비디오 샷 경계면 분할기법 비교를 통한 대표 프레임의 추출

고병철 °, 변혜란
연세대학교 컴퓨터과학과

Key frame Extraction Using Comparison of Video Shot Detection Techniques

Byong-Chul Ko °, Hye-Ran Byun
Dept. of Computer Science Yonsei University

요 약

샷 경계면 분할 기법은 효과적인 비디오 검색 시스템을 구축하기 위한 기본적인 기술이다 따라서 효과적인 비디오 분석을 위해서는 신뢰할만한 샷 경계면 검출 기술이 필요하다 하지만 비디오에는 일반적인 프레임 뿐만 아니라 디졸브, 페이드와 같은 다양한 형태의 특수효과 등이 포함되어 있어 포괄적인 기술 개발이 어렵다. 본 논문에서는 지금까지 알려진 몇가지 기법들의 성능을 비교하고 이를 개선시켜 몇 개의 새로운 알고리즘을 제안하고 있다 샷 경계면 분할 작업이 끝난 뒤에는 비디오 클러스터링을 하기 위한 대표프레임 추출 작업이 필요하다 대표 프레임은 단순하게 각 샷의 첫 번째 혹은 마지막 프레임을 추출할 수도 있지만, 이 경우 각 샷의 동적인 특성들을 제대로 표현할 수 없으므로, 본 논문에서는 샷의 변화량을 측정하여 대표 프레임의 수를 결정하는 방법을 사용하였다

1. 서론

최근 멀티미디어에 대한 관심이 증가하면서 그에 따른 기술 또한 매우 빠른 속도로 증가하고 있다. 특히 비디오 영상 검색 기능에 대한 사용자들의 욕구는 비디오에 대한 수동적인 접근 방식에서, 자신이 원하는 부분만을 선택적으로 검색할 수 있는 보다 편리한 환경을 요구하고 있다 이를 위해서는 대용량의 비디오 데이터를 의미있는 단위로 나누기 위한 비디오 파싱(Parsing)과 클러스터링, 브라우징등을 포함하는 비디오 검색 시스템의 구현이 필요하다. 여기서 샷 경계면 분할 기술은 효율적으로 비디오를 분석하기 위한 가장 기본적인 과정이다[1]

샷은 1대의 카메라로 기록되는 프레임들의 연속을 뜻하는 것으로 샷 경계면이 물리적으로 존재하므로 자동적인 샷 경계면 검출이 가능하다[2]. 하지만 비디오에는 일반적인 프레임 뿐만 아니라 디졸브, 페이드와 같은 다양한 형태의 특수효과 등이 포함되어 있어 포괄적인 기술 개발이 어렵다. 이러한 문제들을 해결하기 위하여 현재 다양한 형태의 샷 경계면 분할 기법이 연구되고 있는데 이를 살펴보면, 영상의 칼리 히스토그램 분포를 이용한 히스토그램 비교법과, 각 영상의 대응되는 픽셀들의 차이값을 이용하는 화소단위 비교법, 카메라 또는 물체의 상대적인 움직임에 의해 발생되는 움직임 벡터를 이용한 방법, 디졸브와 같은 점진적 장면변화를 검출하기 위해 2개의 임계치를 적용하는 이중 비교법등이 있다[2][3][4]

본 논문에서는 샷 경계면 분할 기법중 가장 많이 사용되는 히스토그램 비교법과 화소단위 비교법, 그리고 이 2가지 방법을 혼합한 새로운 하이브리드(Hybrid)방법 및 움직임 벡터를 이용한 방법, 이중 비교법의 단점을 개선한 통계적 이중 비교법들을 제안하고 이들의 성능을 비교한다. 마지막으로는 이렇게 검출된 샷 경계면을 이용하여 클러스터링을 위한 효율적인 대표 프레임 검출 기법을 소개한다

2. 샷 경계면 분할 기법

2.1. 히스토그램 비교법

히스토그램 비교법은 샷 경계면 분할 기법중 가장 일반적으로 사용되는 방법으로서, 비디오 영상에서 이웃하는 프레임들 사이의 색상 히스토그램의 차를 계산하여 장면 변화의 발생 여부를 판단한다[1][2]

$$Diff\,f_{k,k+1} = \sum_{k=1}^{N} Dist(Hist(k), Hist(k+1)) \quad > \quad t \qquad (1)$$

$Diff\,f_{k,k+1}$: k번째 프레임과 k+1번째 프레임간의 색상 히스토그램 차이

히스토그램 비교법은 화소단위 비교법에 비해 물체의 움직임이나

카메라 이동에 덜 민감하지만, 프레임 사이에 섬광이나 선명한 조명이 있을 경우 오류를 발생 시킬수 있다[2][5].

2.2. 화소단위 비교법

연속적인 한쌍의 프레임에서 대응하는 화소의 세기를 비교하여 얼마나 많은 변화가 발생하였는가를 측정하는 방식으로 매우 작은 움직임도 감지해낼 수 있다[5][6].

$$DP_i(k,l) = |p_i(k,l) - p_{i+1}(k,l)| > t \quad (2)$$

2.3. 하이브리드 비교법

히스토그램 비교법과 화소단위 비교법의 단점을 보완하기 위하여 제안하는 방법으로, 히스토그램 비교를 위한 임계값과 화소단위 비교를 위한 2개의 임계값을 사용하여 이 임계값들을 동시에 초과할 경우만을 샷 경계면으로 선언한다.

2.4. 움직임 벡터를 이용한 방법

움직임 벡터는 카메라 또는 물체의 상대적인 움직임에 의해 발생되는데 이웃하는 프레임들 사이의 움직임 벡터를 검출해 냄으로써 샷 경계면뿐만 아니라 카메라 패닝이나 줌밍과 같은 카메라 이동에 의해서 발생되는 장면의 변화도 분석할 수 있다[2][7].

본 논문에서는 움직임 벡터를 계산해 내기 위하여 프레임을 일정한 블록(16*16)으로 나눈 뒤, 후속 프레임의 Search window안에서 가장 근사한 블록을 찾기 위해 MAD(Mean Absolute Difference) 방법을 사용하였다

$$MAD(d_1,d_2) = \frac{1}{MN} \sum_{(n_1,n_2)\in B} |s(n_1,n_2,k) - s(n_1+d_1,n_2+d_2,k+1)| \quad (3)$$

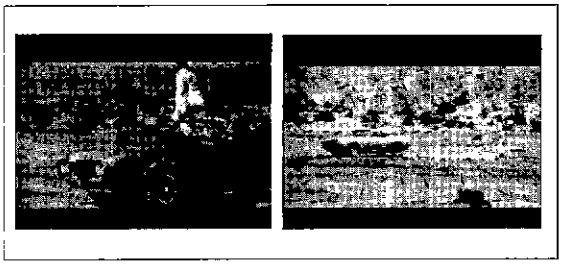

그림1 일반 프레임과 샷 경계면에서의 움직임 벡터

움직임 벡터를 이용하여 샷 경계면을 검색할 경우 프레임 사이의 차이값이 확실하다면 샷 경계면을 쉽게 검출할 수 있지만, 페이드나 디졸브와 같은 점진적 장면전환은 연속된 프레임간의 특징값의 차이가 크지 않기 때문에 잘못된 결과를 나타낸다[7].

2.5. 통계적 이중비교법

디졸브나 페이드와 같은 특수효과가 포함된 영상에서 샷 경계면을 분할하기 위해서는 카메라 브레이크를 검출하기 위한 임계치와 특수효과를 검출하기 위한 2개의 임계치를 사용하는 이중 비교법(Twin-Comparison)이 제안되고 있다[4][5]. 이 방법은 디졸브가 발생할 경우 프레임 사이에 차이값이 일반적으로 증가한다는 사실을 기반으로 특수효과를 위한 작은 임계치를 초과하는 프레임을 점진적 장면전환의 시작으로 간주한 뒤, 프레임 사이의 차가 카메라 브레이크를 위한 큰 임계치를 초과할 경우 점진적 장면전환이라고 선언하고 그렇지 않으면 잠재적 시작점을 제거한다.[5] 하지만 이 방법은 점진적 장면뿐만 아니라 카메라의 이동이나 커다란 물체의 이동에 의한 차이값의 증가도 점진적 장면전환으로 판단하는 오류가 있다 따라서 본 논문에서는 이러한 오류를 해결할 수 있는 새로운 통계적 이중비교법을 제안한다.

그림2 에서 T1,T2는 일반적 샷 경계로 선언하고 D1,D2,D3는 점진적 장면전환으로 선언되지만, 실제로 D3는 점진적 장면전환이 아

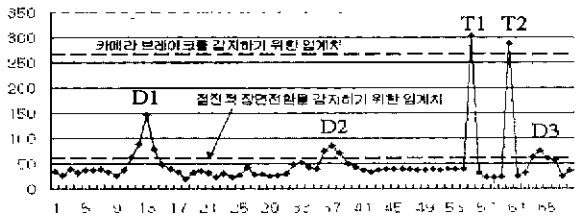

그림 2. 이중비교법을 이용한 디졸브의 검지

니라 커다란 물체의 이동이다 따라서 점진적 장면전환의 후보로 선언된 각 프레임을 N개의 블록으로 나눈 뒤(본 논문에서는 10×10크기의 블록을 영상의 각 부분에 9개로 설정했다) 각 블록의 분산값을 계산한다. 이것은 디졸브 발생시 화면의 전체 픽셀값이 동시에 일정하게 변한다는 사실에 기반하여 평균값이 고른 분포를 보이고 따라서 분산값이 작을 경우 점진적 장면전환으로 판단하고, 그렇지 않을 경우 물체의 움직임 또는 카메라의 빠른 이동으로 판단할수 있다 이 방법을 사용하면 디졸브와 같은 점진적 장면전환 뿐만 아니라, 블록의 차이값을 분석함으로써 물체의 움직임 방향을 추측해낼 수 있다 본 논문에서 제안하는 식은 (4)와 같다.

$$Var = \frac{1}{N} \sum_{i=1}^{N} (Boc_i - Avr)^2 < t \quad (4)$$

Var: 각블록들의분산값, Boc_i: 프레임간의 블록차값

Avr: 블록들의평균값

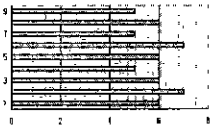

그림 3 디졸브 발생시 각 블록의 차이값 (분산: 1.54)

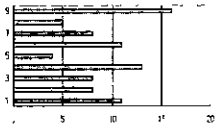

그림 4 물체의 움직임시 각 블록의 차이값 (분산:12.88)

[그림3]은 디졸브가 있는 영상에서 연속적인 프레임 사이의 대응되는 블록의 차이값을 나타낸 것으로 값이 고른 분포를 보이고 분산값도 작지만 [그림4]와 같이 화면상에서 커다란 물체가 이동했을 경우 블록의 차이값이 고르지 못하고 역시 분산도 커지게 된다. 따라서 이

경우 점진적 장면전환의 잠재적 시작점을 제거한다.

2.6. 실험 결과 및 분석

$$\text{Recall(R)} = \frac{D-F}{D-F+M}\ ,\ \text{Precision(P)} = \frac{D-F}{D}$$

Source	방 법	Cut(Missed/False/Detect)(R%,P%)	Dissolve(M/F/D)(R% ,P%)
영 화 (프리웨리-150 frame)	히스토그램법	0/3/15 (100,80)	2/2/4 (50,50)
	화소단위비교법	0/1/13 (100,92)	2/0/0 (0,0)
	하이브리드	0/0/12 (100,100)	2/0/0 (0,0)
	움직임 백터법	1/0/11 (92,100)	2/0/0 (0,0)
	통계적이중비교법	1/0/11 (92,100)	0/0/2 (100,100)
뮤직비디오 (150 frame)	히스토그램법	0/0/9 (100,100)	0/0/0 (,)
	화소단위비교법	2/0/7 (78,100)	0/0/0 (,)
	하이브리드	0/0/9 (100,100)	0/0/0 (,)
	움직임 백터법	5/3/7 (45,45)	0/0/0 (,)
	통계적이중비교법	1/3/10 (89,80)	0/0/0 (,)
광 고 (레귤러-184 frame)	히스토그램법	1/1/6 (83,83)	1/1/1 (0,0)
	화소단위비교법	1/0/6 (81,100)	1/0/0 (0,0)
	하이브리드	1/1/6 (83,83)	1/0/0 (0,0)
	움직임 백터법	2/2/5 (60,60)	1/0/0 (0,0)
	통계적이중비교법	1/0/5 (83/100)	0/0/1 (100,100)
여행홍보 (헌터수노-150 frame)	히스토그램법	1/3/5 (66,40)	1/1/1 (0,0)
	화소단위비교법	2/0/4 (66,100)	1/0/0 (0,0)
	하이브리드	1/0/5 (83,100)	1/0/0 (0,0)
	움직임 백터법	1/4/5 (50,20)	1/0/0 (0,0)
	통계적이중비교법	2/0/4 (66/100)	0/0/1 (100,100)

표 1 샷 경계면 분할 기법 결과 비교

실험결과에서 볼수 있듯이 통계적 이중 비교법을 제외한 다른 방법을 사용했을 경우 점진적 장면전환을 찾아내지 못하거나 잘못된 결과를 나타냈다. 하이브리드 방식의 경우 일반적인 샷 경계면 검출에서는 히스토그램이나 화소단위 비교법보다 성능이 개선된 것을 알수 있다 하지만 움직임 백터를 이용한 경우는 물체의 움직임이나 카메라의 움직임이 적은 영상에서는 좋은 성능을 보여주었지만 뮤직비디오나 카메라 움직임이 많은 여행홍보 비디오에서는 잘못된 샷 경계를 찾음으로써 정확성이 떨어졌다.

3. 대표 프레임 추출

샷 경계면 분할 작업이 끝난 뒤에는 비디오 클러스터링을 하기 위한 대표프레임 추출 작업이 필요하다 대표 프레임은 단순하게 각 샷

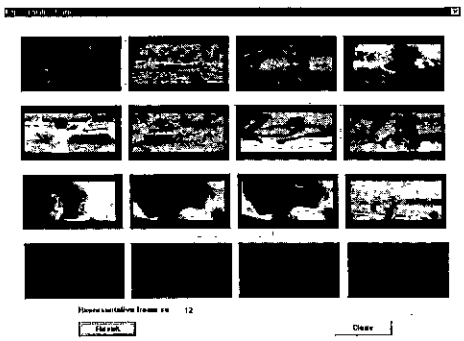

그림 5 대표 프레임 추출 결과

의 첫번째 혹은 마지막 프레임을 추출할 수도 있지만, 각 샷의 동적인 특성들을 제대로 표현할 수 없으므로, 본 논문에서는 샷의 화소값의 변화량을 측정하여 대표 프레임의 수를 결정하는 방법을 사용하였다[8]. 이 경우 변화가 적은 샷일 경우는 1개의 대표 프레임만을 두고, 긴 샷이거나 변화량이 많다면 더 많은 대표프레임을 선택하게 된다. 대표 프레임 추출시 전체 프레임에 대해 화소값을 비교해야 하므로, 본 논문에서는 각 영상의 일정 영역을 선출하여 후속 프레임과 비교하는 방식을 취하였다

4. 결론

지금까지 샷 경계면을 찾기 위한 여러 기지 기법들을 소개하고 몇 기지의 개선된 이론들을 제시한 뒤 각각의 기법들을 이용한 실험결과를 보여줬다. 또한 여기서 분할된 샷들을 이용하여 대표프레임을 추출하였다. 샷 경계면 검출중 디졸브나 페이드와 같은 특수효과가 포함된 영상에 대해서는 기존의 이중 비교법을 개선하여 통계적 특성을 추가시킨 방법을 사용하여 만족스러운 결과를 얻을 수 있었다 또한 히스토그램과 화소단위 비교법을 동시에 이용한 하이브리드 방식의 경우 특수효과가 포함된 프레임을 제외하고는 좋은 결과를 보여주고 있다 이 두가지 방법을 적절히 이용한다면 특수효과 및 일반적인 샷 경계면 검출에도 좋은 결과를 보여줄 것으로 생각된다

참고 문헌

[1] Yong Rui,Thomas S. Huang and Sharad Mehrotra, "Exploring Video Structure Beyond the Shots", IEEE Multimedia System '98, pp237-240, 1998

[2] Thomas S. Huang, Yong Rui, Trausti Kristjansson, Milind Naphande and Yueting Zhuang, " Video Analysis and Representation" , Proposal of University of Illinois at urbana-champaign. 1998

[3] Yukinobu Taniguchi , Akihito Akutsu , Yoshinobu Tonomura, "PanoramaExcerpts' Extracting and Packing Panoramas for Video Browsing", ACM Multimedia 97, 1997

[4] HongJiang Zhang, Chein Yong Low and Stephen W Smoliar, "Video Parsing and Browsing Using Compressed Data", Multimedia Tools and Application, vol 1 ,pp 89-111, 1995

[5] 이재현, 장옥배 " 움직임 백터를 사용한 점진적 장면전환 검출", 정보과학회 논문지(C) 제 3 권 제2호(97.4), 1997

[6] John S Boreczky and Lawrence A. Rowe, " Comparison of Video Shot boundary Detection Techniques" , Ph D dissertion, University of California at Berkeley, 1996

[7] 이미숙, 황본우, 이성환, " 내용기반 영상 및 비디오 검색 기술의 연구 현황", 정보과학회지 제 15권 제9호('97.9) pp 10- 18, 1997

[8] Di Zhong, HongJiang Zhang, Shih-Fu Chang , Clustering Method for Video Browsing and Annotation".