

정규화를 이용한 새로운 샷 경계 검출 방법

신성윤*, 백성은*, 표성배**, 이양원*

New Shot Boundary Detection Method Using Normalization

Seong-Yoon Shin *, Seong-Eun Baik *, Seong-Bae Pyo **, Yang-Won Rhee *

요 약

비디오 분할은 샷 경계 검출 이라고도 하는데, 비디오를 계층적이고 구조적인 형태로 표현하기 위하여 영상, 문자, 오디오와 같은 매체 속에 포함되어 있는 내용들을 특징별로 분석하여 계층별로 분류하는 작업을 말한다. 본 논문에서는 카메라와 객체의 모션에 보다 강건하고 보다 정확한 결과를 산출하여 충분한 공간 정보를 가지는 지역적 X^2 -히스토그램 비교 방법을 이용하여 샷 경계를 검출한다. 또한 영상처리에서 영상의 명암 값 향상을 위하여 사용되는 로그함수와 상수를 변형하여 차이 값에 적용하는 정규화 방법을 제시한다. 그리고 샷 경계 검출 알고리즘을 제시하여 일반적인 샷과 갑작스런 샷의 특징을 기반으로 검출한다.

▶ Keyword : 샷 경계 검출(Shot Boundary Detection), X^2 -히스토그램(X^2 Histogram), 정규화(Normalization), 갑작스런 샷(Abrupt Shot), 로그함수(Log-Formula)

1. 서 론

최근에는 컴퓨터와 통신, 그리고 데이터 압축 기술 및 표준기술의 발달로 비디오를 이용한 다양한 서비스가 가능하게 되었으며, 초고속 통신망의 보급과 멀티미디어 처리 기술로 인해 비디오는 다양한 분야에서 폭 넓게 사용되고 있다. 따라서 과거와는 다르게 일반 사용자들이 언제, 어디서나 쉽게 비디오를 접근할 수 있으며, 이에 따라 비디오를 보다 효율적으로 이용하기 위한 여러 가지 응용 기술에 대한 필요성이 요구되고 있다.

비디오 분할은 샷 경계 검출 이라고도 하는데, 비디오를 계층적이고 구조적인 형태로 표현하기 위하여 영상, 문자, 오디오와 같은 매체 속에 포함되어 있는 내용들을 특징별로 분석하여 계층별로 분류하는 작업을 말한다. 최근의 많은 샷 경계 검출 방법들이 이미 수십 년 전부터 제시되어 왔다[1][2]. 가장 보편적인 샷 경계 검출 방법은 연속적인 프레임들 사이의 주어진 특징에 의해 표현되는 차이값을 계산하여 얻는 방법이다. 비록 정확률이 어느 정도 높다 해도 이들 알고리즘들의 강건성에 대한 제약들은 아직도 문제점으로 남아있다[3].

• 제1저자 : 신성윤

* 군산대학교 컴퓨터정보학과, **인덕대학 컴퓨터소프트웨어과

가샷 경계 검출하기 위한 가장 빠르고 간단한 방법은 어떤 임계치 이상으로 변하는 픽셀의 수(Pixel comparison)를 계산하여 비교하는 방법이다(4)(5)(6). 그러나 이 방법은 작은 영역에서의 큰 변화나 큰 영역에서의 작은 변화를 구별할 수 없는 단점이 있다. 즉 카메라와 객체의 동작, 명암 등에 매우 민감하여 많은 잘못된 검출 결과를 만들어 낸다.

블록기반 비교방법(Block-based comparison)은 카메라와 객체의 동작에 민감하지 않도록 지역적 특성을 강조한 방법이다. 각 프레임을 블록으로 나누고 두 연속되는 프레임의 일치하는 블록들에 대한 픽셀수를 계산하여 장면의 변화를 측정하는 방법이다. Kast 등(7)은 각 블록들에 대하여 평균과 분산정보를 이용한 유사 비율(likelihood ratio)을 측정하여 장면의 변화를 측정하였고 Shah 등(5)은 프레임을 12개의 겹치지 않는 블록으로 나눈 후에 명암 값을 이용하여 이전의 영상과 가장 잘 일치하는 블록을 찾는 방법을 제시하였다. Xion 등(8)은 영상의 일부 영역들만을 계산함으로써 샷 경계를 검출하는 망 비교(net comparison)방법을 제시하였고 Xion 등(9)은 기존의 방법을 확장하여 시간과 공간정보를 활용한 단계-변수 알고리즘(Step-variable algorithm)을 제안하였다. 그러나 이 방법들은 객체나 카메라의 움직임에 매우 민감한 단점을 갖고 있다.

히스토그램 비교방법(Histogram comparison)은 샷 경계의 검출을 위하여 사용되는 가장 보편화된 방법이다. Tono 등(4)(10)(11)은 그레이-레벨의 히스토그램 비교를 통하여 임계치를 기준으로 샷 경계를 추출하는 가장 간단한 방법을 제안하였다. Ueda 등(12)은 장면의 경계를 검출하기 위해서 컬러 히스토그램의 변화 비율을 사용했고, Naga 등(11)은 그레이 레벨과 컬러 히스토그램을 기반으로 한 몇 개의 간단한 통계학적 비교를 수행하였다. Zhan 등(4)은 픽셀 차이, 통계 차이, 그리고 몇 개의 히스토그램 방법을 비교하여 히스토그램 방법이 정확성과 속도사이의 좋은 교환요소를 발견하였다. Naga 등(11)은 두 프레임 사이의 차이 값을 강조할 뿐만 아니라 카메라나 객체의 움직임을 강조할 수 있는 χ^2 -test를 제안하였다. 그러나 χ^2 -test는 Tono 등(10)이 제안한 선형 히스토그램 비교방법보다 전체적인 성능이 더 좋지 않았으며, 계산량이 증가하는 단점을 갖는다. Garg 등(13)은 6개의 다른 컬러 좌표시스템(RGB, HSV, YIQ, $L^*a^*b^*$, $L^*u^*v^*$, Munsell)을 이용하여 3개의 히스토그램 기반방법에 대하여 성능을 평가하였다. Zhan 등(4)은 두 개의 임계치를 이용하여 점진적

장면변화를 검출할 수 있는 twin-comparison 방법을 제안하였으며, Bore 등(14)은 실시간 비디오에 대해서 비디오 분할기술을 비교한 결과 twin-comparison이 간단하며 매우 잘 적용될 수 있음을 발견하였다. 이 방법들은 간단하고 카메라나 객체의 움직임에 민감하지 않은 장점을 가지고 있지만 공간정보를 가지고 있지 않기 때문에 비교적 유사한 두 개의 다른 영상이 주어지면 샷 경계 검출에 실패하는 경우가 발생한다.

최근의 연구에서 보면 컬러 히스토그램과 χ^2 -히스토그램을 합성한 방법(15)을 이용하거나 급진적 장면부터 점진적 장면까지 모두 검출하는 강건하고 복합적인 지역적 χ^2 -테스트를 이용한 방법(16)들이 연구되고 있다.

2. 샷 경계 검출 방법

비디오를 샷 단위로 분할하기 위해서 프레임 차이값을 추출하기위한 적당한 공식을 정의한다. 샷 경계는 값이 주어진 임계치를 초과할 경우에 검출된다. 본 논문에서는 기존의 히스토그램 기반 알고리즘과 비교하여 좋은 성능을 보여주는 χ^2 -히스토그램을 변형하였고, 컬러 값의 세분화 작업에 따른 검출효과를 높이기 위하여 명암도 등급에 따른 가중치를 적용한 지역적 χ^2 -히스토그램을 이용하였다.

전역적 χ^2 -히스토그램은 각 컬러공간에 대한 명암도 등급변환을 위한 가중치를 적용하여 다음 식(2.1)과 같이 정의하여 사용할 수 있다.

$$d_{wr, wg, wb}(f_i, f_j) = \sum_{k=0}^{N-1} \left(|H'_i(k) - H'_j(k)| * \alpha + |H_i^g(k) - H_j^g(k)| * \beta + |H_i^b(k) - H_j^b(k)| * \gamma \right) \dots\dots\dots (2.1)$$

α, β, γ 는 NTSC 표준에 따른 명암도 등급변환을 위한 상수를 나타내며 $\alpha=0.299, \beta=0.587, \gamma=0.114$ 로 정의하여 사용하였다.

본 논문에서는 카메라와 객체 모션에 보다 강건하고 보다 정확한 결과를 산출하여 충분한 공간 정보를 가지는 지역적 χ^2 -히스토그램 비교 방법을 사용한다. 지역적 χ^2 -히스토그램 비교는 식(2.2)와 같다(16).

$$d(f_i, f_j) = \sum_{k=1}^b d_{\chi^2}(f_i, f_j, bl)$$

$$d_{\chi^2}(f_i, f_j, bl) = \sum_{k=1}^{N-1} \left(\frac{(H_i^r(j) - H_j^r(j))^2}{\max(H_i^r(j), H_j^r(j))} \times \alpha \right. \\ \left. + \frac{(H_i^g(j) - H_j^g(j))^2}{\max(H_i^g(j), H_j^g(j))} \times \beta \right. \\ \left. + \frac{(H_i^b(j) - H_j^b(j))^2}{\max(H_i^b(j), H_j^b(j))} \times \gamma \right) \dots \dots \dots (2.2)$$

여기서 bl 은 블록의 총 수이고 $H_i^k(k)$ 는 적색 채널에서 i 번째 프레임의 블록 bl 에 대한 그레이 레벨 k 에서의 히스토그램 차이값이다. α, β 그리고 γ 는 상수이고 이들의 값은 전역적 χ^2 -히스토그램과 같다. 이들 영역에 대한 컬러 히스토그램에 가중된 χ^2 -히스토그램을 사용하여 노이즈, 객체 그리고 카메라의 움직임 효과를 줄이기 위하여 가장 큰 차이값을 버리는 최적의 프레임 차이값은 프레임을 16개의 동일 크기 영역으로 분할하여 얻을 수 있다.

3. 정규화

대부분의 비디오 분할 알고리즘들은 연속적인 프레임들 사이의 유사성 측정에 적당한 임계치 사용에 의존한다. 하지만 임계치가 입력 비디오의 형태에 따라서 매우 민감할 수 있다. 이러한 단점을 프레임 차이값의 정규화를 통해 극복할 수 있다.

제안된 차이 값 추출 알고리즘을 통하여 추출되어지는 차이 값들은 프레임의 특징정보에 따라 편차가 크게 나타나며, 프레임간의 연속적인 연결정보를 얻기가 매우 어렵다. 특히 급진적 전환이나 점진적인 전환을 추출하기 위한 전역 또는 지역 임계치의 결정은 각 차이 값들의 변화에 따라 적응적으로 임계치를 결정해야 하는 어려운 문제점을 갖는다. 따라서 차이 값의 편차를 줄이고 시간적으로 연결된 차이 값들의 변이를 쉽게 파악하고 정보를 얻기 위한 방법이 필요하다.

기존의 프레임에 대한 전체 픽셀수를 이용한 정규화 방법은 차이 값들에 대한 폭을 줄여 예측하는 효과는 있지만 차이 값들의 시간적 연속성과 연관성에 대한 정보는 주지 못한다.

본 논문에서는 주어진 차이 값들을 일정한 값의 범위에 동적으로 압축하여 차이 값들로부터 급격한 장면변화

를 추출할 수 있는 강건한 임계치 결정의 지원은 물론 차이 값들의 시간적 연속성과 연관성을 쉽게 파악하여 점진적인 장면의 변화도 추출할 수 있는 차이 값의 정규화 방법을 제안하고자 한다.

제안된 방법은 영상처리에서 영상의 명암 값 향상을 위하여 사용되는 로그함수와 상수를 변형하여 차이 값에 적용하였다. 제안된 방식은 다음 식(3.1)과 같다.

$$d_{\log} = c \times \log(1 + d^2) \quad c = \frac{\max(d_{\log})}{\max(\log(1 + d^2))} \dots \dots \dots (3.1)$$

여기서 로그함수 d 는 식 (2.2)로부터 추출된 프레임 차이값이고 상수 c 는 d 로부터 계산된 배율상수이다.

로그함수와 상수의 적용은 차이 값의 세부묘사를 두드러지게 증가할 수 있으며, 차이 값들의 시간적인 연속성도 얻을 수 있다. 이렇게 정규화된 비디오들이 다음의 최종적인 샷 경계 검출 알고리즘에 따라 최종적인 샷으로 검출된다.

4. 최종적인 샷 경계 검출 알고리즘

샷 경계 검출은 대개 비디오 처리의 첫 번째 단계로서 샷은 카메라로부터 연속적으로 캡처된 프레임 시퀀스를 대표한다. 따라서 동일 샷에 인접한 프레임들은 시간적 연속성을 갖고 있다. 일반적인 샷과 갑작스런 샷은 플래쉬 라이트, 갑작스런 빛의 변화, 그리고 카메라 모션이나 객체의 움직임의 변화에 따라 생성된다.

본 논문에서는 샷 경계 검출 알고리즘을 일반적인 샷과 갑작스런 샷의 시간적 특징을 기반으로 검출한다. 샷 경계 검출은 <그림 1>과 같으며, 다음 순서에 의해 이뤄진다.

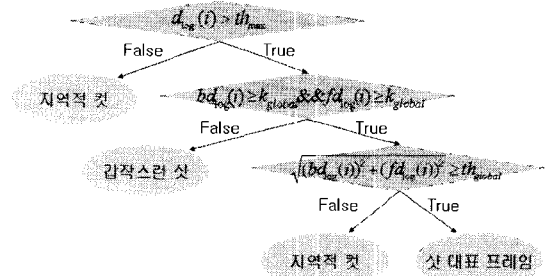


그림 1. 샷 대표프레임 검출

위의 3단계 최종적인 샷 경계 검출 알고리즘들은 (그림 2)에 요약되어 있다.

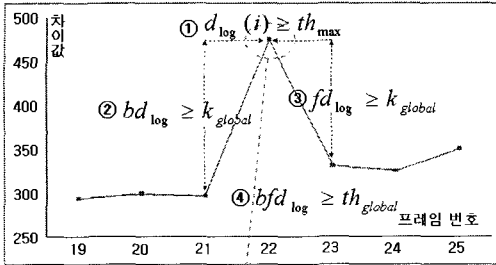
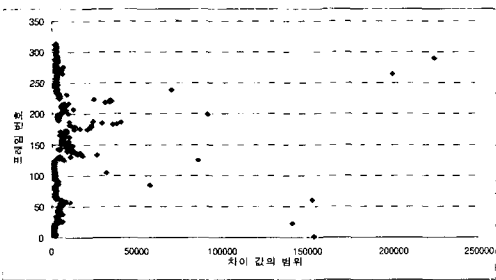


그림 2. 최종적인 샷 경계 검출 알고리즘

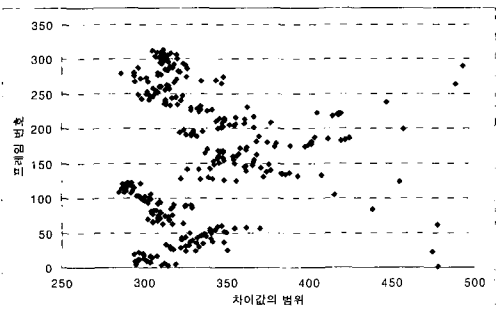
5. 실험 및 결과

실험은 Windows XP 환경에서 MS Visual C++ 6.0과 DirectX 8.1로 수행하였다. 각 비디오 시퀀스는 30frame/sec 속도와 320X240 해상도의 형태로 수행하였다.

(그림 3)는 광고 비디오로부터 추출된 총 프레임 차이값의 분포로서 (a)는 정규화 전의 분포이고 (b)는 정규화 후의 분포를 나타낸다.



(a)



(b)

그림 4. 프레임 차이값의 분포

위의 (그림 3)에서 프레임 차이값 d_{log} 의 분포는 d 의 분포보다 영역에서 보다 광범위하게 분포되어 있고 차이값들은 또한 강조되어 있다.

(표 1)은 뉴스, 스포츠, 그리고 광고 비디오 형태로부터 표현된 최대값, 최소값, 평균, 그리고 표준 편차를 보여준다. 각각의 프레임 차이값 d 와 d_{log} 는 주어진 식 (2)과 식 (3)에 의해서 계산된 값이다.

표 1. d와 d_log값의 비교
Table 1. Comparison of d and d_log

비디오		뉴스	스포츠	광고
최대값	d	223057.9	212158.3	216088.3
	d_{log}	492.6	490.6	491.3
최소값	d	1234.2	703.2	3993.2
	d_{log}	284.7	262.2	331.7
평균	d	9191.3	3740.2	26494.2
	d_{log}	334.2	308.1	391.6
표준편차	d	23778.0	13380.3	30614.5
	d_{log}	38.2	25.1	33.3

위 (표 1)에서 정규화된 값들은 샷 경계를 검출하는데 보다 훨씬 강건하고 신뢰성이 높으며 전체적인 임계치를 정하는데 편리하게 이용된다.

제안된 최종적인 샷 경계 검출알고리즘으로 많은 장면 변화를 가지는 뉴스, 스포츠, 그리고 광고와 같은 비디오 시퀀스를 가지고 실험을 수행하였다. 빠른 객체나 카메라의 변화 및 플래쉬 라이트가 있는 갑작스런 샷과 일반적인 샷의 수로 분류한 실험 결과는 (표 2)와 같다.

표 2. 실험 결과
Table 2. Experimental Result

비디오	프레임 수	갑작스런 샷의 수	일반적인 샷의 수
뉴스 1	2772	33	26
뉴스 2	2665	57	19
스포츠 1	2167	9	22
스포츠 2	665	12	19
광고 1	2975	15	21
광고 2	3096	10	30

(표 2)에서 알 수 있는 것처럼 일반적인 샷의 수는 대체로 고르게 검출되었다. 하지만 뉴스에서는 플래쉬 라이트 및 객체나 카메라의 변화가 많아서 갑작스런 샷이 가장 많이 검출되었고 스포츠와 광고에서는 갑작스런 샷이 적게 검출되었다.

6. 결론

본 논문에서는 지역 X^2 -히스토그램 비교 방법을 이용하여 샷 경계를 검출하는데, 이 방법은 카메라와 객체의 모션에 더욱 강건함은 물론 정확한 결과를 산출한다. 또한 급격하거나 점진적인 장면 변화도 추출할 수 있는 강건한 임계치 결정의 지원은 물론 차이 값들의 시간적 연속성과 연관성을 쉽게 파악하여 점진적인 장면의 변화도 추출할 수 있는 차이 값의 정규화 방법을 제시하였다. 이 정규화 방법은 차이 값의 편차를 줄이고 시간적으로 연결된 차이 값들의 변이를 쉽게 파악하고 쉽게 정보를 얻는 방법이다. 그리고 일반적인 샷과 갑작스런 샷도 모두 검출할 수 있는 샷의 특징을 기반으로 한 샷 경계 검출 알고리즘을 제시하였다. 따라서 플래쉬 라이트, 갑작스런 빛의 변화, 그리고 카메라 모션이나 객체의 움직임의 변화 등도 쉽게 추출할 수 있다.

참고 문헌

- [1] Koprinska and S. Carrato, "Temporal Video Segmentation: A Survey," Signal Processing Image Communication, Elsevier Science 2001.
- [2] G. Ananger, T.D.C. Little, "A survey of technologies for parsing and indexing digital video," Journal of Visual Communication and Image Representation, pp. 28-43, 1996.
- [3] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods," IEEE transaction on circuits and systems for video technology, Vol. 10, No. 1, Feb. 2000.
- [4] Zhang, H. J., Kankanhalli, A., and Smoliar, S. W., "Automatic Partitioning of Full-motion Video," Multimedia Systems, Vol. 1, No. 1, pp. 10-28, 1993.
- [5] Shahraray, B., "Scene Change Detection and Content-Based Sampling of Video Sequences," in Digital Video Compression: Algorithms and Technologies, Proc. SPIE 2419, pp. 2-13, February 1995.
- [6] Hampapur, A., Jain, R., and Weymouth, T., "Digital Video Segmentation," Proc. ACM Multimedia 94, pp. 357-364, October, 1994.
- [7] Kasturi, R. and Jain R., "Dynamic Vision," In Computer Vision: Principles, IEEE Computer Society Press, Washington, 1991.
- [8] Wei Xiong, John Chung-Mong Lee, Ding-Gang Shen, "Net Comparison: An Adaptive and Effective Method for Scene Change Detection" SPIE, 1995.
- [9] John Chung-Mong Lee, Qing Li, and Wei Xiong, "Automatic and Dynamic Video Manipulation" Research and Development in Information Retrieval, 1998.
- [10] Y. Tonomura, "Video handling based on structured information for hypermedia systems, in: Proc. ACM Int. Conf. Multimedia Information Systems, pp.333-344, 1991.
- [11] Nagasaka, A. and Tanaka, Y., "Automatic Video Indexing and Full-Video Search for Object Appearances," in Visual Database Systems II, Knuth, E., Wegner, L., Editors, Elsevier Science Publishers, pp. 113-127, 1992.
- [12] Ueda, H., Miyatake, T., and Yoshizawa, S., "IMPACT: An Interactive Natural-motion-picture Dedicated Multimedia Authoring System," in proceedings of CHI, 1991 ACM, pp. 343-350, New York, 1991.
- [13] U. Gragi, R. Kasturi, S. Antani, "Evaluation of video sequence indexing and hierarchical video indexing," in: Proc. SPIE Conf. Storage and Retrieval in Image and Video Databases, pp. 1522-1530, 1995.
- [14] J. S. Boreczky, L. A. Rowe, "Comparison of Video Shot Boundary Detection Techniques," in Proc. SPIE Conf. Storage and Retrieval in Image and Video Databases. pp. 170-179, 1996.
- [15] 신성운, 표성배, "텔레매틱스에서 효율적인 장면전환 검출기법을 이용한 비디오 브라우징," 한국컴퓨터정보학회논문지 제11권 제4호, pp.147-154, 2006.
- [16] 김영례, 이양원, "지역적 χ^2 -테스트를 이용한 장면전환 검출 기법," 한국컴퓨터정보학회논문지 제11권 제3호, pp. 193-202, 2006.