

3-D DCT를 이용한 비디오 장면 전환 검출

*우석훈, 원치선

동국대학교 전자공학과

e-mail : shwoo@dgu.edu, cswon@dgu.edu

Video Scene Change Detection Using a 3-D DCT

*Seock-Hoon Woo, Chee-Sun Won

Department of Electronic Engineering

Dongguk University

Abstract

In this paper, we propose a simple and effective video scene change detection algorithm using a 3-D DCT. The 3-D DCT that we employ is a $2x2x2$ DCT has simple computations composed only of adding and shifting operations. The simple average values of multiresolution represented video using the $2x2x2$ DCT are used as a detection feature vector.

I. 서론

최근 멀티미디어 데이터 중 디지털 비디오의 양이 증가하고 있다. 이러한 디지털 비디오는 그 양이 방대하여 사용자가 원하는 정보를 얻어내는 것이 쉬운 일이 아니며, 이를 해결하기 위한 검색과 데이터베이스의 구축이 필수적인 요소로 자리 잡고 있다. 그 중, 가장 기본이 되는 과정이 바로 비디오 분할(Video Segmentation)으로, 비디오의 내용이 바뀔 때마다 자동적으로 분류하고, 인덱싱(Indexing)하는 과정을 통하여 비디오 정보를 체계적으로 정리하고, 저장하는 것이다. 이러한 분류는 기본 단위로써, 카메라의 샷(shot)에 의해 구분하는 것이 일반적이며, 이를 장면 전환 검출(Scene Change Detection)이라 한다. 장면 전환 검출의 단위가 되는 샷(shot)은 “카메라로 촬영을 시작하여 끝날 때까지의 연속된 프레임”이라 할 수 있다 [1]. 장면 전환에는 단일 프레임에서 발생하는 급진적 장면의 변화를 나타내는 컷(cut)과, 샷들 사이에서 발생하는 점진적 장면 전환(gradual scene change)이 있다. 이러한 점진적 장면 전환에는 페이딩(fading)과 디졸브(dissolve)등이 대표적이다. 장면 전환 검출은 일반

적으로 비디오 데이터의 인접 프레임들 간의 특징벡터의 거리를 계산하여 설정한 임계치를 초과하였을 때 장면 전환으로 결정하게 된다. 이러한 장면 전환 검출 방법들로는 영상 내 객체들의 이동을 고려한 방법[2], 통계적 특징을 이용한 방법[3], 움직임 특징과 히스토그램 차이를 이용하는 방법[4]등이 제안되었으며, 점진적 장면 전환 검출을 위해 두 가지의 임계치를 사용하는 방법[5]등이 제안되었다. 본 논문에서는 비디오의 장면 전환 검출을 위하여 $2x2x2$ DCT를 이용한 장면 전환 검출 기법을 제안한다. 제안된 장면 전환 검출 기법은 비디오 데이터를 $2x2x2$ DCT를 사용하여 다해상도로 표현한 후 이를 재배열하는 과정에서 얻어지는 각 프레임별 장면 전환 특성 밴드계수의 평균과 분산을 이용하여 급진적 및 점진적 장면 전환 검출을 하게 된다.

본 논문은 2장에서 3-D DCT를 이용한 비디오 데이터의 다해상도 표현에 대하여 기술하고, 3장에서는 다해상도로 표현된 비디오 데이터의 재배열 규칙을 통한 장면 전환 검출 방법에 대하여 설명한다. 4장에서는 제안된 장면 전환 검출 방법의 실험 결과에 대하여 논하고, 5장에서 제안된 검출 방법에 대한 결론을 기술하였다.

II. 3-D DCT를 이용한 다해상도 표현

2.1 $2x2x2$ DCT

비디오 데이터의 다해상도 표현을 위해 그림 1(a)와 같이 $N_x \times N_y \times T$ 크기의 비디오를 겹치지 않는 $2x2x2$ 크기의 블록으로 분할한다. 각 $2x2x2$ 블록에 대해 식 (1)~(8)을 이용하여 $2x2x2$ DCT를 적용한다. 즉, 전체 해상도에서 시작하여($k=1$) (m, n, l)계 블록의 DCT 계수

를 얻을 수 있다. ($m=0,1,\dots, N_t/2^k-1, n=0,1,\dots, N_x/2^k-1, t=0,1,\dots, T/2^k-1, k=1,2,3$)

$$F_{mnt}^k(0,0,0) = [f_{mnt}^k(0,0,0) + f_{mnt}^k(0,0,1) + f_{mnt}^k(0,1,0) + f_{mnt}^k(0,1,1) + f_{mnt}^k(1,0,0) + f_{mnt}^k(1,0,1) + f_{mnt}^k(1,1,0) + f_{mnt}^k(1,1,1)]/8 \quad (1)$$

$$F_{mnt}^k(0,1,0) = [f_{mnt}^k(0,0,0) + f_{mnt}^k(0,0,1) - f_{mnt}^k(0,1,0) - f_{mnt}^k(0,1,1) + f_{mnt}^k(1,0,0) + f_{mnt}^k(1,0,1) - f_{mnt}^k(1,1,0) - f_{mnt}^k(1,1,1)]/8 \quad (2)$$

$$F_{mnt}^k(1,0,0) = [f_{mnt}^k(0,0,0) + f_{mnt}^k(0,0,1) + f_{mnt}^k(0,1,0) + f_{mnt}^k(0,1,1) - f_{mnt}^k(1,0,0) - f_{mnt}^k(1,0,1) - f_{mnt}^k(1,1,0) - f_{mnt}^k(1,1,1)]/8 \quad (3)$$

$$F_{mnt}^k(1,1,0) = [f_{mnt}^k(0,0,0) + f_{mnt}^k(0,0,1) - f_{mnt}^k(0,1,0) - f_{mnt}^k(0,1,1) + f_{mnt}^k(1,0,0) - f_{mnt}^k(1,0,1) + f_{mnt}^k(1,1,0) - f_{mnt}^k(1,1,1)]/8 \quad (4)$$

$$F_{mnt}^k(0,0,1) = [f_{mnt}^k(0,0,0) - f_{mnt}^k(0,0,1) + f_{mnt}^k(0,1,0) - f_{mnt}^k(0,1,1) + f_{mnt}^k(1,0,0) - f_{mnt}^k(1,0,1) + f_{mnt}^k(1,1,0) - f_{mnt}^k(1,1,1)]/8 \quad (5)$$

$$F_{mnt}^k(0,1,1) = [f_{mnt}^k(0,0,0) - f_{mnt}^k(0,0,1) - f_{mnt}^k(0,1,0) + f_{mnt}^k(0,1,1) + f_{mnt}^k(1,0,0) - f_{mnt}^k(1,0,1) - f_{mnt}^k(1,1,0) + f_{mnt}^k(1,1,1)]/8 \quad (6)$$

$$F_{mnt}^k(1,0,1) = [f_{mnt}^k(0,0,0) - f_{mnt}^k(0,0,1) + f_{mnt}^k(0,1,0) - f_{mnt}^k(0,1,1) - f_{mnt}^k(1,0,0) + f_{mnt}^k(1,0,1) - f_{mnt}^k(1,1,0) + f_{mnt}^k(1,1,1)]/8 \quad (7)$$

$$F_{mnt}^k(1,1,1) = [f_{mnt}^k(0,0,0) - f_{mnt}^k(0,0,1) - f_{mnt}^k(0,1,0) + f_{mnt}^k(0,1,1) - f_{mnt}^k(1,0,0) + f_{mnt}^k(1,0,1) + f_{mnt}^k(1,1,0) - f_{mnt}^k(1,1,1)]/8 \quad (8)$$

$F_{mnt}^k(i,j,h)$ 에서 아래첨자 mnt 는 영상공간 내 (m,n,t) 번째 블록 인덱스를 나타내고 (i,j,h)는 그 블록에서 화소의 위치를 나타내며, k 는 다해상도 레벨을 나타낸다. 예를 들어, $F_{000}^k(i,j,h)$ 는 k 번째 다해상도 레벨에서 (0,0,0)번째 블록의 (i,j,h)번째 화소를 나타낸다. 식(1)의 $F_{mnt}^k(0,0,0)$ 는 단지 직류(DC)성분의 집합이므로 항상 $0 \leq F_{mnt}^k(0,0,0) \leq 255$ 를 만족한다. 식 (1)~(8)은 단순한 덧셈과 뺄셈 그리고 쉬프트 연산으로 되어 있으므로 고차 필터를 가지는 웨이브릿 변환에 비하여 계산량이 작다.

2.2 프레임 재배열

모든 $2 \times 2 \times 2$ DCT 계수들을 같은 주파수 성분끼리 모으기 위해 식 (9)와 (10)을 이용해서 모든 계수들을 재배열한다. 식 (9)와 (10)에서 R_i 는 재배열 후 프레임을 나타내고, F_i 는 $2 \times 2 \times 2$ DCT 후 프레임을 나타낸다. N 은 프레임 수(8)를 나타내며, k 는 다해상도 레벨 (1,2,3)을 나타낸다.

$$R_{2i-1} = F_i \quad i=1, \dots, N/2^k \quad (9)$$

$$R_{\frac{N}{2^{k-1}}-i} = F_{\frac{N}{2^{k-1}}-2i} \quad i=1, \dots, N/2^k-1 \quad (10)$$

이러한 $2 \times 2 \times 2$ DCT와 재배열 규칙을 재귀적으로 적용하여 그림 1과 같은 동영상의 옥타브 밴드 표현을 얻을 수 있다.

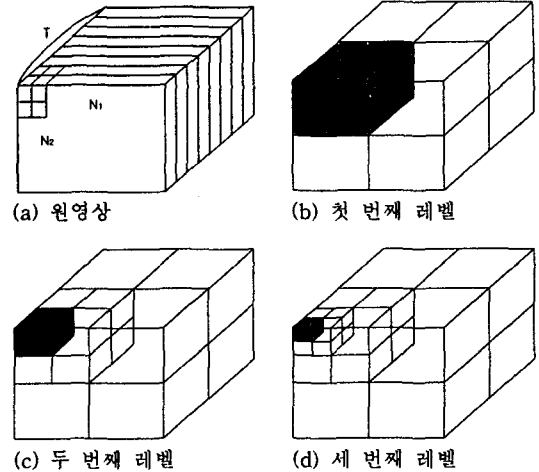


그림 1. $2 \times 2 \times 2$ DCT 계수의 다해상도 표현

그림 2는 8개 프레임에 대해 $2 \times 2 \times 2$ DCT를 이용한 다해상도 표현의 예이다.

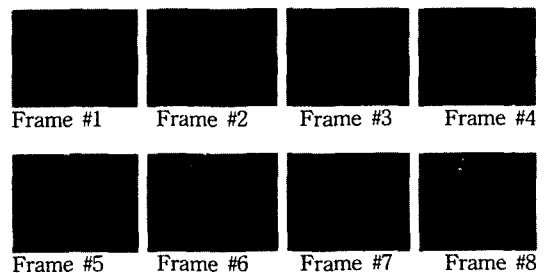
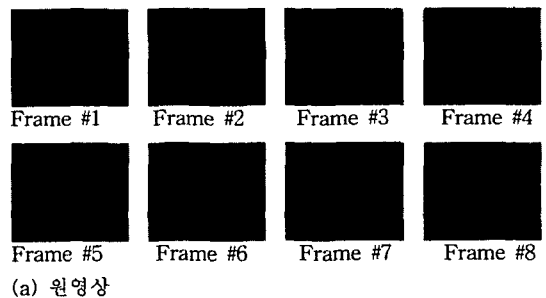


그림 2. $2 \times 2 \times 2$ DCT를 이용한 동영상 다해상도 표현

2.3 2x2x2 IDCT

다해상도로 표현한 비디오 데이터로부터 원영상을 얻기 위한 과정은 가장 높은 레벨(즉, 가장 낮은 해상도)에서부터 재배열과 식 (11)~(18)과 같은 IDCT를 적용하여 원영상을 복원할 수 있다.

$$f_{min}^k(0,0,0) = F_{000}^k(m,n,d) + F_{001}^k(m,n,d) + F_{010}^k(m,n,d) + F_{011}^k(m,n,d) + F_{100}^k(m,n,d) + F_{101}^k(m,n,d) + F_{110}^k(m,n,d) + F_{111}^k(m,n,d) \quad (11)$$

$$f_{min}^k(0,1,0) = F_{000}^k(m,n,d) + F_{001}^k(m,n,d) - F_{010}^k(m,n,d) - F_{011}^k(m,n,d) + F_{100}^k(m,n,d) + F_{101}^k(m,n,d) - F_{110}^k(m,n,d) - F_{111}^k(m,n,d) \quad (12)$$

$$f_{min}^k(1,0,0) = F_{000}^k(m,n,d) + F_{001}^k(m,n,d) + F_{010}^k(m,n,d) + F_{011}^k(m,n,d) - F_{100}^k(m,n,d) - F_{101}^k(m,n,d) - F_{110}^k(m,n,d) - F_{111}^k(m,n,d) \quad (13)$$

$$f_{min}^k(1,1,0) = F_{000}^k(m,n,d) + F_{001}^k(m,n,d) - F_{010}^k(m,n,d) - F_{011}^k(m,n,d) - F_{100}^k(m,n,d) - F_{101}^k(m,n,d) + F_{110}^k(m,n,d) + F_{111}^k(m,n,d) \quad (14)$$

$$f_{min}^k(0,0,1) = F_{000}^k(m,n,d) - F_{001}^k(m,n,d) + F_{010}^k(m,n,d) - F_{011}^k(m,n,d) + F_{100}^k(m,n,d) - F_{101}^k(m,n,d) + F_{110}^k(m,n,d) - F_{111}^k(m,n,d) \quad (15)$$

$$f_{min}^k(0,1,1) = F_{000}^k(m,n,d) - F_{001}^k(m,n,d) - F_{010}^k(m,n,d) + F_{011}^k(m,n,d) + F_{100}^k(m,n,d) - F_{101}^k(m,n,d) - F_{110}^k(m,n,d) + F_{111}^k(m,n,d) \quad (16)$$

$$f_{min}^k(1,0,1) = F_{000}^k(m,n,d) - F_{001}^k(m,n,d) + F_{010}^k(m,n,d) - F_{011}^k(m,n,d) - F_{100}^k(m,n,d) + F_{101}^k(m,n,d) - F_{110}^k(m,n,d) + F_{111}^k(m,n,d) \quad (17)$$

$$f_{min}^k(1,1,1) = F_{000}^k(m,n,d) - F_{001}^k(m,n,d) - F_{010}^k(m,n,d) + F_{011}^k(m,n,d) - F_{100}^k(m,n,d) + F_{101}^k(m,n,d) - F_{110}^k(m,n,d) + F_{111}^k(m,n,d) \quad (18)$$

III. 장면 전환 검출 방법

3.1 재배열 규칙을 이용한 장면 전환 검출 프레임 추출

식 (9)와 (10)의 재배열 공식을 통하여 연산되는 두 인접 프레임의 AC 계수들이 각각 재배열되는 규칙을 표 1과 같이 정의할 수 있다. 따라서, 큐브내 장면 전환이 발생했을 경우 표 2와 같이 장면 전환점에 따른 서로 다른 내용의 영상이 연산된 프레임들을 추출할 수 있다. 예를 들어, 2번 프레임에서 장면 전환이 발생하였을 경우, 다해상도로 표현된 5번 프레임의 계수들을 이용함으로써 2번 프레임이 장면 전환점임을 구할 수 있다.

표 1. 연산 프레임의 AC 계수 재배열

	연산 프레임	재배열된 AC 프레임
다해상도 레벨 1	#1, #2	#5
	#3, #4	#6
	#5, #6	#7
	#7, #8	#8
다해상도 레벨 2	#1, #2	#3
	#3, #4	#4
다해상도 레벨 3	#1, #2	#2

표 2. 장면 전환 점에 따른 재배열 프레임 추출

장면 전환 점	추출된 재배열 프레임
Frame #2	Frame #5
Frame #3	Frame #3
Frame #4	Frame #6
Frame #5	Frame #2
Frame #6	Frame #7
Frame #7	Frame #4
Frame #8	Frame #8

3.2 추출된 재배열 프레임을 이용한 장면 전환 검출

본 논문에서는 비디오의 장면 전환 검출을 수행하기 위하여 두 단계의 검출 과정을 이용한다. 첫 번째 단계에서는 표 2에서 추출된 각 재배열 프레임에 대하여 1개 밴드의 평균값을 계산하여 가장 큰 값을 가지는 재배열 프레임 번호를 이용하여 장면 전환 점을 추출한다. 예를 들어, 가장 큰 평균값을 가지는 재배열 프레임의 번호가 5번이면 두 번째 프레임에서 장면 전환이 발생했음을 알 수 있다. 두 번째 단계에서는 8장 단위의 각 큐브의 DC 밴드의 평균값과 각 재배열 프레임 1개 밴드의 평균값을 계산하여 좀 더 정확한 장면 전환점을 추출한다. 즉, 현재 큐브의 DC 밴드 평균값과 이전 큐브의 DC 밴드 평균값 및 다음 큐브의 DC 밴드 평균값과의 차이와 첫 번째 단계에서 구한 각 재배열 프레임 1개 밴드의 평균값을 이용하여 장면 전환을 검출하게 된다.

IV. 실험 결과

본 논문의 실험을 위해 사용된 비디오 데이터는 급진적 장면 전환과 페이드, 디졸브등의 점진적 장면 전환이 포함된 애니메이션영화, 뉴스, 교양물 비디오를 사용하였다.

그림 3은 제안된 장면 전환 검출 방법을 통해 구한 장면 전환 검출 결과 그래프를 보여 준다. 그림 3과 같이 검출 1 단계와 검출 2 단계를 통하여 더욱 정확한 장면 전환 검출이 구현됨을 알 수 있다.

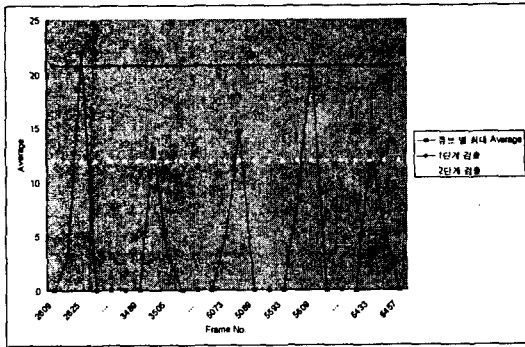


그림 3. 장면 전환 검출 그래프

장면 전환 검출의 성능을 평가하기 위하여 일반적으로 사용되어지는 Recall과 Precision을 계산하였다. Recall은 검출하지 못한 샷이 얼마나 많은가를 보여주며, Precision은 오검출된 샷이 얼마나 많은가를 나타낸다.

$$Recall = \frac{\# \text{ of detected shots}}{\# \text{ of detected shots} + \# \text{ of missed shots}}$$

$$Precision = \frac{\# \text{ of detected shots}}{\# \text{ of detected shots} + \# \text{ of false alarms}}$$

다음 표 3은 제안된 장면 전환 검출 방법의 검출 결과를 Recall과 Precision을 통해 보여 주고 있으며, 간단한 장면 전환 검출 연산 과정을 통해 효율적으로 비디오 장면 전환 검출을 구현함을 알 수 있다.

표 3. 3-D DCT를 이용한 장면 전환 검출 결과

	Detected	Missed	False Alarm	Recall	Precision
애니메이션 (7:09)	116	10	3	92.1%	97.5%
뉴스 (4:27)	23	0	3	100%	88.5%
교양물 (4:27)	24	0	3	100%	88.9%

V. 결론

본 논문에서는 3-D DCT를 이용한 비디오 장면 전환 검출 방법을 제안하였다. 제안된 알고리즘은 장면 전환 특징 벡터를 구하기 위하여 덧셈과 쉬프트 연산을 사용하는 간단한 2x2x2 DCT를 이용하였으며, 특징 벡터 또한 2x2x2 DCT를 통해 표현된 다해상도 영상의 평균값만을 이용하였다. 추출된 특징 벡터들을 장면 전환 검출 시 1, 2 단계를 이용함으로써 좀 더 세밀한 장면 전환 검출이 이루어지도록 하였다.

참고문헌

- [1] Robert B. Allen, Paul England, Comparison of automatic video segmentation algorithms, In SPIE Photon-ics West, volume 2615, pp. 2 - 16, Philadelphia, 1995.
- [2] Nilesh V. Patel and Ishwar K. Sethi, Video Shot Detection and Characterization for Video Databases, Pattern Recognition: Special Issue on multimedia, 1996.
- [3] Ishwar K. Sethi and Nilesh V. Patel, A statistical approach to scene change detection, SPIE Proc., Storage and Retrieval for Image and Video Databases III, Vol. 2420, Feb. 1995.
- [4] Phillips M and Wolf W, Video Segmentation Techniques for News, in Multimedia Storage and Archiving Systems, C. -C. Jay Kuo, Editor, Proc. SPIE 2916, pp. 243-251, 1996.
- [5] Zhang H.J, Kankanhalli A, Smoliar S.W, Automatic Partitioning of Full-motion Video, Multimedia Systems, Vol. 1, No. 1, pp. 10-28, 1993.