

전역적 결정트리를 이용한 샷 경계 검출

신성윤*, 문형윤**, 이양원*

Shot Boundary Detection Using Global Decision Tree

Shin Seong Yoon *, Moon Hyung Yoon **, Rhee Yang Won*

요 약

본 논문에서는 프레임들의 차이값으로부터 카메라 브레이크에 의해 발생하는 큰 변화의 폭을 갖는 경계지점을 추출하는 전역적 결정트리를 이용하여 샷 경계를 검출하는 방법을 제시한다. 먼저 지역적 χ^2 히스토그램과 정규화를 통하여 프레임 간 차이값을 계산하고, 다음으로 차이값간의 거리를 정규화를 통하여 계산한다. 계산된 차이값간의 거리를 바탕으로 전역적 임계치 거리를 계산하여 인접한 두 프레임들에 대한 거리값과 전역적 임계치 거리를 비교하여 샷 경계를 검출한다. 본 논문에서 제시한 전역적 결정트리를 이용하여 객체나 카메라의 움직임과 플래시 라이트와 같은 갑작스런 장면 변화를 쉽게 검출할 수 있다.

Abstract

This paper proposes a method to detect scene change using global decision tree that extract boundary cut that have width of big change that happen by camera brake from difference value of frames. First, calculate frame difference value through regional χ^2 -histogram and normalization, next, calculate distance between difference value using normalization. Shot boundary detection is performed by compare global threshold distance with distance value for two adjacent frames that calculating global threshold distance based on distance between calculated difference value. Global decision tree proposed this paper can detect easily sudden scene change such as motion from object or camera and flashlight.

▶ Keyword : 전역적 결정트리(Global Decision Tree), 샷 경계 검출(Shot Boundary Detection), 정규화(Normalization), 전역적 임계치 거리(Global Threshold Distance)

• 제1저자 : 신성윤

• 접수일 : 2008. 1. 10, 심사일 : 2008. 1. 24, 심사완료일 : 2008. 1. 25.

* 부산대학교 컴퓨터정보공학과 교수 **LG CNS 공공사업팀 차장

I. 서론

비디오는 자체가 비 구조화된 데이터로서 일정한 형식의 규칙을 가지고 정렬하여 사용자에게 보다 쉬운 접근을 허용할 수 있는 구조화된 형태의 데이터 정리가 필요하다. 따라서 비디오의 구조화를 위해서는 비디오를 의미 있는 단위로 분류하는 것이 필요하다.

하나의 카메라에 의해 촬영된 일련의 연속된 프레임들의 묶음을 추출하기 위한 기본적인 비디오 분할 단위를 샷(Shot)이라고 한다.

샷이 발생한 지점의 프레임은 기준으로 카메라나 객체의 움직임과 관련한 잠재적인 장면의 변화를 추출하기 위하여 일련의 유사한 차이 값을 갖는 연속된 프레임들의 묶음 단위를 클러스터(Cluster)라 한다. 이러한 클러스터들의 중심값을 기준으로 군집화 알고리즘에 의하여 그룹화된 클러스터의 모임을 그룹(Group)이라 한다. 클러스터의 첫 번째 프레임은 대표 프레임 추출에 적용되는 기본 단위이다.

본 논문에서는 차이 값들의 정규화 과정을 통하여 이웃하는 프레임과의 차이 값이 일정한 임계치를 만족하고 독립적으로 분포할 때 샷 프레임을 추출할 수 있다. 또한 제안된 정규화 알고리즘을 이용하여 변환된 차이 값들로부터 일정한 임계치를 기준으로 전역적 결정트리를 적용하여 샷의 경계를 검출하고자 한다.

전역적 결정트리(Global Decision Tree)란 연속되는 프레임들의 차이 값으로부터 카메라 브레이크(Camera Break)에 의하여 발생하는 큰 변화의 폭을 갖는 경계지점을 추출하기 위한 방법으로, 전역적 컷(Cut)이 발생하는 샷의 경계를 추출하여 비디오를 샷의 단위로 구분하기 위하여 적용된다. 따라서 전역적 컷이 바로 샷 경계가 되는 것이다.

II. 관련연구

비디오를 분류하기 가장 일반적인 방법은 두 프레임간의 차이 값을 비교하여 컷을 추출하는 것이다. 추출된 컷은 대표 프레임으로 간주되어 샷을 구성하거나 장면을 구성할 때 유용하게 사용되는 정보이다. 그러나 차이 값을 이용한 분류방법은 임계치를 결정하는 문제점이 있다[1][2][3].

클러스터링 기반 구조방법은 상위층의 비디오 계층구조를 얻기 위해 관련이 있는 샷들을 그룹으로 합병하고, 이 합병된 클러스터를 기반으로 브라우징 트리를 구성하여 비디오 구조

를 반영하는 방법이다. 시간적 요소와 시각적 요소를 적용하여 임계치 결정 없이 보여주는 장점이 있지만 의미 차원의 구조를 반영하지 못하는 단점이 있다[4].

장면기반 구조방법은 크게 모델기반 접근과 범용기반 접근 방법으로 나뉜다. 모델기반 접근은 의미차원의 비디오 계층방법으로 특정 모델에 기반을 두어 큰 정확성을 얻을 수 있는 있으나 각 모델에 따른 사전모델의 구성 및 모델구성에 따른 시간, 도메인 지식, 경험 등과 같은 요소들을 필요로 하는 단점을 갖는다. 범용기반 접근 방법은 명확한 도메인을 필요로 하지 않는 일반화된 모델을 따르는 것으로 시간적으로 유사한 샷들을 클러스터로 구성하는 시간제약 조건 클러스터링을 사용한다. 이 클러스터에 기반을 두고 장면전이 그래프가 구성되며, 장면구조를 구성하기 위해 장면들의 흐름을 표시한다[5][6].

또한 최근의 연구에서 보면 컬러 히스토그램과 X2-히스토그램을 합성한 방법[7]을 이용하거나 급진적 장면부터 점진적 장면까지 모두 검출하는 강건하고 복합적인 지역적 X2-테스트를 이용한 방법[8]들이 연구되고 있다.

본 논문에서 사용된 장면 전환 검출 방법은 [7][8]에서 사용한 X2-히스토그램으로서 검출 식은 다음 식 (1)과 같다.

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

$$d_{x^2}(f_i, f_j, bl) = \sum_{k=1}^{N-1} \left(\frac{H'_i(j) - H'_j(j)}{\max(H'_i(j), H'_j(j))} \right)^2 \times \alpha$$

$$+ \frac{(H_i^g(j) - H_j^g(j))^2}{\max(H_i^g(j), H_j^g(j))} \times \beta$$

$$+ \frac{(H_i^b(j) - H_j^b(j))^2}{\max(H_i^b(j), H_j^b(j))} \times \gamma) \dots \dots \dots (1)$$

여기서 bl은 블록의 총 수이고 Hir(k)는 적색 채널에서 i 번째 프레임의 블록 bl에 대한 그레이 레벨 k에서의 히스토그램 차이값이다. α, β 그리고 γ는 NTSC 표준에 따른 명암도 등급변환을 위한 상수를 나타내며 α=0.299, β=0.587, γ=0.114로 정의하여 사용하였다.

또한 정규화를 방법은 영상처리에서 영상의 명암 값 향상을 위하여 사용되는 로그함수와 상수를 변형하여 차이 값에 적용하였다. 제안된 방식은 다음 식 (2)와 같다.

$$d_{\log} = c \times \log(1 + d^2)$$

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

여기서 로그함수 d 는 식 (1)로부터 추출된 프레임 차이값 이고 상수 c 는 d 로부터 계산된 배율상수이다.

III. 전역적 장면영역 결정

전역적 장면영역이란 카메라 브레이크가 일어난 급진적 장면전환이 발생한 지역을 말하며, 전체 비디오를 고려하여 일반적인 컷과 비교하였을 때 전역적으로 일정한 규칙을 가지고 발생하는 영역이기 때문에 전역적 장면영역이라 하였다.

일반적으로 카메라 브레이크가 발생한 샷의 경계 지점은 이전 프레임이나 이후 프레임과 비교하였을 때 큰 차이 값의 편차를 갖는다. 왜냐하면 이전의 프레임들이나 이후의 프레임 들은 어느 정도 연속되는 프레임들의 분포정보를 갖지만 샷의 경계지역은 카메라 브레이크 현상에 의하여 이전프레임과는 크게 구분이 되는 차이 값을 갖는 것이 일반적인 특징이다. 만약 차이 값의 경계가 일정한 임계치를 넘지 않을 경우는 객체나 카메라의 갑작스런 움직임이나 플래시라이트와 같은 장면의 변화에 의해서 발생하는 변화를 예측할 수 있다. 따라서 이러한 카메라에 의해 발생하는 전역적 경계지역을 컷으로 잡아 장면전환 검출을 실시할 수 있다.

(그림 1)에서는 전역적 결정트리의 구조로서 전역적 장면을 결정하기 위하여 사용되는 전체적인 수식구조를 그림으로 나타내고 있다. 우선 $d(f_i, f_{i-1})$ 은 위의 식 (1)의 X2-히스토그램서 검출 식으로서, 먼저 샷 경계를 1차적으로 검출한 결과이다. 그리고 1차적인 샷 경계 검출에 의한 결과로서 나타나는 연속된 (f_i, f_{i-1}) 프레임으로부터 i 번째 프레임에 해당하는 정규화된 차이 값은 $d_{\log}(i)$ 로 나타내며, $d_{\log}(i-1)$ 은 $i-1$ 번째 프레임에 대한 정규화된 차이 값을 나타낸다. $bd_{\log}(i)$ 는 이전 차이 값과의 차이($|d_{\log}(i)-d_{\log}(i-1)|$)를 나타내며, $fd_{\log}(i)$ 는 이후 차이 값과의 차이($|d_{\log}(i)-d_{\log}(i+1)|$)를 나타낸다.

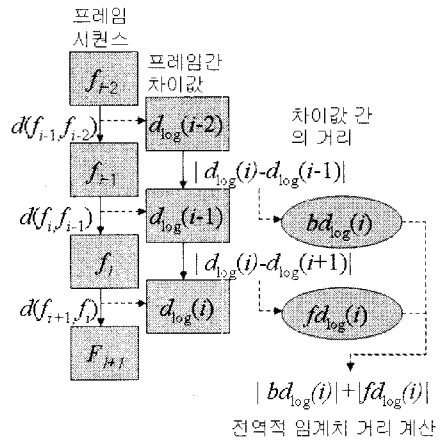


그림 1. 전역적 결정트리의 구조
Fig 1. Global Decision Tree Structure

(그림 2)에서는 전역적 결정트리를 이용할 경우 주어진 조건에 대하여 추출될 수 있는 다양한 장면전환의 예를 보여준다.

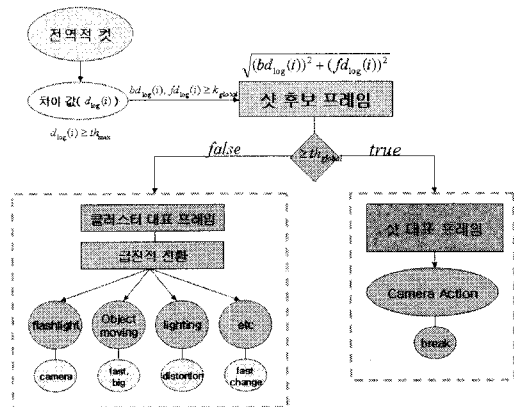


그림 2. 다양한 장면 전환의 예
Fig 2. Examples of Variety Scene Change

현재의 프레임이 카메라 브레이크가 발생한 프레임으로 고려되기 위해서는 가장 먼저 차이 값($d_{\log}(i)$)이 최대 임계치 (th_{max})를 만족해야 하며, 또한 이전 프레임과의 차이 ($bd_{\log}(i)$)나 이후 프레임($fd_{\log}(i)$)과의 차이도 임의의 정해진 상수 값(kg_{global})을 모두 만족해야 한다. 두 개의 조건 중 하나라도 만족하지 못할 경우에는 현재의 프레임이 샷 프레임이 아닌 클러스터로 묶일 수 있는 급진적 장면변환으로 고려되어 클러스터 후보 프레임으로 분류되어 지역적 결정 트리에 의하여 결정된다. 두 인접한 프레임들에 대한 거리 값

$(bfd_{log}(i) = \sqrt{(bd_{log}(i))^2 + (fd_{log}(i))^2})$)은 전역적 임계치 (thglobal)를 만족해야 하며, 이것은 샷 검출을 위한 카메라 브레이크 현상의 비디오 특성에 대한 민감성을 고려하여 각각의 차이 값들에 대한 임계치 조건은 물론 두 값의 합을 통한 거리에도 일정한 임계치의 범위를 두어 보다 적응적으로 결정하기 위한 것이다. 현재의 차이 값이 네 가지 조건을 모두 만족하였을 때 현재의 차이 값에 해당하는 프레임이 샷의 대표 프레임으로 결정된다.

IV. 실험

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

입력되는 비디오로부터 전역적 결정트리를 이용한 샷 검출은 연속되는 프레임으로부터 생성되는 각 차이값들에 대한 거리가 임계치를 만족하도록 해야 한다. <표 1>에서는 광고, 뉴스, 연예의 실험 비디오 데이터를 통하여 추출한 임계치의 최소값, 최대값, 그리고 전체 값들에 대한 전체 평균을 보여 준다. <표 1>에서 kglobal1은 bdlog(i)의 최소 임계치이며, kglobal2는 fdlog(i)의 최소 임계치를 나타낸다.

광고나 연예와 같은 장면의 변화가 많은 비디오는 이웃하는 이전이나 이후프레임과의 거리가 비교적 높게 형성되지만, 장면의 변화가 많지 않고 점진적인 장면의 변화가 많은 뉴스의 경우는 상대적으로 적은 거리 값을 형성하고 있다. 뉴스에서 사용되는 임계치를 광고나 연예에 적용하였을 경우 전역적 컷을 검출하는 데에는 크게 문제가 발생하지 않았다. 따라서 본 논문에서는 전역적 임계값을 두 인접한 차이 값들의 거리로부터 생성하여 비디오의 종류에 민감하지 않은 임계치를 결정하고자 하였다.

표 1. 전역적 결정트리를 적용하기 위한 임계치 범위
Table 1. Threshold Range for Application of Global Decision Tree

	임계값	thmax	kglobal1	kglobal2	thglobal
광고	최소값	415.2	118.1	98.7	162.3
	최대값	488	177.6	166.3	228.1
	전체평균	456.3	145.6	128.6	195.2
뉴스	최소값	422.4	73.4	62.4	96.3
	최대값	487.6	142.4	129.9	179.1
	전체평균	465.5	117.3	102.7	155.4
연예	최소값	447.5	132	88.1	158.7
	최대값	501.2	205.2	208.3	292.4
	전체평균	472.9	172.4	167.8	241.3

<그림 3>는 광고 비디오로부터 전역적 결정트리와 임계치를 이용하여 추출되어진 전역적 컷(shot)을 보여주고 있다.

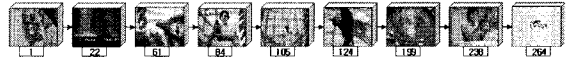


그림 3. 광고 비디오의 추출된 전역적 컷
Fig 3. Extracted Global Cut of Advertisement Video

<표 2>는 추출된 <그림 3>의 광고 비디오의 전역적 컷을 위한 비교 값들을 보여주고 있다.

표 2. 광고 비디오의 전역적 컷 추출을 위한 비교
Table 2. Comparison for Global Cut Extraction of Advertisement Video

	1shot	2shot	3shot	4shot	5shot	6shot	7shot	8shot	9shot
프레임 번호	1	22	51	61	105	124	199	238	264
$d_{log}(i-1)$	0	296.6	346.8	305.6	297	289.5	324.1	312.5	313.6
$d_{log}(i)$	477.7	474.4	477.3	438.2	415.2	454.4	456.7	446.3	488
$d_{log}(i+1)$	314.6	331	311	316	303.8	355.7	344.7	312.3	346.8
$bd_{log}(i)$	477.7	177.6	130.5	132.6	118.1	164.9	132.6	133.8	174.3
$fd_{log}(i)$	163.1	143.1	166.3	122.2	111.3	98.7	112	134	141.2
$bfd_{log}(i)$	504.8	228.1	211.4	180.3	162.3	192.2	173.6	189.4	224.3

광고 비디오는 장면전환의 횟수가 많으며, 카메라의 특수 촬영이나 객체의 빠른 움직임에 의하여 전역적 컷으로 예측될 수 있는 높은 차이 값을 형성하는 경우가 많다.

<그림 4>는 비교적 장면의 변화가 적으며 점진적인 장면의 변화를 많이 가지는 뉴스 비디오의 추출된 샷과 차이 값의 거리를 그림으로 보여준다.

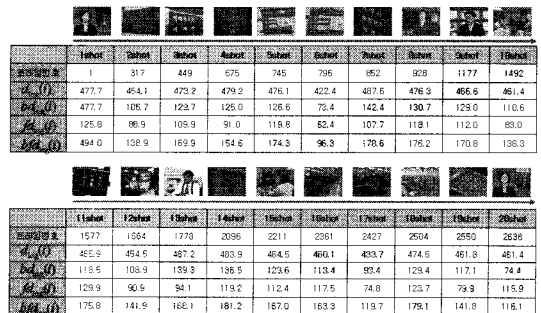


그림 4. 뉴스 비디오의 거리 비교
Fig 4. Distance Comparison of News Video

제한된 결정트리는 뉴스 비디오에서 대표적으로 잘못된 장면전환을 검출하도록 유도하는 플래시 라이트 현상을 제거할 수 있다. 플래시 라이트가 발생하는 프레임은 비교적 높은 차이 값을 형성하여 결정트리에 의하여 샷으로 결정된다.

〈그림 5〉는 연예 비디오로부터 추출되는 전역적 컷과 관련된 각 차이 값들에 대하여 보여준다.

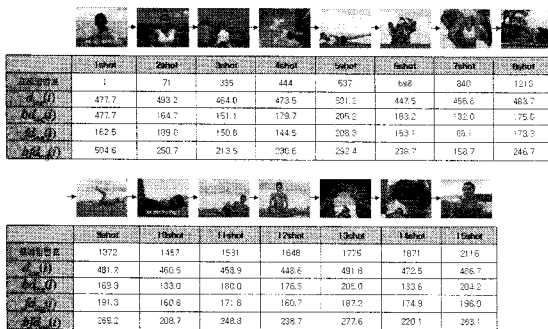


그림 5. 연예 비디오의 거리 비교
Fig 5. Distance Comparison of Entertainment Video

연예 비디오는 비교적 급진적 장면전환을 많이 가지고 있으며, 인터뷰와 같은 장면에서는 많은 플래시라이트를 프레임들에 포함하게 된다. 따라서 임계치의 결정이 민감하게 되고, 장면의 전환점을 찾는 것이 비교적 어렵다. 하지만 제안된 전역적 결정트리는 주어진 임계치를 이용하여 효율적으로 전역적 컷을 추출하고 있다.

플래시라이트는 카메라나 기타 조명의 플래시에 의해 나타나는 현상으로 갑작스런 조명의 변화에 의해 인접된 프레임들의 차이 값에 대한 변화가 매우 크게 나타나는 현상을 말한다. 이러한 플래시라이트는 뉴스나 연예 프로와 곳에서 많이 출현하게 되며 본 논문에서는 전역적 결정트리를 이용하여 추출하는데 〈그림 6〉은 플래시 라이트에 의한 각 프레임의 차이 값들에 대한 변화와 해당 프레임들의 예를 보여준다.

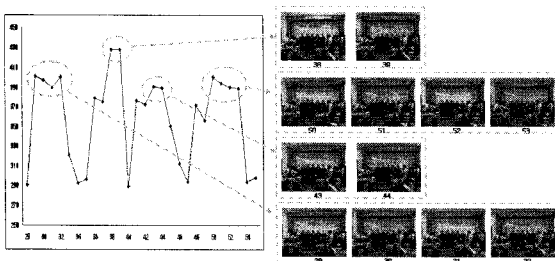


그림. 6 플래시라이트의 예
Fig 6. Example of Flashlight

플래시라이트가 발생한 프레임은 적어도 2개의 프레임이상에서 나타나며, 많이 발생하는 프레임의 경우에도 이 규칙은 예외 없이 적용되고, 연속된 프레임들의 비교에 의하여 검출될 수 있는 특징을 가지고 있다.

이처럼 전역적 결정 트리는 객체나 카메라의 움직임과 플래시 라이트와 같은 갑작스러운 장면 전환을 쉽게 검출할 수 있다.

V. 결론 및 향후 과제

본 논문에서는 카메라 브레이크와 같은 급진적 장면 전환을 전역적 장면영역의 검출 방법으로 전역적 결정트리를 이용하여 검출하였다. 입력된 비디오에 대하여 X2-히스토그램으로 장면 전환을 검출하고 정규화 과정을 통하여 이웃하는 프레임과의 차이값을 만족할 때 샷 프레임으로 추출하였다. 또한 전역적 결정트리를 이용하여 객체나 카메라의 갑작스런 움직임이나 플래시라이트와 같은 장면의 전환을 검출할 수 있었다.

본 논문은 연속된 프레임의 색상정보만을 사용하였으며, 텍스트나 소리와 같은 매체를 이용하지 않았다. 따라서 다양한 매체의 특성을 복합적으로 이용한 장면전환 검출 방법에 대한 필요가 요구된다. 또한 향후 비디오의 구성과 추출된 장면으로부터 장면의 분석 및 표현과 색인, 검색을 지원하는 연구가 요구된다.

참고문헌

- [1] M. Yeung, B. Yeo and B. Liu, "Segmentation of Video by Clustering and Graph Analysis," Computer Vision and Image Understanding, Vol. 71, No. 1, pp. 94-109, 1998.
- [2] B. -L. Yeo, "Efficiency processing of compressed image and video," Technical report, PhD thesis, Princeton University, 1996.
- [3] H. Zhang, C. Y. Low, Y. Gong and S. Smoliar, "Video Parsing Using Compressed Data," In Proc. SPIE Conf. Image and Video Processing II, Vol. 2182, pp. 142-149, 1994.
- [4] D. Zhong and S.F. Chang, "Video Object Model and Segmentation for Content-Based Video Indexing," IEEE International Conference on Circuits and Systems, Jun. 1997

- [5] M.M. Yeung, B.-L. Yeo, W. Wolf, and B. Liu, "Video Browsing Using Clustering and Scene Transition Compressed Sequences," IS&T SPIE, Multimedia Computing and Networking, 1995.
- [6] H. Zhang, S.Y. Tan, S. W. Smoliar, and G. Yihong, "Automatic Parsing and Indexing of News Video," Multimedia Systems, pp. 2:256-266, 1995
- [7] 신성운, 표성배, "텔레매틱스에서 효율적인 장면전환 검출기법을 이용한 비디오 브라우징," 한국컴퓨터정보학회 논문지 제11권 제4호, pp.147-154, 2006.
- [8] 김영례, 이양원, "지역적 χ^2 -테스트를 이용한 장면전환 검출 기법," 한국컴퓨터정보학회논문지 제11권 제3호, pp. 193-202, 2006.

저자소개



신성운

2003년 2월 군산대학교 컴퓨터과학
과 이학박사

2006년~현재 군산대학교 컴퓨터정보
과학과 교수

<관심분야> 비디오 인덱싱, 비디오 요약, 멀티미디어



문형운

1997년 명지대학교 산업기술대학원
컴퓨터공학과 졸업(석사)

2007년 육군본부 전산장교
1997~현재 LG CNS 공공사업본부
차장

<관심분야> 영상처리, 컴퓨터비전



이양원

1994년 8월 숭실대학교 전자계산학
과 공학박사

1986년~현재 군산대학교 컴퓨터정보
과학과 교수

<관심분야> 모바일 프로그래밍, 텔레
매틱스, 가상현실