2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

퓨리에 변환을 이용한 키 프레임 추출

이중용 , 문영식
한양대학교 컴퓨터 공학과

Key frame extraction using Fourier transform

Jung Young Lee, Young Shik Moon
Dept. of Computer Science and Eng., Hanyang Univ.
E-mail : {leejy, ysmoon}@cse.hanyang.ac.kr

Abstract

In this paper, a key frame extraction algorithm for browsing and searching the summary of a video is proposed. Toward this end, important frames representing a shot are selected according to the correlations among frames, by using the Fourier descriptor which is useful for the shot boundary detection. To quantitatively evaluate the importance of selected frames, a new measure based on correlation coefficients of frames is proposed. If there are several frames with a same importance, another criteria Is introduced to break the tie, by computing the partial moment of subframes including each candidate key frame so that the distortion rate is minimized.

Since a key frame extraction algorithm can be evaluated subjectively, the performance of the proposed algorithm has been verified by a statistical test. Experiments show that more than 20% improvement has been obtained by the proposed algorithm compared to existing methods.

I. 서론

근래와 같은 정보화 시대에는 멀티미디어 데이터들이 하루가 다르게 다양한 형태로 쏟아져 나오고 있다. 수없이 많은 멀티미디어 데이터들은 방대한 양의 정보를 포함하고 있기 때문에 전체의 데이터를 다 조사하지 않고 주된 핵심만 설명없이 알아보기란 여간 어려운 일이 아니다. 이러한 문제는 멀티미디어 데이터의 개수가 많아지고, 데이터의 양이 커질수록 더 큰 문제로 나타날 것이다. 또한, 영상 데이터베이스 시스템과 같이 많은 양의 멀티미디어 데이터를 효율적으로 관리하고, 작업하는데 걸리는 시간을 최소화하는 일은 쉬운 일이 아니다. 따라서, 제한된 전송대역폭과 저장용량의 한계로 인해 멀티미디어 데이터는 작은 단위로 분할되어질 필요가 있으며, 동영상에 대한 분석을 통해 체계적인 색인 처리 방법을 모색하여야 한다.

이를 위해서는 동영상 데이터를 시간적 공간적으로 연속적인 프레임들의 집합인 샷 단위로 분할하고 분할된 샷에서 비디오 데이터의 요약과 브라우징, 검색 및 동영상의 유사성을 비교하는데 사용되어지는 대표 프레임을 추출하는 방법이 필요하다. 본 논문에서는 영상 데이터의 요약 및 검색을 위한 기반이 되는 키 프레임을 자동으로 추출하는 방법을 제시한다. 키 프레임이란 샷의 내용을 가장 잘 설명할 수 있는 프레임을 말하며, 마치 영화의 포스터의 같다고 생각 할 수 있을 것이다. 하지만, 현재의 컴퓨터 비젼의 기술은 의미론적인 정보를 포함하는 특징 값을 사용하여 키 프레임을 추출할 수 없기 때문에, 저수준의 특징 값들(Shape, Texture, Color)을 사용하여 키 프레임을 추출하게된다. 또한 키 프레임 추출은 주관적인 작업이기 때문에 사람마다 다를 수 있다는 점이 자동으로 키 프레임을 추출하는 것을 어렵게 하는 요인이라고 할 수가 있다. 키 프레임

추출 시 임의의 한 프레임을 키 프레임으로 선택하면, 선택된 프레임이 샷의 내용을 충분히 설명할 수 있는지에 대한 정량적인 값이 필요하며, 또한 충분히 샷의 내용을 설명하기 위해서는 몇 개의 프레임이 필요한가에 대한 문제도 키 프레임 추출시의 고려사항이라고 할 수 있다.

본 논문에서는 correlation coefficient를 이용하여 프레임의 중요도(importance)를 검출하며 동일한 중요도를 가지는 프레임에 대해서는 변형된 엔트로피(entropy) 수식을 이용하여 키 프레임을 추출한다.

II. 기존 연구

Nagasaka와 Tanka[1]는 샷 경계에 기반한 방법의 가장 단순하며 빠른 키 프레임을 추출 방법을 제시하였으나 샷의 복잡성에도 불구하고 키 프레임의 개수가 각 샷에 한 개로 한정되어져서 충분히 의미론적으로 샷의 내용을 설명하기에 부족하였으며 특히, 선택되어지는 프레임의 충분한 샷의 의미를 포함한다는 정량적인 근거가 없었다. Zhang과 Smoliar[2]는 각 샷의 첫 번째 프레임을 키 프레임으로 선택하는 대신에 여러 시각적인 특징값(visual · feature)들을 이용하여 키 프레임을 추출하는 방법을 제시하였다. 키 프레임의 개수가 한 개로 한정되어 있지 않고 여러 특징값들을 사용함으로써 영상을 표현하므로 보다 세밀한 키 프레임을 추출 방법이라고 할수 있으나 선택되어진 키 프레임이 샷 내의 기여도에 대한 정량적인 기준이 제시되어 있지 못하는 단점이 있다. Wolf[3]는 중요한 프레임의 경우에는 카메라가 정지되어 있을 것이라는 가정 하에 움직임 정보를 분석하여 키 프레임을 추출하고자 하였으며, 움직임 정보를 얻기 위해서 Horn 과 Schunck가 제안한 방법으로 optical flow를 계산하였다. Wolf의 방법은 키 프레임의 개수가 한 개로 한정되어있지 않고, 움직임 분석에 의한 방법이기 때문에 더 세밀한 키 프레임 추출 방법이라고 말 할 수 있으나 optical flow를 계산하는 과정의 소요시간이 많이 걸리며, 전처리 과정이 샷 경계를 검출하는 시간과 키 프레임을 추출하는 시간을 고려한다면 많은 계산시간이 소요된다. 또한, 선택되어진 키 프레임이 전체 샷내의 정보를 설명하는 양에 대한 정량적인 값이 존재하지 않는 단점이 있다. Gresle과 Huang[4]은 샷의 activity에 기반한 접근 방법으로 키 프레임 추출 알고리즘을 제안하였다. 또 Dufaux[5]은 움직임 activity와 공간적인 정보를 이용한 activity를 이용하여 키 프레임을 추출하는 알고리즘을 제안하였다. Dufaux의 방법은 특히 샷 움직임을 가지고 하나의 샷 경계를 검출하고 키 프레임을 추출하는 방법을 제안하고 있는데 히스토그램을 특징값으로 사용하기 때문에 조명의 영향이나 카메라 모션에 민감하다는 단점이 존재한다. Han과 Tewfik[6]은 영상의 고유벡터와 고유값을 얻어 샷 경계를 추출하고 키 프레임을 선택하는 알고리즘을 제안하였다. 하지만 한 개의 키 프레임을 샷 내에서 추출하였으며, 따라서 샷의 복잡성을 고려하지 않은 경우이기 때문에 충분한 내용을 설명한다고 할 수 없고 샷의 내용을 설명하는 정량적인 기준을 제시하지 못하였다.

III. 제안된 방법

본 논문에서는 흔히 동영상 데이터가 가지고 있는 본질적인 특성, 즉, 강조되고자 하는 내용들은 카메라가 많이 주시를 할 것이라는 가정을 가지고 접근한다. 일반적으로 샷내의 프레임들은 유사성이 높은 프레임들로 구성되어져 있을 것이며, 샷의 복잡성이 높다고 한다면 한 개의 프레임만으로는 샷의 의미론적인 내용을 포함하기에는 부족함이 있을 것이다. 이에 본 논문에서는 퓨리에 변환의 스펙트럼 계수를 이용하여 키 프레임을 추출하고자 하며, 기존 연구들에서 보여지지 않았던 키 프레임이 샷에서 차지하는 중요도를 정량적인 값으로 제시함으로써 키 프레임 선택에 타당성을 강조하였다.

우선 샷의 모든 프레임들에 대한 스펙트럼의 유사성을 구한다. 유사성은 (식 1)과 같은 일반적인 correlation 수식을 사용한다.

$$Corr(x,y) = \frac{Cov(x,y)}{SD(x) \times SD(y)} \qquad \text{(식 1)}$$

여기서 Cov(x, y)는 공분산, SD는 표준편차를 의미한다.

(식 1)을 이용하여 각 프레임과 관련성이 높은 프레임의 개수를 계산한다. 다음, 각 프레임과 유사성이 높은 프레임의 개수를 이용하여 (식 2)와 같이 샷에 대한 중요도를 계산한다.

$$Gain(i) = 1 - |P_i \log_2 P_i|$$

$$P_i = \frac{\text{i번째 프레임과 유사성이 높은 프레임의 개수}}{\text{전체 샷내에서의 프레임들의 개수}}$$

(식 2)

위의 수식을 이용하면 샷을 이루는 모든 프레임들에 대한 Gain을 얻을 수가 있다. Gain이 높으면 높을수록 선택되어진 프레임과 상관도가 높은 프레임의 개수가 많은 것이므로, 논문의 가정인 "강조되어지는 내용은 그 내용을 담고 있는 프레임들이 많이 나올 것이다" 라는 가정에 일치하는 것이"된다.

만약에 키 프레임 한 개를 선택할 경우에 동일한 Gain을 가지는 프레임이 존재한다면 우리는 샷의 모양을 최대한 덜 왜곡시키는 프레임을 골라야 한다. 본 논문에서는 모멘트 방정식에 의해서 이를 측정한다. 아래의 수식은 일반적인 모멘트 방정식이다.

(p+q) 차 모멘트 식

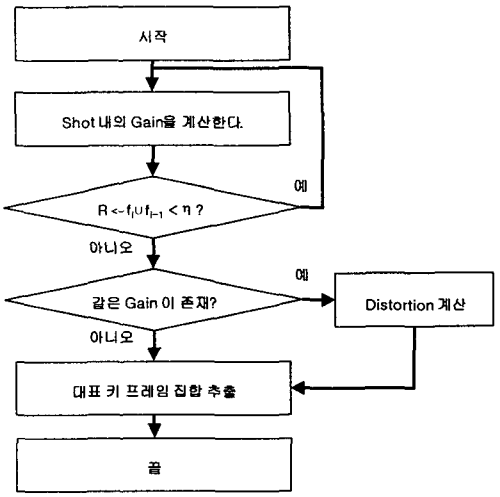
$$m_{pq} = \sum\sum x^p \, y^q f(x, y) \qquad \text{(식 3)}$$

본 논문에서는 중앙모멘트(Central Moment)를 사용하였으며 모멘트 수식에 의해서 최소 왜곡율을 나타내는 프레임을 키 프레임으로 선택한다.

동등한 Gain을 가지는 프레임의 경우 각각의 프레임과 연관되어진 프레임들의 집합 모멘트를 비교하여 왜곡율이 작은 것을 선택하게 된다. 선택 기준이 되는 수식은 (식 4)와 같다.

$$\arg\min\left\{\frac{Moment_{partial}}{Moment_{shot}}\right\} \qquad \text{(식 4)}$$

$Moment_{shot}$은 샷 전체의 모멘트, $Moment_{partial}$은 각 프레임과 연관된 프레임 집합에 대한 모멘트이다. 이상에서 우리는 기존 연구방법에서 정량적으로 제시하지 못한 선택되어진 키 프레임의 샷 기여도 및 중요도를 측정할 수가 있었으며 만약에 선택되어진 키 프레임의 Gain이 특정 임계치 이하일 때는 키 프레임의 개수를 늘려서 추출할 수 있는 근거가 될 수가 있다. 제안된 키 프레임 추출 방법의 개략적인 순서도는 (그림 1)과 같다.



(그림 1) 제안된 알고리즘의 순서도

IV. 실험결과

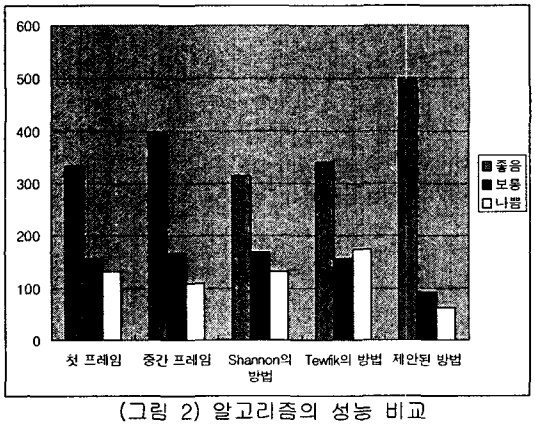
본 논문에서 제안된 알고리즘을 다른 알고리즘과 비교하기 위해서 실험에 사용된 데이터는 <표 1>과 같은 네 가지 종류의 서로 다른 카테고리의 영상이다. 실험은 matlab 5.0으로 PentiumII - 600에서 수행하였다.

<표 1> 실험에 사용된 비디오 데이터

데이터의 종류	샷의 개수	특징
드라마	60	대화형 형식이 많음
영화	39	프레임 개수가 많음
광고	41	조명이 다양함
비디오	36	장면의 복잡성이 큼

첫 번째 프레임을 선택하는 방법, 중간 프레임을 선택하는 방법, Shannon의 방법, 그리고 Tewfik이 제안한

방법과 비교하였으며, 각 키 프레임의 적절성이 주관적인 성향이 강하기 때문에 Web에 출판한 후 좋음, 보통, 나쁨의 세 가지 범주를 두어 실험 참여자들에게 익명으로 기입 받는 조사방법을 채택하였다. (그림 2)는 카테고리 전체에 대한 평균 성능을 나타내고, <표 2>는 카테고리별로 적절한 키 프레임이라고 선택되어진 자료의 알고리즘별 성능 표이다.



(그림 2) 알고리즘의 성능 비교

<표 2> 성능평가 도표 (좋음)

	광고	뮤직비디오	드라마	영화	평균
첫 프레임	51.8%	45.1%	57.5%	36.5%	47.7%
중간 프레임	60.3%	61.1%	63.3%	47.1%	57.9%
Shannon의 방법	56.7%	52.7%	60.4%	47.1%	54.2%
Tewfik의 방법	51.2%	44.4%	60.4%	38.2%	48.5%
제안된 방법	80.4%	71.5%	76.2%	66.6%	73.6%

광고, 드라마, 뮤직 비디오 순으로 성능이 좋음을 실험으로 확인할 수 있으며, 기존의 방법에 비해 15~20% 정도 성능이 향상됨을 알 수 있다.

V. 결론

본 논문에서는 샷 내의 프레임간의 유사성에 근거하여 모멘트와 퓨리에 변환을 이용하여 키 프레임을 추출하는 알고리즘을 제시하였다. 기존의 알고리즘들은 선택된 키 프레임의 샷 내 중요도를 설명하는 정량적인 척도가 없었으며, 또한 선택된 키 프레임을 평가하는 방법이 적절하지 못하였다.

본 논문에서는 샷 내의 프레임의 중요도를 변형된 엔트로피의 식을 이용하여 정량적인 방법으로 수치화하고 선택되어진 프레임의 평가를 실험 참여자에 의한 검증 방법으로 키 프레임의 성능 평가 절차를 행하였다. 주관적인 측면이 포함되어져 있기 때문에 실험에 대한 평가를 통계학적으로 정규분포를 따를 만큼 설문 조사를 수행 하였다. 실험을 통하여 각 카테고리별 성능향상은 기존의 방법에 비해 평균적으로 약 15~20%의 성능향상이 있음을 실험을 통해 확인할 수 있었다.

참고문헌

[1] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," Visual Database Systems II, 1992

[2] H. Zhang, J.Wu, D. Zhong, and S.W.Smoliar, "An integrated system for content-based video retrival and browsing," Pattern Recognition, vol.30, no. 4, pp. 643-658, 1997

[3] W. Wolf, "Key frame selection by motion analysis," Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc., 1996

[4] P. O. Gresle and T. S. Huang, "Gisting of video documents : a key frames selection algorithm using relative acitivity measure." Proc. Conf. on Visual Information Systems, 1997

[5] Dufaux, "Key frame selection to represent video," Proc. International Conference Image Processing, pp 275-278, 2000

[6] J. Han and A. H. Tewfik, "Eigen image based video segmentation and Indexing," Proc. International Conference Image Processing, 1997