

H.264를 위한 적응적인 비트-왜곡 최적화 방법

오관정, 호요성
광주과학기술원 정보통신공학과

An Adaptive Rate-Distortion Optimization Method for H.264 Video Codec

Kwan-Jung Oh, Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
E-mail : {kjoh81, hoyo}@gist.ac.kr

Abstract

Several video coding standards, such as MPEG-4 and H.263, have been investigated to reduce the resulting number of bits while pursuing the maximum video quality. The recent video coding standard, H.264, provides higher coding efficiency than previous coding standards by using the mode decision scheme. For mode decision, H.264 chooses the best macroblock mode among the several candidates using Lagrangian cost function which reflects both the rate and the distortion. H.264 employs only one rate-distortion optimization (RDO) model for all macroblocks. Since the characteristics of each macroblock is different, each macroblock should have its own RDO model. In this paper, we propose an adaptive rate-distortion optimization algorithm for H.264. We regulate the Lagrangian multiplier considering the picture type and characteristics of each macroblock.

I. 서론

최신 비디오 압축 표준인 H.264는 다양한 부호화 기술을 이용하여 기존의 MPEG-4나 H.263에 비해 높은 부호화 효율을 보인다. 이중 대표적인 기술이 가변블록 움직임 보상 기술인데 하나의 매크로블록에 대해 다양

한 매크로블록 모드 중에 비트-왜곡 비용이 가장 작은 모드를 선택하여 해당 매크로블록을 부호화한다.

최적의 매크로블록 모드 선택에 이용되는 비트-왜곡 비용 함수는 Lagrangian 최적화 기술을 이용하는데, 현재 H.264는 모든 화면과 매크로블록에 대해 정형화된 하나의 모델만을 사용한다. 이는 각각의 매크로블록에 대해서는 비트-왜곡 관점에서 최적화되어 있지만, 영상 전체에 대해서는 비트-왜곡 관점에서 손실이 많다. 화면의 타입이나 매크로블록의 특성에 따라 비트와 왜곡에 대한 중요도가 다르기 때문이다. 이러한 손실을 줄이고자 영상의 특성을 고려한 적응적인 비트-왜곡 최적화 방법에 관한 연구가 필요하다 [1].

Tasi가 인간시각 시스템 (human visual system)에 기반한 적응적인 비트-왜곡 최적화 방법에 관한 연구를 했으나, 이 방법의 경우 인간시각 특성을 고려하여 눈에는 좋게 보이나 실제 비트-왜곡 관점에서는 기존의 방법에 비해 오히려 좋지 않은 성능을 보였다 [2].

본 논문에서는 화면 타입과 매크로블록의 특성을 고려한 적응적인 비트-왜곡 최적화 방법을 제안한다. 제안하는 방법은 화면의 타입과 매크로블록에 따라 비트와 왜곡의 중요도를 고려하여 비트-왜곡 최적화 모델의 Lagrangian 계수 값을 조정한다.

본 연구는 광주과학기술원(GIST) 실감방송연구센터(RBRC)를 통한 정보통신부대학 IT 연구센터(ITRC)와 교육인적자원부 두뇌한국 21(BK21) 정보기술사업의 지원으로 이루어졌음.

II. H.264 비트-왜곡 최적화 방법

H.264는 움직임 예측 및 움직임 보상 과정에서 다양한 매크로블록 모드를 이용한다. 이를 통해 기존의 압축 표준에 비해 높은 압축 효율을 얻을 수 있다. 매크로블록 모드는 SKIP, 16×16, 16×8, 8×16, P8×8과 같은 5개의 인터모드와 I16×16과 I4×4과 같은 2개의 인트라모드가 있다. 인터모드는 이전에 부호화된 화면들을 참조하여 움직임 예측 기법을 이용하고, 인트라모드는 이전에 부호화된 인접 화소들을 이용하여 방향성에 기반한 예측을 한다. 그림 1은 H.264의 다양한 매크로블록 모드를 보여주고 있다 [3].

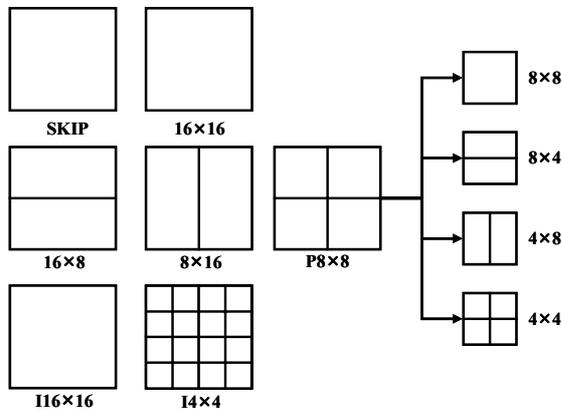


그림 1 다양한 매크로블록 모드

이러한 다양한 매크로블록 모드 중에서 현재 매크로블록에 적합한 모드를 결정하기 위해 Lagrangian 기술에 기반한 비트-왜곡 최적화 방법이 이용된다. 식 (1)은 H.264에 이용되는 비트-왜곡 비용 함수를 보여 주고 있다.

$$J_{MODE}(s,r,MODE | \lambda_{MODE}) = SSD(s,r,MODE) + \lambda_{MODE} \cdot R(s,r,MODE) \quad (1)$$

여기서 s 와 r 는 각각 현재 블록과 참조 블록의 화소 값을 의미하고, $MODE$ 는 매크로블록의 모드를 의미한다. $SSD(s,r,MODE)$ 는 왜곡을 나타내는 값으로 현재 가변블록에 대한 원영상과 복호된 영상의 화소 값의 차이에 대한 제곱의 합으로 식 (2)와 같다. λ_{MODE} 는 Lagrangian 계수 값으로 다수의 실험을 통해 정해졌으며 식 (3)과 같다. $R(s,r,MODE)$ 는 현재 블록을 부호화 하는데 소요되는 비트로 실제 부호화 과정을 통해 부호화 시 필요한 비트를 구한다.

$$SSD(s,r,MODE) = \sum_{x \in H, y \in V} (s(x,y) - r(x - m_x, y - m_y))^2 \quad (2)$$

여기서 H 와 V 는 각각 가변 블록의 넓이와 높이이고, m_x 와 m_y 는 해당 가변 블록에 대한 움직임 벡터이다.

$$\lambda_{MODE} = 0.85 \cdot 2^{(Q-12)/3} \quad (3)$$

여기서 Q 는 양자화 계수이다.

이처럼 H.264는 비트-왜곡 함수를 이용하여 현재 매크로블록에 적합한 최적의 모드를 선택하여 매크로블록을 부호화 한다. H.264의 비트-왜곡 최적화 방법은 전체 영상의 모든 매크로블록에 대해 동일하게 적용된다. 그러나 영상의 화면 타입이나 매크로블록의 특성에 따라 비트-왜곡 최적화 방법은 적응적으로 변해야 한다.

III. 적응적인 비트-왜곡 최적화 방법

본 절에서는 화면 타입이나 매크로블록 특성을 고려한 적응적인 비트-왜곡 최적화 방법을 제안한다. 화면 타입은 I화면과 P화면에 대해 화면의 중요도에 따라 I화면의 경우 P화면의 참조 화면으로 이용되기 때문에 비트 보다는 왜곡에 가중치를 주고, P화면의 경우 I화면의 비트 증가율을 보상하기 위해 왜곡보다는 비트가 적도록 비트에 가중치를 준다. 각 매크로블록에 대해서는 16×16, 16×8, 8×16의 세가지 모드의 왜곡의 분산을 이용하여 매크로블록의 특성을 파악한 후 그 특성을 고려하여 비트-왜곡 모델을 적응적으로 개선한다. 적응적인 비트-왜곡 최적화 방법을 위해 식 (3)의 Lagrangian 계수 값을 화면의 타입과 매크로블록 특성에 따라 변경한다.

3.1 화면 타입에 따른 Lagrangian 계수 값

비디오 압축의 대부분의 효율은 움직임 예측 및 움직임 보상에서 기인한다. 때문에 GOP (group of picture) 구조 또한 압축 효율에 큰 영향을 미친다. GOP는 하나의 I화면과 여러 개의 P화면 혹은 B화면으로 구성된다. 여기서 I화면은 다른 P화면이나 B화면의 참조 영상이 되는 화면으로 발생 비트는 많은 반면에 좋은 화질로 부호화 될 경우 해당 GOP의 화질 향상에 큰 영향을 미친다. 때문에 비트-왜곡 최적화 관점에서 I화면의 경우 비트보다는 왜곡에 가중치를 두어 부호화 하는 것이 전체 부호화 효율의 측면에서 득이 된다. P화면이나 B화

면의 경우 I화면의 비트를 보상하기 위해 비트에 보다 가중치를 두어서 부호화 한다. 표 1은 여러 QCIF 영상에 대해 양자화 계수 (QP)가 28인 경우에 대해 최고의 부호화 효율 향상을 갖는 Lagrangian 계수 값과 그때의 부호화 효율을 보여주고 있다.

표 1 제안하는 Lagrangian 계수 값과 부호화 효율

실험영상	λ_{MODE} (I화면, P화면)	$\Delta PSNR$ (dB)	Δ 비트율 (%)
AKIYO	0.3, 1.0	0.148	-0.541
CARPHONE	0.3, 0.9	0.052	-0.119
CONTAINER	0.3, 1.1	0.084	-2.147
FOREMAN	0.4, 0.9	0.030	-0.960
MOBILE	0.2, 0.9	0.026	-0.915
MOTHER & DAUGHTER	0.2, 1.0	0.027	-0.821
NEWS	0.3, 1.0	0.230	-0.203
SALESMAN	0.3, 1.1	0.220	-0.268
평균	0.2875, 0.975	0.102	-0.747

여러 QP에 대한 위와 같은 실험을 통해 식 (4)와 같이 QP를 고려한 I화면과 P화면에 적합한 Lagrangian 계수 값을 얻어냈다.

$$\begin{aligned}
 &I\text{화면의 Lagrangian 계수 값} \\
 &\lambda_{MODE} = \frac{QP}{\alpha} \cdot 0.85 \cdot 2^{(QP-12)/3} \quad (4) \\
 &P\text{화면의 Lagrangian 계수 값} \\
 &\lambda_{MODE} = \left(1.2 - \frac{QP}{\beta}\right) \cdot 0.85 \cdot 2^{(QP-12)/3}
 \end{aligned}$$

여기서 α 와 β 는 실험적으로 얻어진 값이며, 각각 97과 132이다. 위 식을 통해 I화면에는 0.85보다 작은 값이 P화면에는 0.85보다 큰 값이 할당된다. QP에 따른 변화는 QP가 작을수록 영상이 왜곡에 민감하기 때문에 I화면에 대해 Lagrangian 계수 값이 작게 하였고, QP가 클수록 왜곡에 대한 가중치를 줄였다.

3.2 매크로블록 특성에 따른 Lagrangian 계수 값

이전 절에서 화면 타입에 따라 왜곡과 비트의 중요도를 고려해 Lagrangian 계수 값을 조정하였다. 이번 절에서는 한 화면 내의 매크로블록들에 대해 각 매크로블록의 특성을 고려해 Lagrangian 계수 값을 조절한다. 각 매크로블록의 특성을 나타내는 값으로 16×16 , 16×8 , 8×16 세 모드의 왜곡을 이용하였다. 세 모드를 이용하는 이유는 위의 세 모드의 경우 모드 결정 이전에 움직임 예측을 통해 해당 모드에 대한 왜곡을 미리 알 수 있기 때문이다. 만약 매크로블록이 배경과 같이 단조로

운 영역이면 각 모드의 대한 왜곡 값이 비슷하고, 복잡한 경우는 세분화 된 블록의 왜곡 값이 적다. 이를 이용하여 각 왜곡 값에 대해 식 (5)의 로그 스케일의 왜곡 표준 편차 값 (LSD)을 매크로블록의 특성을 나타내는 값으로 이용한다.

$$LSD = \log \left(\sqrt{\frac{1}{3} \sum_{MODE=0}^2 (M_{distortion} - D_{MODE})^2} \right) \quad (5)$$

여기서 $M_{distortion}$ 은 평균 왜곡 값이며 D_{mode} 는 각 모드에 대한 왜곡 값이다.

각 매크로블록의 LSD를 이용하여 모드 결정 이전 단계에서 식 (6)과 같은 새로운 λ_{NEW} 를 정의한다.

$$\lambda_{NEW} = \chi \cdot \lambda_{MODE} \quad (6)$$

여기서 χ 는 다음과 같이 정의된다.

$$\begin{aligned}
 &if (LSD_{cur} < LSD_{mean}) \\
 &\chi = 1 - \frac{|LSD_{mean} - LSD_{cur}|}{20\delta} \quad (7) \\
 &else \\
 &\chi = 1 + \frac{|LSD_{mean} - LSD_{cur}|}{20\delta}
 \end{aligned}$$

식 (7)에서와 같이 χ 는 LSD를 고려한 가중 값으로 복잡한 매크로블록의 경우 1보다 작은 값을 가지며, 단순한 매크로블록의 경우 1보다 큰 값을 갖는다. LSD_{mean} 과 LSD_{cur} 은 각각 이전 매크로블록까지의 평균 LSD값과 현재의 LSD값을 의미한다. 현재의 LSD가 이전 LSD의 평균값보다 작다면 이 매크로블록은 단순한 영역이라 가정하고 비트에 대한 가중치를 더 주기 위해 χ 값을 1보다 조금 작은 값이 되도록 조정한다. χ 를 1에 근접한 값으로 정의하는 이유는 인접 매크로블록들의 λ_{MODE} 값이 크게 변하여 다음 화면의 움직임 예측의 효율이 떨어지는 것을 막기 위한 것이다. δ 는 각 영상의 특성을 반영하는 값으로 식 (8)과 같이 정의된다.

$$\delta = \frac{I_{bitrate}}{V \times H} + \sqrt{\frac{255^2}{10^{PSNR}}} \quad (8)$$

여기서 $I_{bitrate}$ 는 GOP의 I화면에 대한 발생 비트이고, V 와 H 는 영상의 넓이와 높이이다. $PSNR$ 은 I화면의 $PSNR$ 값으로 결국 δ 는 한 화소에 대한 비트와 그때의 왜곡의 함으로 표현되는 값이다. 영상이 복잡할수록 발생 비트도 많고 왜곡도 커지므로 δ 는 증가하게 되고, 단순한 영상일수록 δ 는 작아지게 된다.

IV. 실험 결과 및 분석

본 논문에서 제안한 적응적인 비트-왜곡 최적화 방법을 검증하기 위해 JM 9.5 [4]를 이용하여 QCIF 영상(AKIYO, FOREMAN, MOBILE, NEWS, SALESMAN) 5개의 초기 120 프레임에 대해 실험하였다. IPPP 구조의 GOP를 사용하였고, 인트라 간격은 30으로 설정하였다. Hadamard 변환, CABAC, 루프 필터를 사용하였고, 움직임 예측에서는 5개의 참조 영상과 ±32의 탐색 범위를 사용하였다. 또한 비트-왜곡 곡선을 유도하기 위해 4개의 양자화 계수 (28, 32, 36, 40)에 대해 실험하였다. 표 2는 H.264에 비교하여 제안한 방법의 비트율과 PSNR의 향상을 보여주고 있다.

표 2 실험 결과 비교

실험영상	△PSNR(dB)	△비트율(%)
AKIYO	0.256	1.995
FOREMAN	-0.002	-1.413
MOBILE	0.077	-1.705
NEWS	0.246	0.681
SALESMAN	0.294	2.479

<표2>에서 알 수 있듯이, 적은 비트율 변화에 비해 PSNR 향상이 큰 경우가 있고, PSNR이 유사한 경우에는 비트율이 줄어드는 것을 알 수 있다. 그림 2는 MOBILE 과 NEWS 영상에 대한 비트-왜곡 곡선이다.

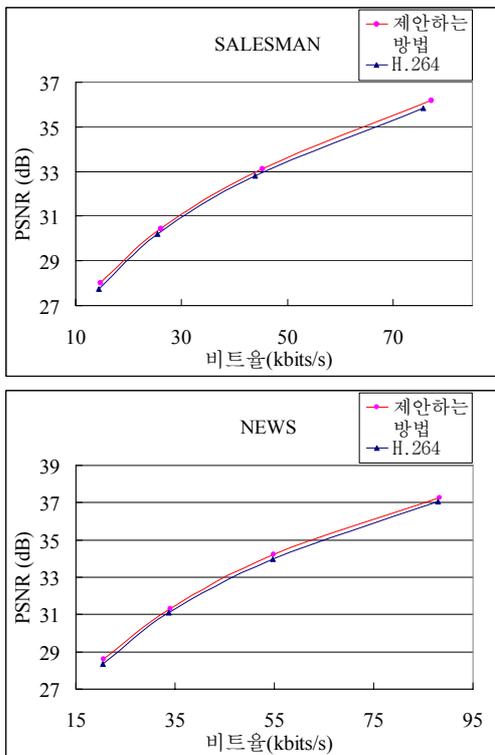


그림 2 비트-왜곡 곡선

V. 결론

본 논문에서는 화면의 타입과 매크로블록의 특성을 고려하여 적응적인 비트-왜곡 최적화 방법을 제안하였다. 제안하는 방법은 기본적으로 복잡하거나 중요한 영역에 대해서는 왜곡에 가중치를 주어 보다 좋은 화질로 부호화 될 수 있게 하였고, 단순하고 상대적으로 중요하지 않은 부분의 경우 다른 모드로 부호화 되더라도 왜곡의 손실이 크지 않으므로 오히려 비트쪽에 가중치를 주어 비트를 줄이게 하였다. 화면 레벨에선 화면의 타입에 따라 I화면의 경우 왜곡에 가중치를 더 주었고, P화면의 경우 비트에 가중치를 더 주었다. 매크로블록 레벨에서는 16×16, 16×8, 8×16의 모드의 왜곡값을 이용하여 로그 스케일의 왜곡 표준 편차 값을 정의하여, 이를 통해 매크로블록의 복잡도에 따라 Lagrangian 계수 값을 변화시켰다. 제안하는 방법은 이를 통해 H.264와 비교하여 평균적으로 0.407% 비트율이 향상 됐으나, 0.174dB 나은 화질로 부호화 할 수 있었다.

참고문헌

- [1] S. Pekka, H. Janne, and S. Olli, "Selection of the Lagrange Multiplier for Block-Based Motion Estimation Criteria", *Proceeding of International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp.325-328, May 2004.
- [2] C. J. Tsai, C. W. Tang, C. H. Chen, and Y. H. Yu, "Adaptive Rate-Distortion Optimization using Perceptual Hints", *Proceeding of 2004 IEEE International Conference on Multimedia and Expo*, pp.667-670, June 2004.
- [3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transaction on Circuits and Systems for Video Technology*, Vol.13, No.7, pp.560-576, July 2003.
- [4] JVT, Reference Software Model Version 9.5, Available at: http://iphome.hhi.de/suehring/tml/download/jm_old