

시공간 활성도를 이용한 적응적 양자화 기반 비디오율 제어

이시웅* · 김차종*

*한밭대학교 정보통신컴퓨터공학부

Video Rate Control Using An Adaptive Quantization Based on Spatio-temporal Activity Measure

Si-Woong Lee* · Cha-Jong Kim*

*Hanbat National University

E-mail : swlee69@hanbat.ac.kr

요 약

본 논문에서는 시각 시스템의 민감도를 고려한 새로운 비디오 율제어 기법을 제안한다. 제안된 방식은 MPEG2 TM5 율제어 기법과 같이 3단계의 구조로 구성되어 있으나, 적응적 양자화를 위한 시각의 시-공간 민감도에 근거한 새로운 매크로블록 활성도 척도를 사용한다. 모의 실험을 통해 TM5 방식에서 사용된 공간 활성도보다, 제안된 시-공간 결합 활성도가 주관적, 객관적 화질 측면에서 우수함을 증명한다.

ABSTRACT

A new algorithm for the rate control of videos considering the sensitivity of the human visual system is presented. The method adopts the three-step structure of MPEG2 TM5 rate control, and a new measure of macroblock activity based on spatio-temporal sensitivity is introduced for the adaptive quantization. Experimental results show that the spatio-temporal activity measure outperforms the spatial activity of TM5 in picture quality.

키워드

비디오 율제어, TM5, 적응적 양자화

1. 서 론

MPEG-2[1][2]와 같은 표준 비디오 부호화 방식들[5]-[7]은 MC-DCT(motion compensation discrete cosine transform) 방식에 의해서 공간 그리고 시간 방향 상관성을 제거하고 VLC(variable length coding)에 의해서 통계적 상관성을 제거한다. 따라서 이 표준 부호화 방식들은 필연적으로 가변 비트율(variable bit rate: VBR)의 비트열을 발생시킨다. 가변 비트율을 갖는 비디오 비트열을 대역이 제한된 채널을 통하여 전송하는 응용에 있어서는 비트열을 평활화하는 비디오 버퍼와 이 버퍼의 넘침(overflow)이나 고갈(underflow)을 방지하기 위한 적합한 비트율 제어 방식이 제공되어야 한다.

비록 비트율 제어 그 자체는 비디오 부호화 방식에 관한 표준으로는 정의되어 있지 않으나, 비트율 제어의 성능은 비디오 부호화의 성능에 큰 영향

을 미친다. 실제적으로 우수한 비트율 제어는 주어진 대역폭 하에서 복호화된 영상의 화질을 향상시킬 수 있다. 따라서 비트율 제어의 성능은 비디오 부호화의 성능을 좌우한다. 비트율 제어 방식을 설계하는데 있어서 고려해야 할 사항은 첫째, 필수적으로 부호기와 복호기 버퍼의 넘침이나 고갈을 방지해야 한다. 만일 이 조건을 만족하지 못하면 비디오 정보의 손실이 발생하고, 결국 복호화된 영상이 크게 훼손된다. 둘째, 주어진 대역폭 하에서 보다 높은 화질뿐만 아니라 일관된 화질을 제공해야 한다. 셋째, 실시간 비디오 서비스에 적용할 경우 비트율 제어로 인한 지연과 복잡도가 낮아야 한다.

율제어 방법은 R-Q (Rate-Quantizer) 모델의 사용 여부에 따라 크게 두 가지로 분류될 수 있다. 모델을 사용하지 않는 대표적인 방법으로는 MPEG-2 Test Model Ver. 5 (이하 TM5)의 율제어 방법이 있고, 모델을 사용하는 대표적인 방법으로

는 H.263 TMN Ver. 10 (이하 TMN10)과 MPEG-4 Verification Model Ver. 8 (이하VM8)의 율제어 방법이 있다.

TM5의 율제어 방법은 수 Mbps 이상의 높은 비트율에 맞도록 설계되었으며, GOP (Group of Picture) 구조의 MPEG-1과 MPEG-2에 적합하다. TM5의 비트율 제어는 비트 할당(bit allocation), 비트율 제어(rate control), 적응 양자화(adaptive quantization)의 세단계로 구성된다. 이중 적응 양자화 단계에서는 인간의 시각 특성을 고려하여 인간의 시각으로 잘 감지되지 않는 복잡한 MB에 대해서는 QP 값을 증가시키고 상대적으로 잘 감지되는 단순한 MB에 대해서는 QP 값을 감소시킨다. 공간 복잡도에 따른 시각의 감지도에 따라서 양자화 스텝의 크기를 조절함으로써 주관적 화질을 개선시키는 것이다. 그러나 시각 특성은 공간 복잡도에 의존적이기도 하지만 움직임 정도에 따라서도 의존적이다. 즉, 빨리 변하는 영역에서는 시각이 둔감하므로 적응 양자화 과정에서 시각의 시간적 특성도 동시에 고려하는 것이 효과적이다. 따라서 본 논문에서는 적응 양자화 과정에서 시각의 시-공간 민감도를 동시에 고려할 수 있는 새로운 마크로블록 활성도 척도를 제안하고, 이를 이용한 비디오 율-제어 알고리즘을 제시하기로 한다.

II. 제안 율제어 알고리즘

제안된 방식은 기본적으로 TM5의 구조를 이용하여 3단계의 구조로 구성하였다. 그러나, 기존 TM5의 단점을 개선하기 위해 첫번째 단계인 비트 할당 단계와 세 번째 단계인 적응적 양자화 단계를 새로 수정하였다. 우선 비트 할당 단계에서는 버퍼 순응 양자화를 위한 버퍼 제한을 도입하였으며, 적응적 양자화 단계에서는 시각의 공간적 특성 뿐 아니라 시간적 특성도 함께 고려하기 위한 시공간 결합 활성도를 이용하는 적응적 양자화를 도입하였다. 전체적인 비트율 제어 과정은 다음과 같다.

2.1 비디오 버퍼 제한 조건을 이용한 목표 비트량 설정

우선, 비트 할당은 TM5와 같이 픽처 레벨로 수행되며, 각각의 I,P,B 프레임에 대한 목표 비트량은 다음식에 의해 결정된다.

$$T_i = \max \left\{ \frac{R}{1 + \frac{N_p X_p}{K_p X_i} + \frac{N_b X_b}{K_b X_i}}, \frac{bit_rate}{8 \times picture_rate} \right\},$$

$$T_p = \max \left\{ \frac{R}{N_p + \frac{N_b K_p X_b}{K_b X_p}}, \frac{bit_rate}{8 \times picture_rate} \right\}, \quad (1)$$

$$T_b = \max \left\{ \frac{R}{N_b + \frac{N_p K_b X_p}{K_p X_b}}, \frac{bit_rate}{8 \times picture_rate} \right\},$$

여기서 K_p 와 K_b 는 I, P, B 프레임의 복잡도와 관련된 상수로서 제안 방식에서는 각각 $K_p=1.0$ 와 $K_b=1.4$ 를 사용하였다. N_p 와 N_b 는 현재의 GOP 내에 남아있는 P 픽처와 B 픽처의 수를 나타내며, X_i , X_p 및 X_b 는 각각 I, P 와 B 타입의 픽처에 대한 복잡도 상수로서 $X_i=(160 \times bit_rate)/115$, $X_p=(60 \times bit_rate)/115$, 및 $X_b=(42 \times bit_rate)/115$ 로 초기화된다. 그리고, R 은 현재 GOP에 대한 잔여 비트수이며, 각 픽처의 부호화 후에 다음과 같이 수정된다.

$$R = R - S_{i,p,b} \quad (2)$$

여기서 $S_{i,p,b}$ 는 바로 직전 픽처의 부호화 과정에서 발생된 비트량이다. GOP의 첫번째 프레임의 부호화 시에는 이전 GOP의 잔여 또는 초과 비트량 R_{prev} 를 고려하여 초기 R 값은 다음처럼 설정된다.

$$R = G + R_{prev}, \quad (3)$$

여기서 $G = bit_rate \times N / picture_rate$ 이며, N 은 GOP 내의 픽처수이다. 첫번째 GOP에서는 $R_{prev} = 0$ 이다.

식(1)에 의해 계산된 현재 픽처의 목표 비트량은 실제 버퍼의 overflow나 underflow를 방지하기 위한 비디오 버퍼 제한 조건을 만족하도록 수정된다. 프레임 주기 당 채널을 통해 전송되는 비트량이 항상 TCBR로 일정한 CBR 전송인 경우, 복호기 버퍼의 넘침과 고갈은 부호기 버퍼의 넘침과 고갈을 방지함으로써 막을 수 있다. 즉 부호기 버퍼 크기 E_{max} 와 복호기 버퍼 크기 D_{max} 를 TCBRL로 설정하면, 부호기 버퍼의 넘침과 고갈이 발생하지 않으면 복호기 버퍼의 넘침과 고갈도 발생하지 않는다. 따라서 CBR 전송의 경우 비트율 제어는 부호기 버퍼의 넘침과 고갈을 방지하면 된다.

부호화된 비트열들이 부호기 버퍼에 저장된 후 프레임 주기 당 전송 비트량이 TCBR인 CBR 채널을 통하여 전송될 경우 부호기 버퍼의 넘침이나 고갈이 발생하지 않을 조건은

$$0 \leq E(i) \leq E^{max}, \quad for \quad \forall i, \quad (4)$$

$$E(i) = E(i-1) + B(i) - T_{CBR},$$

이다. 여기서 E^{max} 는 부호기 버퍼의 크기이며, $E(i)$ 는 i 번째 프레임의 부호화가 수행된 직후의 부호기 버퍼의 총만도이다. 그리고 $B(i)$ 는 i 번째 프레임에서 발생된 비트량이며 T_{CBR} 은 한 프레임 기간동안 전송되는 비트량이다. 위식으로부터 버퍼의 넘침이

나 고갈이 발생하지 않을 $B(i)$ 의 범위는

$$B(i) \in [\min_B^e(i), \max_B^e(i)] \\ = [\max(T_{CBR} - E(i-1), 0), E^{\max} + T_{CBR} - E(i-1)] \quad (5)$$

이 된다. 여기서 $E(i)$ 는 $i \Delta t$ 인 순간의 부호기 버퍼 충만도이다. 따라서 i 번째 프레임에 할당되는 목표 비트량 $BT(i)$ 의 허용 범위는 다음과 같이 설정한다.

$$B^T(i) \in [\min_B^T(i), \max_B^T(i)] \\ = [\min_B^e(i) - B_0^e, \max_B^e(i) - B_0^e]. \quad (6)$$

여기서 B_U^e 와 B_O^e 는 각각 부호기 버퍼 상태가 '0'과 E_{\max} 에 지나치게 근접하는 것을 막기 위한 값으로 본 논문에서는 다음과 같이 설정하였다.

$$B_U^e = B_O^e = 0.05 \times E^{\max} \quad (7)$$

결국 현재 픽처에 대한 목표 비트량은 다음과 같이 계산된다.

$$B^T(i) = \begin{cases} T_i, T_p, T_b, & \text{if } \min_B^T(i) \leq T_i, T_p, T_b \leq \max_B^T(i) \\ \min_B^T(i), & \text{if } \min_B^T(i) \geq T_i, T_p, T_b \\ \max_B^T(i), & \text{if } T_i, T_p, T_b \geq \max_B^T(i) \end{cases} \quad (8)$$

2.2 비트율 제어

이 단계에서는 I, P, 그리고 B-픽처에 대한 가상 버퍼를 설정하고, 각 가상 버퍼의 충만도(fullness)에 따라 양자화 파라미터 QP를 조절함으로써 비트 할당 단계에서 할당된 비트량이 가까운 실제 비트량을 발생시킨다.

j 번째 MB를 부호화할 때 I, P, B-픽처의 가상 버퍼의 충만도 d_j^I, d_j^P, d_j^B 는 다음과 같이 ($j-1$)번째 MB까지 실제로 발생한 비트량의 누적치 B_{j-1} 와 각 MB에 대하여 균등 배분한 목표 비트량 (T_I, T_P, T_B)의 누적치와의 차로 나타낸다.

$$d_j^I = d_0^I + B_{j-1} - \frac{T_I \times (j-1)}{MB_cnt} \\ d_j^P = d_0^P + B_{j-1} - \frac{T_P \times (j-1)}{MB_cnt} \\ d_j^B = d_0^B + B_{j-1} - \frac{T_B \times (j-1)}{MB_cnt} \quad (9)$$

여기서 d_0^I, d_0^P, d_0^B 는 각각 부호화 영상에 대한 가상 버퍼의 초기값이고, MB_cnt 는 한 영상 내에 있는

MB의 수이다.

가상 버퍼의 충만도를 이용하여 기준 양자화 파라미터 Q_j 를 다음과 같이 구한다.

$$Q_j = \frac{d_j \times 31}{r} \quad (10)$$

여기서 r 은 반응 매개 변수로 $2bit_rate/picture_rate$ 의 크기인데, 이 값이 작을수록 가상 버퍼의 충만도에 따른 QP의 변화가 민감하게 반응한다. 결국 가상 버퍼의 충만도가 클수록 QP 값을 크게 하여 발생되는 비트량을 줄이고, 충만도가 적으면 QP 값을 작게 하여 발생되는 비트량을 증가시킨다.

2.3 시공간 결합 활성도를 이용한 적응적 양자화

이 단계에서는 인간의 시각 특성을 고려하여 인간의 시각으로 잘 감지되지 않는 MB에 대해서는 QP 값을 증가시키고 상대적으로 잘 감지되는 MB에 대해서는 QP 값을 감소시킴으로써 주관적 재생화질이 향상되도록 양자화 파라미터를 조절하는 과정이다. TM5에서는 시각의 감지도를 MB의 공간 활성도에 근거하여 추정한다. 그러나, 제안된 방식에서는 시각 감지도의 공간 특성 뿐 아니라 시간 특성도 함께 고려할 수 있도록 시공간 결합 활성도를 이용하는 방안을 새로 제안한다. 이는 시각 감지도가 공간적으로 복잡한 영역일수록 저하되기도 하지만, 움직임이 큰 영역에 대해서도 역시 저하된다는 사실에 입각한 것이다.

적용 양자화를 위하여 식(10)으로부터 얻어진 Q_j 에 다음식과 같이 정규화된 활성도(activity)를 곱하여 최종적으로 현재 MB에 적용할 QP 값을 구한다.

$$mquant_j = Q_j \cdot N_act_j \quad (11)$$

정규화된 활성도 N_act_j 는 다음식과 같이 계산된다.

$$N_act_j = \alpha \cdot N_act_j^s + (1-\alpha) \cdot N_act_j^t \quad (12)$$

윗 식은 TM5와 달리 제안된 방식에서는 공간 활성도 $N_act_j^s$ 와 시간 활성도 $N_act_j^t$ 의 가중 합으로서 현재 MB의 활성도가 계산됨을 나타낸다. 그리고 $N_act_j^s$ 와 $N_act_j^t$ 의 계산식은 다음과 같다.

$$N_act_j^s = \frac{2 \times act_j^s + avg_act^s}{act_j^s + 2 \times avg_act^s} \\ N_act_j^t = \frac{2 \times act_j^t + avg_act^t}{act_j^t + 2 \times avg_act^t} \quad (13)$$

여기서 avg_act^s 및 avg_act^t 는 각각 바로

직전에 부호화된 픽처의 평균 시, 공간 활성도이며, 이들의 초기값으로는 각각 150과 6을 사용하였다. j 번째 MB의 공간 활성도 act_j^s 및 시간 활성도 act_j^t 는 각각 다음식에 따라서 계산된다.

$$act_j^s = 1 + \underset{sbk=1,8}{Min}(var_sbk)$$

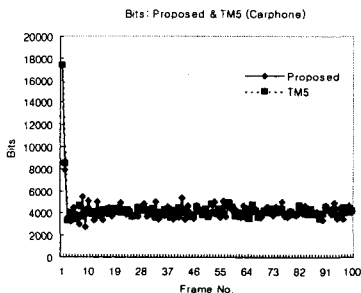
$$act_j^t = 1 + |Mvp(j)| \quad (14)$$

var_sbk 는 각 공간방향 8X8블록의 화소값에 대한 분산이며, $Mvp(j)$ 는 j 번째 MB에 대한 움직임 벡터 예측치이다. 시간 활성도 계산식에서 현재 MB의 움직임 벡터대신 움직임 벡터의 예측치를 사용하는 이유는 비트율 제어 과정에서는 현재 MB의 움직임 벡터가 계산되지 않기 때문이다. 그러나 인접된 MB들 간에 움직임의 유사성이 존재할 경우 움직임 벡터의 예측치가 현재 MB의 움직임 정도를 잘 대변한다고 볼 수 있다.

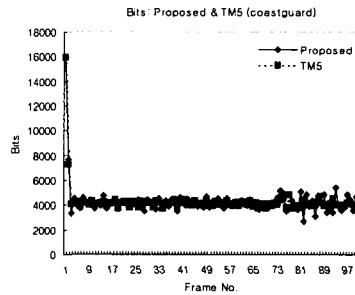
식(12)의 α 는 정규화된 활성도에 대한 공간 활성도의 반영률을 나타낸다. 본 논문의 실험에서는 intra picture인 경우에는 $\alpha = 1.0$ 을, inter picture인 경우에는 $\alpha = 0.5$ 를 사용하였다. $\alpha = 0.5$ 이면 식(4.20)에 따라 시간 활성도와 공간 활성도의 평균값이 현재 MB의 활성도가 되며, 따라서 공간방향으로 복잡한 MB이거나 또는 움직임이 큰 MB일수록 상대적으로 큰 QP가 할당된다. 이는 비교적 시각적으로 둔감한 MB에 적은 비트가 할당되게 함으로써 적응적 양자화를 통해 주관적 화질이 개선됨을 의미한다.

III. 실험결과

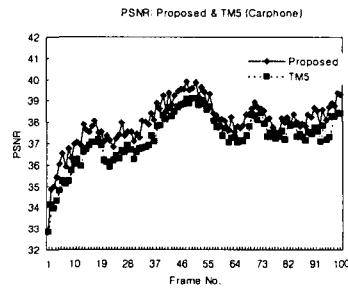
제안된 비트율 제어의 성능을 평가하기 위하여 QCIF(176X144) 해상도를 갖는 4종류의 표준 비디오("carphone", "coastguard", "container", "foreman")를 모의실험에 이용하고 TM5 방식과 성능을 비교하였다. 본 실험에서는 30frames/sec, 4:2:0 format을 이용하였으며 128kbps의 채널 전송율로 부호화 하였으며, 부호화 방식으로는 H.264의 JM을 이용하였다.



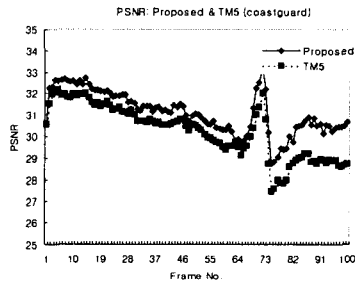
(a) "carphone" sequence



(b) "Coastguard" sequence
그림 1. 발생 비트량



(a) "carphone" sequence



(b) "Coastguard" sequence
그림 2. PSNR

그림 1, 그림 2 및 표1은 시공간 결합 활성도를 이용한 적응 양자화에 대한 결과이다. 이 실험에서는 입력 영상열의 첫번째 프레임만 I-픽처로 부호화하고 그 후의 프레임들은 P-픽처로 부호화 하여 실험하였다. 우선, 그림 1은 각 시퀀스의 100장의 프레임에서 발생한 비트량을 보여준다. 발생 비트량은 TM5 방식이나 제안된 기법이 큰 차이가 나타나지는 않으나, 제안된 기법이 평균 움직임의 크기가 다른 인접 프레임들 간에 있어서 발생 비트량의 변화폭이 TM5에 비해 다소 큼을 알 수 있다. 즉, 이전 프레임에 비해 비교적 큰 움직임을 갖는 프레임에 대해서는 적은 비트량이 할당 되고 시각적으로 민감한 작은 움직임을

갖는 프레임에 대해서는 많은 비트량이 할당됨으로써 움직임 정도에 따른 주관적 화질이 개선될 수 있음을 보여준다. 그림2는 복호화 된 영상의 PSNR을 보여준다. 모든 실험 영상에 있어서 제안된 방식의 PSNR이 TM5 방식보다 우수함을 알 수 있다. 이는 큰 움직임의 MB에서 일반적으로 큰 움직임 보상 오차가 발생하게 되고, 따라서 이런 블록에서의 QP값의 증가에 따른 변환 계수 비트량의 감소가 주관적 화질의 열화 없이 보다 시각적으로 민감한 다수의 MB의 변환계수 부호화에 사용될 수 있기 때문이다. 표 1은 전체 100장의 프레임에 대한 평균 부호화 성능을 보여준다. 제안 방식이나 TM5나 모두 목표 비트율에 상당히 정확한 비트율 제어를 하고 있음을 알 수 있다. 그러나 평균 PSNR은 제안방식이 TM5에 비해 평균 0.7~1.1dB가량 우수함을 알 수 있다. 이 결과로부터 적응적 양자화 시에 시, 공간 결합 활성도를 사용하는 것이 주관적 화질 뿐 아니라 객관적 화질 측면에서도 우수한 방식을 알 수 있다.

[5] ITU-T, Study Group 15, "Draft of recommendation H.263: Video coding for low bitrate communication," 2 May 1996

[6] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustic, Speech, and Signal Processing*, Vol. 36, No. 9, pp. 1445-1453, Sept. 1988

[7] S.-W. Wu and A. Gersho, "Rate constrained optimal block-adaptive coding for digital tape recoding of HDTV," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 1, No. 1, pp. 100-112, March 1991

표 1. 제안방식과 TM5의 평균성능비교 (128kbps)

InputSeq.	TM5			Proposed		
	PSNR	Rate	QP	PSNR	Rate	QP
Carphone	37.26	127.9	13.85	37.99	128.0	14.00
Coastguard	30.23	127.9	20.02	31.11	128.1	19.52
Container	39.15	128.0	8.48	40.19	128.1	8.71
Foreman	35.18	128.0	16.39	35.90	127.9	16.16

IV. 결 론

본 논문에서는 TM5에 기반한 새로운 비트율 제어 알고리즘에 대해 기술하였다. 버퍼 순응 율제어를 위해 비디오 버퍼 제한 조건을 도입하였으며, 효과적인 적응 양자화를 위한 시, 공간 결합 활성도를 이용한 새로운 비트율 제어 방안을 제안하였다. 모의 실험을 통하여 제안 방식이 기존의 TM5에 비해 주관적, 객관적으로 우수한 성능을 가짐을 확인하였다

참고문헌

[1] Siwei ma, Wen Gao, and Yao Lu, "Rate control on JVT standard," JVT-D030, July, 20

[2] ISO-IEC/JTC1/SC29/WG11, "Recommendation H.262 : Generic coding of moving pictures and associated audio information: Video," ISO-IEC 13818-2, Nov. 1994

[3] CCITT, Working Party XV/1, "Draft of recommendation H.261: Video codec for audiovisual services at $P \times 64$ kbit/s," July 1990

[4] ISO-IEC/JTC1/SC29/WG11, "Coding of moving pictures and associated audio," CD 11172-2, Nov. 1991