

MPEG 비디오 시퀀스에서 DC성분의 공간벡터를 이용한 컷 검출

정회원 최인호*, 구동수**, 이대영***

A Cut Detection Algorithm by Using Spatial Vectors of DC Components on MPEG Video Sequence

In-ho Choi*, Dong-soo Koo**, Dae-young Lee*** *Regular Members*

요 약

압축된 비디오 데이터에서 내용기반 컷 검출을 위해 다양한 특징 벡터 추출 방법이 연구되고 있다. 특징 벡터로써 화소값의 히스토그램을 이용한 방법의 경우 화소의 공간적 특성을 고려하지 않아 정확한 컷 검출을 기대하기 어렵다. 그래서 CCV(Color Coherent Vector)나 Color Correlogram 등의 계산량이 복잡한 알고리즘이 많이 사용된다. 그러나 이러한 기법들은 정확한 컷 검출을 가능하게 하나 계산량이 너무 복잡하기 때문에 검출 시간이 많이 소요되는 단점이 있다. 본 논문에서는 MPEG 비디오 시퀀스에서 휘도성분의 DC값들의 공간적 상관도를 이용한 컷 검출 기법을 제안한다. 이 기법은 비교적 간단하기 때문에 처리 시간이 빠르며 또한 개선된 특징차 비교방법을 이용하여 검출율을 더 높일 수 있다.

ABSTRACT

Various techniques extracting feature vectors have been studied for the cut detection in compressed video data. In case of using the histogram of occurrence of pixel's values as a feature vector, the precise detection of cuts would not be expected because of not considering the spatial correlation of pixels. And more sophisticated algorithms such as CCV(Color Coherent Vector) and Correlogram tend to be used. Though these methods can be able to detect cuts rather precisely, they require much more processing time because of a enormous amount of computations. In this paper we propose a method of the cut detection using spatial correlation of DC values of luminance components in MPEG video sequence. This requires less processing time and also It can increase the rates of detecting the correct cuts by using advanced comparative method.

1. 서론

최근 들어 주문형 비디오(VOD), 디지털 라이브러리, 전자상거래, 의료 비디오 검색 시스템, 방송프로그램 편집 시스템, 무인 카메라에 의한 물체 인식 등 멀티 미디어 서비스 요구가 증대되고 있다. 이는

방대한 데이터베이스 시스템 구축을 필요로 하는데 문서 이외 특히 비디오의 색인 검색은 매우 어려운 일이다. 이에 최근 연구 흐름도 비디오 데이터의 내용을 분석하여 인덱스를 구성하고 질의를 주는 내용기반 검색에 대한 연구가 활발히 진행중이다.^{[1][2][3]}

특히 동영상 데이터에 대한 내용기반검색은 MPEG

* 김포대학 전자정보계열(inhochoi@unitel.co.kr),

** 경희대학교 전자공학과(caster@nms.kyunghee.ac.kr),

*** 경희대학교 전자공학과
논문번호: 99323-0814,

접수일자: 1998년 8월 14일

7등에서도 제안되었다.

동영상 데이터를 내용기반의 검색을 하기 위해 제일 먼저 동영상데이터를 계층적으로 내용별 분할해야 한다. 그리고 분할된 동영상 데이터에 색인을 다는 절차가 필요하다. 내용별 분할을 위해 한 장면에서 다른 장면으로 변환되는 점, 즉 컷(cut) 검출이 선행되어야 한다.

컷 검출 알고리즘은 샷(shot)을 분류하는 중요한 처리이다. 일반적인 컷 검출 알고리즘은 이전 프레임과 현재 프레임을 단순히 색상의 히스토그램을 이용한 비교를 하기 때문에 컷을 오검출하는 경우가 많다. 뿐만 아니라 프레임 내 중심사물의 급격한 이동이나 플래시에 의한 밝기변화, 페이드 인(fade in)이나 페이드 아웃(fade out)등 컷의 지점이 불분명할 경우 정확한 컷 검출을 기대하기는 힘들다. 따라서 최근에는 히스토그램을 비롯한 윤곽(edge), 질감(texture), 모양(shape)등 영상을 이루는 여러 특징들을 이용한 컷 검출이 시도되고 있다. 특히 히스토그램에 공간적 정보를 고려하여 특징벡터를 추출하는 연구가 많이 이루어지고 있는데 CCV(Color Coherent Vector)나 Color Correlogram 등이 그것이다.^{[4][5]}

제안하는 특징벡터 추출 알고리즘은 각 화소값의 출현빈도는 물론이고 화소간 거리벡터를 포함하고 있으므로 단순한 히스토그램을 이용한 유사도 비교나 CCV 알고리즘 보다 훨씬 우수한 결과를 도출할 수 있다.

본 논문에서는 특징벡터 추출 전에 선행되어야 할 전처리과정과 휘도 성분에서 공간적 상관관계를 고려한 특징벡터 추출 방법 및 특징차 비교방법을 각각 II장과 III장에서 제안하였고 V장의 모의 실험 및 고찰을 통해 제안 알고리즘의 우수성을 확인하였다.

II. 전처리 과정

본 논문에서는 검색시간 절약을 위해 B 또는 P 픽처를 제외한 I 픽처의 DC계수만을 복호화하여 DC영상을 재구성하는 DC추출 방법을 쓴다.^[6] 일반적으로 TV 프로그램 등의 비디오 시퀀스에서 컷의 발생 빈도는 약 5초 정도이다. 초당 29.97프레임인 NTSC용 MPEG은 하나의 GOP에 보통 15개 이하의 픽처를 포함하고 한 GOP에서 I 픽처가 적어도 1번은 출현하며 각 GOP의 대표 프레임이라 할 수 있기 때문에 I 픽처만으로 컷을 검출하여도 크게 무리는 없다.^[6]

I 픽처에서 8×8 블록의 화소값을 대표하는 DC 계수만을 이용하여 복호화하므로 히스토그램 형태는 매우 첨예한 가우시안 곡선 형태를 가진다. 이러한 형태는 휘도 성분의 양자화 후 특징 벡터를 추출했을 때 대부분의 영상이 비슷한 양자화 값을 가져 특징차를 구해도 큰 차이가 나지 않는다. 이를 보완하기 위해서 각각의 DC영상에 대해 명도대비 스트레칭(contrast stretching) 처리를 한다. 명도대비 스트레칭 방법에는 기본 명도 대비 스트레칭(식 1)과 Ends-in search(식 2)를 이용하는 방법이 있다.

$$new\ pixel = \frac{old\ pixel - low}{high - low} \times 255 \quad (1)$$

$$out\ put(x) = \begin{cases} 0 & (\text{for } x \leq low) \\ 255 \times (x - low) / (high - low) & (\text{for } low \leq x \leq high) \\ 255 & (\text{for } high \leq x) \end{cases} \quad (2)$$

이러한 처리를 통해 중간값에 모인 히스토그램 형태를 다음 그림과 같이 0 에서 부터 255 레벨값까지 분포를 확산시킬 수 있다.

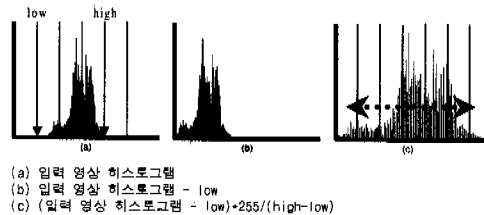


그림 1. 명암 대비 스트레칭

다음과정으로 불필요한 색상의 변화나 잡음을 없애고 계산량과 처리 시간을 줄이기 위해 색상의 블러링(blurring)과 양자화(quantization)를 하게 된다. 이 과정을 통해 급격한 화소값의 변화를 제거할 수 있다.

블러링은 가우시안 함수특성을 갖는 다음식을 이용한다.

$$G[x, y] = \frac{e^{-\frac{(x^2 + y^2)}{2\pi\sigma^2}}}{2\pi\sigma^2} \quad (3)$$

III. 휘도성분의 공간적 상관도

전처리과정 후 그림 2와 같이 각각의 동일한 화소값으로 하나의 계층(layer)을 만든다. 즉 같은 값

을 갖는 화소를 이용해 하나의 이진 영상을 만드는 것이다. 하나의 DC영상이 16레벨로 양자화 되었다면 16개의 계층과 특징 벡터를 구할 수 있다. 계층 분리를 통해 특징 벡터 추출과정에서 계산량을 줄일 수 있다.

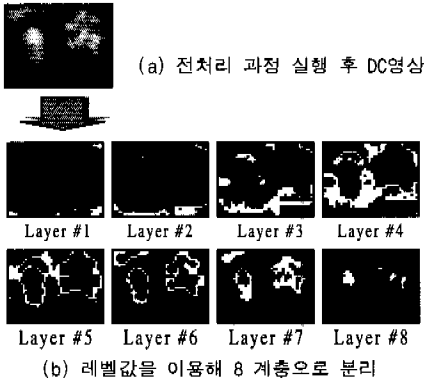


그림 2. 전처리 영상의 계층분리의 예

계층 분리 후 각각의 계층에 대해 특징 벡터 추출을 하게 된다. 특징벡터 추출 없이 이웃 프레임과 각각의 계층에 대해 일대일 대응되는 화소값을 L_1 -Distance를 이용해 두 프레임간 유사도를 측정할 수도 있으나 B 픽처와 P 픽처를 추출하지 않았으므로 재구성된 DC영상간에는 확대나 축소, 회전, 이동 등으로 인한 변화가 일어났을 수도 있다. 따라서 특징 벡터 추출 없이 단순히 L_1 -Distance를 이용하게 되면 컷 지점은 물론이고 확대나 축소, 또는 이동 등의 변화가 일어나는 지점에서 컷으로 오인할 수 있다.

DC영상에서 휘도 성분의 공간적 상관 관계를 특징 벡터로 만들기 위해 다음의 제안 방법으로 처리한다.

행과 열이 각각 $n \times n$ 인 DC영상 τ 는 전처리 과정에서 m 레벨로 양자화 되었으므로 그 값들은 $(y_1, y_2, y_3, \dots, y_m)$ 형태가 된다. τ 내 화소를 p 라 한다면 $p=(x, y) \in \tau$ 로 정의할 수 있다. 다시 $\tau(p)$ 를 화소 p 가 가지는 화소값으로 정의하며 다음과 같이 표현한다.

$$\tau_j = \{p \mid \tau(p) = y\} \quad (4)$$

또한 화소간 거리를 구하기 위해 L_∞ -norm 공식을 이용하며 두 개의 화소가 $p_1=(x_1, y_1)$, $p_2=(x_2, y_2)$

와 같을 때 두 화소간 거리는 다음과 같이 된다.

$$|p_1 - p_2| = \max\{|x_1 - x_2|, |y_1 - y_2|\} \quad (5)$$

또한 $\{1, 2, 3, 4, \dots, m\}$ 을 $[m]$ 으로 표기한다면 앞의 정의들을 바탕으로 DC영상 τ 의 히스토그램은 다음과 같이 표기할 수 있다.

$$h_{y_i}(\tau) \equiv n^2 \cdot \Pr_{p \in \tau} [p \in \tau_{y_i}] \quad (\text{단, } i \in [m]) \quad (6)$$

여기서 $\frac{h_{y_i}(\tau)}{n^2}$ 는 화소값이 y_i 가 될 확률이다. 다시 영상 τ 의 공간적 상관도는 다음과 같이 정의한다.

$$\begin{aligned} \gamma_{y_i, y_j}^{(k)}(\tau) & \equiv \Pr_{p_1 \in \tau_{y_i}, p_2 \in \tau_{y_j}} [|p_1 - p_2| = k] \\ & \quad (\text{단, } i, j \in [m], k \in [d]) \end{aligned} \quad (7)$$

$\gamma_{y_i, y_j}^{(k)}(\tau)$ 는 현재 위치의 화소값(y_i)에서 k 만큼 떨어진 곳에 특정 화소값(y_j)을 가지는 화소가 있을 확률이다. 여기서 최대 거리 d 는 임의로 정해지게 되며 n 보다 크지 않다. 이때 하나의 영상 τ 에 대해 특징 벡터를 추출하기 위한 계산량은 m^2 과 k 의 곱만큼 비례한다.

이럴 경우 너무나 많은 계산시간과 저장공간을 필요로 하므로 실제 영상검색에 적용하기에는 매우 비현실적이다. 따라서 계산상의 문제점을 보완하고 화소간의 공간적 상관 관계를 최대한 이용하고자 DC영상을 계층으로 분리했다. 각각의 계층에 대해 $\gamma_{y_i, y_j}^{(k)}(\tau)$ 을 적용하게 되며 각각의 계층은 이진 영상이므로 다음 식을 최종 도출할 수 있다.

$$\lambda_y^{(k)}(\tau) \equiv \gamma_{y, y}^{(k)}(\tau) \quad (8)$$

$\lambda_y^{(k)}(\tau)$ 는 두 화소가 동일한 화소값을 갖고 화소간의 거리가 k 일 확률이다. 식 8을 이용하면 각각에 계층은 이진 영상이므로 밝은 레벨 부분(그림 3에서 FG)과 어두운 레벨 부분(그림 3에서 BG)에 대해 특징벡터를 얻을 수 있다. 그림 3은 모의 실험과정에서 쓰인 MPEG데이터 중 200번 프레임과 201번 프레임의 1번 계층에 대한 특징벡터를 보여준다.

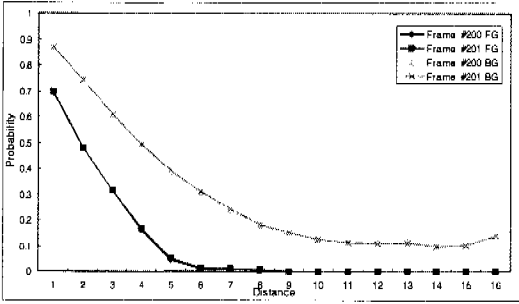


그림 3. 프레임 #200과 #201간의 거리별 특징벡터 값

IV. 특징벡터 비교법 및 특징차 비교법

다음 식은 특징차 비교를 위해 흔히 쓰이는 방법 인 L_1 -distance 방법이다. 여기서 L_s 는 각각의 계층을 뜻하며 s 와 a 는 각각 계층 번호와 최대 계층 수이다.

$$|\tau - \tau'|_{L_s, L_s} = \sum_{y \in [m], k \in [d], s \in [a]} |\lambda_{y, L_s}^{(k)}(\tau) - \lambda_{y, L_s}^{(k)}(\tau')| \tag{9}$$

그러나 가장 계산이 쉬운 반면 영상의 종류에 따라 다양한 특징차를 비교하기 위한 임계치 설정이 쉽지 않으므로 X^2 비교법이 주로 쓰인다.

$$|\tau - \tau'|_{L_s, L_s} = \sum_{y \in [m], k \in [d], s \in [a]} \frac{[\lambda_{y, L_s}^{(k)}(\tau) - \lambda_{y, L_s}^{(k)}(\tau')]^2}{[\lambda_{y, L_s}^{(k)}(\tau) + \lambda_{y, L_s}^{(k)}(\tau')]^2 + 1} \tag{10}$$

분모에 1이 더해진 것은 0으로 나누는 것을 막기 위해서이다.

기존의 X^2 비교법은 특징차가 작을 때는 더 작게, 그리고 클 때는 더욱 크게 해줌으로써 L_1 -distance 비교법보다 임계값을 정하기가 유리하다. 그러나 제안 특징 벡터는 확률단위이므로 매우 작은 값을 가진다. 이렇게 전체적으로 작은 특징차를 가질 때는 작은 값에서 큰 기울기를 가지는 비교법이 임계치 설정면에서 유리하다. 다음에 제안하는 비교법은 본 논문에서 제안한 특징벡터의 특징차를 비교할 때 X^2 비교법보다 임계치 설정에 더욱 효과적이다.

$$|\tau - \tau'|_{L_s, L_s}$$

$$= \sum_{y \in [m], k \in [d], s \in [a]} \frac{[\lambda_{y, L_s}^{(k)}(\tau) - \lambda_{y, L_s}^{(k)}(\tau')]^2}{[\lambda_{y, L_s}^{(k)}(\tau) + \lambda_{y, L_s}^{(k)}(\tau')]^2 + 1} \tag{11}$$

다음은 식 (10)과 식 (11)의 함수 특성을 보여 준다.

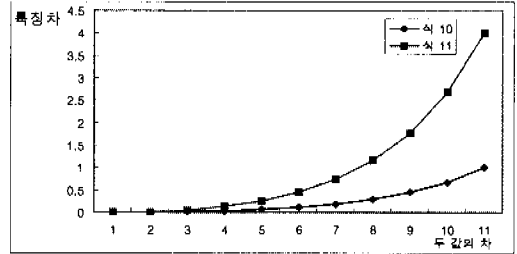


그림 4. 식 10과 식 11의 함수 특성

샷(shot) 내의 프레임 중에서도 물체의 빠른 움직임, 큰 물체의 움직임, 카메라의 움직임, 카메라 후레쉬의 섬광 등은 프레임 내 화소값에 크게 영향을 주므로 DC영상도 많이 변하게 되어 컷을 오 검출할 수 있다. 이를 해결하기 위해 기존에 컷으로 판정된 컷을 컷 예상 지점으로 지정하고 실제의 컷인지 아닌지를 이웃 프레임의 DC영상을 사용하여 재 판정한다.^[8]

두 프레임이 아닌 한 프레임 전이나 한 프레임 후의 프레임과 특징차를 각각 비교하여 평균값을 취하면 해당 프레임들의 평균 특징차가 된다. 이때 이 평균값보다 컷 예상 지점의 특징차가 크다면 다음 그림처럼 실제의 컷으로 판정한다.

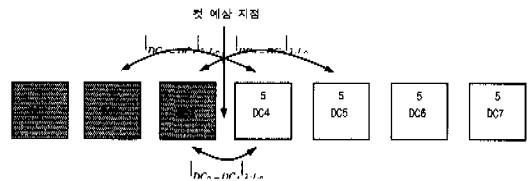


그림 5. 특징차 비교 방법

V. 모의실험 및 고찰

본 논문에서는 Sethi의 MPEG 비디오 시퀀스의 DC추출방법을 이용하였고 제안 알고리즘의 성능 평가를 위해 기존 논문들의 매체(parameter) 비교

알고리즘인 Chi-square Test를 사용하였다.^[7] 블러링은 End-in search 방법을 사용하였고 휘도 성분의 양자화는 16단계로 하였다. 또한 동일 화소의 최대 검색 거리(d)는 16으로 하였다. 실험 대상 비디오 시퀀스는 뉴스와 축구경기과 드라마가 녹화된 비디오 1과 영화가 녹화된 비디오 2이다. 그리고 비디오 1과 2는 해상도가 320×240 이고 각각 3880개와 946개의 프레임을 갖는 MPEG-1 비디오이다. 비디오 1의 시퀀스에는 18개의 급격한 장면 변화와 1개의 점진적인 장면 변화를 가지고 있고 비디오 2의 시퀀스에는 13개의 급격한 장면 변화와 1개의 점진적인 장면 변화가 포함되어 있다.

제안한 특징벡터 추출방법과 이웃 네 개의 프레임들을 비교 대상으로 하는 특징차 비교법을 이용해 컷을 검색 할 경우 Recall과 Precision이 기존 알고리즘에 비해 매우 높다. 뿐만 아니라 제안 알고리즘은 비디오 1과 비디오 2의 내용 차이에 상관없이 일정한 Recall과 Precision값을 가진다. Recall과 Precision은 다음 식과 같다.

$$Recall = \frac{n_c}{n_c + n_f} \quad (12)$$

$$Precision = \frac{n_c}{n_c + n_m}$$

- n_c : 올바르게 검출된 컷 수
- n_m : 검출하지 못한 컷 수
- n_f : 잘못 검출한 컷 수

표 1. 기존 알고리즘과 제안 알고리즘의 검출 결과 비교

	비디오 #1		비디오 #2	
	실제 컷 수	19	실제 컷 수	14
기존	C_t	22	C_t	20
	C_c	15	C_c	12
	C_f	7	C_f	8
	C_m	4	C_m	2
	Recall	0.68	Recall	0.6
	Precision	0.78	Precision	0.86
제안	C_t	20	C_t	15
	C_c	17	C_c	12
	C_f	3	C_f	3
	C_m	2	C_m	2
	Recall	0.89	Recall	0.8
	Precision	0.85	Precision	0.85

- C_t : 총 검출 컷 수
- C_c : 정확히 검출한 컷 수
- C_f : 잘못 검출한 컷 수
- C_m : 검출하지 못한 컷 수

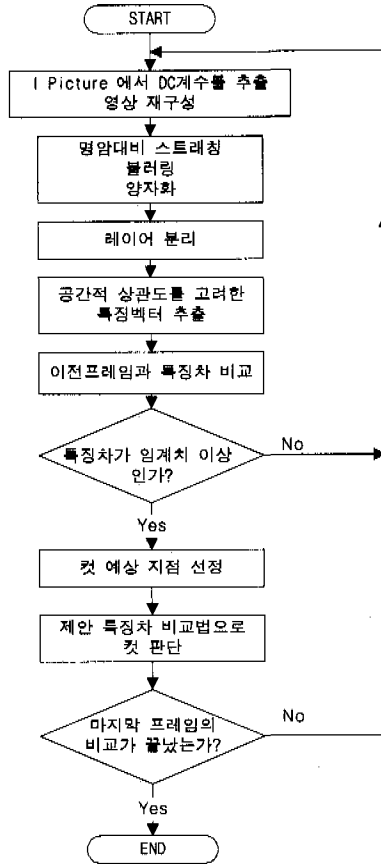


그림 6. 제안 알고리즘 순서도

VI. 결론 및 향후 연구과제

MPEG 비디오 시퀀스에서 DC 계수를 이용한 컷 검출 관련 연구는 아직 뚜렷하게 대표할 만한 것이 없다.

본 논문에서는 MPEG 비디오 시퀀스에서 휘도 성분의 공간적 상관관계를 이용한 특징벡터와 개선된 특징차 비교법으로 컷 검출을 시행하여 기존의 공간적 상관 관계를 포함하지 않고 특징벡터를 추출해 컷을 검출하는 방법보다 우수한 결과를 도출하였다.

향후 연구는 처리 속도저하를 해결하는 방법을

