 영상에서 점진적 장면진환 검출", 내한전자공학회 02 하계종합학술대회 논문집(4), pp 149-152, 2002
- [8] H Kim, J Lee, J Yang, S Sull, W Kim and S M Song, "Visual rhythm and shot verification", Multimedia Tools and Applications, 2001
- [9] Min Gyo Chung, Hycokman Kim, S Moon-Ho Song, "A scene boundary detection method", International Conference on Image Processing (ICIP 2000), Volume 3, 10-13 Sept 2000
- [10] Chong-Wah Ngo, Ting-Chuen Pong, Chin, RT, "Video partitioning by temporal slice coherency", IEEE Transactions on Circuits and Systems for Video Technology, Volume 11, Issue 8, Aug 2001

윤 상 호(Sang-ho Yoon)

준회원



2001년 2월 : 서울산업대학교  
전자공학과 졸업

2003년 3월 ~ 현재 :  
광운대학교 전자공학과 석사과정  
<관심분야> MPEG-7, 데이터  
방송, 영상압축, 영상통신

유 지 상(Ji-sang Yoo)

정회원



1985년 2월 : 서울대학교  
전자공학과 졸업(공학사)

1987년 2월 : 서울대학교  
대학원 전자공학과 졸업  
(공학석사)

1993년 5월 : Purdue 대학교  
전기공학과 졸업(Ph.D.)

1993년 9월 ~ 1994년 8월 : 현대전자산업(주) 산  
전연구소 선임연구원

1994년 9월 ~ 1997년 8월 : 한림대학교  
전자공학과 조교수

1997년 9월 ~ 2001년 8월 : 광운대학교  
전자공학과 조교수

2001년 9월 ~ 현재 : 광운대학교 전자공학과 부교수  
<관심분야> 웨이블릿 기반 영상처리, 영상압축,  
영상인식, 비선형 신호처리