

영상의 통계적 특성을 이용한 급격한 장면전환 검출 알고리즘

곽 대 호 , 박 성 준 , 이 건 호 , 최 은 태 , 송 문 호

고려대학교 전자공학과 멀티미디어 신호처리 연구실

전화 : (02) 927-0560 / 팩스 : (02) 927-0544

Detection of Fast Scene Changes Using a Statistical Technique

Dae-Ho Kwak , Seong-Jun Park , Gun -Ho Lee , Yu-Tae Choi , S. Moon-Ho Song

School of Electrical Engineering in Korea University

E-mail : dhkwak@bessel.korea.ac.kr

Abstract

We propose a statically motivated scene change detection algorithm. As the difference between the neighboring frames will generate peaks at scene boundaries, the problem of detecting fast scene changes is equivalent to detecting peaks in a given sequence.

In this paper, the peak detection is performed via several statistics, namely the sample means and variances. For eliminating flash lights as well as detecting fast scene changes within a small number of frames, we have opted to use a two-stage process for computing the necessary statistics. The results indicate superiority of the proposed algorithm over the previously reported algorithm.

1. 서 론

디지털 비디오는 멀티미디어와 네트워크의 향상으로 인하여 미디어 중 가장 중요한 매체로 인식되어지고 있다. 따라서 이러한 대응량의 비디오를 효율적으로 저장하고 검색하기 위해 장면전환 검출 기술의 필요성은 점차 중요시되고 있다.

장면이라는 것은 논리적으로 연관된 프레임의 하나 또는 그 이상의 집합으로 정의되어지고, 이러한 장면의 경계를 장면전환이라고 한다. 장면전환의 형태는 여러 종류

가 있는데 크게 급격한 장면전환과 점진적인 장면전환으로 나누어진다.

많은 장면전환이 급격한 장면전환에 속한다. 점진적인 장면전환의 경우는 그 종류는 다양하지만 대표적으로 분리를 하면 dissolve, wipe, fade in/out으로 나누어진다.

Dissolve는 이전의 프레임과 이후의 프레임이 선형적으로 증가하거나 감소하는 경우를 말하고, wipe는 이후의 장면에서 있는 pixel들이 이전의 장면에 있는 pixel들을 일정한 행씩을 가지고 대체하는 경우를 말한다. 예를 들어 자동차의 와이퍼와 같이 한 쪽면 혹은 모서리에서부터 비뚤을 움직이는 것과 같은 효과를 말한다. Fade에 의한 효과는 장면이 선형적으로 변한다는 점에서 dissolve의 한 특수한 경우로 생각할 수 있다. Fade in은 이전장면의 색 정보가 모두 검은 색이고, fade out의 경우는 이후 장면의 색 정보가 모두 검은 색인 장면으로 본다면 fade는 dissolve의 한 부분으로 통합할 수 있다.[3]

일반적으로 점진적인 장면전환 알고리즘보다는 급격한 장면전환 검출 알고리즘의 검출률이 높다. 하지만 급격한 장면전환검출 알고리즘은 flash light나 object의 움직임에 의해 잘못된 검출을 하는 경우가 많고 비디오의 대응량으로 인한 계산속도도 문제가 된다. 따라서 본 논문에서는 이러한 급격한 장면전환의 검출율을 높이고 계산속도가 빠른 새로운 알고리즘을 제시하고, 기존의 논문 중에서 비교적 검출률이 우수한 Ven의 알고리즘과 비교한 결과를 나타내었다.

본 논문은 다음과 같은 구성을 갖는다. 2장은 DC 영상 검출 알고리즘의 설명과 급격한 장면전환 알고리즘 설명

한다. 3장에서 그 예 대한 실험결과를 나타낼 것이고, 마지막 4장은 결론으로 앞에 언급했던 내용들을 요약하고 앞으로의 연구과제를 제시 할 것이다.

2.3.2. 저급 저적인 장면 전환 검출 알고리즘

본 논문에서는 장면전환 검출에서의 필수 조건인 실시간 검출과 정확한 검출을 각각 설명하고자 한다. 실시간 검출을 하기 위하여 MPEG 비디오를 직접 사용하여 검출하기 보다는 데이터의 필수 정보들만을 가진 압축된 데이터를 사용하면 알고리즘을 빠르게 구현할 수 있다. DCT 코딩이 되어 있는 디지털 비디오는 8x8 블록 단위로 코딩이 되어 있는데 각각의 블록에서 DC 성분만을 추출한다면 데이터량을 64배 줄일 수 있다. 대부분의 영상정보는 저주파 성분에 포함되어 있으므로 DC 값만을 추출하여 사용하더라도 정보손실이 거의 없이 object의 명암나 픽셀 값 등 전체영상의 정보를 대부분 포함하며 장면전환의 특성 또한 가지게 된다. 따라서 본 논문에서는 실시간 검출을 하기 위해 MPEG 비디오에서 DC 성분만을 추출해낸 thumbnail sequence를 사용하고 있다.

정확한 검출을 위해 논문에서 제안하는 알고리즘은 다음과 같다. 장면전환이 나타날 때 명확한 불연속선이 나타나므로 n번째 프레임의 기준으로 양 옆을 미분하면 장면전환이 발생한 프레임에서 주변 값보다 큰 값이 나타난다.

그림 1은 1차 미분한 DC sequences이고 n번째 프레임의 기준 프레임이라 할 때, 1차 미분



그림 1. sliding window의 사용

한 값은

$$D_n(x, y) = |g_n(x, y) - g_{n-1}(x, y)| \quad (1)$$

이고 이 때의 기준 프레임의 평균 값은

$$\mu_1(n) = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^n \dots \quad (2)$$

이 된다. 이때 좌우 5프레임씩 sliding window (A), (B)를 잡고 각각의 평균과 표준편차를 구한다. (A), (B)의 평균 값을 $\mu_1(n)$, 표준편차 값을 $\sigma_1(n)$ 이라고 하면

$$g(n) > \mu_1(n) + \alpha_1 \cdot \sigma_1(n) \quad (3)$$

일 때 $g(n)$ 을 1차 threshold에 의한 장면전환 프레임으로 선언한다. 하지만 1차 threshold만을 적용하여 검출한 $g(n)$ 을 사용할 경우, flash light나 camera motion에 의해 영향을 많이 받게 되므로 장면전환이지만 잡아내지 못하는 miss가 많이 발생하게 된다. 따라서

$$\mu_2(n) \quad g(n) > \mu_1(n) + \alpha_1 \cdot \sigma_1(n) \quad (4)$$

과 같이 대체를 한 후 다시 2차 threshold를 적용하여

$$g(n) > (\mu_2(n) + \omega) + \alpha_2 \cdot \sigma_2(n) \quad (5)$$

일 때를 장면전환이 일어난 것으로 선언한다. 여기서 ω 는 weighting parameter를 나타낸다. Weighting parameter는 $\mu_2(n)$ 의 픽셀값이 5-60사이의 값을 갖는 데, 이 때의 miss나 false같은 error를 최소화 시키기 위하여 사용되었다.

<표 1> Video test data

Video	Title	Frames	Cut
Movie	Dr. Bullitt	56960	500
	007 series	2430	157
News	CNN 1	6350	41
	CNN 2	7421	42
Document	National-geographic	3509	56
Music Video	M.J History	16313	141
	Crow	2100	147
Commercial	Nike	2351	110
	Boxster	1013	47

3. 실험 결과

위에 제안한 알고리즘을 테스트 하기 위하여 여러 가지 video data를 실험에 적용 하였다. <표 1>은 실험에 사용된 video data의 종류를 나타내고 있다.

낮은 threshold부터 점차 높은 threshold를 적용하여 최적의 recall과 precision을 구하기 위해 1차 threshold에 사용된 α_1 의 값은 1-20까지 변화도록 하고, 2차 threshold에 사용된 α_2 의 값은 1-10까지의 변화를 주었다. 1차, 2차 threshold는 각각 독립적으로 변하기 때문에 전체 threshold는 200번 변하게 된다.

여기서 Recall과 Precision은

$$\text{recall} = \frac{\text{Correct}}{\text{Correct} + \text{Miss}} \quad (6)$$

$$\text{precision} = \frac{\text{Correct}}{\text{Correct} + \text{False}} \quad (7)$$

과 같이 정의한다. Recall과 Precision은 Threshold의 변화에 따라 달라진다. 즉, Threshold가 점차 높아지면 false가 증가하여 precision이 낮아지는 반면에, miss는 감소하여 recall이 증가하게 되는 상관관계를 가지고 있다.

그림 2-4는 movie, music video, commercial 비디오 데이터에 각각 1차, 2차 threshold를 적용하여, 이 논문에서 제안된 알고리즘으로 recall과 precision을 각각 Yeo의 알고리즘으로 얻어진 recall과 precision 값을 비교한

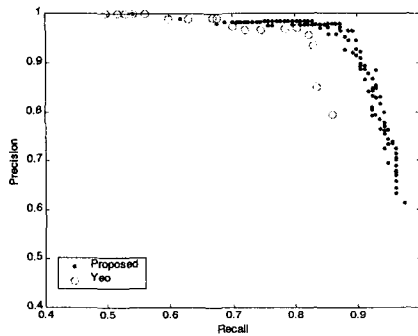


그림 2. movie data를 사용한 recall vs. precision

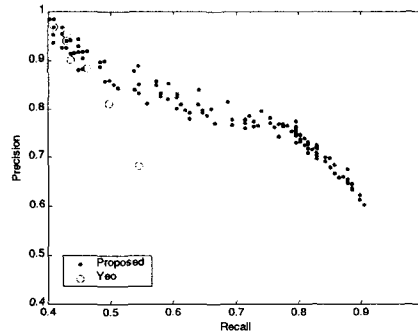


그림 3. music video data를 사용한 recall vs. precision

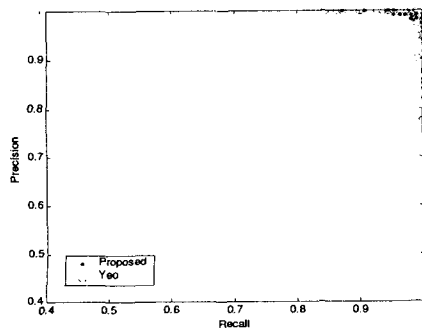


그림 4. commercial data를 사용한 recall vs. precision

여 나타내고 있다. 그림 3에서는 다른 테스트 데이터보다 검출률이 떨어진다. 이것은 music video의 특성상 작은 flash light와 object의 움직임이 많기 때문이다.

Recall과 precision을 나타내는 의미 그림 2-4를 보면 논문에서 제안하는 방법이 Yeo의 방법보다 현저히 우수하다는 것을 알 수 있다.

<표 2>는 제안된 알고리즘과 Yeo의 알고리즘의 검출율을 나타내었다.

4. 결론

본 논문에서는 멀티미디어 데이터베이스를 효율적으로 검색할 수 있는 새로운 알고리즘을 제안하고 있다. 특히 sliding window와 weighting parameter를 사용하여 miss나 false같은 error를 최소화하고, 5 프레임 이내에서 장면전환이 빠르게 일어나는 경우와 flash light가 많은 데이터에서도 우수한 결과가 나타남을 알 수 있다. 기존의 장면전환 검출 방법에는 pixel differences, statistical differences, histogram comparisons, edge differences 등의 방법들이 있는데 본 논문에서는 각 프레임의 mean과 standard deviation을 이용한 statistical 방법을 제안하였다.

제안하는 방법은 전채영상에서 DC 성분만을 추출하여 사용하므로 빠른 계산속도를 얻을 수 있고, sliding window의 사용으로 잘못된 검출을 현저히 감소시키는 효과를 가지고 있다.

본 논문에서 제안된 알고리즘은 점진적인 장면전환 검출에도 사용될 수 있을 것으로 기대된다.

<표 2> Results of recall vs. precision

Title	Proposed		Yeo	
	Recall	Precision	Recall	Precision
Dr. Dollide	0.9966	0.9832	0.9915	0.9898
007 series	0.8537	0.9655	0.8217	0.9556
CNN 1	1	0.9762	0.9756	0.8889
CNN 2	1	0.7925	0.9762	0.7193
National-geographic	0.9821	0.9167	0.9643	0.9153
M.J History	0.9858	0.9858	0.9787	1

Crow	0.7959	0.7748	0.4626	0.8831
Nike	1	0.9910	0.9909	1
Boxster	0.9149	0.9149	0.8723	0.9318

감사의 글

본 논문은 학술진흥재단의 연구비 지원으로 작성되었습니다.
습니다.

Reference

- [1] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques", in *Storage and Retrieval for Image and Video Databases IV, Proc. 157/SPIE 2670*, pp.170-179, 1996.
- [2] B. L. Yeo and B. Liu, "On the extraction of DC sequence from MPEG compressed video", *Image Processing Proceedings, International Conference*, vol. 2, pp. 260-263, 1995.
- [3] S. J. Park, S. M. Song, D. H. Kwak, W. M. Kim, and M. G. Chung, "Automatic detection of dissolving scene change in video", *대한전자공학회 춘계학술대회*, 1999.
- [4] B. L. Yeo and B. Liu, "Rapid scene analysis on compressed video", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp.533-544, 1995.
- [5] A. C. Gonzalez and A. E. Woods, *Digital Image Processing*, Addison Wesley, International editions.