

다중 특징을 포함한 키 프레임 추출에 의한 장면 전환 검출 오류 자동 수정 기법

윤주현, 염성주, 김우생

광운대학교 컴퓨터 공학부

e-mail : jhyoon@cs.kwangwoon.ac.kr

An automatic fault correction technique in the scene change detection by the key frame extraction includes multiple features

Juhyun-Yoon, Sungju-Youm, Woosaeng-Kim

Dept. of Computer Science, Kwangwoon University

요 약

본 논문은 다중 특징을 포함한 대표 키 프레임을 추출을 통해 장면 전환 검출 시 발생할 수 있는 검출 오류를 자동으로 인식하고 수정함으로써 빠르고 신뢰성 있는 장면 분할을 수행하는 새로운 기법을 제안한다. 이를 위해 개선된 고속 장면 전환 검출 기법에 의해 샷을 분할 하고 분할 된 샷 으로부터 대표 키 프레임과 그것에 포함된 후보 키 프레임들의 다중 정보를 포함시킴으로써 샷의 전반에 대한 정보를 보다 잘 표현할 수 있도록 한다. 그리고 다중정보를 포함한 대표 키 프레임의 비교를 통해 샷 검출 오류를 자동으로 인식하여 적절히 수정할 수 있는 기법을 제안하며 실세계 동영상 데이터를 사용한 실험을 통해서 제안하는 기법에 의해 효율적으로 샷이 분할 될 수 있음을 보 인다.

1. 서론

동영상 데이터로부터 컷(cut)이라 불리는 장면전환 지점을 검출 해 내는 일은 내용기반 분석을 위해 먼저 수행되어야 할 기초 작업이라 할 수 있다. 컷에 의해 구분되는 동영상 데이터의 일부분들은 샷(shot)이라 불리게 된다. 일반적으로 장면전환은 갑작스러운 전환(abrupt change)와 점진적인 전환(gradual change)으로 분류 할 수 있다. 컷 검출에 관한 기존의 연구들은 매우 다양한 형태로 진행되어 왔으며 이들은 사용하는 주요 특징값에 따라 여러 가지 형태로 분류될 수 있다[1][2][3]. 이러한 기존의 연구들이 갖는 공통적인 문제점은 정확한 컷 검출을 위해 복잡하고 높은 연산 비용을 필요로 하는 작업들을 반복적으로 하게 되어 긴

분석시간을 갖게 되고 다양한 형태의 검출 오류에 의해 검출된 컷의 정확성이 떨어지게 된다는 점이다. 따라서 보다 빠르고 정확하게 컷 검출을 수행할 수 있는 방안이 절실히 요구된다.

빠르고 정확하게 동영상을 분석하기 위해서는 프레임간의 비교 방법을 개선하여 처리속도를 높이고 검출된 컷들로부터 분할 된 샷을 효율적으로 표현할 수 있는 키 프레임을 생성하는 방안이 있다. 처리 속도를 높이기 위한 방안으로 기존의 연구들은 대부분 비교해야 할 프레임 수를 최대한 줄여 연산 양을 감소시키는 방법을 사용한다. 이진 연산을 사용해 비교 프레임 수를 줄이는 방식[4]이나 비교 구간을 동적으로 늘려가면서 비교 대상을 줄이는 방식[5] 등의 연구가 대표적인 방법이며 분할된 샷이 가지고 있는 정보들을 최대한 잘 표현할 수 있는 키

프레임을 추출하기 위해서는 [6][7]등의 기법들이 제안된 바 있다.

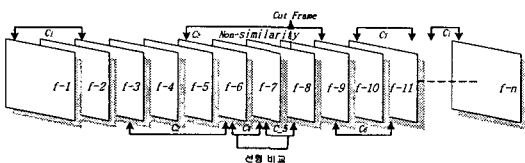
본 논문에서는 검출되는 컷의 신뢰성을 보장하면서 빠른 속도로 처리될 수 있도록 하기 위해 다중 특징값을 포함하는 대표 키 프레임 추출을 통한 샷 분할 기법과 검출 오류를 자동으로 인식하여 정확성을 높일 수 있는 새로운 방법을 제안한다. 제안하는 방법은 빠른 처리와 점진적인 전환 검출에 효율적일 수 초기 샷 분할을 수행하며 후보 키 프레임들로부터 나오는 다중정보를 대표 키 프레임을 중심으로 구성하며 이러한 정보를 통해 검출 오류가 일어나는 경우 자동적으로 감지하고 수정함으로써 보다 정확한 샷 분할이 수행될 수 있도록 하는 방법이라 하겠다.

2. 다중 특징을 포함한 키 프레임 추출 기법

2.1 개선된 고속 장면 전환 검출 기법

적응적 생략 기법[2]은 비교되는 두 프레임이 유사한 경우 비교 대상이 되는 다음 프레임까지의 간격을 증가시키며 탐색해 나가는 방식으로 매 프레임을 비교 대상으로 삼아야 하는 선형 탐색에 비해 비교 대상이 되는 프레임을 생략함으로써 처리 속도를 높일 수 있는 기법이라 하겠다. 그러나 적응적 생략 기법은 디졸브(dissolve)와 같은 점진적인 전환이 자주 나타나는 동영상 데이터에서는 점진적 전환이 나타나는 구간에서 생략 가능한 프레임 수를 계속 재 측정하게 되고 검출 오류가 선형 탐색 방법에 비해 자주 발생하게 되는 단점을 갖는다. 따라서 본 논문에서는 점진적인 전환 검출에 필요한 효율성을 높이기 위해 새로운 비교 기법을 제안한다. 제안하는 비교 기법을 정리하면 다음과 같다.

- 1) 현재 프레임 f 를 $f+n$ 의 위치의 프레임과 비교 (n 의 초기 값은 2)
- 2) 다음 비교에서는 n 의 값을 1 증가시키고 다음 f 는 $f+n$ 으로 해서 비교 구간을 늘려 가면서 비교하지 않아도 되는 프레임은 제외시킴
- 3) 비교되는 두 프레임이 유사하지 않을 경우 n 을 초기 값으로 돌리고 f 의 위치에서 장면 전환이 검출될 때까지 매 프레임 비교
- 4) 컷을 찾으면 현재 프레임 f 는 컷의 다음 프레임으로 하고 n 은 초기 값으로 조정 한 후 1)로 돌아감



<그림 1> 개선된 고속 장면 전환 검출

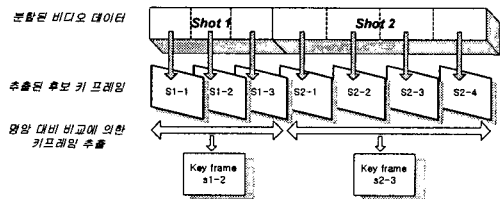
제안하는 기법은 <그림 1>에서와 같이 프레임간의 비교간격을 일정하게 늘려 나가면서 유사하지 않은 비교 구간 C_3 이 나타났을 경우 해당 구간 내에 장면 전환 현상이 있는 것으로 간주하고 비교구간 C_3 의 첫 프레임인 f_5 의 위치에서부터 장면 전환이 일어나는 프레임 f_8 을 발견할 때까지 매 프레임을 비교하는 방식으로 진행 시켜 나가면서 장면전환 유무를 검출하게 된다.

제안하는 기법은 부분적으로 매 프레임 비교를 하게 되므로 적응적 생략 방식에 비해 연산 비용이 높게 되지만 점진적인 변화가 많은 비디오 데이터에서는 프레임 생략을 위한 재 계산 과정 없고 정확한 장면 전환 검출에 필요한 정보를 추출하는데 오히려 효과적이라 할 수 있겠다.

2.2 샷 구간 분할에 의한 대표 키 프레임 추출

샷 구간들이 결정되면 각 샷으로부터 다중 특징 정보[8]를 포함하는 대표 키 프레임을 생성하기 위해서 본 논문에서는 각 샷 구간들로부터 후보 키 프레임들을 추출하고 이들의 특징들을 다시 분석하여 최종적인 대표 키 프레임을 추출하는 과정을 거치게 된다.

먼저 후보 키 프레임을 추출하기 위해서 샷 구간의 총 길이 즉, 해당 구간의 총 프레임 수를 t 라 하고 같은 샷 구간에서 프레임 유사성 측정을 위해 비교된 회수를 c 라 할 때 본 논문에서는 실험을 통해 가장 적당한 것으로 판단된 방법인 $N = \frac{ct}{t} \times 100 / 10$ 에 의하여 얻어진 값 N 으로 해당 샷 구간을 균등 분할하며 분할 되는 각 지점의 프레임들을 후보 키 프레임들로 삼는다. <그림 2>는 샷 1과 2에서 N 의 값이 각각 3과 4가 되는 경우 균등 분할된 각 샷의 각 구간으로부터 후보 키 프레임을 추출하는 예를 보여준다.



<그림 2> 샷 구간 분할에 의한 후보 키 프레임 추출

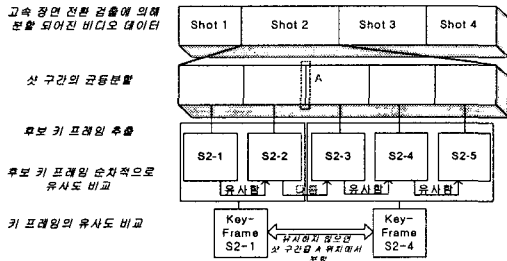
각각의 후보 키 프레임들에서 샷을 대표할 대표 키 프레임을 추출하기 위해서 본 논문에서는 명암도 히스토그램을 사용하여 후보 키 프레임들 중에 가장 명암 대비가 좋은 프레임을 대표 키 프레임으로 삼는 방법을 사용한다. 명암 대비가 가장 좋은 영상의 경우 추후에 사용자가 브라우징을 할 경우에 보다 명확한 영상을 제공할 수 있는 잇점이 있다. 이러한 방법으로

하나의 샷 구간에서 대표 키 프레임이 결정되면 대표 키 프레임은 나머지 후보 키 프레임의 특징 정보들을 링크드 리스트(linked list) 형식으로 저장하도록 하여 대표 키 프레임이 샷의 전체적인 특징을 표현 할 수 있도록 하였다.

3. 장면전환 검출 오류 자동 인식기법

장면 전환 검출에서 검출 오류가 발생하는 경우는 크게 두 가지로 분류되는데 장면 전환이 일어난 사실을 발견하지 못하는 미 검출(miss cut)의 경우와 장면 전환이 발생한 것이 아님에도 이를 장면 전환으로 인식하는 오 검출(fault cut)의 경우로 분류될 수 있다.

장면 전환이 미 검출되는 경우는 대부분 인접 프레임간의 변화율이 갑작스러운 변화를 포함하지 않는 경우 일어나는데 이러한 상황의 대표적인 예로 점진적인 전환 현상을 들 수 있겠다. 이를 해결하고자 본 논문에서는 후보 키 프레임들 간의 비교를 통해 장면 전환이 발생할 가능성이 높은 지점을 자동으로 인식하여 재 분할 할 수 있는 방법을 제안한다.



<그림 3> 후보 키 프레임을 이용한 미 검출 해결 방법

제안하는 방법은 하나의 샷에 속한 후보 키 프레임들이 S_0, S_1, \dots, S_n 이라 할 때 인접한 후보 키 프레임들간에 $S_i \neq S_{i+1} (i=0, \dots, n-1)$ 의 관계가 발견될 때 S_i 와 S_{i+1} 사이의 중간지점을 새로운 샷 경계로 추가하는 방법이다. <그림 3>은 제안하는 방법에 의해 샷 경계가 새롭게 추가되는 예를 보여주는 것으로 후보 키 프레임 S2-2와 S2-3이 서로 유사성이 없는 것으로 판단되어 두 후보 키 프레임의 중간지점을 새로운 샷 경계로 추가한 상황을 보여준다.

장면 전환이 아닌데도 불구하고 잘못 검출된 오 검출 상황 에서는 인접한 두 샷의 대표 키 프레임과 링크드 리스트로 연결된 후보 키 프레임의 특징 정보를 서로 비교함으로써 분할 된 샷이 오 검출에 의한 것 이었음을 발견하는 방안을 제안한다. 일반적으로 오 검출의 경우에는 대부분 한 샷 내에서 비교 프레임 영상의 잡음이나 순간적인 물체의 등장과 소멸에 의한 부분적인 변화에 의해 생기는 경우가 대부분이다. 따라서 인접한 두 샷에 속한 후보

키 프레임들의 정보를 비교하게 될 경우 두 샷의 전체적인 특징을 비교함으로써 순간적인 변화에 의한 장면 분할이 잘못된 것임을 쉽게 인식할 수 있게 된다. 본 논문에서는 대표 키 프레임이 가지고 있는 후보 키 프레임들의 특징 정보로 HSV 64-빈 컬러의 히스토그램을 사용한다. 대표 키 프레임 간의 비교를 위해 대표 키 프레임에 저장된 후보 키 프레임 중 시간적으로 가장 앞선 후보 키 프레임의 히스토그램을 $Ch_b(i)$, 가장 뒤에 있는 후보 키 프레임의 히스토그램을 $Ch_e(i)$ 라 할 때 n번째 샷의 평균 변화율 $DiffH_n$ 는 $DiffH_n = \sum_{i=0}^{64} [Ch_b(i) - Ch_e(i)]$ 로 표현 된다. 그리고 대표 키 프레임의 컬러 히스토그램을 $KH_n(i)$ 대표 키 프레임 n에 연결된 모든 후보 키 프레임들의 평균 컬러 히스토그램을 $AH_n(i)$ 라 할 때 유사도 차이 비율 $DiffRate$ 은 (식 1)와 같이 구할 수 있다. 본 논문에서는 (식 1)에 의해 얻게 되는 $DiffRate$ 이 임계값 θ 에 대해 $DiffRate < \theta$ 를 만족할 때 유사한 것으로 판단하게 된다.

$$DiffKey = \sum_{i=0}^{64} \left(\frac{KH_n(i) + AH_n(i)}{2} - \frac{KH_{n+1}(i) + AH_{n+1}(i)}{2} \right) + (DiffH_n - DiffH_{n+1}) \quad (1)$$

$$DiffRate = \frac{DiffKey}{\text{한 프레임 영상의 총 픽셀 수} \times 2} \times 100$$

4. 실험 결과 분석

본 논문에 구현된 비디오 분석 시스템은 MS-Windows 2000 server 운영체제를 갖는 PC에서 Visual C++ 6.0과 MS Direct Show 6.0 SDK 사용하여 제작되었다. 실험에 사용한 데이터들은 인터넷상에 배포된 다양한 종류의 데이터를 사용 하였는데, 영화 예고, 뮤직 비디오, 광고, 뉴스 등 5000 프레임 이내의 동영상 클립을 사용하였으며 동영상 형식이 데이터마다 약간씩 차이가 있음을 고려하여 각 프레임을 100x100픽셀 크기로 정규화 하여 사용하였다.

<그림 4>는 선형 비교 방식과 제안된 방법에 의해 약 500 프레임 정도의 동영상 데이터의 장면 전환 지점 검출에 사용된 결과를 나타낸 그래프로써 x축은 프레임 수를 y축은 프레임간의 특징값 차이를 나타낸다. <그림 4>의 b 경우 y축 방향으로 표시되지 않은 부분은 2.1 절에서 제안한 방법으로 비교가 생략된 것을 나타낸다. 따라서 본 논문에서 제안된 방법은 비교되는 프레임들의 수를 볼 때 기존의 방법에 비해 현저히 적어져 빠른 처리가 가능 함을 알 수 있다.



a. 선형 비교방식의 경우



b. 제안한 방식의 경우

<그림 4> 프레임간 특징 변화 그래프의 비교

<표. 1>은 제안하는 방법과 기존의 방법들간의 처리속도를 비교한 것이다. 제안된 방법은 점진적인 전환이 많은 뮤직비디오나 광고 같은 경우 적응적 생략 방법과 유사했으며 검출 오류 처리과정을 거칠 경우 적응적 생략 방법보다는 느려지지만 검출 오류를 크게 줄일 수 있기 때문에 훨씬 효율적이며 오류 처리과정을 거치더라도 선형 비교방식에 비해 빠른 처리가 가능한 장점을 갖는다.

<표. 1> 제안된 방법의 수행시간 비교

비디오 클립	총 프레임 수	선형 비교	적응적 생략방법	제안된 방법	검출 오류 처리
스포츠 뉴스	1650	4'12"	48"	1'7"	2'30"
Music Video	1160	3'3"	32"	46"	1'25"
광고	890	2'31"	21"	37"	1'2"

<표. 2>은 제안한 방법에 의해 오 검출 및 미 검출된 것을 자동 인식한 결과를 precision을 사용하여 평가한 결과를 나타낸 것이다. 이 표에서 기존의 방식이란 본 논문에서 제안한 검출 오류 자동 인식을 적용하지 않은 결과를 보여준다. 이 결과에서 알 수 있듯이 제안한 기법을 사용하게 될 경우 미 검출과 오 검출이 자동으로 인식되어 수정될 수 있기 때문에 전체적인 precision이 매우 향상된 것을 알 수 있다.

<표. 2> 검출 오류가 발생하는 경우 자동 인식 결과

데이터	기존의 방식		검출오류 자동인식		Precision(%)	
	FALSE	MISS	오검출 인식	미검출 인식	오검출 인식	미검출 인식
스포츠 뉴스	7	4	7	3	100	75
드라마 1	9	7	8	7	88.8	100
드라마 2	6	5	6	5	100	100
CF 1	7	2	5	2	71.4	100
CF 2	8	4	8	3	100	75
MusicVideo 1	9	6	9	6	100	100
MusicVideo 2	8	5	8	4	100	80
MusicVideo 3	12	9	11	7	91.6	77.7

5. 결론

동영상 데이터의 컷 검출과 샷 분할 작업은 내용기반 분석에 필요한 기초가 되는 작업으로 매우 중요하다고 할 수 있겠다. 본 논문에서는 이 두 가지 조건을 만족할 수 있는 새로운 샷 분할 방법을 제안했다. 빠른 처리속도를 위해서 먼저 기존에 제안된 적응적 생략 기법을 개선한 방법을 제안했다. 제안한 방법은 점진적인 전환이 자주 일어날 경우 적응적 생략기법이 오히려 비 효율적인 점에 착안하여 해당 구간에서 순차 처리를 하도록 하여 오버헤드를 감소시킬 수 있었다. 또한 후보 키 프레임 추출과정을 거쳐 대표 키 프레임을 추출하도록 하고 대표 키 프레임을 중심으로 한 샷의 전체적인 정보를 저장하도록 구성하여 샷과 샷 또는 프레임과 프레임간의 유사성 비교에서 보다 정확한 비교가 될 수 있도록 하였다. 이러한 기법을 미 검출과 오 검출을 자동으로 인식하고 수정할 수 있는 방안도 제안했다. 제안된 기법들은 실세계 동영상 데이터를 사용한 실험을 통해서 처리 속도 향상과 정확성 모두를 만족함을 알 수 있었다. 앞으로 서로 다른 특징값을 사용하여 효율을 더욱 높이는 방안에 대한 연구를 진행하게 될 것이다.

참고문헌

- [1] John S. Boreczky, Lawrence A. Rowe, "Comparison of video shot boundary detection techniques", in Storage & Retrieval for Image and Video Database IV, Proc. of SPIE 2670, pp. 170-179, 1996
- [2] C. W. Ngo, T. C. Pong, R. T. Chin, "Detection of Gradual Transitions through Temporal Slice Analysis," Proc. IEEE CVPR99, pp.36-41, June, 1999.
- [3] Dalong Li, Hanqing Lu, "Model Based Video Segmentation," The IEEE Workshop on Signal Processing Systems, pp.120-129, 2000.
- [4] 김성철, 오일근, 장종환 "동영상의 고속 장면분할을 위한 이진검색 알고리즘" 정보처리학회지 07, 2000, Vol 07, No 04, pp. 1044-1049.
- [5] Kien A. Hua, JungWhan Oh, "Detecting Video Shot Boundaries Up to 16 Times Faster" 385-387, ACM Multimedia 2000: Marina del Rey, Los Angeles, CA, USA
- [6] X. Sun, M. S. Kankanhalli, Y. Zhu, and j. Wu, "Content-based Representative Frame Extraction for Digital Video" Intl. Conf. on Multimedia Computing and Systems, Austin Texas, pp. 237-240, June 1998.
- [7] Girsensohn, A.; Boreczky, J. "Time-Constrained Key-Frame Selection Technique" Multimedia Computing and Systems, IEEE International Conference on, Volume: 1, 1999 pp. 756 -761 vol.1
- [8] Byongchul Ko, Hae-sung Lee, Hyeran Byun, "Content-Based Retrieval System Design for Image and Video using Multiple Fetures" Journal of KISS(B) VOL. 26, NO. 12, DECEMBER 1999.