

# 동적 임계값을 이용한 컷 검출

윤인구, 김우생

광운대학교 전자계산학과

서울시 노원구 월계동 447-1 광운대학교

Phone: +82-02-940-5217 Fax: +82-02-909-0998

E-mail: ikyoon@cs.kwangwoon.ac.kr, woosaeng@cs.kwangwoon.ac.kr

## Cut Detection with Dynamic Threshold

Inku Yoon, Woosaeng Kim

Department Of Computer Science, Kwangwoon Univ.

447-1, Wolgye-dong, Nowon-gu, SEOUL, KOREA

Phone: +82-02-940-5217 Fax: +82-02-909-0998

E-mail: ikyoon@cs.kwangwoon.ac.kr, woosaeng@cs.kwangwoon.ac.kr

### Abstract

A content-based search method is required for video which has an unformatted and huge size of data. The index technique is necessary for the content based search of the video data. The first step of the video indexing is a cut detection. We propose a dynamic threshold method which change a threshold value during the cut detection process. We demonstrate that the proposed method is more efficient than the existing methods.

### 1. 서론

컴퓨터와 통신 그리고 데이터 압축기술의 발달로 동영상 데이터를 이용한 다양한 서비스가 가능하게 되었다. 영상, 음성, 분사 정보 등 다양한 정보를 갖고면서 데이터 양이 방대한 동영상 데이터로부터 사용자가 필요로 하는 정보를 찾기 위해서는 기존의 키워드 기반의 검색은 한계에 도달한 상황이기 때문에 사용자가 원하는 정보를 내용에 기반하여 검색할 수 있는 방법들이 요구되고 있다[1]. 이러한 최근의 기술 발전 추세 및 시장 요구를 바탕으로 하여, 국제 표준화 기구인 ISO와 IEC의 연합기술위원회 산하의 MPEG에서는 MPEG-7이라는 이름으로 멀티미디어 데이터의 내용기반 검색을 위한 내용 표현 방식에 관한 국제 표준화 작업을 시작하였다[3].

동영상인 비디오를 내용기반 검색을 하려면 동영상 데이터를 색인하기 위한 비디오 파싱기법이 필요한데, 일반적으로 비디오 색인을 하기 위한 출발점은 컷 검출이며 현재까지 컷 검출을 위한 다양한 방법이 연구되어져 왔다. 이러한 방법들의 기본적인 원리는 임계값을 이용하는 것이다. 컷을 검출하기 위해 프레임간의 특징값의 차이를 구해

이 값이 미리 정해진 임계값보다 큰 경우 장면 전환이 이루어졌다고 판단하게 된다. 따라서 정확한 컷을 검출하기 위해서는 적절한 임계값이 선택되어야 함에도 불구하고 기존의 연구에서는 임계값에 대한 특별한 연구나 결과를 발표한 경우가 드물다.

기존의 컷 검출 방식에서는 실험을 통한 적절한 고정된 임계값을 사용해왔다. 그러나 서로 다른 동영상에 대하여 고정된 임계값을 사용하는 것은 문제가 있을 수 있다. 그 이유는 히스토그램 분포도와 같은 특징값들이 동영상마다 다르기 때문이며, 또한 하나의 영상이라도 많은 장면과 색상 변화가 있기 때문에 같은 임계값을 지속적으로 적용하는 것은 문제가 될 수 있다. 본 논문에서는 이러한 문제점을 해결하여 컷 검출의 효율성을 향상시키기 위해서 컷 검출 과정 중 임계값이 능동적으로 변화하는 동적 임계값 기법을 제안한다. 본 연구에서 제안하는 동적 임계값 기법은 기존의 임계값을 사용하는 대부분의 컷 검출 기법에 적용하여 사용할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 컷 검출 기법에 관하여 설명하고 3장에서는 제안하는 동적 임계값 기법에 대하여 설명한다. 4장에서는 실험 및 결과 분석을 하고 마지막으로 5장에서는

결론을 내린다.

## 2. 컷 검출 기법

비디오를 구성하는 최소단위는 필름 한 장에 해당하며 하나의 영상을 나타내는 프레임이다. 비디오에서 장면의 전환이 이루어지는 부분을 컷(cut)이라고 하고, 컷으로 구분되며 하나의 카메라 동작에 의해 촬영된 작은 비디오 단위를 샷(shot). 논리적인 내용이 같은 연속된 샷으로 이루어진 단위를 에피소드(episode)라 한다. 비디오를 샷으로 구분하는 작업을 비디오 분할이라고 하며, 비디오 분할을 위해 장면의 전환점인 컷을 검출하는 작업을 컷검출이라고 한다.

지금까지 컷검출을 위한 다양한 알고리즘이 연구되었다. 컷을 검출하기 위한 방법으로는 히스토그램의 차이 비교[4], 화소간의 차이 비교[5,6], 유사율 측정법[7], 에지 변화 비교[8], 압축 상관 계수 비교[9], 그리고 움직임 벡터 비교 등이 있다 [10]. 기존의 대부분의 컷 검출 기법에서는 고정된 임계값을 사용하였는데 반해 최근에 몇몇 연구에서는 다른 방법들이 시도되고 있다. [11]에서는 현재 조사하려는 장면을 장면전환이 일어난 부류와 장면전환이 일어나지 않은 부류 중 하나로 결정하는 패턴인식 문제로 모델링하고 신경망 모델의 하나인 학습 가능한 퍼셉트론 패턴인식기를 이용하여 보다 정확한 장면전환을 검출하기 위한 방법을 제안하였다. [12]에서는 은닉마코프 모델(HMM)을 사용하여 비디오 세그멘테이션을 하는 방법을 시도하였다. 컷 검출 방법은 HMM을 이용하여 음성정보를 포함한 여러 가지의 특징값들을 사용하며, 훈련된 HMM이 임계값을 대체하기 때문에 기존의 임계값이 필요하지 않다.

## 3. 동적 임계값을 이용한 컷 검출 기법

기존의 컷 검출 기법에서의 임계값은 대부분의 경우 실험을 거쳐 얻은 고정된 임계값을 사용한다. 그러나 서로 다른 동영상은 물론 하나의 동영상에서도 장면이나 색상의 변화가 많기 때문에 고정된 임계값이 모든 장면에 좋은 결과를 나타 내기는 어렵다. 동영상을 조사해보면 동영상의 종류에 따라 약간씩 다르긴 하지만 대부분의 컷들 사이의 간격이 최소한의 프레임 수는 넘은 것을 알 수 있다. 따라서 컷이 발생한 후 바로 다음 프레임에서 컷이 다시 발생할 확률은 매우 낮고 시간이 지날수록 컷이 발생할 확률은 높아진다. 고정된 임계값을 사용한 기존의 컷 검출 기법들은 이러한 점을 고려하지 않기 때문에 컷이 발생한 후 몇 프레임이 지나지 않아서 컷이 아닌데도 불구하고 잡음이나 빛, 카메라 동작 등의 영향으로

인하여 컷으로 잘못 검출하는 경우가 많다. 더욱이 임계치를 낮게두면 (일반적으로 임계치를 낮추는 것이 높이는 것보다 컷을 놓치는 경우를 줄일 수 있어서 바람직하다) 이러한 현상은 더욱 두드러지게 나타난다.

일반적으로 임계치가 낮을수록 제대로 된 컷을 발견할 확률이 높으나 잘못된 컷을 발견할 확률 또한 높아진다. 반면에 임계치가 너무 높으면 잘못된 컷을 발견하는 경우는 적어지나 제대로 된 컷을 발견하는 경우도 줄어들게 된다. 따라서 본 연구는 이러한 두 가지 점을 보완하여 컷이 발생할 가능성이 많은 구역에서는 임계값을 낮추고 컷이 일어날 가능성이 적은 구역에서는 임계값을 높임으로써 실제의 컷은 검출하되 잘못된 컷은 가급적 줄어지는 데 목적이 있다. 동적 임계값 기법에서의 동적 임계값  $T_{dyn}$ 은 다음과 같다.

$$T_{dyn} = T_{high} - \frac{(T_{high} - T_{lim}) \times d_t}{d_{max}}$$

(단,  $0 \leq d_t \leq d_{max}$ )

여기서 초기 임계값  $T_{lim}$ 이란 기존의 임계값과 동일한 것으로 동영상들을 테스트해 구할 수 있는 적절한 임계값이며,  $T_{high}$ 란 컷이 일어난 직후의 높아진 임계값을 말한다. 이 방식에서 동적 임계값  $T_{dyn}$ 은 컷이 발생한 직후는  $T_{high}$ 값과 같다가 시간이 경과되면서 임계값이 자아지다가 결국은  $T_{lim}$ 값과 같아지게 된다.  $d_t$ 란 시간에 따른 거리의 의미하며 0부터  $d_{max}$ 까지의 값을 갖는다.

$T_{high}$ 와  $d_{max}$ 값을 어떻게 선정하는가에 따라 여러 가지 방식이 나올 수 있으나, 본 논문에서 제안하는 동적 임계값 방식은 다음의 기호화 식으로 정의한다. 초기 임계치로 컷을 검출할 때  $n$ 개의 컷이 검출된다고 가정할 때  $i$ 번째 컷으로 판단되는 프레임을  $F_i$ , 샷 경계를 발견하기 위해  $F_i$ 와 비교된 이전 프레임은  $F_j$ 라하고  $T_{high}^i$ 를 컷으로 검출된 프레임  $F_i$ 의  $T_{high}$ 값, 함수  $V(a)$ 는 프레임  $a$ 의 특징점 값, 함수  $D(a,b)$ 는 두 프레임  $a,b$ 사이의 거리라고 할 때  $T_{high}$ 와  $d_{max}$ 값은 다음과 같다.

$$T_{high}^i = |V(F_i) - V(F_j)|$$

$$d_{max} = \frac{\sum_{i=1}^n D(F_{i+1}, F_i)}{n-1}$$

동적 임계값의 장점은 영상의 종류에 따라 또는 같은 영상이라도 장면 변화에 따라 임계값이 달라진다는 것이다. 예를 들어 프레임간의 특징점 차이가 크고 장면 변화가 느린 동영상의 경우 자동적으로 동적 임계값이 높아지고 이를 적용하는 기간도 길게 설정하게 된다. 동적 임계값을 적용한 컷 검출 기법은 다음과 같다.

- 1) 초기 임계값을 적용하여 컷을 검출한다.

- 2) 컷이 검출되면 동적 임계값을 적용하여 컷을 조사한다.
- 3) 컷이 검출되면 (2)로 돌아가고, 만약  $d_{max}$  동안 컷이 검출되지 않을 경우 (1)로 돌아간다.

#### 4. 실험 및 결과 분석

본 연구에서는 제안하는 동적 임계값 기법의 성능을 평가하기 위해 동영상 데이터로 영화와 뉴스를 사용하였다. 영화 데이터는 초당 15 프레임이고 각 프레임은 320 \* 240 해상도의 RGB 컬러 영상이며 뉴스 데이터는 초당 10 프레임이고 각 프레임은 240 \* 180 해상도의 RGB 컬러 영상으로 구성되어 있다. 각 비디오 데이터는 사람의 눈에 의해 정확한 컷의 위치를 발견하였으며 이 데이터를 실제 컷의 근거로 삼았다. 영화의 경우 눈으로 확인한 실제와 컷 개수는 61개였으며, 뉴스의 경우는 64개였다. 사용된 데이터 중 뉴스는 샷의 전환이 비교적 확실한 반면 영화는 줌, 패닝 등의 카메라 기법을 통한 점진적인 장면 변환 등으로 샷의 전환이 확실하지 않은 경우가 많았다.

##### 4.1 사용한 컷 검출 기법

본 연구에서는 동적 임계값 기법을 조사하기 위해 히스토그램 컷 검출 기법을 사용하였다. 기존의 히스토그램 컷 검출 기법은 잡음이나 밝기의 미세한 변화와 카메라 움직임 등에 비교적 강한 장점을 갖고 있으며, 또한 가장 보편적인 방법이라서 채택하였다. 히스토그램 사용시 본 연구에서는 RGB 값을 그레이 값으로 변환한 후 히스토그램 구간(bin)을 64개로 설정하여 성능 평가를 하였다. 히스토그램 기법에서 컷 검출은 다음의 식처럼 비교되는 두 프레임간의 히스토그램의 차이가 주어진 임계값  $T$ 보다 크면 컷으로 판단한다. 여기서  $n$ 은 그레이 레벨의 개수,  $i$ 는 프레임 번호,  $j$ 는 그레이 레벨,  $H_i(j)$ 는  $i$ 번째 프레임의  $j$ 번째 그레이 레벨에서의 히스토그램 값이다.

$$\sum_{j=1}^n |H_i(j) - H_{i+1}(j)| > T$$

##### 4.2 동적 임계값 기법의 실험 결과

초기 임계치가 성능에 미치는 영향을 조사하기 위해서 초기 임계치를 20%, 30%, 그리고 40%로 각각 설정하여 조사하였다. 본 논문에서 히스토그램 기법의 임계치 20%란 비교되는 두개의 프레임에서 변경된 픽셀의 수와 프레임에 있는 총 픽셀의 수를 비교하여 20% 이상일 경우 서로 다른 프레임으로, 그 외의 경우에는 같은 프레임으로 간주함을 의미한다. 또한 컷 검출의 시간 단축과

정규화를 위하여 원래 영상을 108 \* 81 크기로 축소하여 사용하였다.

다음은 영화와 뉴스 비디오를 대상으로 히스토그램 컷 검출 기법에 기존의 정적 임계값과 동적 임계값을 적용했을 경우의 테스트 결과이다. <표 1>은 초기 임계값을 20%로 사용한 경우이다. 동영상의 종류에 상관없이 정적 임계값을 사용하였을 경우, 초기 임계값이 낮기 때문에 실제 컷을 제대로 검출 하지만 컷이 아닌 프레임은 컷으로 잘못 발견하는 경우도 많음을 알 수 있다. 반면에 동적 임계값을 사용하였을 경우 잘못 발견되는 컷의 수가 많이 줄어드는 것을 볼 수 있다.

| 동영상        | 임계값 기법 | 발견 | 못발견 | 잘못발견 |
|------------|--------|----|-----|------|
| 영화<br>(61) | 정적     | 61 | 0   | 45   |
|            | 동적     | 60 | 1   | 11   |
| 뉴스<br>(64) | 정적     | 64 | 0   | 34   |
|            | 동적     | 64 | 0   | 13   |

<표 1> 초기임계값 = 20%

<표 2>는 초기 임계값을 30%로 사용한 경우이다. <표 1>의 결과와 비교해 보았을 때, 모든 기법들이 실제의 컷을 적게 발견하는 반면 잘못된 컷을 발견하는 경우도 줄어들었음을 알 수 있다. 이것은 초기 임계값이 높아졌기 때문에 생기는 당연한 결과이다. 동적 임계값 방법이 기존의 방법보다 전체적으로 좋은 결과를 나타내기는 하나, 초기 임계값이 높은 상태에서 다시 동적으로 임계값을 높임으로써 실제 컷을 놓치는 경우도 발생함을 알 수 있다.

| 동영상        | 임계값 기법 | 발견 | 못발견 | 잘못발견 |
|------------|--------|----|-----|------|
| 영화<br>(61) | 정적     | 38 | 23  | 26   |
|            | 동적     | 35 | 26  | 8    |
| 뉴스<br>(64) | 정적     | 54 | 10  | 7    |
|            | 동적     | 52 | 12  | 2    |

<표 2> 초기임계값 = 30%

<표 3>은 초기 임계값을 40%로 사용한 경우이다. 임계값이 높아져서 실제의 컷이나 잘못된 컷을 제대로 발견하지 못함을 알 수 있다. 이러한 현상은 장면 변환이 불규칙한 영화에서 더욱 두드러짐을 볼 수 있다. 이 경우에도 동적 임계값 기법이 정적 임계값 기법에 비하여 잘못 발견하는 컷의 수를 상당히 줄임을 알 수 있다. 그러나 한편으로는 결과에서 보듯 초기 임계값이 높을수록 동적임계값 방식에서 실제의 컷을 놓치는 경우가 늘어나는데 이는  $d_{max}$  값이 필요 이상으로 커졌기 때문이다. 동적 임계값 기법에서  $d_{max}$  값은 초기 임계값으로 일단 검출된 컷 사이의 평균 거리로 구해지기 때문에 초기 임계값이 커지면 발견하는 실제의 컷의 수가 줄어들게 되어 상대적

으로 컷의 간격이 벌어져  $d_{max}$  값이 커지게 된다. 이는 초기 임계값이 잘못 선정되어 나온 필연적인 결과로 볼 수 있다.

| 동영상        | 임계값 기법 | 발견 | 못발견 | 잘못발견 |
|------------|--------|----|-----|------|
| 영화<br>(61) | 정적     | 24 | 37  | 13   |
|            | 동적     | 20 | 41  | 4    |
| 뉴스<br>(64) | 정적     | 33 | 31  | 1    |
|            | 동적     | 30 | 34  | 0    |

<표 3> 초기임계값 = 40%

전체적인 실험 결과를 볼 때 임계값뿐만 아니라 영상의 종류가 결과에 영향을 미치는 것을 알 수 있다. 특히, 영화와 같은 장면 변환이 불규칙한 동영상의 경우에는 초기 임계값이 많은 영향을 미침을 알 수 있다. 동영상의 종류에 상관없이 초기 임계값을 잘만 설정해주면 컷(실제 컷이든 잘못된 컷이든)을 제대로 검출 해준다는 것을 알 수 있다. 그러나 동적 임계값 기법이 정적 임계값보다 잘못 발견하는 컷의 수를 상당량 줄임을 알 수 있다. 반면에 동적 임계값 기법은 정적 임계값보다 임계값이 커지는 경우가 생가므로 초기 임계값이 높을 경우 상대적으로 기존의 정적 임계값 방식이 발견한 컷을 검출하지 못하는 경우가 발생함을 알 수 있다. 이 연구에서 제안하는 동적 임계 방식은 기존의 방식에서 발견하지 못하는 컷들을 더 발견하기 보다는 잘못 발견한 컷들을 최소한으로 줄이는데 그 목적이 있다. 따라서 초기 임계치를 적절히 낮추면서 이와 동시에 동적 임계값 기법을 사용할 때 가장 좋은 결과가 나타남을 알 수 있다.

### 5. 결론

비디오 데이터에 대한 내용기반 검색을 위한 많은 컷 검출 기법들이 제안되었다. 그러나 서로 다른 종류의 동영상은 물론 하나의 동영상에서도 장면이나 색상의 변화가 많기 때문에 기존의 고정된 임계값이 모든 경우에 좋은 결과를 나타내기는 어렵다고 본다.

이러한 문제를 보완하기 위하여 본 연구에서는 임계값이 영상의 종류에 따라 동적으로 변하는 동적 임계값 기법을 제안하였다. 구현과 실험을 통하여 제안하는 방식이 기존의 정적 임계값 방식보다 좋은 결과를 나타냄을 보였다. 특히 실제 컷 발견을 제대로 하기위해 초기 임계값을 그 목적에 맞추어 설정하였을때 제안하는 동적 임계값 기법의 효과는 더욱 크다고 할 수 있다.

### 참고문헌

- [1] A. and M. Witbrock, "Informedia News-On-Demand: Using Speech Recognition to Create a Digital Video Library", AAAI-97 Spring Symposium on Intelligent Integration and Use of Text, Image, Video and Audio Corpora, pp.24-26, March 1997.
- [2] 조현철, 윤인구, 김우생, "영상과 음성정보를 이용한 비디오 편집 및 검색 시스템", 98 가을 한국정보과학회 학술발표논문집, October 1998.
- [3] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Requirements Document V.5", MPEG98/N2208, Tokyo, March 1998.
- [4] M.J. Swain, D.H. Ballard, "Color Indexing", International journal of computer vision No. 1, pp. 11-32, November 1991.
- [5] Zhang, H.J., Kankanhalli, A., Smoliar, S.W., "Automatic partitioning of full-motion video", Multimedia Systems, pp. 10-28, 1993.
- [6] Shahraray, B., "Scene change detection and content-based sampling of video sequences", in Digital Video Compression: Algorithms and Technologies, Arturo Rodriguez, Robert Safranek, Edward Delp, Editors, Proc. SPIE 2419, pp. 2-13, February, 1995.
- [7] Kasturi, R., Jain R., "Dynamic vision", In Computer vision:Principles, Kasturi R., Jain R., Editors, IEEE Computer Society Press, Washington, 1991.
- [8] Ramin Zabih, Justin Miller, Kevin Mai, "A feature-based algorithm for detecting and classifying scene breaks", Proc. ACM Multimedia 95, pp. 189-200, November, 1995.
- [9] F. Arman, et. al, "Feature Management for Large Video Database," Proc. of SPIE-Storage and Retrieval for Image and Video Databases II, Vol. 2185, pp. 140-149, 1994
- [10] Q. Wei and Y. Z. Zhong, "Content-Based Parsing in Video Database," Proc. of the First International Conference on Multimodal Interface, Beijing, pp. 93-96, 1996.
- [11] 이충훈, 이홍규, "패턴인식을 이용한 MPEG 비디오 스트림상에서의 장면전환 검출", 98 봄 한국정보과학회 학술발표논문집, April 1998.
- [12] J. S. Boreczky, L. D. Wilcox, "A hidden markov model framework for video segmentation using audio and image features", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, pp. 3741-3744, Seattle, WA, 1998.