

대역별 웨이블릿 계수특성을 이용한 장면전환점 검출기법

문영호*, 노정진*, 정희원 유지상*

Cut Detection Algorithm Using the Characteristics Of Wavelet Coefficients in Each Subband

Young-ho Moon*, Jung-jin Noh*, Ji-sang Yoo* *Regular Members*

요약

본 논문에서는 동영상의 장면전환점 중 급진적인 장면전환점인 컷(cut)과 점진적인 장면전환점인 페이드(fade)와 디졸브(dissolve) 구간을 웨이블릿 변환영역에서 검출하는 알고리즘을 제안한다. 웨이블릿 변환을 이용한 기존의 연구들은 공간영역과 변환영역 각각의 특징을 이용하여 장면전환점을 검출한다. 그러나 본 논문은 입력된 컬러 영상을 먼저 YUV 공간으로 변환하고, Y 성분에 대해 리프팅기법을 적용하여 2 레벨 웨이블릿 변환 후, 변환영역에서 공간영역의 특징이 유지되는 저주파 부대역을 히스토그램 비교하고, 나머지 고주파 부대역에서 추출된 에지 정보를 전체(global), 부분(semi-global), 국부(local) 영역으로 정의하여 웨이블릿 에지 히스토그램 비교를 한다. 모의실험 결과 기존의 방법보다 recall에서는 약 17%, precision에서는 약 18%의 성능향상을 보였으며 점진적인 장면 전환점인 페이드와 디졸브 구간 검출에도 좋은 성능을 나타내었다.

Key Words Cut, Fade, Dissolve, Lifting, Wavelet Edge Histogram

ABSTRACT

In this paper, an algorithm using wavelet transform for detecting a cut that is a radical scene transition point, and fade and dissolve that are gradual scene transition points is proposed. The conventional methods using wavelet transform for this purpose is using features in both spatial and frequency domain. But in the proposed algorithm, the color space of an input image is converted to YUV and then luminance component Y is transformed in frequency domain using 2-level lifting. Then, the histogram of only low frequency subband that may contain some spatial domain features is compared with the previous one. Edges obtained from other higher bands can be divided into global, semi-global and local regions and the histogram of each edge region is compared. The experimental results show the performance improvement of about 17% in recall and 18% in precision and also show a good performance in fade and dissolve detection.

I. 서론

최근 초고속 통신망 및 대용량의 저장매체 등의 발달로 멀티미디어 정보의 전송, 저장, 관리, 검색,

색인 기술이 중요한 연구분야로 떠오르고 있다^{[1][2]}. 특히 비디오 정보는 동영상과 오디오, 텍스트 정보를 포함하는 데이터로서 중요성이 점차 증가되고 있다. 프레임(frame), 샷(shot), 장면(scene), 시퀀스

* 광운대학교 전자공학과 디지털 미디어 연구실(moonsky707@image.gwu.ac.kr, jungjin@image.gwu.ac.kr, jsyoo@daisy.kw.ac.kr),
논문번호 #KICS2004-05-022, 접수일자 2004년 05월 31일
*본 연구는 교내학술연구 지원 결과로 수행되었음

(sequence)의 구조로 된 동영상을 효과적으로 관리, 검색하기 위해 유사데이터를 찾아서 구분하는 작업이 필요하며, 공간정보와 시간정보를 분석하여 두 정보의 관계성을 분석해야 한다^[5].

구분된 비디오 데이터는 주석기반 검색방법(annotation-based retrieval technique)과 내용기반 검색방법(content-based retrieval technique)으로 분류가 되어 활용이 된다 주석기반 검색방법은 사람이 비디오 정보에 대한 텍스트를 입력하는 방법으로 많은 시간이 필요하다는 단점이 있다 내용기반 검색방법은 이러한 단점을 보완하기 위해서 시간적인 내용을 자동으로 추출하여 검색을 수행하는 방법으로 활용이 된다. 내용기반 검색을 위해서는 동영상의 구조 중에 카메라의 동작이 끝나고 다른 카메라로 넘어가기까지의 프레임 집합의 경계인 샷을 찾아 장면전환점을 추출하는 기법이 선행되어야 한다^[4]

기존에 연구되었던 장면전환 검출기법으로는 히스토그램 기반의 방법이 가장 대표적이다 Swenberg^[5]와 Nagasaka^[6]는 히스토그램의 차이와 히스토그램의 통계적 특징을 이용하여 장면전환점 검출을 수행하였다 또한 Zhang^[7]은 픽셀 밝기차이에 의한 장면전환점 검출방법을 제안하였다 Zabih, Miller, Mai는 복원영상의 에지픽셀의 분포를 이용하여 점진적인 장면전환점을 검출하였다^[8]. 동적임계값을 이용한 방법이나^[9] MPEG-7 디스크립터를 이용한 방법^[10]등도 장면전환 검출 기법에 많이 사용되고 있다

본 논문에서는 리프팅 기법을 사용한 웨이블릿 변환에 기초를 두어 급진적인 장면전환점과 점진적인 장면전환점을 검출한다 비연속성을 지닌 급진적인 장면전환점인 컷(cut)은 RGB 컬러영상을 YUV 성분으로 변환하여 Y성분만을 2 레벨 웨이블릿 변환 영역에서 최저주파 부대역 영역을 히스토그램 비교를 통해 검출한다. 연속적인 성질을 가진 점진적인 장면전환점, 즉 페이드(fade)는 검은색으로 표현되는 블랭크(blank) 프레임을 정의하여, 블랭크 프레임 앞, 뒤의 40개 프레임의 집합을 이용해서 검출을 한다. 디졸브(dissolve)는 웨이블릿 변환 영역에서의 고주파 부대역들의 계수들의 에지 히스토그램을 정의하여 이의 평균과 평균의 변화량을 이용하여 검출을 한다

본 논문의 구성은 다음과 같다. 2절에서는 기존의 연구와 웨이블릿 변환을 기술하고, 3절에서는 제안하는 장면전환점 검출 알고리즘을 설명한다 4절에서는 실험을 통하여 기존의 장면전환 검출방법과 비

교를 하여 효율성을 평가한다. 마지막으로 5절에서는 제안하는 방법의 결론과 향후 연구 방향을 기술한다.

II 장면전환점 검출 기법과 리프팅 기법

1 히스토그램 비교방법

히스토그램 비교방법은 동일한 샷으로 분류되는 프레임들의 색상분포는 유사하다는 성질을 이용하여 컷을 검출한다. 이 방법은 각 프레임의 컬러 히스토그램 차이를 계산한 후, 식 (1)과 같이 그 차이 값이 임계값을 넘으면 컷으로 판단한다.

$$\sum_{j=1}^N |H_i(j) - H_{i+1}(j)| > T \quad (1)$$

여기서 $H_i(j)$ 는 i 번째 프레임에서 j 레벨의 히스토그램 값을 나타내고, N 은 히스토그램의 레벨 수를 나타낸다 인접한 프레임들의 차이를 정해놓은 임계값 T 와 비교함으로써 컷을 검출한다.

히스토그램 비교방법은 구현이 쉽고 연산이 적어 효율적이며 움직임이 있는 장면에서도 전체적인 명암도는 크게 변하지 않으므로 물체의 움직임이나 카메라 이동에 덜 민감하다. 그러나 플래시 효과처럼 밝기가 순간적으로 변화하는 경우에는 대처하기 곤란하고, 다른 샷이지만 색상분포가 유사할 때에는 오검출이 발생하는 단점이 있다.

2 화소단위 비교방법

화소단위 비교 방법은 동일한 샷 내에서는 화소값의 변화가 적다는 성질을 이용하는 것으로, 식 (2)와 같이 인접하는 프레임의 대응하는 화소값의 차가 임계값 t 를 초과하는 경우 1로 된다

$$DP_i(x, y) = 1, \text{ if } |F_i(x, y) - F_{i+1}(x, y)| > t \quad (2)$$

$$0, \text{ otherwise}$$

여기서 $F_i(x, y)$ 는 i 번째 프레임에서 (x, y) 의 화소 값을 의미한다 여기서 DP는 화소값의 차이를 의미한다 전체 화소에 대해 1로 된 화소의 비율이 특정 임계값 T_p 를 넘으면, 컷으로 간주한다 이 과정을 식 (3)에 정의하였다.

$$\frac{\sum_{x,y=1}^{X,Y} DP_i(x, y)}{X \times Y} \times 100 > T_p \quad (3)$$

여기서 X, Y는 이미지의 최대 높이와 넓이를 표시한다. 이 방법은 구현이 간단한 반면, 카메라 연산과 객체 움직임에 민감하여, 샷의 경계가 아님에도 불구하고 경계로 잘못 판정될 가능성이 크다.

3. 리프팅 기법

웨이블릿 변환은 인간의 지각 관점을 가장 잘 표현하는 변환으로써, 시간 및 주파수에 대하여 국부성을 가진다. 웨이블릿 변환을 적용한 영상의 다해상도 표현에서는 주파수 공간의 계수들을 분석하여 처리를 하며, 이것은 영상의 시공간적인 중복성을 내포하고 있는 화소를 효과적으로 제거함으로써 적은 양의 데이터 정보로도 영상의 특징을 전달할 수가 있어 다양한 응용이 가능한 유용한 변환이다^[11]. 그림 1에 2차원 영상에 대한 웨이블릿 변환을 나타내었다.

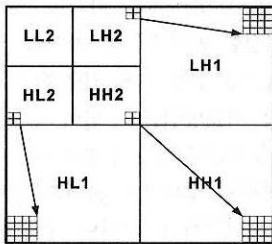


그림 1. 2차원 영상에 대한 2레벨 웨이블릿 변환

본 논문에서는 RGB 컬러 영상에 대해 YUV 변환을 한 후에 Y 성분에만 2 레벨 웨이블릿 변환을 적용하였으며, 이때 리프팅(lifting)이라 불리는 고속 웨이블릿 변환 방법을 사용한다. 리프팅은 기존의 필터뱅크(filter bank)방식이 아닌 사다리(ladder) 구조로 변환을 수행하며, Mallat과 Meyer에 의해 소개가 되었다^[12].

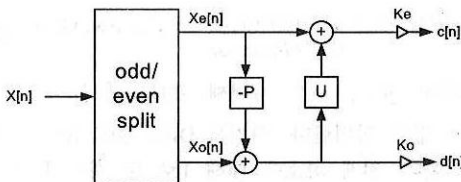


그림 2. 리프팅 과정 : 분할, 예측, 갱신

리프팅은 원래 복잡한 기하학적 배열들과(complex geometries), 불규칙한 샘플링에 대한 웨이블릿 변환을 조정하기 위해 발전되었다. 리프팅의 주된 특징은 주파수 영역 구조와 다르게 전체 공간영역에

대한 해석을 제공하는 것이다. 이는 크게 분할(split), 예측(predict), 갱신(update)의 3가지 과정으로 수행이 된다. 그림 2에 리프팅의 정방향 과정을 나타내었다. 그림 3의 (a)와 (b)는 각각 원 영상과 리프팅 과정을 통해서 얻어낸 영상을 나타낸다.



그림 3. (a) Lena 원영상(256×256) (b) (a) 영상에 대한 N레벨 리프팅 결과

III. 제안하는 장면 전환 검출기법

1. 컷 검출(Cut Detection)

본 논문에서는 급진적인 장면전환점인 컷을 식 (1)에서 정의한 히스토그램 비교 방법을 이용하여 검출한다. 하지만 기존의 컬러 히스토그램 사용법과는 다르게, 입력되는 RGB 컬러영상에 대하여 YUV 컬러로 색공간을 변환하고, 여기서 얻은 Y 성분만을 2 레벨 리프팅 변환하여 최저주파수 대역인 LL 대역만을 가지고 식 (1)을 적용하게 된다. 이 방법은 기존의 컬러 히스토그램 비교 방법보다 비교하는 인자가 1/3로 줄어들어 연산속도가 증가하게 된다. 그러나 히스토그램 비교법의 단점인 명암 변환에 민감한 성질은 여전히 유지하게 된다. 식 (4)에서 Y_{diff} 라는 변수를 정의하는데 이것은 현재와 과거 프레임간의 Y성분만의 히스토그램 차이 값을 나타낸다. Y_{diff} 값이 고정 임계값을 넘었을 경우 컷으로 판별을 하게 된다.

$$Y_{diff} = \sum_{j=1}^N |H_i(j) - H_{i+1}(j)| \quad (4)$$

2. 페이드 검출(Fade Detection)

그림 4는 페이드 검출기법의 구성도를 보여주고 있다. 그림에서 보는 것과 같이 페이드는 블랭크(blank) 프레임이라는 검은색 프레임을 정의하여 검출하게 된다. 본 논문에서는 입력되는 영상을

256×256 크기로 스케일링한 후 리프팅을 수행하게 되고, LL 대역만을 가지고 히스토그램을 구하기 때문에 64×64 크기의 LL 대역의 모든 히스토그램 레벨의 총 합은 4096이 된다. 블랭크 프레임은 히스토그램의 레벨 중에 0부터 9까지의 10개의 어두운 성분 히스토그램 값의 총합이 임계값 T_{blank} 이상일 경우에 검출된다. 식 (5)에서 블랭크 프레임을 정의하였다.

$$f(i) = \begin{cases} BLANK, & \sum_{j=0}^9 |H_i(j)| > T_{blank} \\ NORMAL, & otherwise \end{cases} \quad (5)$$

여기서 $f(i)$ 는 현재프레임의 결과를 나타내고, $H_i(j)$ 는 i 번째 프레임에서 j 레벨의 히스토그램 값을 나타내고, j 는 0부터 9까지의 값을 갖는다. T_{blank} 는 3800으로 하여 4096개의 최소 값 중에 92.7%가 0에서 9사이에 있을 경우에 블랭크 프레임으로 정의하게 된다. 또한 블랭크 프레임이 검출되면 연속되는 블랭크 프레임이 검출되지 않도록 앞, 뒤 20프레임 내에서는 블랭크 프레임이 없다고 가정한다.

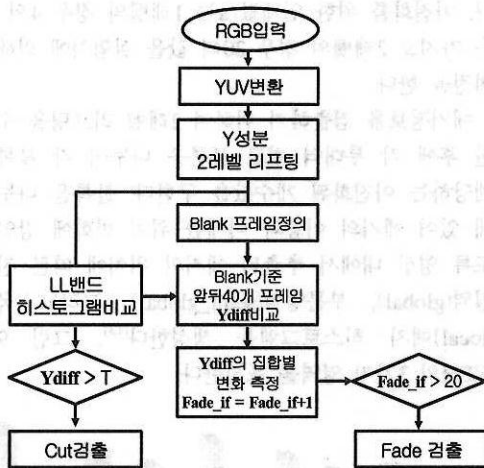


그림 4. 페이드 검출의 구성도

이렇게 블랭크 프레임이 정의가 되면 식 (4)의 컷 검출에서 사용하였던 차이값 Y_{diff} 를 이용하여 페이드의 특성을 이용한 검출을 하게 된다. 그림 5는 페이드 구간을 나타낸다.

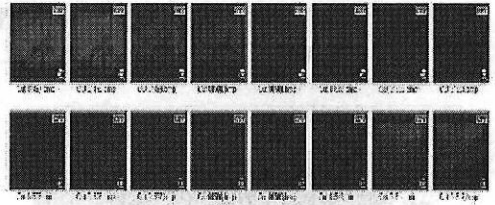


그림 5. 페이드 구간의 예

여기서 좌측 맨 위에서 우측 아래로 9번째 프레임이 블랭크 프레임으로 검출되면 뒤의 40프레임과 앞의 40프레임을 각각의 집합으로 분류하여 Y_{diff} 값의 변화되는 부호를 측정하게 된다.

$$Set A = \{x | (n-40) < x < n, n = blank\}$$

$$Fade_{if} = \begin{cases} Fade_{if} + 1, & Set A(n) - Set A(n+1) < 0 \\ Fade_{if}, & otherwise \end{cases} \quad (6)$$

$$Set B = \{x | n < x < (n+40), n = blank\}$$

$$Fade_{if} = \begin{cases} Fade_{if} + 1, & Set B(n) - Set B(n+1) > 0 \\ Fade_{if}, & otherwise \end{cases} \quad (7)$$

식 (6)과 (7)에서 $set A$ 는 블랭크 프레임 기준으로 과거 프레임집합이고, $set B$ 는 미래 프레임 집합이다. x 는 현재프레임을 나타내고 n 은 블랭크 프레임을 나타낸다. 또한 $Fade_{if}$ 라는 변수를 만들어 $Set A$ 에서 현재 프레임과 이전프레임의 히스토그램 차이값이 음수인 경우가 20개 이상이면 페이드로 검출하게 된다. 또한 $Set B$ 에서도 현재프레임과 이전프레임의 히스토그램 차이값이 양수인 경우가 20개 이상이면 페이드로 검출하게 된다. 식 (8)은 현재 프레임이 페이드 구간으로 검출되는 조건을 나타내며 이때 20이라는 임계값은 실험에 의한 것으로 조건을 만족하지 않을 때는 일반적인 프레임으로 결정이 된다.

$$f(i) = \begin{cases} Fade, & Fade_{if} \geq 20 \\ Normal, & Fade_{if} < 20 \end{cases} \quad (8)$$

페이드를 검출하는 부분에서 미래 40개의 프레임 집합을 구하기 위해 지연이 발생하며 컷과 디졸브를 검출하는 데는 영향을 미치지 않는다. 그림 8은 페이드 구간의 Y_{diff} 값의 변화를 나타낸다. 도표의 가

로축의 프레임 번호 41번이 그림 6의 9번째 블랙 프레임임을 나타내고, 이를 기준으로 앞, 뒤 두 집합의 변화를 누적하게 된다.

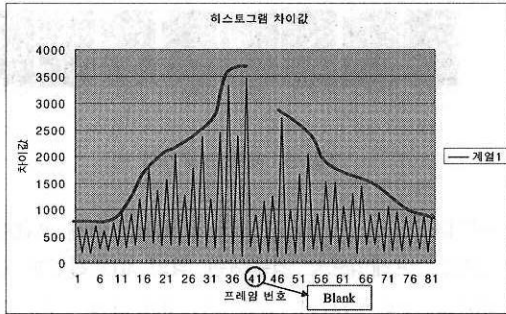


그림 6. 페이드 구간의 Ydiff값의 변화

3. 디졸브 검출(Dissolve Detection)

본 논문에서는 디졸브를 검출하는데 있어 에지정보를 이용하기 위하여 웨이블릿 에지 히스토그램이라는 지표를 정의하고 WEH(Wavelet Edge Histogram)의 평균과 평균의 변화량을 이용하여 디졸브를 검출하게 된다. 그림 7는 디졸브를 검출하는 구성도를 나타낸다.

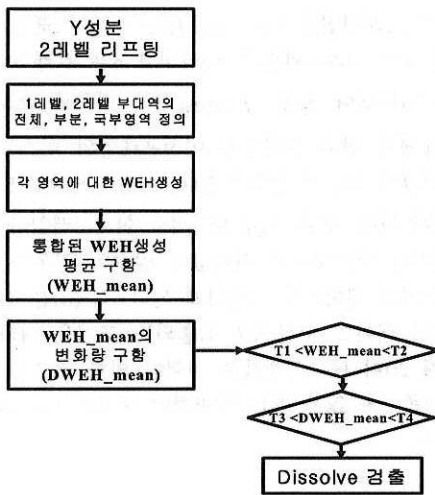
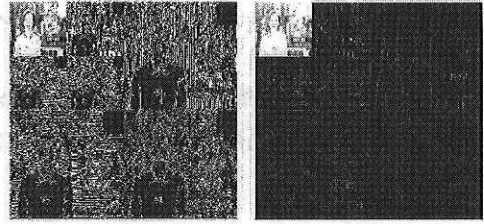


그림 7. 디졸브 검출의 구성도

2레벨 리프팅된 영상은 많은 양의 에지 정보를 가지고 있으므로 우선 0과 255를 기준으로 보정 (saturation)이 되고, LL대역을 제외한 1레벨의 3개의 부대역들(LH1, HL1, HH1)과 2레벨의 3개의 부대역들(LH2, HL2, HH2)에 대해 실험에 의한 임계

값을 설정하여 이진화를 수행한 후에 WEH를 구하게 된다. 그림 8은 보정된 영상을 나타낸다.



(a) 원 영상 (b) 보정된 영상

그림 8. 원영상과 0과 255를 기준으로 보정(saturation)된 영상

식 (9)은 2레벨 LH 부대역에 대해 이진화를 수행하는 과정을 나타낸다.

$$\begin{cases} LH2[i][j] = 1, & Lifting[i][j] \geq a \\ LH2[i][j] = 0, & otherwise \end{cases} \quad (9)$$

2레벨의 부대역들은 256×256크기의 영상에 대해 64×64의 크기를 가지므로 [i],[j]는 0부터 63까지의 범위를 가지며 1레벨의 부대역들은 128×128의 크기를 가지므로 [i],[j]는 0부터 127까지의 범위를 가진다. 이진화를 위한 임계값 a는 1레벨의 경우 4의 값을 가지고 2레벨의 경우 20의 값을 실험치에 의해서 설정을 한다.

에지정보를 검출하기 위하여 2레벨 리프팅을 수행한 후에 각 부대역 별로 블록을 나누어 각 블록에 해당하는 이진화된 계수값을 구한다. 블록을 나누는데 있어 에지의 이동과 미세한 위치 변화에 강인하도록 영상 내에서 추출된 에지의 위치에 따른 전체영역(global), 부분영역(semi_global), 그리고 국부(local)에지 히스토그램을 생성한다^[13]. 그림 9는 WEH의 3가지 영역을 표시한다.

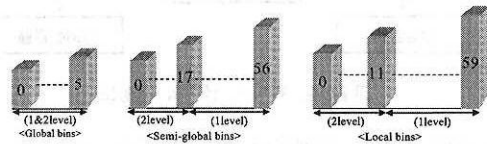


그림 9. 3 종류의 웨이블릿 에지 히스토그램

전체영역(global)히스토그램은 영상 전체에 각 방향성 에지량을 나타내고, 국부(local) 히스토그램은 영상 내 객체의 위치를 고려한 국부적인 에지 분포를 나타낸다. 그림 10은 전체영상에서 전체영역

(global) 히스토그램을 그리기 위한 부대역들의 표시를 나타낸다.

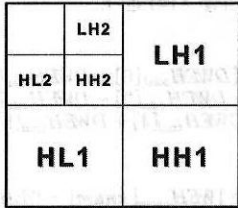


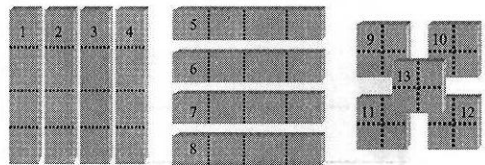
그림 10. WEH중 Global 영역에 해당하는 6개 부대역

6개 대역의 대역별 이진화된 값들을 더해서 WEH_global을 구한다. 식 (10)에 WEH_global을 정의하였다.

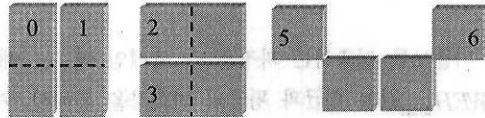
$$\begin{aligned}
 WEH_{global}[0] &= \sum_{y=0}^{63} \sum_{x=0}^{127} LH_2(x, y) \\
 WEH_{global}[1] &= \sum_{y=64}^{127} \sum_{x=0}^{63} HL_2(x, y) \\
 WEH_{global}[2] &= \sum_{y=64}^{127} \sum_{x=64}^{127} HH_2(x, y) \quad (10) \\
 WEH_{global}[3] &= \sum_{y=0}^{127} \sum_{x=128}^{255} LH_1(x, y) \\
 WEH_{global}[4] &= \sum_{y=128}^{255} \sum_{x=0}^{127} HL_1(x, y) \\
 WEH_{global}[5] &= \sum_{y=128}^{255} \sum_{x=128}^{255} HH_1(x, y)
 \end{aligned}$$

여기서 x, y는 리프팅된 256×256영상의 화소의 위치를 나타내고 LH2, HL2, HH2는 2레벨의 부대역들을 나타내며 LH1, HL1, HH1은 1레벨의 부대역들을 나타낸다.

부분영역(semi-global) 히스토그램은 1레벨 부대역들에 가로, 세로 방향에 대한 8종류(가로 4, 세로 4)의 부분 영역과 4분할, 중심 영역에 대한 5종류(영역분할 4, 중심 1)의 부분 영역을 정의하고, 각 부분 영역에 대하여 에지 종류별 누적 값을 사용하여 39개의 특징 정보를 생성한다. 2레벨 부대역들에는 가로, 세로 방향에 대한 4종류(가로 2, 세로 2)의 부분영역과 대각선 방향의 대한 2종류의 부분 영역을 정의하고, 각 부분 영역에 대해 동일하게 누적값을 통한 18개의 특징 정보를 생성한다. 그림 11(a)는 1레벨 부대역들에 적용되는 블록 마스크를 나타내고, 그림 11(B)는 2레벨 부대역들에 적용되는 블록 마스크를 나타낸다.



(a) 1레벨



(b) 2레벨

그림 11. 부분영역에 대한 블록 마스크 (a) 1레벨 (b) 2레벨

국부영역(local) 히스토그램은 LL대역을 제외한 2레벨 리프팅된 6개의 부대역(LH1, HL1, HH1, LH2, HL2, HH2)들에 대해 1레벨 부대역 48개, 2레벨 부대역은 12개의 부분영역들로 나누어 종류별 누적 값을 사용하여 총60개의 특징 정보를 생성한다. 그림 12는 국부 히스토그램에 적용되는 60개의 부분영역들을 나타낸다.

		1	2	13	14	17	18
		3	4	15	16	19	20
5	6	9	10	21	22	25	26
7	8	11	12	23	24	27	28
29	30	33	34	45	46	49	50
31	32	35	36	47	48	51	52
37	38	41	42	53	54	57	58
39	40	43	44	55	56	59	60

그림 12. 국부히스토그램에 사용되는 영역들

이렇게 전체 6개, 부분 57개, 국부 60개의 영역에 대한 히스토그램을 통합하여서 각 프레임에 대한 총 웨이블릿 에지 히스토그램을 얻게된다. 그림 13은 통합된 123개의 웨이블릿 에지 히스토그램을 나타내며 식 (11)에서 이것을 정의하였다.

$$WEH_{total}[i] = \sum_{i=0}^5 Globa[i] + \sum_{i=6}^{62} Semi_global[i] + \sum_{i=63}^{122} Local[i] \quad (11)$$

여기서 i는 123개의 통합된 WEH의 색인(index)을 나타낸다.

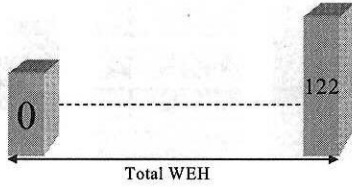


그림 13. 통합된 웨이블릿 에지 히스토그램

디졸브를 검출하는 과정은 각 프레임 별로 구해진 WEH_{total} 값의 평균과 평균의 변화량을 통해서 수행이 된다. 우선 WEH_{total} 의 평균은 7개의 프레임을 하나의 집합으로 하여 구하여진다. 이는 디졸브의 특성을 파악하는데 있어 프레임별로 비교를 하는 것보다 구간을 정하여 비교를 함으로써 점진적인 장면 전환에 대한 특성을 효율적으로 파악하는데 목적이 있다. 즉 6번째 프레임이 입력이 되면 0번과 1번의 WEH_{total} 값의 차이를 구하고 순차적으로 0번과 2번의 차이, 0번과 3번의 차이 등을 구하게 된다. 이렇게 비교된 6개의 값의 평균을 구해 이 평균값들의 변화량을 파악하게 된다. 그림 14는 일반적인 디졸브 구간을 나타낸다.



그림 14. 디졸브 구간의 예

식 (12)은 7개 프레임 구간의 WEH_{total} 을 비교하는 방법을 정의하고 있다.

$$\begin{aligned}
 DWEH_{total}[0] &= WEH_{total}[0] - WEH_{total}[6] \\
 DWEH_{total}[1] &= WEH_{total}[0] - WEH_{total}[5] \\
 DWEH_{total}[2] &= WEH_{total}[0] - WEH_{total}[4] \\
 DWEH_{total}[3] &= WEH_{total}[0] - WEH_{total}[3] \\
 DWEH_{total}[4] &= WEH_{total}[0] - WEH_{total}[2] \\
 DWEH_{total}[5] &= WEH_{total}[0] - WEH_{total}[1]
 \end{aligned} \tag{12}$$

여기서 $DWEH_{total}[f]$ 은 6번의 비교를 한 각각의 값이 들어가게 된다. 대괄호 안에 있는 숫자는 프레임

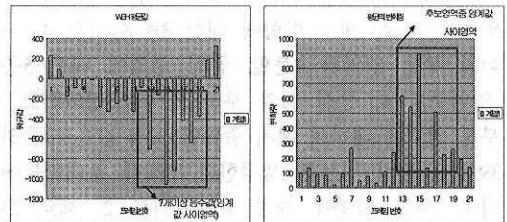
색인을 나타낸다. 이렇게 구한 $DWEH_{total}[f]$ 의 평균을 구하면 식 (13)과 같고, 식 (14)는 평균의 변화량을 계산하는 식을 나타낸다.

$$WEH_{mean}[f] = \frac{(DWEH_{total}[0] + DWEH_{total}[1] + DWEH_{total}[2] + DWEH_{total}[3] + DWEH_{total}[4] + DWEH_{total}[5])}{6} \tag{13}$$

$$DWEH_{mean}[f] = |WEH_{mean}[hnum] - WEH_{mean}[pnum]| \tag{14}$$

여기서 $DWEH_{mean}[f]$ 는 현재 프레임의 WEH의 평균값과 바로전 프레임의 WEH의 평균값의 차이의 절대값을 의미한다. 즉 변화량을 의미하며 부호는 없다. 여기서 $hnum$ 은 현재 프레임을 $pnum$ 은 바로전 프레임을 나타낸다.

이렇게 현재 프레임 f 에 대해 WEH_{total} 의 평균과 변화량을 구해 평균값이 연속해서 음수값으로 7프레임 이상 나오고 그 음수값의 범위가 일정한 임계값 범위안에 있으면 우선 디졸브구간의 후보 영역으로 추출을 한다. 이때 7개 프레임 집합을 사용하는 이유는 디졸브 특성을 나타내는데 가장 효율적인 집합단위로, 실험에 의한 수치를 사용하였다. 이는 디졸브 구간에서 WEH의 7개프레임의 평균값을 비교하여 평균값의 변화가 7개의 비교된 프레임 집합 안에서 나오는 특성이며 이렇게 추출된 후보영역에 대해 평균값의 변화량을 측정하여 일정한 임계값의 범위를 동시에 만족하면 디졸브 구간으로 추출을 하게된다. 그림 15는 그림 14의 디졸브 구간에 대한 평균값과 평균값의 변화량을 나타낸다.



(a) WEH 평균값 분포 (b) WEH 평균의 변화량 분포

그림 15. 디졸브 구간의 WEH_{total} 의 평균과 변화량

본 논문에서는 점진적인 장면전환점인 페이드와 디졸브에 대해 각 프레임의 특성을 파악하는 기존의 방법들과는 다르게 구간별 평균과 변화량을 사용하

여 최대한 페이드와 디졸브의 특성에 맞는 방법을 제안하였으며 이때 웨이블릿 변환의 장점인 인간의 시각 판정에 근거를 두어 장면전환 검출방법을 제안하고 있다. 그림 16에 제안하는 기법의 구성도를 보였다.

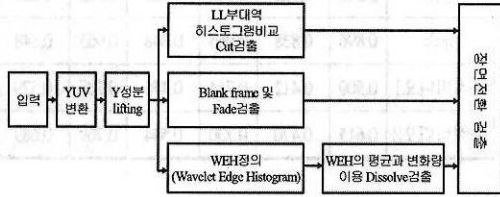


그림 16. 제안하는 시스템 구성도

IV. 실험결과 및 분석

본 논문에서는 표 1에서와 같이 다양한 형식의 동영상을 사용하여 실험을 수행하였다. 뮤직비디오 경우는 페이드, 디졸브 같은 점진적인 장면전환점이 많고, 뉴스 같은 경우는 앵커구간과 내용구간으로 뚜렷하게 나누어지는 특징이 있다. 또한 광고영상의 경우는 조명변화와 빠른 장면전환의 특징을 보인다.

표 1. 실험에 사용한 동영상 데이터

	광고	뉴스	뮤직비디오1	뮤직비디오2
크기 (pixel)	352×240	352×240	304×240	352×240
프레임수 (# frames)	1832	2670	2023	2005
재생시간 (mm:ss)	1:01	1:29	1:07	1:06
영상포맷 (codec)	AVI	MPEG	AVI	MPG

표 2. 제안된 기법과 기존 방법들의 성능비교

	광고			뉴스			뮤직비디오1			뮤직비디오2		
	히스토그램	화소단위	새기법	히스토그램	화소단위	새기법	히스토그램	화소단위	새기법	히스토그램	화소단위	새기법
실제컷수 (cut+dissolve+fade)	39 (37+1+1)			29 (29+0+0)			14 (9+3+2)			26 (17+5+4)		
찾은컷수	42	36	41	31	33	33	17	21	19	34	32	29
잘못찾은컷수	9	10	7 (7+0+0)	5	5	5 (5+0+0)	10	11	8 (8+0+0)	18	13	9 (8+1+0)
정확히찾은컷수 (cut+dissolve+fade)	33 (33+0+0)	26 (25+0+1)	34 (33+1+0)	26 (26+0+0)	28 (28+0+0)	28 (28+0+0)	7 (6+1+0)	10 (9+0+1)	11 (8+2+1)	16 (14+1+1)	19 (16+2+1)	20 (14+3+3)
못찾은컷수 (cut+dissolve+fade)	6 (4+1+1)	13 (12+1+0)	5 (4+0+1)	3 (3+0+0)	1 (1+0+0)	1 (1+0+0)	7 (3+2+2)	4 (0+3+1)	3 (1+1+1)	10 (3+4+3)	7 (1+3+3)	6 (3+2+1)

실험환경은 펜티엄4 1.6GHz프로세서, 512MByte 메모리 상에서 Visual C++ 6.0과 Directshow 라이브러리를 이용하여 구현하였다. 일반 동영상압력 방법과는 달리 본 실험에서는 Directshow 라이브러리를 사용함으로써 다양한 동영상 포맷에 대한 실험을 수행할 수 있고, 실시간 처리를 구현하므로 원하는 시간과 원하는 부분에서 동영상의 장면전환점을 추출할 수 있다.

실험에서 컷을 판별하는 임계값은 고정임계값을 사용하였다. 제안하는 기법을 통해 비교된 값들이 임계값을 넘었을 경우에 컷으로 판정을 하였다. 이 값은 실험에 의해 최적화된 임계값으로 설정하였다. 성능 비교를 위해 기존의 방법들도 구현하였다. 표 2에 제안하는 기법과 기존의 히스토그램 비교법과 화소단위 비교법의 성능을 비교하였다. 실제 컷을 표현하는데 있어서 컷과 디졸브 및 페이드를 구분하여서 표시하였다. 표에서 보듯이 제안된 기법은 기존의 히스토그램 비교법의 컷 검출보다 성능이 개선되었고, 특히 점진적인 장면전환인 페이드와 디졸브에 대해서도 기존의 히스토그램 비교법과 화소단위 비교법보다 높은 검출 결과를 보였다.

점진적인 장면전환에 대해 실험영상별로 페이드와 디졸브구간을 구분하여 검출의 유무를 측정해 보았다. 뉴스영상은 점진적인 장면전환이 없으므로 결과에서 제외하였고, 페이드와 디졸브가 있는 광고와 2개의 뮤직비디오 영상들에 대해서 검출 결과를 표3에 나타내었다. 표에서 볼 수 있듯이 점진적인 장면전환이 많은 뮤직비디오영상들에서 제안하는 방법이 좋은 검출 결과를 보인다.

본 논문에서 제안한 컷 검출의 효율성을 평가하고 다른 컷 검출 방법과의 성능비교를 위해, 일반적으로 사용되는 Recall과 Precision이라는 성능평가의 척도를 사용하였다. Recall은 실제 컷 중에서 정확히 검출된 컷의 비율로서 컷 감지율을 나타내고, Precision은 검출된 컷들 중에서 정확하게 검출된 컷의 비율로서 컷 검출의 정확도를 의미한다. 식 (15)과 식 (16)에서 Recall과 Precision을 정의하였다.

$$Recall = \frac{N_c}{N_c + N_m} \quad (15)$$

$$Precision = \frac{N_c}{N_c + N_f} \quad (16)$$

여기서 N_c 는 정확히 검출된 컷 수를 가리키고, N_m 는 못 찾은 컷 수, N_f 는 잘못 찾은 컷 수를 나타낸다. Recall과 Precision은 임계값을 높게 설정할수록 컷을 검출하는 능력인 Recall은 낮아지고, 검출된 컷의 정확도인 Precision은 높아진다. 반면, 임계값을 낮게 설정할수록 Recall이 높아지고 Precision은 낮아진다. 따라서 Recall과 Precision간에 적절한 타협을 필요로 한다.

표 4. 실험데이터의 Recall과 Precision

	히스토그램		화소단위		제안하는 방법	
	recall	precision	recall	precision	recall	precision
광고	0.846	0.786	0.666	0.722	0.872	0.829
뉴스	0.896	0.838	0.965	0.848	0.965	0.848
뮤직비디오1	0.500	0.412	0.714	0.476	0.785	0.579
뮤직비디오2	0.615	0.470	0.730	0.594	0.769	0.689

표 4는 성능평가 척도를 기준으로 하여, 본 논문에서 제안한 방법과 기존의 검출 방법인 히스토그램 비교법, 화소단위 비교법의 결과를 비교한 것이다. 본 논문에서는 점진적인 장면전환점의 검출률 향상에 목적을 두고 제안하는 방법을 사용한 결과 히스토그램 비교법 보다 Recall은 평균 23%, Precision은 평균 23%의 성능향상을 보였으며, 화소단위 비교법 보다 Recall은 평균 11%, Precision은 평균 13%의 성능향상을 보였다. 표 5에 기존의 검출방법과 제안하는 검출방법의 성능향상에 대한 결과를 나타내었다.

표 3. 실험영상의 점진적인 장면전환 검출결과 (O=검출, x=검출하지 않음)

	장면전환종류 [구간]	히스토그램 비교법	화소단위 비교법	제안하는 방법		장면전환종류 [구간]	히스토그램 비교법	화소단위 비교법	제안하는 방법
	dissolve [1090~1110]	x	x	o	fade [987~1009]	x	x	o	
					fade [1580~1602]	x	x	x	
뮤직 비디오1 영상	fade [70~96]	x	o	o	뮤직 비디오2 영상	fade [1992~2005]	x	x	o
	fade [1487~1522]	x	x	x		dissolve [208~233]	x	o	o
	dissolve [1216~1233]	o	x	o		dissolve [470~495]	x	x	x
	dissolve [1295~1316]	x	x	x		dissolve [962~972]	o	o	o
	dissolve [1420~1437]	x	x	o		dissolve [1217~1229]	x	x	o
						dissolve [1715~1729]	x	x	x

표 5. 제안하는 방법의 성능향상률

	성능향상(%)			
	Recall		Precision	
	vs. 히스토그램	vs. 화소단위	vs. 히스토그램	vs. 화소단위
광고	3.07	30.93	5.47	14.82
뉴스	7.70	0	1.19	0
뮤직비디오1	57.00	9.94	40.53	21.63
뮤직비디오2	25.04	5.34	46.59	15.99
성능향상 평균	23.20	11.55	23.44	13.11
	17.425		18.275	

표 4와 표 5에서 볼 수 있듯이 웨이블릿 변환과 대역별 에지 히스토그램을 이용한 본 논문의 실험결과가 기존의 검출방법보다 좋은 성능을 나타내었으며, 특히 점진적인 장면전환점이 많은 뮤직비디오 영상들에서 높은 성능 향상률을 보여주고 있다.

V. 결론

본 논문에서는 효율적인 동영상 검색과 브라우징을 위한 전 단계로 영상의 전환점인 컷 검출 기법을 제안하고 그 성능을 평가해 보았다. 제안된 기법에서는 급진적인 장면전환과 점진적인 장면전환의 특징을 분석하고 블랭크 프레임과 웨이블릿 변환에 기초한 대역별 웨이블릿 계수의 에지 히스토그램을 정의하여 동영상의 장면전환점을 검출하고 있다. 히스토그램 비교에 기초를 두고 컷을 판별하였으며, 페이드는 블랭크 프레임을 정의하여 과거와 미래의 일정구간의 히스토그램 차이값을 비교하여 검출을 하였다. 디졸브는 웨이블릿 에지 히스토그램을 정의하여 이것의 평균과 평균의 변화량을 측정하여 디졸브구간의 특성을 파악하여 디졸브를 검출하였다. 기존의 방법과 제안하는 방법에서 사용한 임계값은 실험을 통하여 가장 좋은 성능을 내는 고정임계값을 사용하였고, 실험에 쓰인 영상들도 동영상의 여러 특징들을 가진 데이터를 가지고 실험을 하였다.

Recall과 Precision이라는 성능평가의 척도를 이용하여 제안한 방법의 성능을 측정한 결과 4종류의 동영상 데이터에 대해 기존의 검출방법들 보다 Recall은 평균 약 17%, precision은 평균 약 18%의 검출 성능의 향상을 보였다. 이는 기존의 기법보다 우수한 결과이다. 특히 웨이블릿 에지 히스토그램에 의해 디졸브검출에 대해서 우수한 성능을 보였다.

참고 문헌

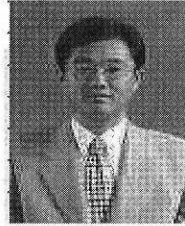
- [1] Hong Heather Yu, Wayne Wolf, "A Hierarchical Multiresolution Video Shot Transition Detection Scheme", *Computer Vision and Image Understanding*, Vol 75, Nos. 1/2, pp.196-213, Jul/Aug, 1999.
- [2] J. Meng, Y. Juan and S. F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence", *Digital Video Compression ; Algorithms and Technologies*, Vol. SPEI-2419, pp.14-25, Feb, 1995.
- [3] Yoshinobu Tonomura, Akihito Akutsu, Yukinobu Taniguchi and Gen Suzuki, "Structured Video Computing", *IEEE Multimedia*, pp.34-43, Fall, 1994.
- [4] G. Lupatini, C. Saraceno, R. Leonardi, "Scene Break Detection ; A Comparison", *In Int. Workshop On Research Issues In Data Engineering*, pp.33-41, 1998.
- [5] D. Swanberg, C.F. Shu, R. Jain, "Knowledge Guided Parsing In Video Database", *In Proc. Of SPIE Symposium On Electronic Imaging : Science And Technology*, pp.12-24, San Jose. CA., 1993.
- [6] A. Nagasaka, Y. Tanaka, "Automatic Video indexing And Full-Video Search For Object Appearances", *Visual Database Systems II Eds. E. Knuth and L. M. Wegner, Elsevier Science Publishers B. V., IFIP*, pp.113-127, 1992.
- [7] Stephen W. Smoliar and Hong jiang Zhang, "Content-Based Video Indexing And Retrieval", *IEEE Multimedia*, pp.62-72, Summer, 1994.
- [8] R. Zabih, J. Miller and K. Mai, "A Feature-Based Algorithm For Detecting And Classifying Scene Breaks", *ACM International Conference On Multimedia*, pp.189-200, 1995.
- [9] 윤인구, 김우생, "동적임계값을 이용한 컷 검출", *한국정보과학회 논문집*, 7월, 1999.
- [10] S. J. Park, D. K. Park, C. S. Won, "Core Experiments On MPEG-7 Edge Histogram Descriptor", *MPEG Document M.5984*,

Geneva, May, 2000.

- [11] Changliang Wang, Kap Luk Chan, Stan Z. Li, "Spatial-Frequency Analysis For Color Image Indexing And Retrieval", *ICARCV* 98, Vol. 2, pp.1461-1465, 1998.
- [12] S. G. Mallat, "Multi-Frequency Channel Decompositions Of Images And Wavelet Model", *IEEE Trans. Acoust. Speech Signal Process*, Vol. 37, No. 12, pp.2091-2110, 1989.
- [13] 박수준, 박성희, 김현진, 장명길, 원치선, "이미지 검색을 위한 에지 히스토그램", *The Proceedings of Human Computer Interaction 2001*, Feb. 2001.

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) 산전연구소 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

문 영 호(Young-ho Moon)

준회원

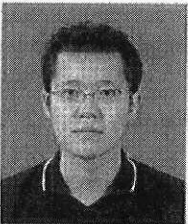


2004년 2월 : 광운대학교 전자공학과 졸업
 2004년 3월~현재 : 광운대학교 전자공학과 석사과정

<관심분야> MPEG-7, 데이터 방송, 영상압축

노 정 진(Jung-jin Noh)

준회원



2004년 2월 : 광운대학교 전자공학과 졸업
 2004년 3월~현재 : 광운대학교 전자공학과 석사과정

<관심분야> MPEG-7, 데이터 방송, 영상압축