

동적 분할 기법을 이용한 효율적인 대표키 프레임 추출

김 영 희* 이 순 희** 최 운 중* 박 장 춘*
*건국 대학교 컴퓨터 공학과 **안산 공과 대학

Efficient Representative-Key Frame Extraction Using Dynamic Segmentation Method

Young-Hee Kim* Soon-Hee Lee** Woon-Jong Choi* Chang-Chun Park*
*Dept. of Computer Engineering, Konkuk University **Ansan College of Technology

요 약

본 논문에서는 내용 기반 검색을 위한 방법의 핵심 기술 중의 하나인 효율적인 대표키 프레임 추출을 위해 동적인 분할 기법을 제안하고, 다양한 장르의 비디오 데이터에 대하여 그 성능을 평가해 본다. 제안된 대표키 프레임 추출법은 기존의 균등 분할 방법에서 필요 이상의 분할이나 적은 분할로 인하여 중복 추출되거나 추출 대상에서 제외되었던 대표키를 효율적으로 검출할 수 있는 방법으로서 이는 분할 기준을 영상의 각 장르별 특징에 맞도록 동적으로 적용함으로써 획일적인 값을 적용한 기존의 연구와는 차별을 가진다. 제안된 알고리즘으로 대표키 프레임 추출을 위하여 뉴스, 영화, 뮤직 드라마, 광고와 같은 다양한 장르별로 실험한 결과, 제안된 대표키 프레임 검출 방법이 효율적임을 확인할 수 있었다.

1. 서 론

최근 영상 매체의 발달과 함께 영상 관련 종사자들이 증가하고 있고, 오락성과 대중성을 가진 비디오 데이터에 대한 VOD와 같은 다양한 서비스가 제공되고 있다. 이런 서비스는 컴퓨터와 통신, 그리고 영상 데이터 압축 기술의 발달에 힘입어 급속히 발전하고 있다. 동영상 데이터의 특징은 영상 정보 뿐 아니라 음성 정보와 문자 정보를 비롯한 여러 가지의 의미있는 정보들을 가지고 있으며, 시간적, 복합적, 비정형적, 대용량적 특징을 동시에 가지고 있는 점이다. 이러한 동영상과 비디오가 가진 방대한 양의 데이터로부터 사용자가 필요로 하는 정보를 찾는 일은 기존의 키워드나 인덱스 기반의 검색 방법으로 검색하기에는 이미 한계에 부딪치고 있는 실정이다. 그러므로 동영상 비디오 데이터가 담고 있는 내용을 표현하고 검색하기 위한 방법으로 내용 기반 이미지 검색에 대한 연구는 꼭 필요하다. 내용 기반 비디오 검색에는 비디오 데이터를 색인하기 위한 비디오 파싱(비디오 파싱은 비디오 분할과 비디오 색인을 포함), 제한된 저장 공간에 대용량 비디오 및 이에 관련된 메타 데이터를 효율적으로 저장하기 위한 비디오 압축 및 저장, 그리고 사용자가 원하는 비디오를 검색할 수 있는 환경울. 제공하기 위한 비디오 검색 및 브라우징 기술이 핵심 요소 기술로 지적되고 있다.[1] 내용 기반 비디오 검색의 핵심 기술 중의 하나인 비디오 분할은 컷이나 점진적 장면 전환 효과로 나타나는 샷(shot)이나 장면 사이의 경계를 검출하는 기술인데, 비디오 스트림의 인접한 프레임 간의 측정된 특징 차이를 비교하여 임계값 이상일 경우에 샷(shot)의 경계 지점으로 검출한다.

기존에는 대표키를 추출하기 위한 방법에서 사용된 균등 분할 방법에서는 필요 이상의 분할이나 적은 분할로 인하여

중복 추출되거나 추출 대상에서 제외되는 대표키가 생길 수 있는 단점이 있다. 이런 단점을 개선한 동적 분할 방법은 분할 기준을 영상의 각 장르별 특징에 맞도록 동적으로 적용함으로써 모든 대표키 프레임을 추출할 수 있다. 제안된 동적 분할 기법을 적용한 대표키 프레임 추출 방법을 뉴스, 영화, 드라마, 상업용 광고와 같은 다양한 장르별로 실험한 결과, 제안된 방법이 효율적임을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 언급하고, 3장에서는 동적 분할 기법을 적용한 대표키 프레임 추출 방법을 설명한다. 제 4장에서는 기존 방식과의 비교를 통하여 제안한 방식의 효율성과 정확성을 보여주며 마지막으로 5장에서는 결론을 내린다.

2. 관련 연구

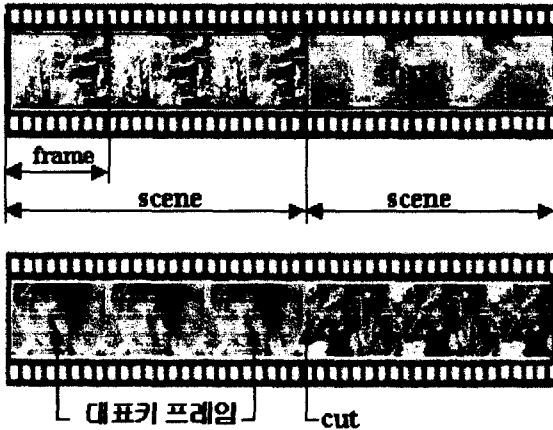
지금까지 컷 검출을 위한 다양한 알고리즘이 연구되었다. 비디오 프레임 간의 장면 전환에서 발생하는 샷(shot)은 기존의 연구에서 샷(shot)의 경계 지점을 검출하는데 관심이 모아졌으나 최근의 연구에서는 샷(shot) 뿐 아니라 씬(scene)이나 이와 관련된 대표키 프레임들을 검출하는 기법들이 등장하고 있다. 먼저 샷(shot) 경계를 찾는 전형적인 방법은 비디오 내의 인접한 두 프레임 간에 특징 차이를 구한 후 비교 값이 일정한 임계값 보다 클 경우 장면 전환이 이루어 졌다고 본다. 이러한 연구들은 화소간의 차이 비교[2,3], 유사율 측정법 [4], 히스토그램 분포 차이 비교[5,6], 에지 변화 비교[7], 압축 상관 계수 비교[8], 그리고 움직임 벡터 비교[9] 등에 기초를 두고 있다. 키 프레임 추출 방법 중 가장 간단한 방법은 각 샷(shot)마다 임의의 한 프레임을 추출하여 이를 키 프레임으로 삼는 것이다. [10] 또한 [11]에서는 샷(shot) 경계의 검출 없이 비디오 영상의 정보만 가지고 대표키 프레임들을 찾는 방

법을 제안한 방법이다. 이 방법은 우선 프레임들을 일정 간격으로 나누는 후 각 간격의 특징 차이를 구한 다음 차이가 큰 그룹과 작은 그룹으로 나눈다. 차가 작은 그룹에 있는 프레임들 중 가장 앞과 뒤의 프레임만 남기고 나머지 프레임들은 삭제한 후 다시 두 그룹을 합친 후 이들을 다시 그룹화와 삭제의 작업을 반복 수행하여 사용자가 원하는 대표키 프레임들을 만드는 방법이다. 또한 기존의 일반적인 방식처럼 비디오 컷들을 검출하지 않고도 대표키 프레임들을 찾는 방법으로는 내용 기반 적응 클러스터링 방식이 있다. [12]

그리고 계승된 특징 차이를 이용한 연구는 비디오 분할 과정에서 프레임들 간의 특징 차이를 계승하여 비디오의 주요 프레임들을 짧은 시간 안에 검출할 수 있는 기법이다. [13] 이 기법은 계승된 특징 차이를 이용하여 n 개의 대표키 프레임들을 검출하는 방법을 먼저 비디오 전체를 균등 분할하여 m ($m > n$)개의 프레임들을 추출한 다음, 계승된 특징 차이가 가장 작은 두 개의 프레임 F_i 와 F_j 를 선택하고, 이 두 프레임들을 해당 프레임 그룹의 중간 프레임인 $F((i+j)/2)$ 로 대체한다. 마지막으로 남은 프레임 수가 n 이면 두 번째인 계승된 특징 차이가 가장 작은 두 개의 프레임 선택으로 돌아간다. [13]

3. 동적 분할 기법을 적용한 대표키 프레임 추출

3.1 비디오 데이터 구조



(그림 1) 비디오 데이터의 구조
(Fig.1) The structure of video data

비디오 데이터는 (그림 1)과 같이 하나의 장면을 나타내는 프레임의 연속된 집합으로 구성되고, 연속적인 카메라 동작에 의해 촬영된 비디오 단위의 샷(shot)은 비디오 분할의 기본 단위면서 몇 개의 프레임으로 구성되어진다. 또한 하나 혹은 그 이상의 샷(shot)들의 집합으로 씬(scene)이 구성되는데, 이는 하나의 스토리를 나타낸다. 일반적으로 주인공이나 특정한 장소와 같이 하나의 대상을 연속하여 촬영한 영상을 나타낸다. 하나의 씬(scene)을 대표하는 프레임들을 대표키 프레임이라 하는데, 만일 연속된 프레임들의 집합에서 프레임의 첫번째 프레임과 마지막 프레임이 대표키 프레임이고, 이들의 특징 차이가 비슷하다면, 이는 하나의 씬(scene)을 나타내고 있다고 할 수 있다. [14] 그리고 동영상 간의 내용 비교를 비롯하여 각종 내용 기반의 검색에서 전체 자료를 대신하여 이용 가능한 프레임은 키 프레임이라고 한다.

이러한 대표키 프레임들을 추출에 대한 최근 연구로는 계승된 특징 차이를 이용한 방법[13]이 있다. 이 방법에서는 사용자가 요구한 대표키 프레임 수 n 의 p 배수(개략 $p=3$ 정도)로 균등 분할하였다. 이 방법은 사용자가 영상에 대한 내용을 이미 파악하여 개략적인 씬(scene)의 개수를 파악하고 이를 근거로 대표키 프레임 수를 요구했다면 매우 우수한 결과를 기대할 수 있을 것이다. 그러나 이는 이미 내용을 파악한 경우에 제한적으로 적용 가능한 일이며, 내용을 파악하지 못한 모든 비디오 데이터에 적용하기엔 무리가 있을 것이다.

이에 본 연구에서는 [13]에서 계승된 특징 차이를 이용한 대표키 프레임 추출 방법에서 적용한 균등 분할 방법을 개선하여 동적 분할 방법을 제시한다.

3.2 동적 분할 방법

동적 분할 기법이란 대표키 프레임 추출 시 해당 비디오를 사용자가 요구하는 대표키 프레임 수를 기준으로 프레임들을 분할하는 것이 아니라, 비디오 데이터의 장르별 내용에 따라 분할하는 프레임 수를 동적으로 적용하는 것이다. 이는 사용자가 사전에 비디오 데이터의 장르별 분야만 구분할 수 있다면, 그 내용을 파악하지 않고도 쉽게 전체에 대한 대표키 프레임들을 추출할 수 있는 방법이다.

<표 1> 동영상 예제
<Table1> Examples of video data

Video Sequence	#frames	#shots	Min:sec	Frames per average shot
Commercials	876	19	0:30	47
Movie Clip	908	52	1:00	26
Music Video	6227	162	3:43	38
News	21673	210	14:26	107

<표 1>의 분야별 '샷(shot)당 평균 프레임 수'에서 제시한 바와 같이 비디오 영상을 이루고 있는 샷(shot)당 평균 프레임 수를 보면 분야별 특징이 있음을 알 수 있다. 즉 비디오 데이터는 분야별로 담고자 하는 내용이 다르기 때문이다. 이를 근거로 분할하는 기준을 달리 적용하는 방법을 제시한다. 30 초 분량의 상업용 광고의 경우 프레임이 876, 샷(shot)이 19일 때 평균 샷당 프레임 수는 47개 프레임이 된다. 마찬가지로 뉴스는 14분 26초 분량일 때 프레임이 21673개, 샷(shot)이 210개일 때 평균 샷(shot)당 프레임 수는 107개 프레임이다. 이와 같이 상업용 광고라는 장르별 특성상 샷(shot)당 프레임 수는 뉴스보다는 많은 점을 적용해 대표키 프레임들을 추출할 경우 동적으로 비디오를 분할하게 된다.

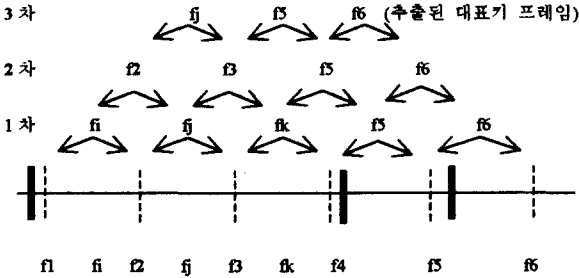
본 연구에서 제안하는 동적 분할 기법을 적용한 계승된 특징 차이를 이용한 대표키 프레임 추출 단계는 다음과 같다.

- (Step.1) 비디오 데이터의 내용의 분야를 구분한다.
- (Step.2) Step.1)을 근거로 비디오를 동적 분할 (분할 수 $m =$ 분야별 샷(shot)당 평균 프레임 수 / p)하여 m 개의 프레임들을 추출한다.
- (Step.3) 계승된 특징 차이가 가장 작은 두개의 프레임 F_i 와 F_j 를 선택한다.[13]
- (Step.4) 두 프레임 F_i, F_j 를 해당 프레임 그룹의 중간 프레

임인 $F(i+j)/2$ 로 대체한다.

(Step.5) 더 이상 특징 차이가 임계값보다 적은 것이 없으면 멈추고, 그렇지 않으면 단계 (Step.3)으로 돌아간다.

이 방법을 적용할 경우 사용자가 임의로 지정한 대표키 프레임 수가 비디오 전체의 대표 프레임 수 보다 적을 경우에, 처음부터 추출 대상에서 제외되는 대표키 프레임을 추출할 수 있으며, 또한 사용자가 미리 비디오의 전체 내용을 판단할 필요 없이 비디오 내용의 분야만 구분함으로써 모든 대표키 프레임을 추출할 수 있다.



(그림 2) 동적 분할 기법을 이용한 대표키 프레임 추출 과정
(Fig. 2) The process of Representative-Key frame Extraction Using Dynamic Segmentation Method

4. 실험 및 평가

본 연구에서는 방송용 비디오 데이터인 상업용 광고, 영화 클립, 뮤직 비디오, 뉴스를 대상으로 다음과 같이 성능을 평가해 하였다.

<표 2> 씬(scene) 안에 있는 대표키 프레임 수

<Table2> The number of Representative Key-Frame in the scene

	Commerci als		Movie Clip		Music Video		News	
	c.d	f.d	c.d	f.d	c.d	f.d	c.d	f.d
균등 분할	6	4	7	4	5	3	5	2
클러스 터링	4	2	6	3	5	2	4	3
특징 차이	4	1	6	3	4	2	5	3
동적 분할	6	1	7	2	6	2	7	3

(c.d = correct detection / f.d = false detection)

(표 2)에서와 같이 제안된 동적 분할 방법을 적용한 대표키 프레임을 찾는 방법에서는 초기에 대표키 프레임 수를 지정하지 않고도 각 비디오 데이터의 장르별 분야에 따라 원하는 대표키 프레임을 추출할 수 있어서 다른 방법보다 우수함을 알 수 있다. 즉 제안된 방법으로는 실제 눈으로 확인된 거의 모든 대표키 프레임을 찾을 수 있었으나 계승된 특징 차이를 이용한 방법에서는 균등 분할 방법만을 사용하였기 때문에 사용자가 원하는 대표키 프레임 수가 다를 경우 대표키가 아니거나 혹은 대표키 임에도 불구하고 추출되지 않은 경우가

발생한다. 또한 동적 분할 방법의 효율성을 증명하기 위하여 다양한 장르에 걸쳐 실험을 해서 그 성능을 평가해 볼 수 있었다.

5. 결론

본 연구에서는 비디오의 대표키 프레임 추출에 있어 동적 분할 방법에 의한 대표키 프레임의 추출 방법이 기존의 균등 분할 방법 보다는 매우 효율적임을 증명하였다. 그러나 본 연구에서 아쉬웠던 점은 이런 동적 분할 방법을 하기 위한 비디오 데이터의 자동 판별 부분을 추가할 수 없었던 것이다.

향후 연구할 사항으로는 비디오 데이터의 자동 판별 부분을 추가할 것이며, 분야별로 분할 기준을 적용해도 부분적으로 포함된 아주 짧은 씬(scene)에 대해서는 검출하지 못하고 지나쳐 버리는 경우가 발생하는 단점을 개선하는 일일 것이다. 이런 점이 개선된다면, 비디오 데이터 중에서 모든 대표키를 추출할 수 있으리라 기대되며, 내용 기반 비디오 검색을 위한 비디오 파싱을 좀더 효과적으로 할 수 있으리라 기대한다.

6. 참고문헌

- [1] 이미숙, 황본우, 이성환 "내용기반 영상 및 비디오 검색 기술의 연구 현황" 1997, 9 정보 과학회지 제 15 권 제 9 호 p.13
- [2] H. Zhang , A. Kankanhalli and S. W. Smoliar, "Automatic Portioning of Full-motion Video." Multi Systems, Vol. 1, No. 10-28, 1993
- [3] R.Zabih, J Miller, and K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks", Proc. of ACM Multimedia '95, San Francisco, CA, pp. 189-200, 1995
- [4] R. Kasturi, R. Jain, "Dynamic vision", In Computer vision : Principles, IEEE Computer Society Press, Washington, 1991
- [5] H. Ueda, T. Miyatake, and S. Yoshizawa, "IMPACT : An Interactive Natural-motion-picture Dedicated Multimedia Authoring System", in proceedings CHI, New York, pp.343-350, 1991
- [6] A. Nakasaka and Y.Tanaka, "Automatic video Indexing and Full video Search for Object Appearances," Proc. of 2 nd Working Conf. Visual Database Systems, pp.119-133 Oct.1991
- [7] Ramin Zabih, Justin Miller, Kevin Mai, "A feature-based algorithm for detection and classifying scene breaks", Proc. of ACM Multimedia '95, San Francisco, CA, pp. 189-200, 1995
- [8] F. Idris and S. Panchanathan, "Storage and Retrieval of Compressed Image," IEEE Transaction on Consumer Electronics, Vol. 41, pp.937-041, August 1995
- [9] D. Swanberg, C.F. Shu and R. Jain, "Knowledge Guided Parsing in Video Database," Proc. of SPIE-Storage and Retrieval for Image and Video Database, San Jose, CA, Vol 1908 pp.13-24, 1993
- [10] K.Otsuji and Y.Tonomura, "Projection-detecting filter for video cut detection", Multimedia System, 1:pp.205-210, 1994
- [11] X. Sun, M. S. Kankanhalli, Y. Zhu, and J. Wu, "Content-Based Representative Frame Extraction for Digital Video", Intl. Conf. On Multimedia Computing and System , Austin Texas, pp. 237-240, June, 1998
- [12] 김영민, 이성환, "효율적인 비디오 검색 및 브라우징을 위한 대표 프레임 군집화", 한국정보과학회 98 가을 학술 발표 논문집, 제 25 권 2 호, pp.553-555, Oct, 1998
- [13] 김우생, "계승된 특징 차이를 이용한 효율적인 대표키 프레임과 관련 키 프레임 추출 기법" 한국정보과학회 논문지(B) 제 26 권 제 5 호 pp.567-664