MSE를 이용한 프레임의 컷 검출시스템 설계 및 구현

박석남*, 정창렬*, 윤홍상**, 고진광*
*순천대학교 컴퓨터과학과
**광주대학교 정보통신과
e-mail:seognam@naver.com, kjg@sunchon.ac.kr

Design and Implement on the Cut Detection Retrieval System of Frame Using MSE

Seog-Nam Park*, Chang-Ryul Jung*, Hong-Sang Yoon**, Jin-Gwang Gho*
*Dept. of Computer Science. Sunchon National University
**Div. of Computer Science. Kwangju University

요약

동영상 검색 시스템은 사용자가 전체 동영상 정보를 한눈에 파악할 수 있어야 한다. 필요한 경우 원하는 지점부터 직접 재생할 수 있도록 해야 한다. 이에 대한 많은 연구가 수행되고 있다. 그러나 데이터를 처리하는데 시간이 오래 걸리는 등의 몇가지 문제점이 있다.

따라서 본 논문은 내용기반 색인에 기초가 될 동영상의 장면전환점 검출에 관한 효과적인 방법을 제안하였다. 제안된 방법은 프레임의 수평방향, 수직방향 또는 대각선 방향으로 나누어 일정한 픽셀 값을 추출하여 정지영상으로 변환하였다. 각각의 프레임에서 추출한 픽셀의 칼라값을 저장하고 MSE (Mean Square Error)를 이용하여 일정한 임계값에 도달하지 못한 프레임을 장면전환점으로 검출하였다. 검출된 장면전환점을 이용하여 변화된 프레임의 장면전환을 검색하는 시스템을 구현하였다.

1. 서론

기존의 멀티미디어 검색 방법은 미리 입력된 키워드를 사용한 텍스트를 기반으로 하는 질의 및 접근 방법이 주류를 이루고 있다. 이러한 접근 방식은 텍스트 주석이 각 영상에 첨가되고 텍스트를 검색 질의어로 사용하기 때문에 사용자는 사전에 영상을 기술하는 키워드 범위를 알고 있어야 한다. 또한, 나라마다 다양한 언어와 문화적 특성에 종속되는 문제점을 가지고 있다. 따라서 현재 텍스트를 기반으로 하는 영상 검색 시스템의 문제를 극복하기 위한 방법으로 영상의 내용(content)을 특징으로 사용하는 여러 방법들이 연구되고 있다.

일반적으로 동영상 검색 시스템에서는 사용자가 전체 동영상 정보를 한눈에 파악하고, 필요한 경우 동영상의 원하는 지점부터 직접 재생할 수 있도록 하기 위하여 전체 동영상의 내용을 요약해 놓은 대표 프레임 리스트를 제공한다. 대표 프레임 리스트를 작성하기 위하여 장면전환을 정확하게 검출할 필요성이 발생한다. 동영상의 장면 분할을 위한 장면전환 알고리즘으로 여러 가지 방법들이 연구된 바 있다. 그러나 히스토그램의 임계값이 영상의 값의 분포에 따라 변경되는 문제점, 그리고 MPEG로 압축된 동영상 데이터에 먼저 디코딩 한 다음 데이터를 처리하는데 시간이 오래 걸리고 메모리 양이 많이 필요한 문제점이 있다.

본 연구에서는 장시간에 걸친 동영상 데이터를 한눈에 알아볼 수 있도록 각 프레임 별로 영상데이터의 특정영역을 샘플링하여 대용량 동영상 데이터를 소용량의 정지영상으로 전환한다. 연속된 각각의 프레임들을 수평방향, 수직방향 또는 대각선 방향의 일정한 픽셀 값을 추출하여 정지영상으로 샘플링하여 한눈에 장면전환점을 나타내는 프레임의 컷을 볼 수 있도록 컷을 검출하는 시스템을 구현하였다.

2. 관련연구

2.1 픽셀차이(pixel difference)를 이용한 방법

이 방법은 두 프레임의 동일한 위치의 픽셀 강도 값의 차이들을 비교하여 두 프레임 사이의 변화를 검출한다. 이 알고리즘은 변화된 픽셀의 수를 카운트하여 만약 전체 픽셀에 대한 변화된 픽셀의 퍼센트가 임계값을 넘으면 장면변환이 발생하는 것으로 보고 프레임의 컷을 추출한다.[1][2][3] 픽셀의 차이와 임계값 계산을 수학적으로 식(1)과 (2)와 같이 나타낼 수 있다. 식 (1)에서 $F_i(x,y)$는 프레임 i의 좌표점 (x,y)에서 픽셀의 강도 값으로 만약 두 개의 연속적인 프레임에서 대응하는 픽셀사이의 차이가 최소 강도 값보다 크다면 다른 그림이다 라고 판단하고 $DP_i(x,y)$는 1로 설정한다. 식(2)에서 두 프레임의 픽셀 사이의 퍼센트 차이는 프레임에서 픽셀값 차이의 합을 전체 픽셀수로 나눈 것에 의해 계산한다.

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

$$\frac{\sum_{x,y=1}^{X,Y} DP_i(x,y)}{X * Y} * 100 > t \quad \text{-------- (2)}$$

이러한 픽셀값 비교는 카메라의 움직임과 물체의 움직임에 매우 민감하다는 단점을 가지고 있어 장면전환 검색에 많은 어려움이 있다.

2.2 명도히스토그램(histogram)을 이용한 방법

카메라와 움직이는 물체의 감도는 두 프레임의 명도 히스토그램의 비교에 의해 감소되는데.[1][2] 이것은 두 프레임의 배경에 많은 차이가 없어 약간의 물체 움직임은 거의 유사한 히스토그램을 갖는다는 사실에 기인한다. 히스토그램은 프레임에서 각각의 명도값에 해당하는 픽셀들의 수를 나타낸 것이다.

$$\sum_{j=0}^{g} |H_i(j) - H_{i+1}(j)| > t \quad \text{---------- (3)}$$

식(3)에서 g는 명도값의 개수이고, j는 명도값, i는 프레임 번호, Hi(j)는 명도값 j에 대한 히스토그램을 나타내고 있다. 만약 합계가 주어진 임계값 t보다 크다면 변화가 있는 것으로 인식한다.

X2-test 칼라 히스토그램 비교법은 Nagasaka와 Tanaka에 의해 제안되었다.[1] 이 식은 두 개의 히스토그램간의 차이를 더욱 확실히 하기 위해 제곱근을 이용한다.

$$\sum_{j=0}^{g} \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)} > t \quad \text{--------- (4)}$$

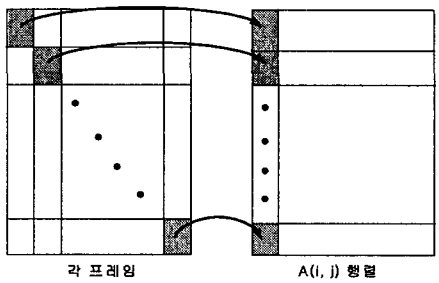
X2-test는 카메라 정지와 카메라와 물체의 이동에 의한 작은 변화에서도 차이를 확장시킨다.[2] 그러므로 이 방법은 그레이 또는 칼라 히스토그램 비교 기법보다 효율적이지 못할 수도 있다는 문제점을 가지고 있다.

3. 제안한 컷 검출시스템

3.1 동영상 데이터의 특정영역 샘플링

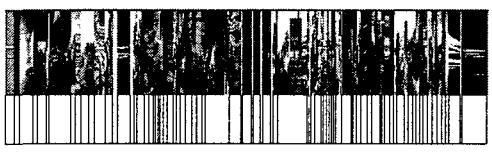
동영상은 장시간에 연속적인 프레임들이 서로 유사한 배경 구성과 이미지들의 변형으로 이루어진다. 따라서 동영상 데이터는 연속성이 끊어지는 부분이 검색 대상이 되며 그 영상이 장면전환점이 될 수 있다. 본 연구에서는 일반적인 장면 전환 기법과 장면 전환의 특수효과인 점진적인 장면 전환 기법을 검출하는데 중점을 둔다. 이를 위하여 동영상 데이터 A를 정의하는데 i×j행렬(i는 프레임의 수, j는 프레임 영상의 높이)로 나타낸다. 각 프레임의 대각선 대표값들로 샘플링된 행렬 A의 i열과 i+1열의 차이가 큰 i열이 장면 전환이 일어나는 곳이다.

각 프레임 별로 동영상 데이터의 특정영역을 샘플링하여 대용량 동영상 데이터를 소용량의 정지영상으로 전환하여 행렬로 저장한 내용을 그림으로 도식화하면 (그림 1)과 같다.



(그림 1) 프레임의 추출한 픽셀 값 저장

동영상의 각 프레임에서 추출된 데이터를 소용량의 정지영상으로 샘플링한 결과는 (그림 2)와 같다.



(그림 2) 대각선 방향의 픽셀 값을 샘플링 한 영상

샘플링된 영상에서 프레임 컷은 수직선상의 픽셀 값들의 확연한 변화가 나타나게 된다. 점진적인 장면전환점은 픽셀 값들의 점진적인 변화로 나타나게 된다. 이러한 영상 데이터는 여러 프레임간의 간격으로 샘플링을 함으로써 그 크기를 줄일 수가 있다. 또한 동영상 데이터의 전체 구조를 파악할 수 있다는 특징을 가지고 있다.

3.2 동영상 데이터의 평균오차 추출 알고리즘

행렬A의 i열과 i+1열의 차이를 정의하기 위하여 MSE를 이용한다.

$$E_1=\sqrt{(a_{11}-a_{12})^2+(a_{21}-a_{22})^2+\cdots+(a_{i1}-a_{i2})^2} \quad \text{---- (6)}$$

식 (6)에서 E1은 영상 데이터 행렬 A의 1열과 2열의 MSE이다.

$$E_k=\sqrt{(a_{1k}-a_{1(k+1)})^2+(a_{2k}-a_{2(k+1)})^2+\cdots+(a_{ik}-a_{i(k+1)})^2} \quad \text{---- (7)}$$

식 (7)에서 E_k는 영상 데이터 행렬 A의 k-1열과 k열의 MSE이다. 이렇게 MSE를 이용하여 식(7)에서 얻어진 E_k을 이용하여 프레임의 컷 f를 찾을 수가 있다.

$$|E_{f-1}-E_f| \geq t \quad \text{---------- (8)}$$

식(8)에서 t는 장면전환점인 프레임 컷을 찾는 임계값이고, E_f는 전환이 일어나는 프레임 번호를 나타낸다.

다음은 본 연구에서 제안한 검색 알고리즘이다.

```
; 추출한 데이터의 MSE 구함
for(f=1 to frame_end) {
   for(g=1 to frame_end) {
      hap=hap+(StoreA[g][f]-StoreA[g][f+1])^2
   }
   MSE_Tab[f]=SQRT(hap); // Ef의 MSE를 구
하여 저장;
}
```

4. 실험 및 결과

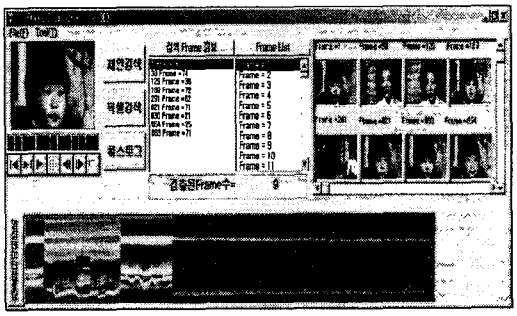
본 연구에서 사용된 시스템 CPU는 PentiumⅢ 700Mhz이고 RAM은 512MB, VGA는 GeForce2MX이다. 사용되는 운영체제는 Windows 2000 Server에 Visual C++ 6.0을 이용하였다.

현 사회에서 접할 수 있는 동영상은 무수히 많다. 이렇게 많은 동영상 중에서 광고, 뉴스, 영화, 드라마를 본 연구에서는 이용하였다. 이용된 네가지 데이터는 동일한 환경에서의 성능 측정을 하였고, 데이터의 동영상 종류별 프레임크기, 수, 장면전환점의 수는 (표 1)과 같다.

(표 1) 장면전환점 검색에 이용한 동영상 데이터

동영상 종 류	프레임 크 기	프레임 수	프레임 컷의 개수	파일 포맷
광고	320*320	480	95	AVI
뉴스	320*320	625	196	
영화	320*240	805	170	
드라마	320*240	908	13	

제안한 검색 시스템의 임계값은 검출된 장면전환점의 수가 바뀌므로 인하여 장르에 따라 최적화 된 임계값으로 변화시켜 사용하면 더 나은 결과를 얻을 수 있다. 실험에서 사용된 임계값은 최대 크기를 1로 정규화 시킨 수치이다. 동영상의 자료 중 드라마 데이터는 TV에서 방영중인 여인천하의 일부분으로 총 908개의 프레임으로 시각적으로 확인한 장면전환 프레임은 13개이다.



(그림 3) 제안한 검색 시스템

제안한 검색 시스템에서 추출한 결과 화면은 (그림 3)에서 나타내고 있다. 그리고 장면전환점 검색 작업이 끝나게 되면 추출된 프레임 리스트가 화면에 나타나게 된다. 이때 프레임 리스트에서 하나의 프레임을 클릭하면 선택된 프레임부터 좌측 창에 로드되어 그 프레임부터 재생할 수 있게 하였다.

본 연구에서는 실험한 네 가지 유형의 동영상 데이터인 광고, 뉴스, 영화, 드라마 각기 화면 구성에 특성을 지니고 있다. 때문에 본 연구가 모든 유형의 동영상에 효율적인지 판단하기 위하여 다양한 데이터를 이용하여 먼저 각각의 동영상데이터에 대한 프레임 컷을 검색하여 비교해 본 후 정확도와 재현율까지 비교하였다.

$$Precision(\text{정확도}) = \frac{\text{검출된프레임컷}}{\text{총존재하는프레임컷}}$$

<표 3> 정확도 비교

정확도	픽셀 값	히스토그램	MSE
광고	63.16(60/95)	61.05(58/95)	84.21(80/95)
뉴스	80.61(158/196)	88.78(174/196)	91.84(180/196)
영화	82.95(180/217)	88.94(193/217)	90.78(197/217)
드라마	82.95(10/13)	88.94(10/13)	95.29(11/13)

$$Recall(\text{재현율}) = \frac{\text{정확하게검출된프레임컷}}{\text{검출된프레임컷}}$$

<표 4> 재현율 비교

재현율	픽셀 값	히스토그램	MSE
광고	81.67(49/60)	77.59(45/58)	88.75(71/80)
뉴스	89.24(141/158)	77.01(134/174)	97.22(175/180)
영화	83.89(151/180)	83.94(162/193)	95.93(189/197)
드라마	60.00(6/10)	70.00(7/10)	81.81(9/11)

본 연구에서 비교의 대상으로 선정한 픽셀에 의한 검출, 히스토그램에 의한 검출은 현재 장면전환점 검출에 보편적으로 이용하는 방법이기 때문에 본 연구에서도 비교에 대상으로 선정하였다. 결과에서 확인할 수 있듯이 광고영상에서는 다른 방법에 비해 7%정도 우수한 검출율을 보여주고 있다. 단지 영화 화면에서의 검출율이 다른 방법과 비슷한 결과를 보이고 있다.

전체적으로 보았을 때 기존의 방법(픽셀값에 의한 방법, 히스토그램에 의한 검출법)과 비교하여 검출율에서 20%정도의 효율적인 결과를 보이고 있다. 그리고 정확도면에서 10%정도 향상됨을 알 수 있다. 또한 본 연구에서 임계값은 검출된 프레임 컷의 수가 바뀌므로 인하여 동영상 데이터의 종류에 따라 최적화된 임계값으로 변화시켜 사용함으로 인해 점진적인 장면 변화에도 민감하게 반응하고 있음을 알 수 있다.

5. 결론

현재의 비디오는 특정한 부분을 찾고자 할 때 반복적인 검색과정을 해야 한다. 하지만 디지털화된 동영상은 특정 부분으로 한번에 이동을 할 수 있으므로 장면이 바뀌는 프레임들을 추출하여 멀티미디어 데이터베이스에 대한 색인으로 구성할 필요가 있다.

본 연구에서는 각각의 프레임을 대각선 방향으로 영상 데이터를 검출하여 대용량의 동영상 데이터를 정지영상으로 샘플링하였다. 동영상의 전체 구조를 파악할 수 있고 장면이 전환된 점을 나타내는 프레임 컷을 한눈에 확인할 수 있다. 따라서 각각의 프레임에서 추출한 픽셀의 칼라 값은 i×j행렬로 저장하고 MSE를 사용해 일정의 임계값을 도달하지 못할 경우 이 프레임을 장면이 전환하는 점인 컷으로 검출하였다. 기존의 방법과 비교해 보았을 때 검출율에서 20%정도의 효율적인 결과를 보였다. 또한 정확도면에서 10%정도 향상됨을 확인할 수 있었다.

의료 영상에 적용하면 상당히 효과를 볼 수 있을 것이다. 앞으로 영상데이터의 특수효과를 내는 카메라 등의 이미지와 같은 복잡한 프레임 변화에도 검출되도록 하는 연구가 계속되어야 할 것이다. 또한 씬 단위로 분류할 수 있는 알고리즘의 개발이 요구된다.

참고문헌

[1] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full Motion Search for Object Appearance", Visual Database System, II, Eds, E. Knuth, and L.M Wegner, Elsevier Science Publishers B.V.,IFIP, pp. 113-127, 1992.

[2] H. J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic Partitioning of Full-Motion Video", ACM/Springer Multimedia Systems, Vol. 1, No. 1, pp. 10-28, 1993.

[3] S. Berretti, A. Del Bimbo, P. Pala, "Sensations and Psychological Effects in Color Image Database", IEEE International Conference on Images Processing, pp.560-563, 1997.

[4] W. Niblack et al. "The Qbic project Querying Images, by content using color, texture, and shape In SPIE 1908", Storage and Retrieval for Image and Video Databases, Feb. 1993.

[5] A. Hampapur, R. Jain, and T. Weymouth, "Digital Video Segmentation", Proc. ACM Multimedia 94, San Francisco, CA, pp.357-364, October, 1994,.