영역 흐름 및 칼라 정보를 이용한 MPEG 데이타의 내용 기반 셧 경계 검출

(Content-based Shot Boundary Detection from MPEG Data using Region Flow and Color Information)

강 행 봉 †

(Hang-Bong Kang)

요 약 비디오 데이타에서 셧 경계 검출은 비디오 인덱싱 및 검색에 있어서 매우 중요한 단계이다. 셧 경계 검출 방법은 현재 많은 연구가 진행되어 왔으나 아직까지 셧에 존재하는 내용을 기반으로 한 일반적인 방법이 제시되지 못하고 있다. 주로 연속된 비디오 프레임 사이의 칼라 히스토그램 변화나 DCT 계수의 편차에 의해 검출하는 방식이 주로 사용되고 있다. 하지만, 이러한 방식들은 비디오 데이타의 내용이나 의미있는 특징들을 사용하고 있지 않아서 사용자가 원하는 고도의 정보를 처리하는데는 어려운 점이 많다. 더욱이, 이러한 내용 기반의 처리가 압축 데이타에서 직접 수행할 수 있다면, 시간 및 오버헤드를 줄일 수 있어 바람직하다. 본 논문에서는 중요한 특징인 영역 정보 및 칼라 정보를 이용하여 MPEG 데이타를 원 영상으로 복호화하지 않고, 셧 경계를 검출하는 방법을 제안한다. 먼저, 압축비디오 데이타로부터 DC영상들을 복원하여 이 영상들에 존재하는 영역 정보 흐름 및 양자화 된 HSV 칼라 정보의 중요한 특징들을 추출하여, 영역 흐름의 불연속점이나 칼라 히스토그램의 값이 급격히 변화하는 프레임을 찾아 이를 제안된 알고리즘에 따라 비디오 데이타의 셧 경계로 검출한다.

Abstract *It is an important step in video indexing and retrieval to detect shot boundaries on video data. Some approaches are proposed to detect shot changes by computing color histogram differences or the variances of DCT coefficients. However, these approaches do not consider the content or meaningful features in the image data which are useful in high level video processing. In particular, it is desirable to detect these features from compressed video data because this requires less processing overhead. In this paper, we propose a new method to detect shot boundaries from MPEG data using region flow and color information. First, we reconstruct DC images and compute region flow information and color histogram differences from HSV quantized images. Then, we compute the points at which region flow has discontinuities or color histogram differences are high. Finally, we decide those points as shot boundaries according to our proposed algorithm.*

1. 서 론

최근들어 인터넷의 발전으로 인하여 멀티미디어 정보의 사용이 급격히 증가하는 추세에 있다. 텍스트, 오디오, 비디오 및 컴퓨터 그래픽등의 다양한 미디어로 구성된 멀티미디어 정보 중에서도 비디오 데이타는 다른 모든 미디어 정보를 하나의 데이타스트림 (datastream)으로 결합할 수 있기 때문에 매우 중요한 위치에 있다. 특히, 저장 장치의 가격이 급속히 인하되는 추세에 있고, 초고속 전송 및 압축 기술의 발전에 힘입어, 디지털 비디오는 도처에서 이용 가능한 미디어가 되었다.

하지만, 이러한 비디오 데이타는 일반적으로 대용량이고 또 비구조화(unstructured)된 특성을 갖고 있어서 비디오 인덱싱 및 검색에 있어서 효율적인 처리가 매우 까다롭다. 이를 해결하기 위해, 수년 전부터 비디오 데이타의 구조화(video structuring)에 관한 다양한 연구

· 본 연구는 1997-98년도 정보통신부 대학 기초 연구 지원사업에 의한 것임.

† 종신회원 : 가톨릭대학교 컴퓨터전자공학부 교수
hbkang@www.cuk.ac.kr

논문접수 : 1999년 7월 22일
심사완료 : 2000년 1월 25일

가 진행되어 왔다[1]. 이러한 구조화 작업은 비디오 데이타를 연속된 카메라 동작으로 구성된 셧(shot), 이러한 셧들의 그룹으로 구성되어 일관된 내용을 나타내는 장면(scene) 및 여러 개의 장면으로 이루어진 이야기 단위(episode)들로 분할하는 것이다. 따라서, 비디오 데이타의 구조화를 위해서는 그 첫 단계로서 비디오 데이타에 존재하는 셧 경계 검출(shot boundary detection)이 정확하게 이루어져야 한다.

셧 경계 검출 방법은 현재 많은 연구가 진행되어 왔으나 아직까지 셧에 존재하는 내용을 기반으로 한 일반적인 방법이 제시되지 못하고 있다. 주로 연속된 비디오 프레임 사이의 칼라 히스토그램 변화에 의한 방식이 주로 사용되고 있다. 하지만, 사용자가 원하는 정보를 자유자재로 처리하기 위해서는 비디오 데이타에 존재하는 의미있는 특징(meaningful features)들을 추출하는 것이 필요하다. 이렇게 추출된 특징들을 바탕으로 비디오 데이타의 분석 및 구조화를 수행하는 것은 이러한 정보를 바탕으로 고급(high level) 개념인 비디오 데이타의 의미(semantics)를 해석할 수 있는 단서를 제공해 줄 수 있기 때문에 매우 중요한 작업이다. 더욱이, 이러한 내용 기반의 처리를 압축 데이타에서 직접 수행할 수 있다면, 시간 및 오버헤드를 줄일 수 있어 바람직하다.

본 논문에서는 비디오 데이타에서 중요한 특징들인 영역 정보 및 칼라 정보를 이용하여 MPEG 데이타를 원 영상으로 복호화하지 않고, 셧 경계를 검출하는 방법을 제안한다. 2 장에서는 셧 경계 검출에 관한 지금까지의 수행된 연구를 비교하고, 3 장에서는 의미있는 특징들인 영역 정보 및 칼라 정보를 추출하는 방법을 기술하고, 4 장에서는 이런 특징들을 이용한 셧 경계 검출 알고리즘을 설명하며, 5 장에서는 실험 결과를 기술한다.

2. 기존의 셧 경계 검출 연구

효율적인 비디오 인덱싱을 위해 셧 경계 검출을 위한 다양한 연구가 진행되어 왔다. 비디오 데이타의 셧 경계는 하나의 셧이 끝난 후 바로 다음 셧이 시작되는 급격한 셧 전환(abrupt shot transition)과 페이드 인/아웃이나 디졸브(dissolve)와 같은 점진적 셧 전환(gradual shot transition)의 두 가지로 구성되어 있다. 이러한 셧 경계 검출에 대한 연구는 화소 영역(pixel domain or uncompressed domain)과 압축 데이타 영역(compressed domain)의 두 개의 분야에서 연구되어 왔으나, 향후의 비디오 데이타는 주로 압축된 형태로 저장되므로 압축 비디오 데이타에서 직접적으로 처리할 수 있는 알고리즘을 개발하는 것이 바람직하다.

먼저, 화소 영역에서의 셧 경계 검출에는 연속된 두 프레임 간의 대응되는 화소 사이의 차이의 합이 임계값 이상이 되는 프레임을 셧의 경계로 정하는 방법이다[2]. 또, 점진적인 셧 전환을 검출하기 위해서는 두 개의 임계값을 정하여 이중 비교 방식(twin-comparison)을 사용하였다. 이런 화소 차이를 계산하는 방식은 간단한 것이 장점이나 잡음(noise)이나 오브젝트(object)의 이동에 민감하지 못한 단점이 있다. 이를 해결하기 위한 방법으로 Otsuji와 Tonomura[3]는 연속된 프레임간의 급격한 히스토그램 차이를 이용한 셧 경계 검출 방법을 제안하였고, Nagasaka와 Tanaka[4]는 한 개의 프레임을 여러 개의 겹치지 않는 영역으로 나누어 밝기 및 칼라 히스토그램이나 템프렛 매칭(template matching)을 이용하여 셧의 경계를 구하는 방법을 제안하였다. Zabih *et al.*[5]은 새로 발생하는 에지와 사라지는 에지의 변화 비율을 계산하여 디졸브를 검출하는 방법을 제시하였다. Naphade *et al.*[6]은 칼라 히스토그램 차이와 대응하는 화소 차이 값들을 이용하여 k-means 클러스터링 기법을 사용하여 셧 경계를 검출하였다. unsupervised learning에 의해 임계값을 사용하지 않고 검출하는 것이 장점이지만, 처리해야 할 프레임 개수가 많지 않을 때는 정확도가 많이 떨어지는 단점이 있다.

압축 영역에서의 셧 경계 검출 방법은 여러 가지 장점을 갖는다. 첫째로, 압축 영역에서는 처리할 데이타 양이 화소 영역보다 많이 감소하므로 알고리즘을 구현하는데 있어서의 오버헤드가 감소하고, 둘째로, 압축 영역에서는 일반적으로 작은 양의 데이타를 처리하기 때문에 시간을 절약할 수 있다. 끝으로, 압축된 데이타를 완전히 복호화할 필요가 없기 때문에 복호화하는데 걸리는 시간을 줄일 수 있는 잇점이 있다. 압축 영역의 셧 경계 검출 방법으로서는 Arman *et al.*[7,8]이 JPEG으로 압축된 데이타의 선택된 영역에서 DCT 계수를 비교하여 셧 경계를 검출하였고, Zhang *et al.*[9]은 P 픽쳐 또는 B 픽쳐의 프레임에서 타당한 움직임 벡터의 개수를 계산하였다. 이러한 움직임 벡터의 개수가 임계값보다 적어 영(zero)에 가까우면, 카메라 움직임의 경계는 P 픽쳐 또는 B 픽쳐 형태의 프레임 전후에 존재한다고 제안하였다. Meng *et al.*[10,11]은 급격한 셧 경계 검출을 위해 움직임 벡터의 분포를 이용하였고, 점진적 셧 경계 검출을 위해서는 DCT의 DC계수의 편차를 이용하였다. Yeo and Liu[12]는 시간축 상으로 슬라이딩 윈도우(sliding window)를 이용하여 지역적인 활동성

(local activity)을 계산하여 셧 경계 검출에 이용하였다. Kobla 와 Doermann[13]은 연속된 프레임간의 유사도를 측정하기 위해 매크로 블록 단위의 움직임 보상 방법을 사용하였다.

3. 압축 비디오로부터 의미있는 특징 추출

기존의 셧 경계 검출에서는 주로 화소의 차이나 칼라 히스토그램 변화, DCT 계수 및 움직임 벡터등의 신호적인 특성을 이용하였으나, 내용 기반의 비디오 인덱싱을 수행하기 위해서는 비디오 프레임 안에 존재하는 중요한 특징을 추출하여 이용하는 것이 바람직하다. 본 논문에서는 MPEG 데이타로부터 직접 영역 정보 및 칼라 정보 등의 의미 있는 특징들을 추출하여, 이를 바탕으로 셧 경계 검출 방법을 제안한다. 즉, MPEG 데이타로부터 DC 영상을 복원하여서, 밝기 영상(Intensity image)은 필터를 사용하여 단순화하여 주 영역(dominant region)들을 추출하고, 이런 주 영역들의 시간 축 상의 불 연속점을 찾고, 칼라 영상은 인간 시각에 근거한 양자화 과정을 거쳐 시간 축 상의 칼라 히스토그램 변화를 계산하여, 이 두 가지 정보를 바탕으로 셧을 검출한다. 그림 1은 본 논문에서 제안한 셧 경계 검출 방법을 보여주고 있다.

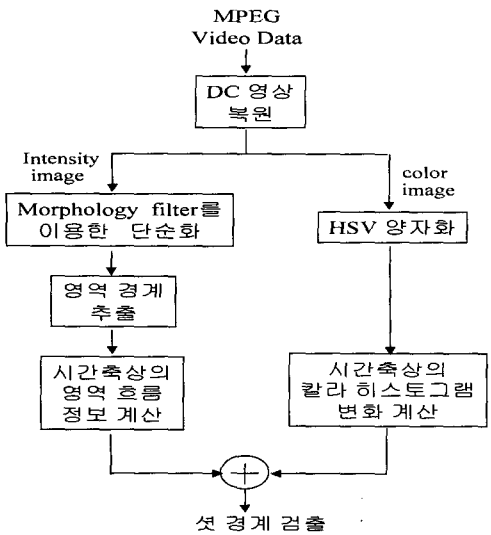


그림 1 셧 경계 검출 방법

3.1 DC 영상 구성

MPEG 비디오 데이타로부터 원 영상(original image)을 복호화하지 않고, 압축 데이타에 포함되어 있는 DCT 계수의 DC 항을 이용하여 구성한 영상을 DC 영상이라 한다. 이 DC 영상은 크기가 원 영상의 1/64로 축소되어 작지만, 원 영상이 가지고 있는 중요한 특징들을 가지고 있어서 영상의 분석에 유용하게 사용할 수 있다. DC 영상을 사용하는데 있어서 장점은 원 영상에 비해 빠른 처리가 가능한 것이다. 하지만, 크기가 작으므로 중요한 특징들을 처리하는데 있어서 민감하게 반응하는 단점이 있다. MPEG 데이타로부터 DC 영상의 복원은 픽쳐 타입에 따라 다르다. I 픽쳐 프레임으로부터 DC 영상은 간단히 복원하지만, P 픽쳐 프레임이나 B 픽쳐 프레임에서 DC 영상의 복원은 쉽지 않다. 여기서는 Yeo and Liu[12]의 알고리즘을 바탕으로 다양한 픽쳐 타입으로 구성된 MPEG 영상으로부터 DC 영상을 복원하였다. 그림 2는 원 영상과 1/64로 축소되었지만 원 영상의 중요한 특징들을 가지고 있는 DC 영상을 보여주고 있다.

3.2 DC 영상으로부터 의미있는 특징 추출

DC 영상에서 추출할 수 있는 의미있는 특징들로서는 오브젝트(object), 칼라 정보, 움직임 정보 및 텍스츄어(texture) 등을 들 수 있다. 일반적으로 오브젝트는 한 개 또는 여러 개의 영역으로 구성되어 있는 의미있는 단위이므로, 셧 경계 검출을 위해서는 오브젝트라는 의미 단위보다 이를 구성하고 있는 영역 정보 및 이에 관련된 칼라 정보를 추출하여도 충분하다. 또, 프레임의 특징을 모든 영역들을 이용하여 표현하기 보다는 몇 개의 커다란 영역을 가지고 표현하여도 셧 경계 검출에는 영향이 없으므로, 커다란 영역들을 주 영역(dominant region)으로 간주하고, 이들 주 영역들의 속성들을 비교함으로서 계산량을 줄일 수 있고, 효율적으로 셧의 경계를 검출할 수 있다. 이 절에서는 영역 정보 및 칼라 정보 추출에 관해 기술한다.

그림 2 원 영상과 DC 영상

3.2.1 영역 정보 추출

밝기 값으로 이루어진 DC 영상(DC intensity image)

으로부터 영역 정보를 추출하기 위해서는 DC 영상을 단순화하는 것이 필요하다. 단순화하는 방법에는 모폴로지 필터(morphology filter)를 사용하는 경우와 DC 영상 값을 양자화하는 두 가지가 가능하다. 첫 번째 방법으로, DC 영상에 모폴로지 필터를 사용하여 부분적인 재구축에 의한 opening 동작(opening by partial reconstruction)을 실행하면 많은 부분들이 단순화되지만 영역의 경계를 보존하고 있으므로 영역을 찾기가 쉬어진다[14].

$$\textit{Opening by partial reconstruction} = O^{rec}(E(I), O_k(I)) \quad (1)$$

이 작업은 영상 I 에 erosion$(E(I))$을 실행하여 필터 구조화 요소(structuring element)보다 작은 밝은 부분들을 제거하고, 제거되지 않은 부분들의 경계를 재 구축 과정$(O^{rec}(,))$을 통하여 보존하는 것으로서, 이때 기준 영상으로는 작은 구조화 요소로 수행한 opening의 결과인 $O_k(I)$를 사용함으로서 부분적인 재구축 과정이 이루어진다. 이러한 부분적 재구축 과정은 원래 영상인 I 를 기준 영상으로 하는 완전한 재구축 과정에 비해 3차원 신호의 처리에 있어서 영역의 경계를 잘 보존하는 성질을 가지고 있다[14].

또 하나의 단순화 방법은 밝기 값을 인간 시각 시스템에 근거하여 균등하지 않게 분할하여 양자화하는 것이다. 실제로 인간이 밝기 영상을 인식할 때, 밝기 값이 0에서 50사이의 변화 정도나 200 에서 250 사이의 변화 정도는 50 에서 200 사이의 변화 정도와 많은 차이가 있기 때문에, 밝기 변화 단계를 변화가 심하게 느끼는 영역에서는 작게 분할하고, 밝기의 변화가 적게 느껴지는 영역에서는 크게 분할하는 것이 바람직하다. 이렇게 양자화하면 원 영상의 급격한 밝기 변화가 단순화되는 영상을 얻을 수 있다. 이런 방식은 모폴로지 필터를 사용하는 경우보다 계산이 간단한 잇점이 있으나 복잡한 영상의 경우 중요한 특징을 추출하는데 어려움이 있다.

본 논문에서는 모폴로지 필터로 얻어진 단순화된 영상으로부터 먼저 편평한 영역(flat region)을 찾고 이러한 편평한 영역으로부터 watershed 알고리즘을 이용하여 영역의 경계를 찾는다[15]. 본 논문에서는 정확한 영역의 경계를 구하는 것 보다 중요한 영역의 흐름을 구하는 것이 목적이므로 watershed 알고리즘을 수정하여 사용하였다. 그림 3에서처럼 먼저 단순화된 영상으로부터 영상 값의 차이가 임계값(Th_{flat}) 이하인 편평한 영역을 찾기 위해 시드 화소(seed pixel)로부터 4-connectivity를 바탕으로 연결된 영역을 구한다. 이렇게 얻어진 편평한 영역들로부터 인접한 영역이 존재할 때에는, 두 인접한 영역의 밝기 값 차이가 임계값(Th_{adj}) 이하일 때 두 인접한 영역을 합친다. 그림 3(a)에서 A, B, C 및 D의 4개의 영역은 인접한 영역을 고려할 때 그림 3(b)처럼 "1", "2", 및 "3" 등의 세 개의 영역으로 라벨을 붙인다. 이렇게 함으로서 영상의 단순화 과정에서 모폴로지 필터 구조화 요소의 사이즈로 인해 발생하는 결점을 보상할 수 있다.

다음 단계로는 얻어진 편평한 영역을 시드(seed) 영역으로 시작하여 단계적으로 물을 채우기 시작하되 그 경사도가 가파르거나 다른 편평한 영역을 침범할 때에는 댐을 쌓아 더 이상 그 영역의 증가를 멈춘다(그림 3(c) 참조). 이렇게 하여 각각의 프레임에서 영역의 경계를 구한다(그림 3(d) 참조). 각각의 영역에 대해 밝기 값의 평균, 사이즈 및 영역의 중심 좌표등을 계산하여 영역 정보로서 저장한다.

$$\textit{Region} = \{\textit{Average Intensity, Size, Center Point}\} \quad (2)$$

수정된 watershed 알고리즘을 수행한 결과로 많은 영역들이 얻어지나, 연속된 프레임간의 오브젝트의 변화를 계산하기 위해서는 몇 개의 주 영역(dominant regions)들의 변화로도 충분하다. 본 논문에서는 각 프레임에서의 영역 정보로 영역의 사이즈의 크기에 의해 5개의 주 영역을 선정하였다.

$$\textit{Frame}[i] = \{\textstyle\sum \textit{Dominant Region}[k]\},$$
$$\textit{where } i = \textit{frame number}, \; k = 0,1,2,3,4 \quad (3)$$

3.2.2 칼라 정보 추출

칼라 정보도 비디오 데이타를 분석하는데 있어서 매우 중요하다. 칼라 정보를 처리하는데 있어서 RGB 칼라 스페이스를 사용할 수 있으나, RGB 칼라 스페이스에서는 일반적으로 두 개의 칼라의 근접도(proximity)가 두 칼라의 유사도(similarity)를 나타내지 않기 때문에, 칼라 정보 유사도 측정에 어려운 점이 많다. 더욱이, 24 비트의 RGB 칼라 정보를 사용한다면, 얻는 성과에 비해 계산하는데 많은 시간 및 자원을 소모하기 때문이다. 이러한 단점
을 보완하기 위해서 본 논문에서는 HSV(Hue, Saturation, Value) 칼라 스페이스를 사용하였다. 이 칼라 스페이스는 자연적이어서, 다른 조명 조건하에서도 색상(Hue)이 변치 않는 장점을 가지고 있다. HSV 의 값은 보통 RGB의 비선형 변환으로 구할 수 있다[16].

그 요소들은 다음과 같이 계산할 수 있다.

$$H=\arccos\left\{\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right\}$$
$$V=\max(R,G,B) \quad (4)$$
$$S=\max(R,G,B)-\min(R,G,B)$$

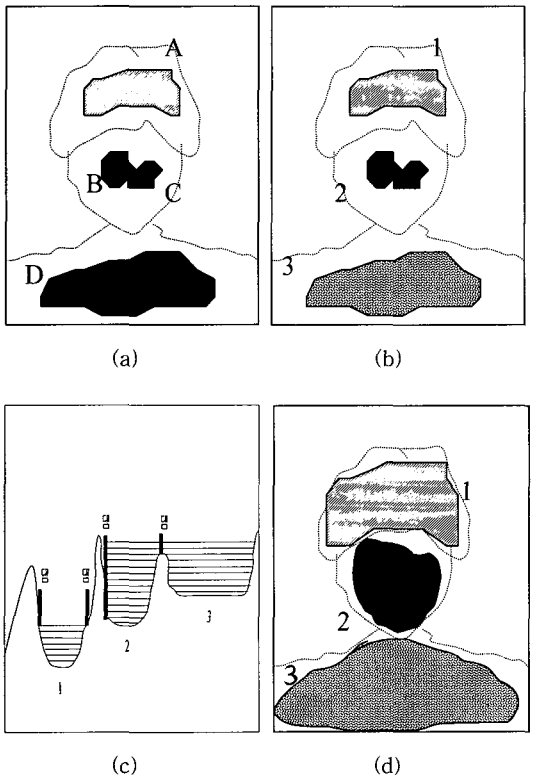


그림 3 (a) 편평한 영역 검출, (b) 인접도에 의한 편평한 영역 추출, (c) 수정된 watershed 알고리즘, (d) 영역 검출

이런 HSV 칼라 스페이스에서의 칼라 영상은 저 색상 영역(low chromatic intensity region), 고 색상 영역(high chromatic intensity region) 및 무 색상 영역(achromatic region)으로 나눌 수 있다[17]. 저 색상 영역에서는 채도 레벨이 낮고 색상(hue) 값이 불안정하다. 하지만, 고 색상 영역에서는 채도 레벨이 높고, 색상 값이 안정적인 성질을 갖고 있다. 또, 무 색상 영역에서는 색상 값이 정의되지 않는다. 이러한 특성을 이용하여 본 논문에서는 HSV 칼라스페이스를 인간의 지각에 근거하여 분할하였다. 먼저, V 및 S 값을 비 균등하게 나누고, 각각의 V 및 S 값에 따라, 저 색상 영역에서는 H(hue) 값을 크게 분할하여 양자화 하였고, 고 색상 영역에서는 H 값을 세분하여 양자화 하였다. 그림 4는 인간 시각 시스템에 근거한 HSV 칼라 양자화를 보여주고 있다. 이런 방식으로 우리는 DC 영상의 칼라 정보를 양자화를 통해 97개의 상자(bin)로 저장한다.

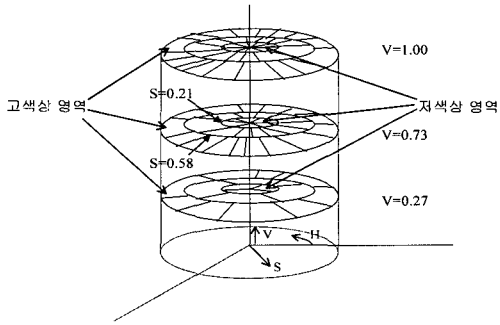


그림 4 인간 시각 시스템에 근거하여 양자화된 HSV 칼라 스페이스

4. 영역 정보 흐름 및 칼라 정보를 이용한 셧 경계 추출

셧 경계란 일반적으로 장면이 바뀌는 곳으로 시간 축 상에서의 영역의 흐름이 불연속한 곳이나, 칼라 정보의 급격한 변화가 생기는 곳에서 발생한다. 이 장에서는 앞 장에서 기술한 영역 정보 및 칼라 정보를 이용하여 내용 기반의 비디오 셧 경계 검출 방법에 대해 기술한다.

4.1 영역 정보 흐름

공간적으로 분할된 영역을 바탕으로 시간 축 상의 영역의 흐름을 찾는 것이 셧 경계 검출에 매우 중요하다. 시간 축 상의 영역의 흐름을 찾기 위해서는 먼저 기준 프레임의 각각의 영역에 대해 다음 프레임의 영역들에 대한 대응 관계(correspondence)를 찾아야 한다. 일반적으로 대응 관계를 찾는 것은 매우 어려운 문제이지만, 비디오 데이타에서는 프레임간에 오브젝트의 움직임이나 사이즈의 변화가 크지 않고, 또 밝기 값의 변화도 크지 않는 특성을 가지고 있으므로, 기준 프레임의 영역을 다음 프레임에 투사하여 해당되는 영역부근에서 대응 영역을 찾는 것이 바람직하다. 다시 말해서, 현재 프레임의 경역이 다음 프레임에서도 비슷한 위치에 있고, 또 영역의 밝기나 사이즈에서 값의 차이가 임계 값 범위 안에 드는 영역을 찾는다면 대응되는 영역으로 간주하여 기준 프레임의 영역과 같은 라벨을 붙인다. 만약, 밝

기 정도나 사이즈가 임계값을 벗어나면 새로운 영역이 생긴 것으로 간주하여 새로운 라벨을 붙인다. 그림 5는 이러한 과정을 보여 준다.

대응하는 영역을 구하는 데 있어서 문제점은 하나의 영역이 다음 프레임에서 두 개 또는 그 이상의 영역들과 대응할 경우(1-to-N)와, 두 개 또는 여러 개의 영역이 하나의 커다란 영역과 대응(N-to-1)할 경우이다(그림 6 참조). 먼저, 1-to-N인 경우에는 그림 6(a)처럼 인접한 영역에는 같은 라벨을 붙이고, 영역의 정보도 공유한다. 하지만, 분리된 경우는 큰 영역에 같은 라벨을 붙이고, 그 외의 작은 영역들에게는 새로운 라벨을 붙인다. 왜냐하면, 영역의 흐름은 주 영역을 찾아서 계산하므로 작은 영역들의 흐름은 무시할 수 있기 때문이다. 두 번째의 경우로, 그림 6(b)처럼 N-to-1로 대응할 경우에는 같은 이유로 기준 프레임의 커다란 영역의 라벨을 다음 프레임의 대응 영역에 붙인다. 이런 과정을 반복하여 비디오 데이타로부터 시간 축 상으로 주된 영역의 흐름을 구한다.

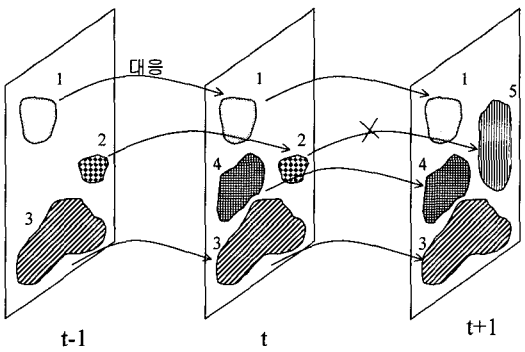


그림 5 연속된 프레임간의 대응 영역 찾기

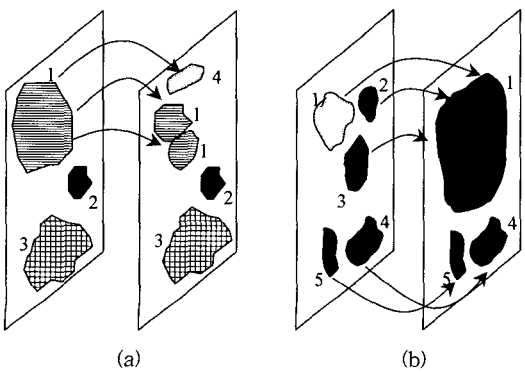


그림 6 (a) 1-to-N 대응, (b) N-to-1 대응

주 영역의 흐름에 관한 정보는 시간 축 상에서 시작 프레임의 번호와 영역이 끝나는 프레임의 번호를 저장하고 있다.

$$Region_Flow[i] = (start\ frame\ number,\ end\ frame\ number) \quad (5)$$

각각의 주 영역들에 대해 영역 흐름 정보를 이용하여 불연속한 곳을 찾으면 셧 경계를 검출 할 수 있다.

4.2 양자화된 칼라 정보

계산량을 줄이기 위해 또, 잡음에 의한 오류를 줄이기 위해 양자화 방법이 사용되어 왔다. 특히, HSV 칼라 스페이스 경우 Smith *et al.*[18]등은 H(hue) 값을 20 도씩 구분해서 18개로 나누었고, S(Saturation) 및 V(Value) 값도 같은 간격으로 균등하게 3 단계로 나누었다. 그리고, R, G, B의 세 개의 값이 모두 같아서 H 값이 정의되지 않는 경우에는 4단계로 나누어서 166개(18x3x3 +4)로 구분하였다. 하지만 이 방법은 계산하기는 간단하지만, 인간의 시각 인지도를 고려하지 않았기 때문에, 인간이 민감하게 느끼는 부분과, 민감하게 느끼지 못하는 부분을 구별하지 못하였다. 이러한 단점을 보완하기 위해서 우리는 HSV 칼라스페이스를 3.2.2에서 기술하였듯이 비 균등하게 양자화 하였다. 실험에 의해서 97개의 값으로 양자화하는 것이 DC 영상을 구별하는데 적합하다는 것을 알았다. 이러한 방법은 R, G, B의 각각의 256개의 칼라에 대해 총 768개의 칼라 히스토그램을 비교한 것이나[6], 166개의 균등화된 양자화 비교[18] 보다도 훨씬 적은 정보를 사용하여 구별할 수 있는 장점을 가지고 있다. 양자화된 칼라 정보의 히스토그램 변화는 두 방식으로 계산한다. 첫 번째 방식은 셧 경계를 검출하기 위한 것으로서 연속된 두프레임 사이의 히스토그램의 차이를 각각의 상자(bin)에 해당되는 화소의 개수 차이로 계산하고(h1), 다음으로 점진적인 셧 전환을 찾기위해 기준 프레임과 현재 프레임과의 히스토그램 값의 차이(h2)를 계산한다.

$$h1 = \Sigma\ diff(h_j[i], h_{j+1}[i]) \quad where\ i = 0,....96,\ j = frame\ number, \quad (6)$$
$$h2 = \Sigma\ diff(h_{ref}[i], h_{curr}[i]) \quad where\ ref = reference\ frame,\ curr = current\ frame.$$

여기서, $h2$의 계산에 있어서 기준 프레임 h_{ref}에 대한 초기 값은 시작 프레임이고, 셧 경계가 검출되면 새로운 셧의 첫 번째 프레임으로 정한다.

4.3 셧 경계 검출 방법

앞에서 기술한 영역 흐름 정보와 양자화된 HSV 칼라의 히스토그램 변화를 이용하여 셧 경계를 검출할 수 있다. 먼저, 각각의 프레임에서 주 영역 5개를 선정하고, 이 영역들의 시간 축 상의 변화도를 측정한다. 주 영역들이 시간 축 상으로 연속성을 가지고 있으면 장면 변화가 존재하지 않고, 이 영역들의 흐름이 끊어질 때가 장면 변화가 생길 가능성이 높다. 이때, HSV 칼라 히스토그램 변화를 참조하여 점진적이거나 급격한 셧 경계를 결정한다. 그림 7은 셧 경계 검출 방법을 보여주고 있다.

그림 7에서 보면 주 영역들이 시간 축 상의 연속성을 가질 때에는 셧 경계가 존재하지 않는다. 하지만, 현재 프레임의 첫 번째 주 영역이 다음 프레임의 첫 번째 주 영역과 대응되지 않거나, 대응되는 영역이 존재하지 않아서 불연속일 때에는, 현재 프레임의 두 번째 주 영역과 세 번째 주 영역의 연속성을 체크한다. 만일 이 두 영역들이 연속성을 가지면 첫 번째 주 영역의 불연속으로 인해 점진적인 셧의 경계가 존재할 가능성이 있으므로, 칼라 히스토그램의 변화를 검사한다. 칼라 히스토그램의 변화(**h2**)가 점진적인 변화의 임계값 ***dissolve_th*** 보다 작으면, 첫 번째 영역의 불연속은 잡음이나 조명의 변화 때문일 가능성이 크므로, 셧 경계를 검출하지 않는다. 이 칼라 히스토그램의 변화가 임계값 ***dissolve_th*** 보다 크면 점진적인 셧 경계를 검출한다. 여기서 임계값 ***dissolve_th***는 급격한 셧 경계 임계값 ***cut_th***보다 작은 값이다. 만일, 첫 번째 주 영역이 불연속이고, 두 번째 또는 세 번째 영역 중 하나만 연속성을 가질 때에는, 세 개의 영역 중 두 개의 영역이 불연속이므로, 셧의 경계가 존재하는 것으로 간주하고, 프레임간의 변화가 많은 지를 검사하기 위해, 현재 프레임과 다음 프레임사이의 주 영역간의 대응되는 화소 개수와 현재 프레임의 주 영역들의 화소 개수의 합과의 차이가 임계값 ***num_th***와 비교하여 급진적인 셧 경계 또는 점진적인 셧 경계를 결정한다. 즉, 5개의 주 영역 중 두 개의 영역이 불연속 하더라도 대응되는 영역의 화소 수가 크게 차이가 나지 않으면 칼라 히스토그램(**h1**)을 비교하여 변화량이 임계값 ***cut_th***보다 적으면 점진적인 셧 경계이고, 크면 급진적인 셧 경계로 판단한다. 만일 대응되는 영역의 화소 수의 차이가 많이 나면 다른 주 영역(네번째 주 영역 및 다섯 번째 주 영역)들도 불연속이므로 급진적인 셧 경계가 검출된다.

만일, 첫 번째 주 영역이 연속성을 가질 때에는, 두 번째 또는 세 번째 주 영역중 하나라도 연속성을 가지면, 점진적인 셧 경계의 가능성이 있으므로 칼라 히스토그램에 따라 결정한다. 또, 첫 번째 주 영역이 연속이지만, 두 번째 주 영역과 세 번째 주 영역 둘 다 불연속일 때는 현재 프레임과 다음 프레임과의 대응 영역의 화소의 수를 계산하여 현재 프레임의 주 영역들의 화소 수의 합과의 차이가 임계값 ***num_th*** 보다 작을 때에는, 칼라 히스토그램 변화량(**h1**)에 따라 점진적인 셧과 급진적인 셧 경계를 검출한다. 이런 조건을 만족하지 않을 때에는 셧 경계가 존재한다. 그림 7은 이와 같은 셧 경계 검출 방법을 보여주고 있다.

5. 실험 결과

앞 장에서 제안한 알고리즘을 이용하여 MPEG 비디오 데이타를 원 영상으로 복호화하지 않고 직접 셧 경계를 검출하였다. 먼저, MPEG 비디오 데이타로부터 DC 영상만을 복원하여, 모폴로지 필터의 opening작업(opening by partial reconstruction)을 통하여 DC 영상을 단순화 시켰다. 단순화된 영상으로부터 수정된 watershed 알고리즘을 이용하여 주 영역들을 계산하였다. 각 프레임 당 5개의 주 영역들과 이들의 밝기 값 평균, 사이즈, 그리고 중심 좌표등을 계산하였다. 이런 공간적인 분할을 통하여 얻어진 영역을 다음 프레임에 투사하여 시간 축 상의 영역들간의 대응 관계를 계산하였다. 계산된 대응 관계를 바탕으로 영역 흐름 정보를 계산하였다. 그림 8은 두 셧 경계에서의 영역 정보의 흐름을 보여 주고 있다. 그림 8(a)는 mjackson 비디오 시컨스 #10, #11, 및 #12의 원 영상을 보여주고 있고, 그림 8(b)는 압축 데이타로부터 구축한 DC 영상들을 보여주고 있으며, 그림 8(c)는 주 영역으로 분할 된 영상을 보여주고 있다. #10 프레임과 #11 프레임은 영역들이 연속적이고, #12 프레임에는 불연속이어서 셧의 경계가 존재하고 있음을 알 수 있다.

칼라 정보는 RGB 칼라 스페이스에서 비선형 변환을 통해 HSV 값을 계산하고 이를 인간 시각시스템에 근거하여 양자화하여, 각각의 프레임간의 칼라 히스토그램 변화를 계산하였다. 양자화된 HSV 영상의 칼라 히스토그램의 변화는 급격한 셧 경계 및 점진적 셧 경계를 구하기 위해 식 (6)을 이용하여 두 가지 방식으로 계산하였다. 이렇게 계산된 영역 흐름 정보 및 칼라 히스토그램 변화를 바탕으로 그림 7과 같은 알고리즘에 의해 셧 경계를 검출하였다.

제안한 검출 방법의 시뮬레이션을 위해 다양한 장면들로 구성된 뮤직 비디오인 "mjackson.mpg", 줌 인/아웃으로 구성된 에니메이션 "wg_cs_5.mpg", 디졸브 효

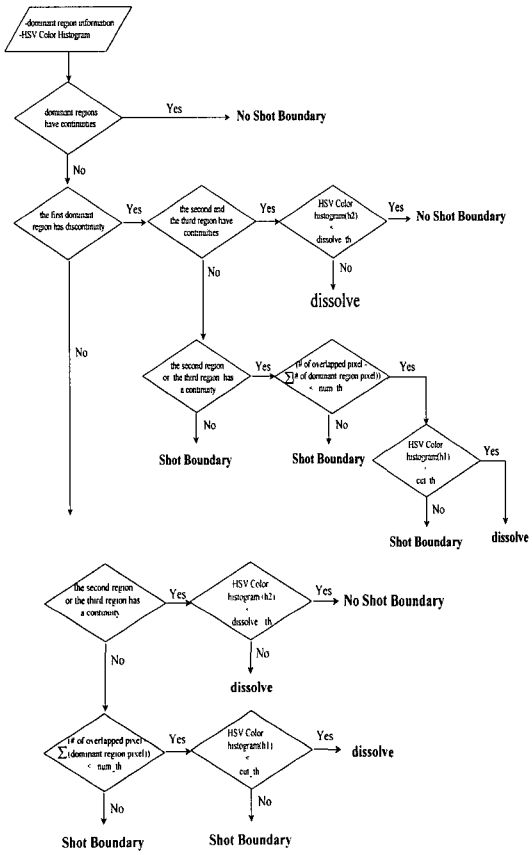


그림 7 셧 경계 검출 방법

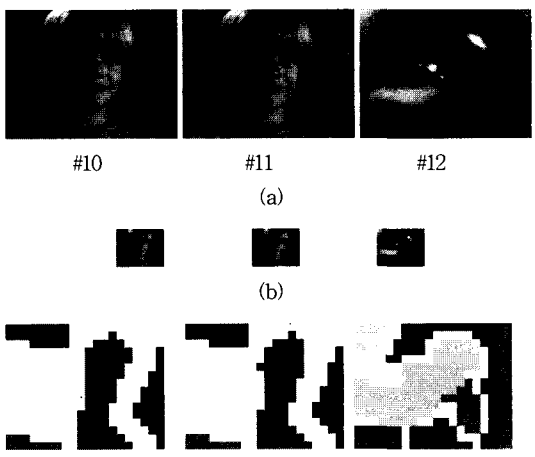
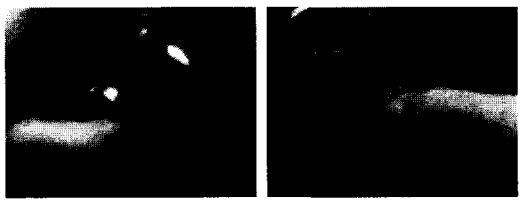


그림 8 셧 경계검출: (a) 원영상, (b) dc 영상, (c) 영역 정보 흐름

과를 갖고 있는 "clapton.mpg" 및 셧 경계 없이 카메라의 움직임만 있는 "flower.mpg"등을 이용하여 셧 경계를 검출하였다. 표 1은 검출 결과를 보여준다. 여기서 recall은 실제 장면 전환이 일어난 것 중에서 검출된 장면 전환에 대한 비율이고, precision은 검출 된 것 중 실제 장면 전환인 것의 비율을 나타낸다.

표 1 셧 경계 검출 결과

비디오데이타	Detection	Misdetection	Fault	Recall	Precision
mjackson1.mpg (180 frames)	21	1	1	0.95	0.95
wg_cs_5.mpg (270 frames)	5	0	1	1.0	0.83
clapton.mpg (150 frames)	2(dissolve)	0	0	1.0	1.0
flower.mpg (100 frames)	0	0	0	1.0	1.0



(a) (b)

그림 9 (a) mjacskon1 #14 프레임, (b) mjacskon1 #15 프레임

$$\textbf{\textit{recall:}}\ \frac{d}{d+m} \tag{7}$$

$$\textbf{\textit{precision:}}\ \frac{d}{d+f},$$

여기서 d(detection): 실제 셧 전환이 일어났고 또 검출도 된 개수,
m(misdetection): 실제 셧 전환이 일어났지만 검출되지 않은 것,
f (fault): 실제 셧 전환이 일어나지 않았지만, 검출 된 것이다.

본 논문에서는 칼라 정보이외에 영역 정보를 사용함으로서 칼라 히스토그램이 가지고 있는 단점도 보완할 수 있다. 예를 들어, 그림 9는 "mjackson1.mpg"의 #14

와 #15 프레임으로 칼라 히스토그램은 같으나 영역의 흐름이 다르므로, 본 논문에서 제안한 방식에 의해 셧 경계가 검출된다. 더욱이, MPEG 비디오 데이타를 원 영상으로 복호화하지 않고, DC 영상만을 추출함으로서 복호화 과정의 IDCT 모듈을 수행하지 않을 뿐더러, DC 영상들을 처리함으로서, 셧 경계 검출을 신속히 할 수 있다. 예를들어, "mjackson1.mpg"의 경우 1296 kbit(160 x 120 x 180 프레임) 데이타를 57.6kbit(20 x 16 x 180 프레임) 데이타의 처리로 가능함으로 1/60 정도로 데이타 처리 시간을 줄일 수 있다.

6. 결 론

본 논문에서는 영역 흐름 정보 및 칼라 정보를 이용하여 압축 비디오 데이타로부터 내용 기반의 셧 경계 검출 방법에 관한 알고리즘을 제안하였다. 제안된 방법은 압축 데이타를 원 영상으로 복호화하지 않고, DC 영상들을 구성하여 이들로부터 중요한 영역들의 흐름 및 칼라 정보 변화를 계산하여 셧 경계를 검출하였다. 제안된 방법은 비디오 데이타의 각 프레임에서 중요한 영역들에 대한 정보를 저장하고 있으므로, 이들을 바탕으로 주석(annotation)을 붙이는데도 사용할 수 있고 고급 특징들과의 결합을 통한 영상의 의미적인 해석(semantic meaning interpretation)에도 이용할 수 있다. 아울러 사이즈가 약 1/64 정도로 축소된 DC 영상에서 처리되므로, 획기적으로 시간을 절약할 수 있다. 이러한 방식은 향후 의미있는 오브젝트의 추출 및 변형(manipulation)을 수행하거나 이를 이용한 인덱싱에 있어 중요한 역할을 할 수 있으며, 비디오 단락의 대표 프레임 선정에도 유용한 결과를 제공할 수 있다.

참 고 문 헌

[1] W. Grosky, R. Jain and R. Mehrotra, "The Handbook of Multimedia Information Management," Prentice Hall PTR, 1997.

[2] H. Zhang, A. Kankanhali, S. Smoliar and S. Tan, "Automatic Partitioning of Full Motion Video," *ACM Multimedia Systems,* Vol. 1., No. 1, pp 10-28, 1993.

[3] K. Otsuji and Y. Tonomura, "Projection Detecting Filter for Video Cut Detection," In Proc. 1st ACM Conf. Multimedia, 1993.

[4] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-video Search for Object Appearances," In 2nd Working Conf. Visual Database System, Oct. , pp. 119-133, 1991.

[5] R. Zabih, J. Miller and K. Mai, "Video Browsing Using Edges and Motion," *IEEE Conf. CVPR '96,* Sanfranciso, Jun., pp.439-446, 1996.

[6] M. Naphade et. al,"A High Performance Shot Boundary Detection Algorithm Using Multiple Cues," IEEE Int. Conf. Image Processing, Chicago, 1998.

[7] F. Arman, A. Hsu and M. -Y. Chiu, "Feature Management for Large Video Databases," in Storage and Retrieval for Image and Video Databases, Vol. SPIE 1908, pp. 2-12, 1993.

[8] F. Arman, A. Hsu and M. -Y. Chiu, "Image processing on Compressed Data fro Large Video Databases," *Proc. ACM Multimedia'93,* Jun., pp. 267-272, 1993.

[9] H. J. Zhang, C. Y. Low and S. W. Smolia, "Video Parsing and browsing using compressed data," Multimedia Tools and Applications, Vol. 1, No.1, Mar., pp. 89-111, 1995.

[10] J. Meng, Y. Juan, and S.-F Chang, "Scene change detection in a MPEG compressed video sequence," *Proc. IS&T/SPIE Symposium in Electronic Imaging: Science and Technology,* San Jose, CA, Feb. 1995.

[11] J. Meng and S. -F. Chang, "CVEPS- A Compressed Video Editing and Parsing System," ACM Multimedia'96, Nov., pp.43-53, 1996.

[12] B. Yeo and B. Liu, "On the Extraction of DC sequence from MPEG compressed video," *Proc. Int. Conf. on Image Processing,* Oct, pp.260-263, 1995.

[13] V. Kobla and D. Doermann, "Compressed domain video indexing techniques using DCT and Motion Vector information in MPEG video," SPIE Vol. 3022, pp.200-212, 1997.

[14] M. Pardas and P. Salembier, " 3D Morphological Segmentation and Motion Estimation for Image Sequences," *Signal Processing,* Vol. 38, pp. 31-43, 1994.

[15] L. Vincent and P. Sollie, "Watersheds in Digital Spaces: An Efficient Algorithms based on Immersion Simulation," *IEEE Trans. Pattern and Mach. Intell.,* Vol. 13, No. 6, pp. 583-598, June 1991.

[16] M.D. Fairchild, Color Appearance Models, Addison Wesley, 1998.

[17] S. Ishibashi and F. Kishino, "Color/Texture Analysis and Synthesis for Model-based Human Image Coding," SPIE VCIP'91: Visual Communication, Vol. 1605, pp. 242-249,1991.

[18] J. Smith and S. Chang, "Tools and techniques for color image retrieval," In sym. Electronic Imaging: Storage & Retrieval for Image and Video Databases IV, Vol. 2670, pp. 426-437, 1996.

강 행 봉

1980년 한양대학교 전자공학과 졸업(학사). 1986년 한양대학교 대학원 전자공학과 졸업(석사). 1989년 미국 Ohio State University 컴퓨터 공학 석사. 1993년 미국 Rensselaer Polytechnic Institute 컴퓨터 공학 박사. 1994년 ~ 1997년 삼성 종합기술원 수석연구원. 1997년 ~ 현재 가톨릭대학교 컴퓨터전자공학부 조교수. 관심분야는 컴퓨터 비전, 멀티미디어 시스템, 인공지능 및 인터넷 응용 등