변형에 강인한 내용기반 동영상 검색방법

Modification-robust contents based motion picture searching method

최갑근, Gab-Keun Choi 김순협 Soon-Hyob Kim

서울특별시 노원구 월계동 447-1 광운대학교 컴퓨터공학과
Wolgye-dong 447-1 Nowon-gu, Seoul Korea
Kwangwoon Univ. Computer Engineering Dept.

요약 동영상 내용검색을 위해서 가장 많이 사용되고 있는 기술은 컷 추출에 의한 내용비교 방법이다. 그러나 컷 추출을 위해 사용되는 CHD(Color Histogram Difference)나 ECR(Edge Change Ratio)등은 영상물의 Cropping, Resizing, Low bit rate등의 변화에 대해 대단히 취약하다. 본 방법은 이러한 변형에 강인하도록 상대적으로 변형이 적은 오디오정보를 이용하여 Indexing과 Searching을 수행하였다. 특히 변형에 강인한 Searching을 위해 오디오의 장면(Scene)을 검출하였고 장면을 중심으로 Time-frequency domain에서 각각의 Frequency bin.에 대한 스펙트럴 파워를 파워임계값을 중심으로 이진화(Binary)하였다. 제안된 방법으로 Cropping, clipping, Lowbit rate, Additive Frame 등의 변형본에 대한 검색을 시도한 결과 False positive Error 와 True Negative Error에 대해 각각 1 %미만의 오탐지 결과를 얻었다.

Abstract The most widely used method for searching contents of motion picture compares contents by extracted cuts. The cut extraction methods, such as CHD(Color Histogram Difference) or ECR(Edge Change Ratio), are very weak at modifications such as cropping, resizing and low bit rate. The suggested method uses audio contents for indexing and searching to make search be robust against these modification. Scenes of audio contents are extracted for modification-robust search. And based on these scenes, make spectral powers binary on each frequency bin. in the time-frequency domain. The suggested method shows failure rate less than 1% on the false positive error and the true negative error to the modified(using cropping, clipping, row bit rate, addtive frame) contents.

핵심어: 내용기반 동영상 검색, 오디오 장면검출, 오디오인덱싱, 컷디텍션.

1. 서론

내용기반 동영상 검색을 위해서는 데이터양이 많은 동영상의 특성을 고려하여 각각의 프레임을 비교하지 않고 동영상에 대한 요약본을 만든 후 비교한다. 요약본은 전체 프레임을 샷 바운더리 기준으로 세그멘테이션 한다. 이때 샷 바운더리의 기준이 되는 컷 프레임의 검출을 위해 CHD, ECR, Standard Deviation of Pixel Intensities[1]등을 이용하여 컷 프레임을 검출한다. 하지만 영상영역에서의 이 방법들은 Cropping, Lowbit rate, Resizing등의 변형이 일어나게 되면 원본과 동일한 위치의 컷 프레임을 얻기 어렵다. 본 논문에서는 이와 같은 영상영역에서의 문제점을 해결하기 위해 동영상의 오디오 영역에서 특징을 추출하고 Indexing과 Searching 속도를 개선하기 위해 각각의 특징 벡터에 대해 바이너리화 하였다.

2. 동영상 내용 검색 시스템을 위한 Video Segmentation 방법

동영상 요약본을 위한 샷 바운더리를 위한 컷 프레임 검출 알고리듬은 동영상의 모든 프레임에 대해 수행된다. 따라서 컷 프레임을 검출하는 성능과 검출을 위한 연산속도 모두 중요하지만 일반적으로 특징의 변화 표현이 상대적으로 유리한 CHD(Color Histogram Difference)를 많이 사용한다.

2.1 CHD(Color Histogram Difference)

칼라 히스토그램에 기반한 샷 경계 검출 알고리듬은 샷과 샷이 바뀌는 경계영역에 있는 프레임들에서 급격하게 칼라가 변하는 특성을 이용한 것이다. [1]

$$CHD_i = \frac{1}{N}\sum_{r=0}^{2^B-1}\sum_{g=0}^{2^B-1}\sum_{b=0}^{2^B-1}|p_i(r,g,b) - p_{i-1}(r,g,b)| \qquad (2.1)$$

2.2 ECR(Edge Change Ratio)

에지 변화율(Edge Change Ratio)은 다음과 같이 정의된다. 수식 2.2에서 σ_n은 n 프레임에서 에지 픽셀의 개수이며, X_n^{in}과 X_{n-1}^{in}은 프레임 n과 n-1의 입력과 기존 에지 픽셀의 수가된다. 그러면 주어진 에지변화율 ECR_n은 n-1과 n 프레임사이가 되며 이 것의 범위는 0과 1사이가 된다. 에지는 Canny Edge Detector[2]를 사용하여 계산되어 진다.

$$ECR_n = \max(X_n^{\in}/\sigma_n, X_n^{\in}/\sigma_{n-1}) \qquad (2.2)$$

3. Audio Spectral Image를 이용한 내용기반 동영상 검색

3.1 오디오 인덱싱을 위한 전처리

3.1.1 Short-time Energy

동영상에서 음성 신호의 진폭은 시간에 따라 변화한다. 무성음 segments의 진폭은 일반적으로 유성음 segments의 진폭에 비해 훨씬 작다. 음성 신호의 단구간(Short-time) 에너지는 이러한 진폭의 변화를 쉽게 나타낸다. 따라서 각 구간 마다 구해진 에너지의 크기를 이용하여 특징 데이터를 만들 수 있다. 본 논문에서는 각 구간의 에너지를 4비트로 인코딩하여 구간 특징데이타로 사용하였으며 구간 에너지는 수식 3.1과 같이 얻을 수 있다.

$$E_n = \sum_{m=-\infty}^{\infty}|x(m)|w(n-m) \qquad (3.1)$$

3.1.2 Audio Cut

오디오 컷은 동영상 내용 검색시 유사도 비교를 위한 탐색공간을 줄이고 클리핑(Cliffing)된 오디오 클립의 비교를 위해 사용한다. 오디오 컷의 추출은 각각의 FED(Frame Energy Difference), 프레임 에너지 변화량을 임계값을 기준으로 구분하며 n프레임과 n-1프레임에서의 에너지 변화량의 차이를 기준으로 추출하며 식 3.2와 같다.

$$FED_i = |E_{n-1} - E_n| \qquad (3.2)$$

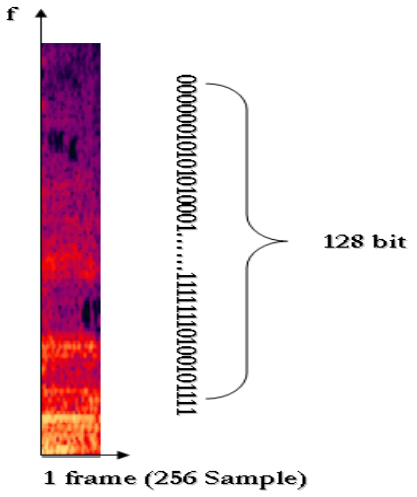
3.2 오디오 특징 벡터 추출

동영상은 비디오와 오디오 정보를 모두 갖고 있다. 본 논문에서는 여러 유형의 영상영역에서의 변형에 강인하도록 상대적으로 변형에 강한 오디오영역에서 특징 데이타를 추출하여 사용하였다. 특징데이타는 입력오디오 트랙에 대해 Sampling Frequency = 11.025KHz, 양자화는 16bit로 정규화한다. 정규화된 입력오디오를 256 sample 구간씩 FFT하며 128개의 각각의 frequency bin. 에 대한 파워값을 특징 벡터로 정의한다.

3.3 Binary Audio Fingerprint

3.3.1 Binary Audio Fingerprint threshold

바이너리 오디오 핑거프린트는 Indexing 성능과 Searching 성능향상을 모두 고려하기 위해 사용하며 각각의 프레임에서 구해진 Frequency bin. 파워 값에 대해 이진연산 수행만으로도 고속연산이 가능하도록 임계값을 기준으로 파워값을 이진화 시킨다. 임계값은 각 프레임의 평균에너지가 되며 평균에너지 보다 크면 1로 평균에너지 보다 작으면 0으로 인코딩되며 인코딩 과정은 그림과 같다.



[그림 1 Binary Audio Fingerprint]

3.4 BAF(Binary Audio Fingerprint)를 이용한 유사도 비교

동영상 탐색을 위해 입력 쿼리와 원본데이타의 유사도 비교를 위해 Reference DB의 Record에 대해 각각의 BER값을 구한다. 구해진 각각의 BER값은 0에 가까울수록 원본과의 유사도가 높은 것으로 판정하며 가장 적은 Error 비율을 출력하는 Record를 동일 영상으로 판정한다. 고속연산을 위해 비트 레벨에서 식(3.3)과 같이 XOR연산을 수행한 후 BER(Bit Error Rate)을 구한다.

$$Error\,bit = (Query\,Frame)\ X\,O\ R(Source\,Frame)$$
$$BER = \frac{Total\,Error\,Bit}{Total\,Bit} \qquad (3.3)$$

4. 실험 및 결론

본 논문에서는 실험을 위해 약 1시간 분량의 드라마 데이터 1000편을 DB로 구축하고 Clipping, Resizing, Cropping등으로 변형된 Query를 수행해 제안된 방법을 검증하였고 실험 결과는 [표1]과 같다. 제안된 BAF(Binary Audio Fingerprint)는 Searching 성능은 우수하나 트랜스코딩에 의한 변형, Lowpass filtering, Resampling등과 같은 변형에서 다소 미흡한 것으로 확인됐다.

실험조건 :

Database

DB Size : 1000 hour

Processing time : 228sec/min

4 Genre (Drama, Movie, Talk show, News)

Format : wave files mono 8 KHz 16 bits

Parameterization

Window = 256 samples

BAF= 128 bit

	Clipping	Add Frame	UCC	D-Data
Total Record	173	11	58	168
True Positive	153	11	41	165
False Negative	18	0	14	3
False Positive	0	0	0	0
True Negative	2	0	3	0

[표 1 변형본에 대한 탐색결과]

참고문헌

참고문헌

[1] J. S. Boreczky and L. A. Rowe. Comparison of Video Shot Boundary Detection Techniques. In Storage and Retrieval for Still Image and Video Databases IV, Proc. SPIE 2664, pp. 170-179, Jan. 1996.

[2] J. Canny. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp. 34-43, Nov. 1986.