

DCT 기반 압축 영상에서 자막 영역을 검출하기 위한 효율적인 방법

장현성, 강경옥
한국전자통신연구원 방송미디어연구부

An Efficient Scheme for Detecting Caption Regions in DCT-Compressed Images

Hyun Sung Chang and Kyeongok Kang
Broadcasting Media Research Department, Electronics and Telecommunications Research Institute
{chs,kokang}@etri.re.kr

요약

동영상 장면에서 자막은 흔히 중요한 의미 정보를 나타내기 때문에 영상으로부터 자막 영역을 검출하는 것은 동영상에 대한 의미적인 분석 및 색인 등 다양한 범위에서 널리 응용될 수 있다. 본 논문에서는 DCT 기반으로 압축된 영상에서 자막 후보 영역을 검출하기 위한 고속의 방법을 제안한다. 제안하는 방법은 자막 영역에서 나타나는 수평 밝기값의 교대 패턴을 검출하기 위하여 각 블록 별로 1회의 1-D IDCT 과정을 필요로 한다.

1. 서론

이산여현변환(DCT)은 영상 내 공간적인 상관도를 제거하는데 있어서 비교적 우수한 성능을 보이기 때문에 JPEG, MPEG-1, 2 및 H.26x 등의 영상 압축 표준에 채택되어 널리 사용되고 있다. 가장 전형적인 경우를 예로 들면 정지영상 혹은 동영상 내의 각 프레임을 8×8 화소 블록의 2차원(2-D) 배열로 간주하고 각 블록 별로 2-D DCT를 적용한 후, DCT 계수에 대한 비균일 양자화 및 가변 길이 부호화 과정을 통하여 압축 효율을 더욱 높게 된다. 이러한 표준이 널리 사용됨에 따라 인터넷을 비롯하여 여러 디지털 매체를 통하여 산재하는 영상의 상당수는 이러한 방식으로 압축 부호화된 상태로 존재하고 있다.

한편, 영상으로부터 자막 영역을 검출하는 작업은 동영상 내의 대표 프레임 선정이나, 의미적인 관점에서 비디오 요약 및 동영상 검색을 위한 키워드 색인 등 다양한 범위에 널리 이용될 수 있다[2],[5],[6]. 이를 위하여 화소 영역에서 자막 영역을 검출하기 위한 방법은 오래 전부터 활발히 연구되어 왔다. 그러나, 이미 상당수의 영상이 DCT 기반으로 압축되어 있음을 고려하면 화소 영역에서의 접근 방법은 압축된 영상을 복호화하는 과정을 필요로 하기 때문에 시간이 많이 소요된다. 최근에는 압축 영역에서 자막 영역을 검출하기 위한 몇몇

방법이 제안되기도 하였으나, 대개의 경우는 발견적인(heuristic) 방법에 의존하며 화소 영역에서의 예측가능한 동작으로 해석하기 어렵다[3],[7].

본 논문에서는 화소 영역에서의 해석이 용이하면서 DCT 압축 영역으로부터의 복원을 최소로 한 고속의 알고리즘을 제안하고, 몇 가지 방송 영상에 적용한 실험 결과를 제시한다.

2. 자막 영역의 수평 밝기값 교대 패턴

자막 영역에서 나타나는 뚜렷한 특징 중의 하나는 수평 밝기값의 변화가 심하게 나타난다는 것이다. 이러한 수평 밝기값의 변화를 화소 영역에서 관찰하면 그림 1과 같이 밝기값이 자막 영역에서 교대하는 패턴으로 나타난다.

그림 1은 자막 영역을 포함하고 있는 $240(\text{높이}) \times 352(\text{너비})$ 크기의 동영상 프레임과 선 ℓ 을 따라 측정된 각 화소의 밝기값에 대한 그래프이다. 그림에서 실제 자막 영역은 위로부터 196 ~ 213인 화소에 위치하고 있다. 그림 1(b)에서 보듯이 이 구간에서 밝기값은 명암이 크게 교대하는 패턴을 형성하고 있음을 볼 수 있다.

[7]에서 Zhong *et al.*은 이러한 특징을 DCT 계수로부터 검출하는 방식을 제안하고 있다. [7]에서 제안된 방법에 의하면 영상의 각 블록 별로 자막 영역의 블록인지 그렇지 않은 블록인지 구분하기 위하여 수직 공간 주파수 AC 성분의 DCT 계수 크기가 큰 블록을 자막 후보 영역으로 선정한다. 이후에 자막 후보 영역에 대한 수평 방향의 균집화 과정 등 후처리를 통하여 최종 자막 영역을 확정하는 방식을 취하고 있다. 이러한 방법에 따르면 그림 1(b)에서 나타나는 교대 패턴을 각 블록 단위로 분할할 때 수직 공간 주파수의 AC 성분으로 모호화하고 있는 것으로 볼 수 있다. 그러나, 화소 영역에서의 교대 패턴이 곧바로 주파수의 AC 성분 구간으로 모호화되는 데에 대한 근거가 부족하고 실험결과 자막 영역 이외의 다른 블록에서도 수직 공간 주파수의 AC 성분이 비교적 자주 나타나 오검출되는 경우

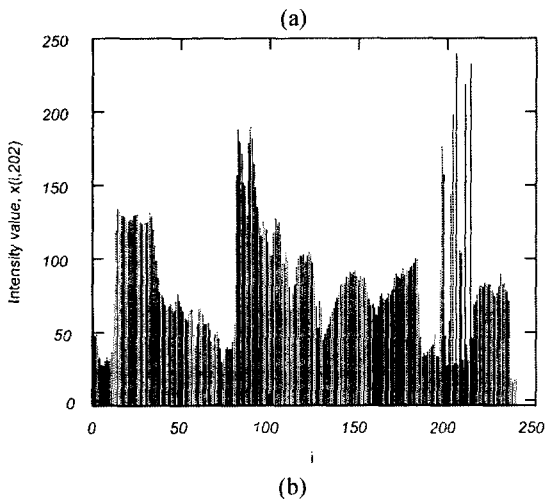
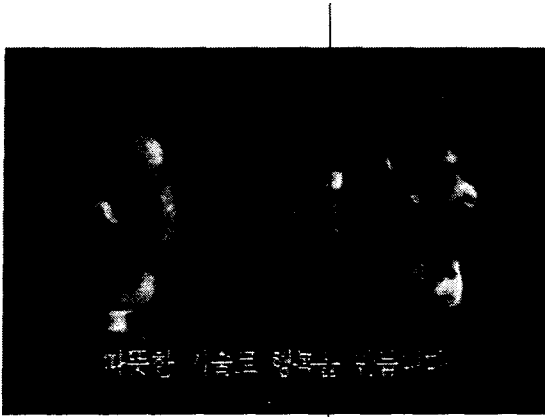


그림 1. (a) 자막 영역을 포함하고 있는 동영상 프레임, (b) 선 l 을 따라 측정된 화소의 밝기값 그래프 (가로축의 i 는 선 l 을 따라 위에서 아래의 방향으로 매겨진 것임.)

가 많음을 확인할 수 있었다.

3. 제안하는 방법

본 논문에서는 화소 영역에서의 해석이 용이하면서 DCT 압축 영역으로부터의 복원을 최소로 한 고속의 알고리즘을 제안한다. 먼저, 2절에서 기술한 수평 밝기값 교대 패턴을 이용하여 화소 영역에서의 간단한 자막 영역 검출 방법을 제안하고, 제안하는 방법이 DCT 계수에 대하여 매우 적은 연산을 통하여 수행될 수 있음을 보이기로 한다. 제안하는 방법은 블록의 크기에 상관없이 적용될 수 있으나, 본 논문에서는 편의상 블록의 크기를 8×8 화소로 한정하여 설명한다.

우선, 각 행과 열의 표기를 위하여 그림 2에 도시한 표기법을 따른다. 크기가 $h \times w$ 인 전체 영상에서 (k, l) 의 위치에 있는 화소가 (i, j) 블록 내에 포함되며, 블록 내에서의 좌표는 (k', l') 임을 가정한다. (i, j) 번째 블록 내 각 화소의 밝기값은

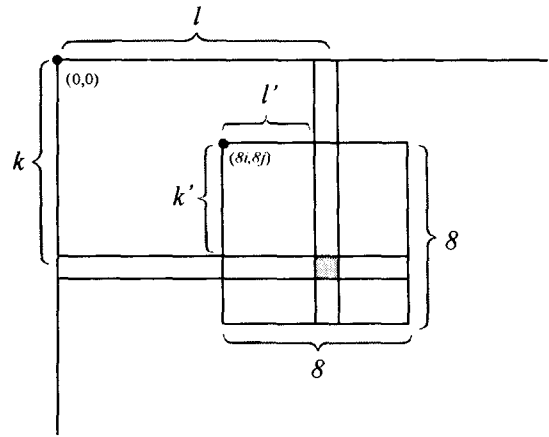


그림 2. 전체 영상 및 해당 블록에서 특정 화소의 위치표기에 대한 정의

$$x_{i,j}(k', l'), \quad k', l' = 0, 1, \dots, 7$$

로 표기하기로 한다.

그림 1(b)와 같이 영상을 화소 단위로 복원하고, 각 열별로 이러한 교대 패턴을 검출하도록 할 수 있으나, 이는 복원의 복잡도 뿐만 아니라 자막을 이루는 문자의 수평적인 연결성이 낮은 경우 검출하지 못하는 등의 문제점을 내포한다. 따라서, 아래와 같이 각 블록의 수평 밝기값의 평균을 그 화소의 밝기값 대신 사용하여 이러한 문제점을 해결하고자 한다.

$0 \leq j < w/8$ 인 각 j 에 대하여, 아래의 단계 1, 2를 반복하여 자막 후보 영역을 찾는다.

단계 1: $0 \leq i < h/8$ 인 각 i 에 대하여,

$$I_{i,j}(k') = \frac{1}{8} \sum_{l=0}^7 x_{i,j}(k', l'), \quad k' = 0, 1, \dots, 7 \quad (1)$$

를 구하고, $k = 8i + k'$ 인 k 에 대하여 $I_j(k) = I_{i,j}(k')$ 로 정의한다. ($0 \leq k < h$)

단계 2: 밝기값의 변화 상태를 각 k 에 대한 $\Delta I_j(k) = I_j(k+1) - I_j(k)$ 값에 따라 다음과 같이 'RISING', 'FALLING', 'FLAT'으로 구분한다.

$\Delta I_j(k) > \delta_{TH}$ 인 경우 'RISING'; $\Delta I_j(k) < -\delta_{TH}$ 인 경우 'FALLING'; 이외의 경우 'FLAT'.

H_c 화소 범위 내에 상태의 교대가 발생한 경우 이들을 동일한 자막 후보 영역으로 간주하고 그 영역 내에서 발생한 상태의 교대가 N_{thr} 개 이상인 경우 이를 자막 후보 영역으로 확정한다.

제안하는 방법의 제 1단계에서 사용한 식 (1)은 블록 전체에 대한 IDCT 과정을 거치지 않고 훨씬 효율적으로 계산이 가능하다. 식 (1)은 블록의 각 행에 대한 DC 값이기 때문에 잘 알려진 바대로 2-D DCT가 1-D DCT를 각 행과 열로 분리한 형태로 적용된다는 사실을 감안하면 블록 첫번째 열에 대한 DCT 계수를 1-D IDCT

하면 그 값을 얻을 수 있다. 이를 수식적으로 표현하면 다음과 같다.

C_{16}^n 와 $e(n)$ 을 각각

$$C_{16}^n = \cos \frac{n\pi}{16}$$

와

$$e(n) = \begin{cases} 1/\sqrt{2}, & \text{if } n=0 \\ 1, & \text{elsewhere} \end{cases}$$

로 정의하면, 8×8 블록에 대한 2-D DCT는

$$X_{i,j}(m',n') = \frac{e(m')e(n')}{16} \sum_{k'=0}^7 \sum_{l'=0}^7 x_{i,j}(k',l') C_{16}^{(2k'+1)m'} C_{16}^{(2l'+1)n'}$$

$$m', n' = 0, 1, \dots, 7 \quad (2)$$

와 같이 정의될 수 있으며, 그 역은

$$x_{i,j}(k',l') = \sum_{m'=0}^7 \sum_{n'=0}^7 e(m')e(n') X_{i,j}(m',n') C_{16}^{(2k'+1)m'} C_{16}^{(2l'+1)n'}$$

$$k', l' = 0, 1, \dots, 7 \quad (3)$$

로 주어진다.

식 (3)을 식 (1)에 대입하여 정리하면,

$$I_{i,j}(k') = \frac{1}{\sqrt{2}} \sum_{l'=0}^7 e(m') X_{i,j}(m',0) C_{16}^{(2k'+1)m'} \quad (4)$$

와 같고, 이는 1-D IDCT 결과에 $1/\sqrt{2}$ 를 곱한 값이 된다. 식 (4)의 IDCT 과정은 고속의 알고리즘을 이용하면 11개의 곱셈과 29개의 덧셈을 통하여 계산될 수 있으며 [4], 앞단의 DCT 계수 역양자화 과정과 연계하여 구현될 경우 곱셈의 수를 5개로 줄일 수 있다[1].

4. 실험결과

모의 실험을 위하여 뉴스와 광고 비디오로부터 I 프레임 추출 후 [7]의 방법과 제안한 방법을 적용하여 자막 후보 영역을 검출하였다. 실험에 사용한 [7]의 방법은 수평 밝기 값의 변화 패턴을 이용한 첫 번째 단계까지만을 사용하였으며, 수평 방향의 자막 후보 영역 군집화 및 기타 후처리 과정은 생략하였다. 제안한 방법을 적용함에 있어서의 파라미터 값은 $\delta_{TH} = 20$, $H_c = 6$, $N_{alt} = 4$ 로 하였다.

그림 3, 4에서 보는 바와 같이, 두 가지 방법이 자막 후보 영역을 검출하는 정밀도 면에서는 크게 차이가 없으나, 오검출율에 있어서는 제안하는 방법이 상대적으로 개선된 성능을 보이고 있음을 확인할 수 있다. 본 절에서 비교된 자막 후보 영역 검출 결과는 비록 후처리를 하기 전 단계에서 얻어진 결과이기는 하지만, 자막 후보 영역의 오검출이 적기 때문에 보다 정밀한 후처리 과정에 대한 시간을 단축하거나 보완하는 성질이 있다.

5. 결론

본 논문에서는 블록 단위 DCT 기반으로 압축된 영상에서 자막 후보 영역을 검출하기 위한 고속 알고리즘을 제안하였다. 제안한 방법은 자막 영역에서 나타나는 수평 밝기 값의 교대 패턴을 검출하기 위하여 최소한의 DCT 계수 복원 과정을 거침으로써 산술 연산의 수를 감소시킬 수 있었으며, 기존의 방법에 비하여 비슷한 정밀도에서 오검출율이 상대적으로 낮은 효과를 얻을 수 있었다.

감사의 글

본 논문은 정보통신부에서 지원하는 지능형 통합정보방송(SmarTV) 기술개발 과제 일부로서 수행된 연구 결과입니다.

참고문헌

- [1] Y. Arai, T. Agui, and M. Nakajima, "A fast DCT-SQ scheme for images," *Trans. IEICE*, vol. E71, no. 11, pp. 1095—1097, Nov. 1988.
- [2] H. S. Chang, J.-G. Kim, K. Kang, and J. Kim, "Context-based detection and highlights in a video sequence," in *Proc. Picture Coding Symp.*, Apr. 2001, pp. 295—298.
- [3] H. Kato, Y. Nakajima, and M. Sugano, "A fast caption detection from MPEG coded data," in *Proc. Picture Coding Symp.*, Apr. 2001, pp. 215—218.
- [4] C. Loeffler, A. Ligtenberg, and G. Moschytz, "Practical fast 1-D DCT algorithm with 11 multiplications," in *Proc. IEEE ICASSP*, May 1989, pp. 988—991.
- [5] T. Sato, T. Kanade, E. K. Hughes, M. A. Smith, and S. Satoh, "Video OCR: indexing digital news libraries by recognition of superimposed captions," *Multimedia Syst.*, vol. 7, no. 5, pp. 385—395, Sept. 1999.
- [6] D. Zhang, R. K. Rajendran, and S.-F. Chang, "General and domain-specific techniques for detecting and recognizing superimposed text in video," in *Proc. IEEE ICIP*, Sept. 2002, pp. 593—596.
- [7] Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 4, pp. 385—392, Apr. 2000.

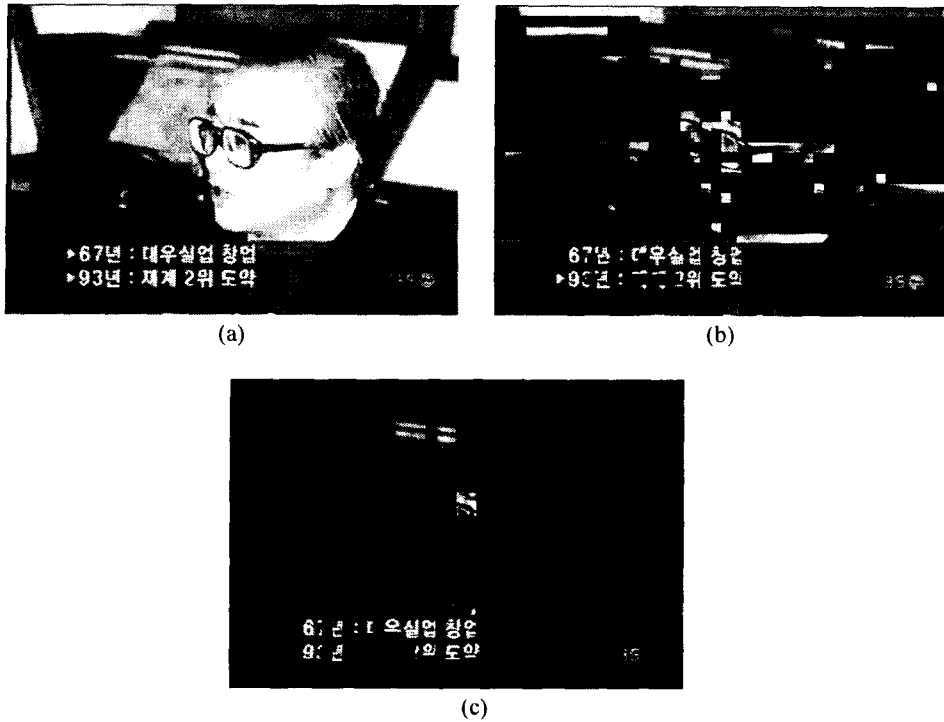


그림 3. (a) 뉴스 동영상 프레임, (b) [7]의 방법에 의한 자막 후보 영역 검출 결과, (c) 제안한 방법에 의한 자막 후보 영역 검출 결과

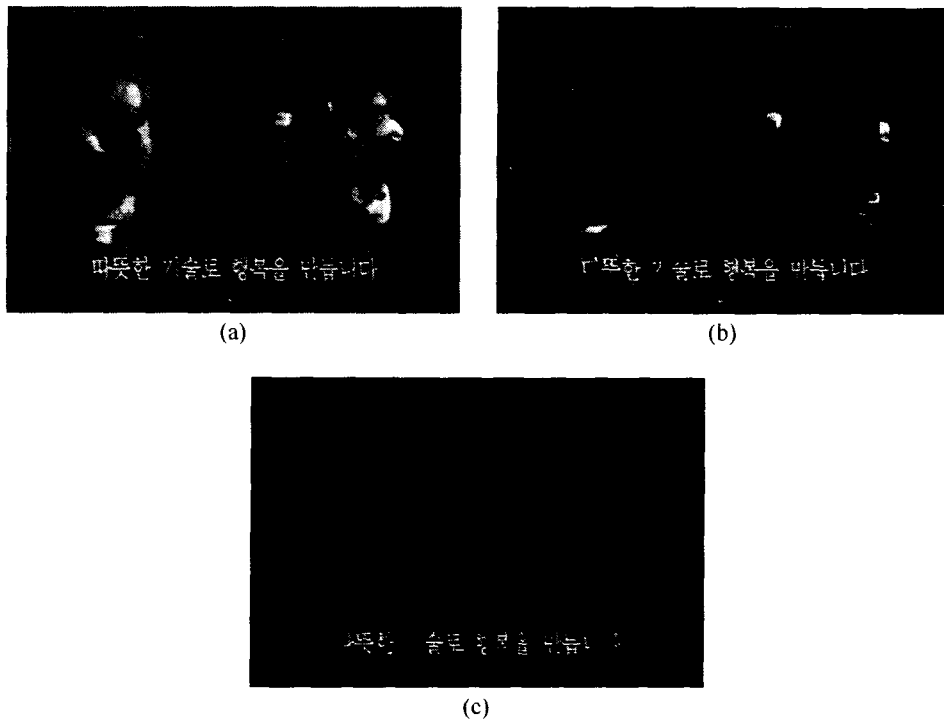


그림 4. (a) 광고 동영상 프레임, (b) [7]의 방법에 의한 자막 후보 영역 검출 결과, (c) 제안한 방법에 의한 자막 후보 영역 검출 결과