

자기상관과 에지 히스토그램을 이용한 동영상 전환점 검출

노 정 진*, 문 영 호*, 정회원 유지 상*

Detection of Video Cut Using Autocorrelation Function and Edge Histogram

Jungjin Noh*, Youngho Moon*, Jisang Yoo* *Regular Members*

요 약

본 논문에서는 자기상관 함수와 에지 히스토그램 비교법을 이용하여 동영상 전환점을 자동으로 추출하는 기법을 제안한다. 기존에는 컬러 히스토그램 비교법을 많이 사용하였으나, 급격한 조명 변화에 민감하고 컬러 히스토그램 분포가 비슷한 부분의 장면 전환을 놓칠 수 있다는 단점이 있다. 본 논문에서 제안하는 자기상관과 에지 히스토그램 비교법을 사용하면 조명 변화에 의한 장면전환이 잘못 검출되는 것을 방지 할 수 있다. 제안한 기법의 성능평가를 위하여 광고, 뉴스, 드라마, 애니메이션 등 다양한 비디오 데이터에 가지고 실험한 결과, 컬러 히스토그램 비교법보다 Recall에서는 약 10%, Precision 에서는 약 17% 성능이 향상되었다. 특히 카메라 기법 중의 Fade 와 Zoon In/out에 좋은 성능을 보인다.

Key Words : cut; autocorrelation; edge histogram

ABSTRACT

While the management of digital contents is getting more and more important, many researchers have studied about scene change detection algorithms to reduce similar scenes in the video contents and to efficiently summarize video data. The algorithms using histogram and pixel information are found out as being sensitive to light changes and motion. Therefore, visual rhythm gets used in recent work to solve this problem, which shows some characteristics of scenes and requires even less computational power. In this paper, a new scene detection algorithm using visual rhythm by direction is proposed. The proposed algorithm needs less computational power and is able to keep good performance even in the scenes with motion. Experimental results show the performance improvement of about 30% comparing with conventional methods with histogram. They also show that the proposed algorithm is able to keep the same performance even to music video contents with lots of motion.

1. 서 론

멀티미디어 기기의 발전과 더불어 사용자의 요구 사항 증가로 동영상 데이터를 빠르고 효과적으로 저장, 색인, 검색, 관리하는 연구가 활발히 진행되고

있다. 그러나 동영상은 비정형 데이터이고 그 양이 방대하여 정보의 검색 및 브라우징이 힘들다는 문제점을 가지고 있다. 또한 텍스트 키워드를 이용한 검색 기술이 한계에 도달하였기 때문에 영상자체의 정보와 내용에 기반을 둔 검색기술이 최근에 연구

* 광운대학교 전자공학과 디지털 미디어 연구실(jungjin@image.gwu.ac.kr, moonsky707@image.gwu.ac.kr, jsyoo@daisy.kw.ac.kr),
논문번호 : 040137-0331, 접수일자 : 2004년 03월 31일

※본 연구는 대학 IT 연구센터 육성·지원 사업의 연구 결과로 수행되었음.

되고 있다[1].

동영상 데이터를 보다 효율적으로 관리하기 위해서는 검색이나 특징을 추출하는 기본적인 단위로 영상을 분할하여 구조화할 필요가 있다. 동영상의 구조는 크게 4개의 계층으로 구성되는데, 최하위는 개개의 정지영상을 가리키는 프레임(frame)이고, 다음은 샷(shot)으로 한 카메라의 연속적인 촬영에 의한 프레임의 모임을 말한다. 그리고 내용상 관련이 있는 인접한 샷의 집합을 장면(scene)이라 하고, 최상위 계층으로 연속적인 장면들의 집합인 시퀀스(sequence)로 구성이 된다. 동영상에서 급격한 장면 전환이 이루어지는 부분을 컷(cut)이라 하며, 컷과 컷 사이의 프레임 집합이 샷이 된다.

최근에는 ISO/IEC의 SC29의 MPEG-7이라는 이름으로 멀티미디어 데이터의 내용표현을 위한 국제 표준화 작업이 완성되어 멀티미디어 각각의 내용을 표현하고 특징을 정의하므로, 이를 이용하여 보다 다양한 정보의 검색을 효율적으로 할 수 있는 기반이 마련되었다. 하지만 크기가 큰 영상 데이터의 검색 시에 프레임마다 메타데이터를 첨부시키는 것은 비효율적이므로 동영상의 내용검색을 위하여 물리적인 장면 전환이 이루어지는 부분에 대한 장면전환 검출이 선행되어야 할 필요성이 생기게 되었다.

장면전환 검출을 위해 일반적으로 쓰이는 방법은 컬러 히스토그램 비교와 색상 차를 이용한 방법이 있다[2-5]. 히스토그램 비교 방법은 동일한 샷 내의 프레임들은 서로 유사한 색상분포를 가진다는 특성을 이용한 가장 보편적인 검출방법으로 동영상에서 인접한 프레임들의 히스토그램 차이를 계산하여 주어진 임계값과 비교함으로써 컷을 검출하게 된다. 이러한 방법은 화소 단위 비교 방법과 비교하여 물체의 움직임이나 카메라 이동 등에 덜 민감하지만, 공간정보를 포함할 수 없고, 갑작스런 조명변화에 잘못된 장면전환을 검출할 수 있는 단점을 가지고 있다. 즉 빛의 영향에 민감하여 다른 샷이지만 색상 분포가 같은 경우를 검출해 내지 못한다는 단점을 가지고 있다.

화소단위 비교 방법은 동일한 샷 내에서는 화소값의 변화가 적다는 성질을 이용하는 것으로, 연속하는 한 쌍의 프레임에서 대응하는 화소 값을 비교하여 얼마나 많은 변화가 발생하였는가를 측정한다. 이 방법은 구현이 간단한 반면, 카메라 움직임에 민감한 반응을 일으킨다. 화소단위 비교방법은 화소단위로 하기 때문에 움직임이 많은 영상이면 장면전환이 아닌 부분을 장면전환으로 검출하는 결과가

나오게 된다.

이러한 기존 기법들이 가지고 있는 문제점들을 해결하기 위하여 본 논문에서는 히스토그램 비교법과 자기상관도를 이용하여 동영상의 장면전환을 검출하는 기법을 제안하고 있다. 제안한 기법에서는 히스토그램 비교를 통해 과거 차이 값을 저장하여서 그 값을 가지고 자기상관(autocorrelation)값을 구하고, 이 자기상관 값과 현재프레임과 이전프레임의 히스토그램 비교법을 사용한 결과 값의 차이를 구하게 된다. 이때 임계값은 실험에 의하여 설정이 되고 위에서 구한 자기상관과 히스토그램 비교법의 차이 값이 임계값보다 크면 다음 단계로 넘어가게 된다.

다음단계에서 임계 값보다 큰 차이 값을 가진 프레임에 대해 에지 히스토그램(edge histogram)을 구하고 이전 프레임의 에지 히스토그램 값을 구한다. 다음으로 에지 히스토그램간에 차이를 다시 한번 구하여 그 값이 임계값보다 크면 장면전환으로 판별하게 된다. 에지 히스토그램에서 사용되는 임계값도 실험치에 의해 설정되며, 자기상관과 히스토그램을 비교할 때 쓰인 임계값과는 다른 값으로 설정된다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 장면전환 검출방법과 배경 이론에 대해 기술하고 3장에서는 제안하는 장면전환 검출기법에 대한 연구 내용을 기술한다. 4장에서는 실험결과 및 평가에 대해 기술하며 마지막으로 5장에서 결론을 맺었다.

II. 기존의 장면전환 검출 기법 및 배경 이론

서론에서 언급한 바와 같이 대표적인 장면전환 검출 기법으로는 히스토그램을 이용한 비교 방법, 화소단위 비교 방법, 통계적인 차를 이용한 비교 방법, 그리고 변환 영역 특징을 이용한 방법 등이 있다. 특히 동영상 촬영기법의 발전으로 인해 컷(cut), 페이드(fade), 디졸브(dissolve), 와이프(wipe)등을 찾기 위해 Miller와 Mai는 에지(edge)를 사용했으며, Zabih는 복원 영상의 에지 화소의 분포를 이용하였다[6]. 그림 1은 영상의 장면전환의 효과들을 나타내고 있다. 기존 기법에서 사용된 평가 기준 중 일부가 본 논문에서 제안하는 새로운 장면전환 검출 기법에도 사용됨으로 본 절에서는 이에 대한 기본적인 배경이론을 설명하고자 한다.



그림 5. 영상의 여러 효과들
(위부터 각 행 : 페이딩/컷/디졸브/와이프 효과)

2.1 화소단위 비교법과 히스토그램 비교법의 계산
화소단위 비교 방법은 동일한 샷 내에서는 화소 값의 변화가 적다는 성질을 이용하는 것으로, 식 (1)에서 $F_i(x,y)$ 는 i 번째 프레임에서 (x, y) 의 화소 값이고, 이 때 인접하는 프레임의 대응하는 화소 값의 차가 임계값 t 를 초과하는 경우 1로 된다. 식 (2)에서 전체 화소에 대해 1로 된 화소의 비율이 특정 임계값 T_p 를 넘으면, 컷으로 간주한다. X, Y 는 이미지의 최대 높이와 넓이를 표시한다 [2-5].

$$DP_i(x, y) = \begin{cases} 1 & \text{if } |F_i(x, y) - F_{i+1}(x, y)| > t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$\frac{\sum_{x,y=1}^{X,Y} DP_i(x, y)}{X \times Y} \times 100 > T_p \quad (2)$$

이러한 화소 단위 비교 방법은 카메라의 움직임과 물체의 움직임에 따라 잘못된 장면전환을 많이 검출한다. Fade와 zoom in/out과 같은 카메라 이동, 그리고 물체의 이동은 많은 화소의 변화를 가지고 오게 되고, 결국 잘못된 장면전환을 검출하는 결과를 보이게 된다.

히스토그램 비교 방법은 동일한 샷 내의 프레임들은 서로 유사한 색상분포를 가진다는 특성을 이용한 가장 보편적인 검출방법으로, 식 (3)과 같이 동영상에서 인접한 프레임들의 히스토그램 차이를 계산하여 주어진 임계값 T_h 와 비교함으로써, 장면전환을 검출하게 된다. 여기서 i 는 프레임 번호를 나타내고, j 는 히스토그램 상의 색상 값을 나타낸다. 또한 H 는 주어진 색상 값에 대한 빈도 수를 나

타낸다.

$$D_i = \sum_{j=1}^N |H_i(j) - H_{i+1}(j)| > T_h \quad (3)$$

이러한 히스토그램 비교 방법은 빠르게 물체가 움직인다 해도 전체의 히스토그램의 변화는 크게 변하지 않기 때문에 앞의 화소단위 비교방법보다 움직임에는 정확한 장면전환을 검출할 수 있다. 그러나 히스토그램 비교 방법은 밝기 변화에 잘못된 장면전환을 검출할 수 있다. 즉, 갑작스런 조명변화나 섬광, 비슷한 배경이나 분위기에서 잘못된 장면전환을 검출할 수 있다. 그리고 fade나 zoom in/out 같은 카메라 기법에도 잘못된 장면전환을 검출할 수 있다.

2.2 엔트로피를 이용한 비교방법

영상 정보의 복잡도를 확률에 기반한 엔트로피로 나타내면 식 (4)와 같다. 이 엔트로피를 이용하면 영상의 복잡도를 측정할 수 있다[7].

$$E = - \sum_{j=0}^K P(a_j) \log P(a_j) \quad (4)$$

여기서 a_j 는 j 번째 화소 값의 개수이며, $P(a_j)$ 는 a_j 의 확률이다.

동영상에서는 연속되는 두 프레임간의 엔트로피 차이를 측정하여 프레임 차이를 판별한다. 갑작스런 조명변화가 생겼을 때 컬러 히스토그램의 분포는 크게 차이가 난다. 그러나 물체나 배경, 혹은 영상 전체의 밝기 변화가 생기더라도 그 영상의 복잡도는 크게 변하지 않기 때문에 엔트로피 값이 크게 변하지 않는다. 따라서 연속되는 두 프레임간의 엔트로피 차이를 사용하면 조명 변화로 인한 잘못된 장면전환 검출을 막을 수 있다. 그래서 컬러 히스토그램과 엔트로피 비교 방법을 함께 사용하여 두 프레임간의 컬러 히스토그램과 엔트로피의 차이를 구하여, D_{total} 이 임계값을 넘을 때에 장면전환으로 간주한다.

식 (5)는 연속되는 두 프레임간의 엔트로피 차이를 나타내며, 식 (6)은 컬러 히스토그램 차이를 D_h 라 했을 때 식 (5)의 엔트로피 차이 D_e 를 결합한 전체 차이를 나타낸다. 이는 두 프레임간의 차

이가 칼라 히스토그램의 차이와 엔트로피 차이의 합으로 나타낼 수 있음을 뜻한다.

$$D_e = E_m - E_{m+1} \quad (5)$$

$$D_{total} = D_h + wD_e \quad (6)$$

$$D_{total} > T_e \quad (7)$$

여기서 w 는 가중치이며, 두 프레임의 차이가 식 (7)과 같이 어느 임계값을 넘으면 장면전환으로 간주한다. 엔트로피 방법은 갑작스런 조명 변화에 잘 못된 장면전환을 검출하지 않지만 fade처럼 점점 어두워지던가 밝아지면 장면전환으로 판별하게 된다.

2.3 에지 히스토그램(Edge Histogram)

MPEG-7의 시각 표현자 중 하나인 에지 히스토그램 디스크립터는 영상의 에지 정보를 효과적으로 표현하기 위하여 영상에서 추출된 에지 정보를 전체(global), 부분영역(semi-global), 국부영역(local)에 대한 에지 히스토그램으로 구분하고 에지 히스토그램의 구조를 표현한다[8].

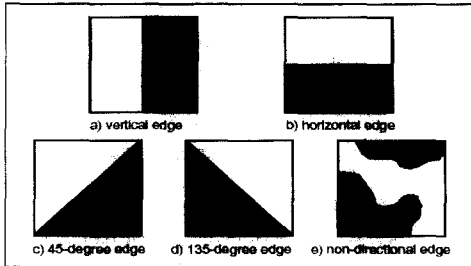


그림 6. 5가지 종류의 에지

에지 정보 추출은 영상 내 화소들의 집합인 이미지 블록에서 특징을 추출하는 방법을 사용한다. 즉 이미지 블록을 4개의 서브 이미지 블록으로 나누고 각 서브 이미지 블록 내 평균 밝기 값의 차이를 사용하여 서브 이미지 블록간의 에지를 검출하게 된다. 그림 2에서 에지 히스토그램에는 4개의 방향성 에지와 1개의 비방향성 에지를 정의한다. 방향성 에지로는 수평, 수직, 45도 및 135도 대각선 방향이며 이들 에지는 이미지 블록으로부터 추출된다. 만일 방향성이 없는 에지가 추출되면 비방향성 에지로 간주한다.

각 블록의 평균값들은 그림 3과 같은 계수를 갖는 방향성 필터를 사용하여 식 (8) ~ (12)과 같이 각 방향성 에지의 강도를 구하고, 5개의 에지 강도 값 중 가장 큰 값이 임계값보다 크면 그 블록이 해

당 에지를 갖는 것으로 판정한다.

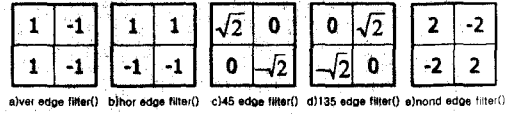


그림 7. 에지 검출을 위한 필터

$$edge0 = \sum_{i=0}^3 |mean_{sub_block}(i) \times edge0_filter(i)| \quad (8)$$

$$edge45 = \sum_{i=0}^3 |mean_{sub_block}(i) \times edge45_filter(i)| \quad (9)$$

$$edge90 = \sum_{i=0}^3 |mean_{sub_block}(i) \times edge90_filter(i)| \quad (10)$$

$$edge135 = \sum_{i=0}^3 |mean_{sub_block}(i) \times edge135_filter(i)| \quad (11)$$

$$com_edge0 = \sum_{i=0}^3 |mean_sub_block(i) \times com_edge_filter(i)| \quad (12)$$

여기서 mean_sub_block은 에지를 구할 영상 공간이다. edge0_filter, edge45_filter, edge90_filter, edge135_filter 그리고 com_edge_filter는 그림 3의 edge filter의 계수 값이다. edge0, edge45, edge90, edge135 그리고 com_edge는 각각의 에지 성분 값이다.

블록 단위로 에지 정보를 추출하기 위해서 전체 영상 공간을 4×4=16개의 서브 이미지로 분할한다. 그러나 16개의 국부영역에서의 특징 정보만을 이용하여 히스토그램을 만들 경우 영상 내의 상하, 좌우의 물체 움직임에 대응할 수 없다. 그래서 에지의 이동과 미세한 위치 변화에 강인하도록 영상 내에서 추출된 에지의 위치에 따른 전체(global), 부분영역(semi-global), 그리고 국부(local)에지 히스토그램을 생성한다.

그림 4는 전체적인 히스토그램을 크게 3가지로 구분하여 나타낸다.

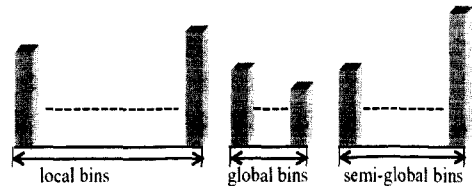


그림 4. 히스토그램 구조

국부(local) 히스토그램은 16개의 부분영상에 대해 각각 5개의 에지 종류별 누적 값이다. 부분 영역

(semi-local) 히스토그램은 16개의 부분 영상에 대해 화면의 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 특징 정보를 생성한다.

전체(global)히스토그램은 영상 전체에 각 방향성 에지량을 나타내고, 국부 히스토그램은 영상 내 객체의 위치를 고려한 국부적인 에지 분포를 나타낸다. 그림 5은 부분 영역(semi-global) 히스토그램 생성 방법을 나타낸다.

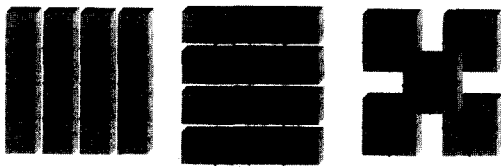


그림 5. 부분 영역 히스토그램 생성을 위한 서브 영상의 집합

식 (13)은 컷을 검출하기 위해서 이전 프레임의 에지 히스토그램과 현재 프레임의 에지 히스토그램 차이를 구해 임계값보다 크면 컷으로 판별하는 과정을 나타낸다. 여기서 전체(global)끼리의 차이에서 는 가중치가 사용되었다[9].

$$D(A, B) = \sum_{i=1}^{79} |Local_A[i] - Local_B[i]| + 5 \times \sum_{i=1}^4 |Global_A[i] - Global_B[i]| + \sum_{i=1}^{64} |Semi_Global_A[i] - Semi_Global_B[i]| > K \quad (13)$$

여기서 Local_A와 Local_B는 국부영역에서 현재 프레임의 히스토그램과 이전 프레임의 히스토그램이다. 마찬가지로 Global_A와 Global_B는 전체영역에서 현재 프레임과 이전 프레임의 히스토그램이고 Semi_Global_A와 Semi_Global_B도 부분영역의 현재 프레임과 이전 프레임의 히스토그램이다.

III. 제안하는 장면전환 검출기법

본 논문에서는 장면전환 검출을 위해 크게 두 부분의 비교과정을 통해 동영상의 장면전환을 검출하고 있다. 첫 번째는 자기상관 함수를 이용하여 장면전환을 검출하고, 두 번째는 에지 히스토그램을 이용하여 장면전환을 검출하게 된다. 자기상관 함수를

이용하게 되면 급진적인 컷뿐만 아니라 fade와 zoom in/out과 같은 카메라 기법에도 정확히 장면전환을 찾을 수 있다. 그리고 에지 히스토그램을 사용함으로써 갑작스런 조명변화나 섬광, 비슷한 배경이나 분위기에서 잘못된 컷을 검출하지 않게 된다.

3.1 자기상관 함수(Autocorrelation Function)를 이용한 방법

자기상관 함수는 어떤 시간에서의 신호 값과 다른 시간에서의 신호 값과의 상관성을 나타내는 것으로 m에서의 신호 값 x(m)과 k 만큼 지연된 m+k에서의 신호 값 x(m+k)의 곱으로 정의되며 일반적인 식은 식 (14)과 같다.

$$\psi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \quad (14)$$

자기상관을 계산하는 과정에서 적당한 윈도우를 씌워서 연산 구간을 설정하게 되는데 본 논문에서는 해닝 윈도우(Hanning window)를 기반으로 실험값에 의해 가장 좋은 결과를 보이는 윈도우를 식 (15)와 같이 새로 설계하였다. 그림 6에 본 논문에서 설계된 윈도우를 보였다.

$$h(n) = \cos\left[\left(\frac{\pi}{2N}\right)n\right] \quad (15)$$



그림 6. 사용된 윈도우 함수

이러한 윈도우를 설계한 이유는 바로 이전의 신호가 현재의 신호 값에 가장 큰 영향을 미치기 때문이다. 영상에서 컷이 발생한 후 바로 다음 프레임에서 컷이 발생할 확률이 매우 낮고, 시간이 지남에 따라 컷이 발생할 확률이 커지기 때문에 위와 같은 윈도우를 사용하게 되었다.

우선 식 (16)은 식 (3)에서 구한 과거 N개의 히스토그램 차이를 구한 것에서, 식 (15)에서 정의한 윈도우를 곱한다. 즉 Dw_k 는 히스토그램 차이와 윈도우를 곱한 값이며 식 (14)에서 x(m)에 해당한다. i는 현재의 프레임 번호이고 윈도우 크기가

N 이면 k 는 0부터 N 까지의 값이다. 그리고 식 (17)과 같이 윈도우를 곱한 결과 값을 가지고 자기상관(Ac)을 구한다.

$$Dw_k = D_{i-k-1} \times h(k) \quad (0 \leq k \leq N) \quad (16)$$

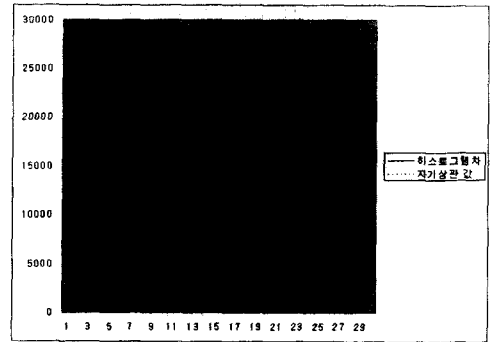
$$Ac = \frac{1}{\alpha} \sum_{k=0}^N Dw_k \times Dw_{k+1} \quad (17)$$

여기서 α 는 가중치이다.

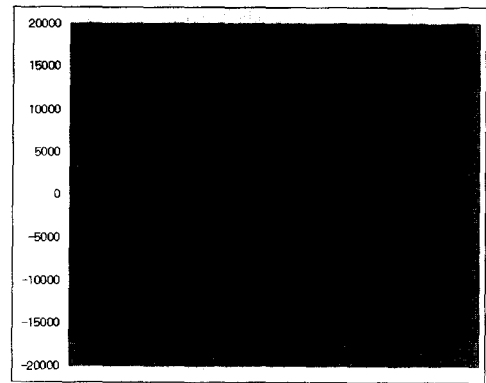
장면전환 검출은 식 (18)처럼 현재 구한 프레임의 히스토그램 차이 값과 과거 신호를 가지고 구한 자기상관 Ac 의 차이 값이 임계값 T_{ac} 보다 크면 장면전환으로 판단한다.

$$D_i - Ac > T_{ac} \quad (18)$$

과거의 히스토그램 차이를 이용하여 장면전환을 구하기 위해 자기상관을 사용하였다. 영상이 변화가 없는 경우 히스토그램의 차가 거의 변화가 없다. 즉 같은 scene인 경우에 영상의 히스토그램은 변화가 없다. 여기서 자기상관을 구한 값은 현재 히스토그램 차와 비교하면 비슷한 값이 나온다. 그러나 갑작스런 영상이 변할 경우 자기상관을 구한 값과 현재 히스토그램 차와 비교하면 값이 크게 변한다. 그리고 그림 7(a)은 fade in이 되는 부분의 히스토그램 차와 자기상관 값이다. 처음의 점은 영상이었다가 서서히 밝아지면서 영상이 뚜렷해진다. 그림 7(a)에서 보는바와 같이 히스토그램의 차를 보면 크게 증가했다가 서서히 떨어지는 것을 볼 수 있다. 이때의 자기상관 값도 증가하였다가 서서히 떨어진다. 이것을 이용하여 현재의 히스토그램의 차와 자기상관 값을 빼면 그림 7(b)와 같이 된다. 히스토그램 비교법처럼 히스토그램의 차만 가지고 장면전환을 구하면 잘못 검출되는 프레임이 많아진다. 그러나 자기상관을 가지고 장면전환을 구하게 되면 잘못된 검출을 줄일 수 있게 된다. 마찬가지로 zoom in/out 같이 영상이 변하면 히스토그램 차가 점점 증가한다. 그때의 자기상관 값도 역시 같이 증가한다. 그 두 개의 값을 비교하면 차이가 크게 나지 않기 때문에 zoom in/out으로 인한 잘못된 장면전환 검출을 막을 수 있다.



(a)



(b)

(a) fade in 구간에서 히스토그램 차(D)와 자기상관 값(Ac)
(b) D와 Ac의 차이

그림 7. fade in 구간에서의 히스토그램과 자기상관 값

자기상관을 이용한 방법의 전체구조는 그림 8과 같다. 우선 공통적으로 동영상 데이터의 RGB성분을 YCbCr성분으로 변환하여서 히스토그램 비교를 위한 기본설정을 한다. YCbCr성분으로 변환된 성분 중에 Y성분만을 히스토그램 비교를 한다. 그 후에 구해진 히스토그램 차이 값 N 개를 가지고 자기상관을 구한 후, 현재의 히스토그램 차값과 자기상관 값과의 차이를 구한다. 그 후에 차이 값이 임계값을 넘을 경우 두 번째 비교인 에지 히스토그램 비교를 위한 대기상태가 된다.

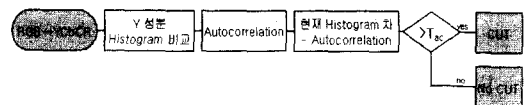


그림 8. 자기상관을 통한 비교

이러한 자기 상관을 통한 비교 방법은 히스토그

램 비교 방법을 쓰기 때문에 히스토그램 비교 방법의 장점인 움직임에 잘못된 장면전환을 검출하지 않는다. 제안하는 자기상관을 사용함으로써 히스토그램의 단점 중에서 fade나 zoom in/out과 같은 장면변환에서 잘못된 장면전환을 검출하지 않게 된다. 그러나 갑작스런 조명변화에는 잘못된 장면전환을 검출하기 때문에 아래와 같이 에지 히스토그램 방법을 사용하게 된다.

3.2 에지 히스토그램을 이용한 방법

앞에서 설명한 에지 히스토그램을 이용하여 장면전환을 검출하게 된다. 에지 히스토그램을 이용한 장면전환 검출 방법의 구조는 그림 9과 같다. YCbCr 성분으로 변환된 성분 중에 Y 성분만을 에지 히스토그램을 구한 후에 이 값을 가지고 이전 에지 히스토그램과 현재 에지 히스토그램과의 차이 값을 비교하여 임계값을 통해 최종적으로 장면전환으로 검출하는 방법을 제안한다.

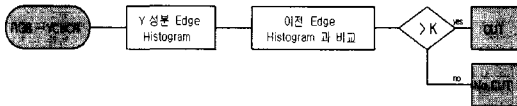


그림 9. 에지 히스토그램을 통한 비교

영상의 에지는 움직임에는 잘못된 장면전환을 검출하지만 갑작스런 조명변화나 섬광, 비슷한 배경이나 분위기에서 잘못된 장면전환을 검출하지 않게 된다.

3.3 자기상관과 에지 히스토그램을 이용한 방법

위의 자기상관을 이용한 방법과 에지 히스토그램을 이용한 방법을 통해 컷을 판별하는 방법을 통합하여 그림 10에 제안하는 기법의 흐름도를 나타내었다. 그림에서 보여지듯이 자기상관을 통한 차이값과 에지 히스토그램 비교를 통한 비교 값을 저장하여 두 가지 다른 임계값과 비교를 하여 컷을 검출하게 된다.

임계값을 넘어서 컷으로 판정이 된 프레임은 영상의 재생순서대로 저장되며, 임계값을 넘지 못한 프레임은 저장하는 순서를 거치지 않고 처리를 마치게 된다.

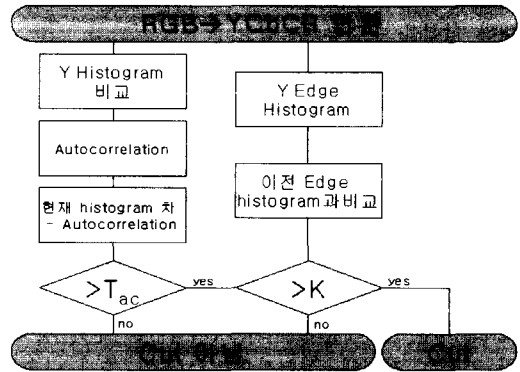


그림 10. 제안하는 장면전환 검출 기법

처음의 자기상관 통한 장면전환 검출 기법을 사용하여 컷을 검출 한 후 검출된 컷들 중에서 에지 히스토그램을 사용함으로써 자기상관 방법의 단점인 갑작스런 조명변화나 섬광 등에 인한 잘못된 장면전환 검출을 줄일 수 있다. 이러한 두 가지 방법을 사용함으로써 움직임과 갑작스런 조명변화나 섬광 그리고 카메라 기법 중에 특히 fade나 zoom in/out에서 잘못된 장면전환 검출을 줄일 수 있다.

IV. 실험결과 및 분석

본 논문에서는 표 1에서와 같이 다양한 형식의 동영상을 사용하여 실험을 수행하였다. 여기서 광고와 뮤직동영상 같은 경우는 빠른 장면전환 장면과 fade in/out 효과가 많고, 뉴스 같은 경우는 앵커 구간과 내용구간으로 뚜렷하게 나누어지는 특징이 있다. 또한 드라마의 경우는 조명변화와 조명반사 장면이 많고, 애니메이션은 뚜렷한 색채와 zoom in/out, 플래쉬 효과 등의 많은 특징을 보인다.

표 1. 실험에 사용한 동영상 데이터

	size	length(sec)	frames	format
광고	352×240	1:01	1833	MPEG
뉴스	352×240	1:29	2670	MPEG
드라마	512×384	7:50	14101	AVI
뮤직동영상	512×384	4:01	7241	WMV
애니메이션	512×384	1:33	1755	AVI

실험환경은 펜티엄4 2.4GHz 프로세서, 512MByte 메모리 상에서 Visual C++ 6.0과 Directshow 라이브러리를 이용하여 구현하였다. 동영상 샘플에 따라 성능이 많이 좌우되는 경향이 있지만 Directshow

라이브러리를 사용함으로써 다양한 동영상 포맷에 대한 실험을 수행할 수 있었다.

실험에서 장면전환을 판별하는 임계값은 고정임계값을 사용하였다. 제안하는 기법을 통해 비교된 값들이 임계값을 넘었을 경우에 장면전환으로 판정을 하였다. 그림 10에서 사용한 임계값 Tac 는 500으로 하였고, K 는 5000으로 고정임계값을 사용하였다. 이 값은 실험에 의해 구하였다. 식 (17)의 α 는 실험치로써 3500으로 하였다. 그리고 N 은 과거히스토그램을 몇 개를 가지고 자기상관을 구할 것인지 정하는 값인데, 실험적으로 성능이 좋게 나오는 값인 5로 하였다.

우선 성능 비교를 위해 기존의 방법들을 구현하였다. 표 2는 히스토그램 비교법을 사용하여 나온 결과이다. 표 3은 화소단위 비교법의 결과이고, 표 4는 엔트로피를 사용하여 나온 결과이다. 이때의 임계값 역시 실험에 의해 구하였다.

표 2. 히스토그램 비교법(임계값 5000)

	광고	뉴스	드라마	뮤직동영상	애니메이션
실제장면전환	38	29	91	80	35
찾은장면전환	35	32	112	112	59
잘못찾은장면전환	9	5	34	44	25
정확히찾은장면전환	26	27	78	68	34
못찾은장면전환	12	2	13	12	1

표 3. 화소단위 비교법(임계값 1200)

	광고	뉴스	드라마	뮤직동영상	애니메이션
실제장면전환	38	29	91	80	35
찾은장면전환	42	32	149	49	50
잘못찾은장면전환	12	3	71	4	17
정확히찾은장면전환	30	29	78	45	33
못찾은장면전환	8	0	13	35	2

표 4. 엔트로피(임계값 4000)

	광고	뉴스	드라마	뮤직동영상	애니메이션
실제장면전환	38	29	91	80	35
찾은장면전환	35	30	104	93	70
잘못찾은장면전환	9	3	26	37	36
정확히찾은장면전환	26	27	78	56	34
못찾은장면전환	12	2	13	34	1

표 5. 제안하는 기법(자기상관, 에지 히스토그램, 임계값 500, 5000)

	광고	뉴스	드라마	뮤직동영상	애니메이션
실제장면전환	38	29	91	80	35
찾은장면전환	34	29	94	74	38
잘못찾은장면전환	0	0	6	1	3
정확히찾은장면전환	34	29	88	73	35
못찾은장면전환	4	0	3	7	0

표 6. 히스토그램 비교법(장면전환 종류별)

	광 고			뮤직동영상		
	Cut	fade in/out	dissolve	Cut	fade in/out	dissolve
실제장면전환	36	1	1	73	3	4
찾은장면전환	26	9	0	92	20	0
잘못찾은장면전환	1	8	0	26	18	0
정확히찾은장면전환	25	1	0	66	2	0
못찾은장면전환	11	0	1	7	1	4

표 7. 제안하는 기법(장면전환 종류별)

	광 고			뮤직동영상		
	Cut	fade in/out	dissolve	Cut	fade in/out	dissolve
실제장면전환	36	1	1	73	3	4
찾은장면전환	33	1	0	72	2	0
잘못찾은장면전환	0	0	0	1	0	0
정확히찾은장면전환	33	1	0	71	2	0
못찾은장면전환	3	0	1	2	1	4

본 논문에서 제안하는 방법을 사용하여 나온 결과는 표 5에 보였다. 표 6은 히스토그램 비교법에서 장면전환 종류별로 나타낸 결과이다. 사용한 실험 동영상에서 뉴스와 애니메이션은 fade나 dissolve같은 장면전환이 없기 때문에 장면전환 종류별로 구별을 하지 않았다. 표 7은 제안하는 기법을 사용하여 얻은 결과인데 표 6과 표 7을 비교해 보면 fade in/out에서 제안하는 기법은 히스토그램 비교법 보다 잘못 찾은 장면전환이 거의 없음을 알 수 있다. 본 논문에서 제안한 장면전환 검출의 효율성을 평가하고 다른 장면전환 검출 방법과의 성능비교를 위해, 일반적으로 사용되는 Recall과 Precision이라는 성능평가의 척도를 사용하였다. Recall은 실제 장면전환 중에서 정확히 검출된 장면전환의 비율로서 장면전환 감지 율을 나타내고, Precision은 검출된 장면전환들 중에서 정확하게 검출된 장면전환의 비율로서 장면전환 검출의 정확도를 의미한다. 식 (19)과 식 (20)에서 Recall과 Precision을 정의하였다.

$$Recall = \frac{N_c}{N_c + N_m} \quad (19)$$

$$Precision = \frac{N_c}{N_c + N_f} \quad (20)$$

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

표 8은 성능평가 척도를 기준으로 하여, 본 논문에서 제안한 방법과 기존의 검출 방법인 히스토그램 비교법, 색차비교법, 그리고 엔트로피 기법의 결과를 비교한 것이다. 표 8에서 보는 것과 같이 히스토그램 비교법, 색상차, 그리고 엔트로피 이용해 장면전환을 검출한 결과는 각각 Recall에서 84%, 83%, 84%를 보였고, Precision에서 68%, 74%, 69%의 성능을 보였다. 또한 Precision이 약 10%씩 적어서 Recall에 비해 성능이 좋지 않음을 확인할 수 있었다. 본 논문에서는 정확한 장면전환 검출에 목적을 두고 제안하는 방법을 사용한 결과 Recall은 94%, Precision은 97%로 Recall은 기존의 방법보다 10%정도의 성능향상을 가져왔다. 그리고 Precision도 약 17%정도의 성능향상을 보였다.

따라서 본 논문의 실험 결과는 자기상관을 이용하여 히스토그램 비교법만을 사용한 것보다 더 많

은 장면전환을 검출할 수 있다는 것을 보여주고 있다. 자기상관을 사용함으로써 fade나 zoom in/out같은 장면전환에도 잘못된 장면전환을 줄일 수 있었다. 에지를 사용함으로써 히스토그램의 단점인 조명의 변화에 개선된 성능을 얻을 수 있었다. Precision도 개선된 성능을 얻을 수 있음을 보였다.

V. 결론

본 논문에서는 효율적인 동영상 검색과 브라우징을 위한 전 단계로 영상의 전환점인 장면전환을 검출 기법을 제안하고 그 성능을 평가해 보았다. 제안된 기법에서는 히스토그램 비교법과 자기상관도를 이용하여 동영상의 장면전환을 검출하고 있다. 먼저 히스토그램 비교를 통해 과거 5개의 차이 값을 저장하여서 그 값을 가지고 자기상관 값을 구하고, 이 자기상관 값과 현재프레임과 이전프레임의 히스토그램 비교법을 사용한 결과 값의 차이를 구하게 된다. 이때 임계 값보다 큰 차이 값을 가진 프레임에 대해 에지 히스토그램(edge histogram)을 구하고 이전 프레임의 에지 히스토그램 값을 구한다. 최종적으로 에지 히스토그램 간에 차이를 다시 한번 구하여 그 값이 임계값보다 크면 장면전환으로 판별하게 된다.

Recall과 Precision이라는 성능평가의 척도를 이용하여 제안한 방법의 성능을 측정한 결과 5종류의 주제가 다른 동영상 데이터에 대해, Recall은 약 94%, precision은 약 97%의 검출 성능을 보였다. 이는 기존 기법보다 우수한 결과이다. 특히 자기상관에 의해 fade와 zoom in/out이 많은 드라마 영상에 대해서 잘못된 장면전환 검출을 줄일 수 있었고, 에지 히스토그램을 사용함으로써 조명변화에도 잘못된 장면전환 검출을 줄일 수 있었다.

표 8. 실험데이터의 Recall과 Precision

	히스토그램차		색상차		엔트로피		제안하는 기법	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
광고	0.743	0.743	0.789	0.714	0.743	0.743	0.859	1.000
뉴스	0.931	0.844	1.000	0.906	0.931	0.900	1.000	1.000
드라마	0.857	0.658	0.857	0.523	0.857	0.750	0.967	0.936
뮤직동영상	0.850	0.607	0.563	0.918	0.700	0.602	0.913	0.986
애니메이션	0.971	0.576	0.942	0.660	0.971	0.485	1.000	0.932
검출률	87.04%	68.24%	83.02%	74.42%	84.04%	69.90%	94.78%	97.08%

VI. 참고 문헌

[1] H. J. Zhang, A Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video", *ACM Multimedia System*, Vol 1, No 1, pp.10-28, 1993.

[2] Nilesh V. Patel and Ishwar K. Sethi, "Video shot detection and characterization for video databases", *Pattern Recognition*, Vol 30, No 4, pp.583-592, April, 1997.

[3] 염성주, 김우생, "동적 영역 히스토그램을 이용한 동영상 데이터의 컷검출 방법", *정보과학회 논문지(A)*, Vol 30, No 3, pp.221-230, Mar, 1998.

[4] 정해준, 이우선, 정성환, "칼라 히스토그램과 DC 계수를 이용한 비디오 세그먼트이션", *정보처리학회, CD-paper #43*, Oct, 1999.

[5] 이명주, 김형균, 김태성, "칼라 임계값을 이용한 동영상의 컷 검출에 관한 연구", *한국멀티미디어 공학회 춘계학술대회*, Vol 3, No 1, 2000.

[6] R. Zabih, J. Miller and K. Mai, "A feature-based algorithm for detecting and classfying scene breaks", *ACM International Conference on Multimedia*, pp.189-200, 1995.

[7] 송현석, 안강식, 조석제 "컬러 히스토그램과 엔트로피를 이용한 동영상 컷 검출", *한국정보처리 시스템학회 학술대회 논문집*, Vol 2, No 1, 2001.

[8] 박수준, 박성희, 원치선 "이미지 검색을 위한 에지 히스토그램", *HCI 학술대회*, Oct, 2001.

[9] B. S. Manjunath, Philippe Salembier, Thomas Sijora "Introduction to MPEG-7", *WILEY*, pp.223-227, 2002

노 정 진(Jung-jin Noh)

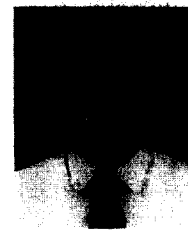
준회원



2004년 2월 : 광운대학교 전자공학과 졸업
 2004년 3월~현재 : 광운대학교 전자공학과 석사과정
 <관심분야> MPEG-7, 데이터방송, 영상압축

문 영 호(Young-ho Moon)

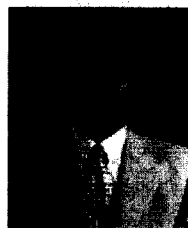
준회원



2004년 2월 : 광운대학교 전자공학과 졸업
 2004년 3월~현재 : 광운대학교 전자공학과 석사과정
 <관심분야> MPEG-7, 데이터방송, 영상압축

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