

웨이블렛 변환을 적용한 장면전환의 cut과 fade검출

이명은*, 박종현*, 박순영*, 방만원*, 조완현**

*목포대학교 전자공학과, **전남대학교 통계학과

Cut and Fade Detection of Scene Change Using Wavelet transform

MyungEun Lee*, JongHyun Park*, SoonYoung Park*, ManWon Bang*, WanHyun Cho**

*Dept. of Electronics Engineering, Mokpo National University

**Dept. of Statistics, Chonnam National University

E-mail : melee@apollo.mokpo.ac.kr, {jhpark, sypark, mwbang}@chungkye.mokpo.ac.kr
whcho@chonnam.chonnam.ac.kr

요약

본 논문에서는 신호를 해석하는데 유용한 웨이블렛 변환을 적용하여 장면전환 요소 중 cut과 fade를 검출하는 알고리즘을 제안한다. 제안된 방법은 웨이블렛 저대역 부밴드로부터 각 프레임의 히스토그램을 구한 후 이전 프레임과 현재 프레임사이의 히스토그램 차를 구하여 이 값이 임계값 이상이면 급격한 장면전환(abrut shot transition)인 cut으로 분류한다. 다음으로 페이드인(fade in)이나 페이드아웃(fade out)등 컷의 지점이 불분명한 점진적 장면전환(gradual scene transition)을 검출하기 위하여 고대역 부밴드에서 추출한 에지성분에 모멘트를 계산하여 인접한 프레임 사이의 변동율을 분석하여 값이 증가하면 페이드인을 검출하고 반면에 감소하면 페이드아웃을 검출하게된다. 성능평가를 위하여 실제의 비디오 분할에 적용한 결과 웨이블렛 적용 방법론이 매우 높은 Precision을 갖는다는 것을 알 수 있으며 윤곽정보에 모멘트 정보를 더함으로써 기존의 방법보다 정확한 페이드(fade) 구간을 검출할 수 있었다.

I. 서론

디지털 영상과 비디오 데이터의 양이 증가함에 따라 장면전환 검출은 비디오 색인(video indexing), 주문형 비디오(VOD), 비디오 편집(video editing), 검색(retrieval) 등에 응용되면서 최근 많은 연구가 이루어지고 있다 [1][2]. 또한 사용자의 요구도 다양해짐에 따라 비디오

내의 장면단위로 접근 할 수 있도록 하는 다양한 서비스를 제공할 수 있어야 한다. 이러한 디지털화된 비디오 정보를 데이터베이스로 구축하고자 할 때는 먼저 비디오를 내용별로 색인하여 사용자가 원하는 비디오 정보에 대해 빠른 시간에 접근할 수 있도록 하고 내용기반 검색을 가능하게 하는 것이다. 따라서 한 편의 비디오를 내용별로 색인하기 위해서는 해당 비디오 데이터를 장면의 내용이 바뀌는 시점 단위로 분할해야 되며, 비디오 정보의 데이터베이스화에 있어서 장면전환 검출은 필수적으로 선행되어야 하는 기술이다. 위에서 기술한 장면전환이란 한 대의 카메라가 연속적으로 찍은 장면이 끝나고 다른 장면으로 바뀌는 것을 의미하며, 종류는 인접한 두 프레임사이에서 급격하게 장면 전환이 일어나는 급격한 장면 전환과 여러 프레임에 걸쳐서 서서히 일어나는 점진적인 장면전환 두 가지로 나뉘어진다.

기존의 장면전환 접근방법에는 비디오 영상에 히스토그램차, 화소차, 모션벡터 분석, Neural network등을 적용한 방법이 있으며, Miller와 Mai는 컷, 페이드, 디졸브(dissolve), 와이프(wipe) 등 여러 가지 장면전환을 찾기 위해 에지를 사용했으며[1], Zabih는 복원 영상의 에지 픽셀의 분포를 이용하여 디졸브, 페이드 및 와이프를 검출하였다[5].

본 논문에서는 웨이블렛 변환을 적용하여 장면전환 요소 중 컷을 검출하고, 또한 고대역 부밴드의 에지 정보에 모멘트를 계산하여 페이드 인/아웃을 검출하는 알고리즘을 제안한다. 논문의 구성은 II장에서 비디오의 각 프레임에 wavelet 변환을 이용한 장면전환검출에 대하여 설명하였으며, III장에서는 제안된 방법의 실험결

과에 대하여 논하였다. 마지막으로 IV장에서는 제안된 방법의 결론에 대하여 기술하였다.

II. 웨이블릿 변환을 이용한 장면전환 검출

2.1 웨이블릿 변환(Wavelet Transform)

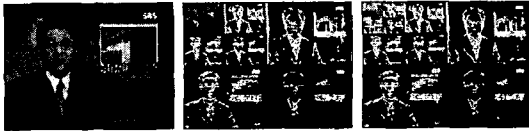
웨이블릿 변환은 인간의 시각 관점을 가장 잘 나타내는 변환으로써 시간 및 주파수에 대하여 국부성을 가지며, wavelet 변환을 적용한 영상의 다해상도 표현에서 주파수 공간의 계수들을 분석하여 처리할 수가 있는데 이것은 영상의 시공간적인 중복성을 내포하고 있는 화소를 효과적으로 제거함으로써 적은 양의 데이터 정보로도 영상의 특징을 전달할 수가 있어 응용이 가능한 유용한 변환이다. 본 논문에서는 그레이 영상 $G(x, y)$ 에 대하여 wavelet 변환을 적용하였으며 2-D wavelet 변환은 다음과 같다[3].

$$W_f(a, b_x, b_y) = \int_y \int_x G(x, y) \psi_{a, b_x, b_y}(x, y) dx dy \quad (1)$$

여기서 $\psi_{a, b_x, b_y}(x, y)$ 는 wavelet의 기저 함수가 된다. 기저함수는 일반적으로 orthogonal하며 mother wavelet ψ 은 전위(translation)와 팽창(dilation)에 의하여 얻어진다.

$$\psi_{a, b_x, b_y}(x, y) = \frac{1}{|a|} \psi\left(\frac{x-b_x}{a}, \frac{y-b_y}{a}\right) \quad (2)$$

그림 1은 실험에 사용된 일부 영상의 wavelet 변환을 적용했을 때 영상의 저대역 영역과 고대역 영역의 다해상도 구조를 보여주고 있다.



(a)original image (b) 2-level (c) 3-level

그림 1. wavelet 변환을 적용한 다해상도 구조

2.2 컷(cut) 검출

웨이블릿 변환에 의해 얻어지는 저대역 영역은 원래 영상의 압축되고 스무딩된 정보를 포함한 것으로 모든 웨이블릿 변환된 부밴드 중에서 가장 중요한 정보를 포함하고 있다. 따라서 비디오 시퀀스의 내용별 분할을 위해 한 장면(shot)에서 다른 장면(shot)으로 변환되는 점, 즉 컷(cut) 검출이 선행되어야 하는데 본 논문에서는 웨이블릿 영역 저대역 부밴드로부터 각 프레임의 히스토그램을 구한 후 이전 프레임과 현재 프레임의 히스토그램 차를 구하여 이 값이 임계값 이상이면 급격한

장면전환인 cut으로 분류하였다.

2.3 모멘트를 이용한 페이드(fade in/out) 검출

웨이블릿 변환을 한 후 고대역 부밴드는 수직, 수평, 대각선 방향에 대한 에지성분으로 구성되어있다. 따라서 여러 프레임에 걸쳐 장면전환이 일어나는 점진적인 장면전환 검출을 할 경우 많이 이용된다. 페이드 인, 페이드 아웃 등 컷의 지점이 불분명한 점진적 장면전환을 검출하기 위하여 페이드 구간에 나타나는 특성중의 하나인 시각적으로 윤곽은 뚜렷하여지거나 희미해지는 특징을 활용한다. 즉 수식으로 표현하면 특성함수 $E(t)$ 는 다음과 같이 나타낼 수 있다[1].

$$E_{fade-in}(t) = S(t)(1 - \eta(t)) + C \quad (3)$$

$$E_{fade-out}(t) = S(t)\eta(t) + C \quad (4)$$

여기서 $S(t)$ 는 image sequence, $\eta(t)$ 는 decreasing function, C는 constant 이다.

그러나 예지만을 가지고 페이드를 검출할 경우는 잘못 검출하거나 찾지못한경우가 많이 있기 때문에 본 논문에서 제안한 고대역 부밴드에서 추출한 에지성분에 각각의 고대역 부밴드에 대한 모멘트를 계산하여 인접한 프레임 사이의 변동율을 분석하여 페이드를 검출하는 방법을 제안한다.

Moments는 영상의 형태를 특정 짓는데 매우 유용한 계수들로서 이진화 함수 $f(x, y)$ 에 대한 모멘트는 다음 수식과 같이 정의된다[4].

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (5)$$

여기서 p, q 는 차수를 나타낸다.

1차 central moment μ_{00} 는 m_{00} 에 의하여 표시할 수 있으며, 차수 $(p+q)$ 의 central moments는 다음 식으로부터 정의된다.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (6)$$

여기서 $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$ 이다.

식(6)에 의해 2, 3차 central moments를 계산하여 요약하면 아래 식(7)과 같다.

$$\begin{aligned} \mu_{00} &= m_{00}, & \mu_{10} &= 0, & \mu_{01} &= 0 \\ \mu_{20} &= m_{20} - \bar{x}m_{10}, & \mu_{02} &= m_{02} - \bar{y}m_{01} \\ \mu_{11} &= m_{11} - \bar{y}m_{10} \\ \mu_{30} &= m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2m_{10} \\ \mu_{12} &= m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2m_{01} \\ \mu_{21} &= m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2m_{01} \\ \mu_{03} &= m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2m_{01} \end{aligned} \quad (7)$$

따라서 본 논문에서는 각 고대역 부밴드에 해당하는 2차, 3차 모두 7개의 central moments 계수를 구하여 $\mu(1), \mu(2), \mu(3) \dots \mu(6), \mu(7)$ 라 표시한다.

점진적 장면전환 검출을 위하여 수직, 수평, 대각선 방향의 고대역 부밴드에서 구한 모멘트의 프레임간 차를 다음과 같이 구한다.

$$\begin{aligned} D_i^V &= \frac{1}{7} \sum_{k=1}^7 [\mu_{i+r}^V(k) - \mu_i^V(k)] \\ D_i^H &= \frac{1}{7} \sum_{k=1}^7 [\mu_{i+r}^H(k) - \mu_i^H(k)] \\ D_i^D &= \frac{1}{7} \sum_{k=1}^7 [\mu_{i+r}^D(k) - \mu_i^D(k)] \end{aligned} \quad (8)$$

여기서 $\mu_i^V(\cdot), \mu_i^H(\cdot), \mu_i^D(\cdot)$ 은 i 번째 프레임의 고대역 부밴드 중에서 수직방향, 수평방향, 대각선 방향에 해당되는 모멘트들이다.

다음으로 각 방향에 대한 모멘트 차의 평균값을 식(9)와 같이 구하여 값들이 점진적으로 증가되는 구간을 페이드 인으로 정의하고, 값들이 점진적으로 감소되는 구간은 페이드 아웃으로 정의하였다.

$$D_i = \frac{1}{3} (D_i^V + D_i^H + D_i^D) \quad (9)$$

III. 실험 결과 및 고찰

본 논문에서 제안된 방법으로 웨이블릿 변환 후 저대역 부분에서의 히스토그램차를 이용한 컷 검출과 고대역 부분의 에지정보에 모멘트 정보까지 포함한 페이드 검출 알고리즘을 제안했다. 실험에 사용한 영상의 크기는 352X240이며 급격한 장면이 많은 뮤직비디오, 영화, 뉴스 등의 영상과 페이드 인/아웃이 포함되어 있는 두 종류의 뮤직비디오 영상을 이용하였다. 실험에 사용한 프레임은 최대 1200프레임에서 최소 500프레임이며 총 프레임들 중 일정한 간격으로 두 프레임씩 선택하여 알고리즘을 수행하였다.

1. 급격한 장면 전환 cut 검출

웨이블릿을 적용한 영상의 저대역 부밴드에서 프레임간 히스토그램의 차를 적용하여 급격한 장면전환인 cut을 검출하였다. 장면전환의 시점은 연속되는 두 장면이 급격하게 달라지는 시점으로 정하였다.

precision값은 식(10)과 같이 표현 할 수 있다.

$$precision = \frac{n_c}{n_c + n_m} \quad (10)$$

여기서 n_c 는 올바르게 검출된 컷의 수를 나타내며, n_m 은 검출하지 못한 컷 수, 그리고 n_f 는 잘못 검출한

컷 수를 나타낸다.

표1 컷(cut) 검출결과

실험영상	원래 컷수	n_c	n_f	n_m	precision
뮤직비디오 (1100Frames)	30	27	2	1	96.4%
영화 (500Frames)	15	10	3	2	83.3%
뉴스 (1070Frames)	27	20	5	2	90.9%

표 1은 제안된 알고리즘의 수행결과를 보여주고 있는데 뮤직비디오나 뉴스와 같은 영상의 경우 장면전환 검출 결과가 우수했으나 움직임이 복잡한 장면이 반복된 영화 영상에서는 다소 검출 정확도가 떨어졌다.

그림 2는 555번째 프레임에서 급격한 장면 전환을 갖는 뮤직비디오 영상의 장면전환이 일어나는 부근의 프레임들의 예를 나타낸 것이다.



그림 2. 급격한 장면전환이 일어나는 영상의 예

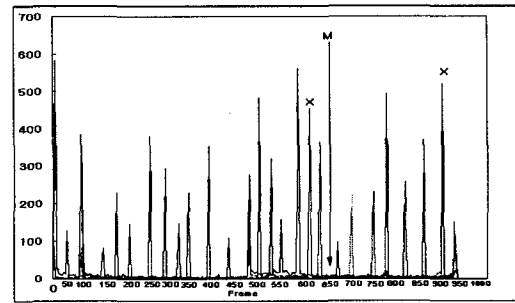


그림 3. 히스토그램 결과값(뮤직비디오)

그림 3에서 "X" 표시는 잘못 검출한 부분이며, "M" 표시는 놓친 부분을 나타내고 있다.

제안한 웨이블릿 변환을 적용하여 인접 프레임의 히스토그램값으로 컷을 검색할 경우 precision이 우수했다. precision이 높다는 것은 잘못 검출된 샷 경계가 많지 않다는 것을 의미한다.

2. 점진적인 장면 전환 fade 검출

웨이블릿 변환된 고대역 부밴드에서의 페이드 영상을 그림 4에 나타내었다. (a)는 120~135프레임 구간사이에서 페이드 아웃을 (b)는 290~300프레임 구간사이

에서 페이드 인을 나타내는 점진적인 장면 전환의 뮤직비디오 영상을 나타낸 것이다.

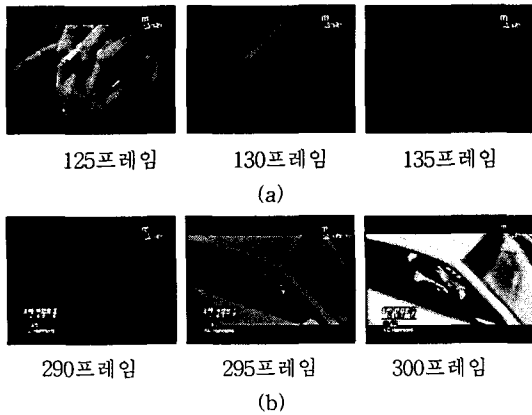


그림 4. 페이드 인/아웃이 나타나는 영상의 예
(a) fade out이 나타나는 영상 예,
(b) fade in이 나타나는 영상 예

표 2. 제안한 방법에 의한 페이드 구간 검출 결과표

Sample 영상	Fade In		Fade Out	
	실제 구간	검출된 구간	실제 구간	검출된 구간
뮤직 비디오 1	139-155	138-155	120-135	122-135
	180-195	180-195	165-176	165-175
	231-250	230-250	220-230	218-230
	288-310	289-310	260-285	260-285
뮤직 비디오 2	85-130	85-128	221-235	221-235
	458-525	458-520	337-352	335-352

표 2는 실험에 사용했던 2개의 뮤직비디오 영상에서 제안한 알고리즘을 통한 페이드 검출 결과를 나타낸 것이다.

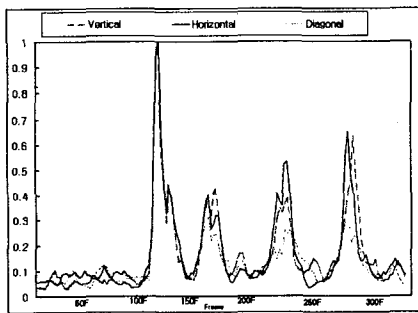


그림 5. fade in/out 검출 결과
[고대역 밴드의 V(수직), H(수평), D(대각선) 검출결과]

그림 5는 페이드 검출 실험에 사용된 뮤직비디오 1의 실험결과를 그래프로 나타낸 것이며, 값이 증가하는 구간이 페이드 인이며, 반대로 감소하는 구간이 페이드 아웃이다. 페이드는 여러 프레임에 걸쳐서 나타나기 때

문에 모든 프레임 하나하나를 조사하는 것이 아니라, 몇 개의 프레임만을 조사해도 페이드 구간이 나타나는 곳을 알 수 있다. 페이드 구간은 정확하게 시작하고 끝나는 부분을 한 프레임의 오차도 없이 가리는 것이 중요한 것이 아니라 페이드 구간 자체를 찾는가 못찾는가가 더 중요한 문제이다. 따라서 본 논문에서 제안한 고대역 부밴드의 에지정보에 수직, 수평, 대각선 모멘트정보를 평균을 내어 변동율을 분석함으로써 에지만을 사용하여 구한 기존의 방법보다 더 정확한 페이드 구간을 검출할 수 있었다.

IV. 결론

본 논문에서는 정확한 검색을 위한 방법으로 영상에 웨이블릿 변환을 적용하여 영상을 공간-주파수 영역에서 분석하였으며 변환된 저대역 부밴드에서 프레임간 히스토그램의 차를 적용하여 급격한 장면전환인 cut을 검출하였다. 고대역 부밴드에서는 수직, 수평, 대각선 밴드의 에지성분에 모멘트를 구하여 인접 프레임간의 차이를 평균을 내어 페이드 인과 페이드 아웃을 검출하였다. 페이드 검출에서는 에지만을 사용하여 구한 방법보다 제안한 방법이 더 정확한 페이드 구간을 검출할 수 있었다. 따라서 제안된 알고리즘은 여러 가지 장면 전환 검출과 영상의 키프레임 추출시 유용하게 응용될 수 있을 것이다.

참고문헌

- [1] Hong Heather Yu and Wayne Wolf, "A Hierarchical Multiresolution Video Shot Transition Detection Scheme", *Computer Vision and Image Understanding*, vol. 75, Nos. 1/2, July/August, pp. 196-213, 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. SPIE-2419, pp. 14-25, Feb. 1995.
- [3] Changliang Wang, Kap Luk Chan, Stan Z Li, "Spatial-Frequency Analysis for Color Image Indexing and Retrieval", *ICARCV 98*, vol. 2, pp. 1461-1465, Apr. 1998.
- [4] Xiaolong Dai, Siamak Khorrarn, "A Feature-Based Image Registration Algorithm Using Improved Chain-Code Representation Combined with Invariant Moments", *IEEE Vol. 37*, pp. 2351-2362, Sept. 1999.
- [5] 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.