

JND 기반 인지품질 향상 지향 비트 할당 방법 및 이를 이용한 HEVC 기반 인지 비디오 부호화

김대은, *김문철
한국과학기술원

kimde@kaist.ac.kr, *mkim@ee.kaist.ac.kr

HEVC based Perceptual Video Coding using JND based Bit Assignment toward Perceptual Quality Enhancement

Dae Eun Kim and *Munchurl Kim
Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 HEVC 기반 비디오 부호화에 있어 CTU 단위의 시각 민감도에 따라 CTU 별로 QP 를 조절하여 주관적 화질을 향상시키는 방법을 제안한다. 시각 민감도를 측정하는 방법으로서 화소 영역에서의 최소가지차(JND, just noticeable distortion)를 계산하여 이용하였고, 이를 HM 12.0 참조 소프트웨어에서 이용되는 R-λ 모델 기반의 울 제어 모듈에 결합하여 시각 민감도에 따라 QP 를 제어할 수 있도록 하였다. 시각 민감도가 큰 영상의 영역에 대해서는 상대적으로 작은 QP 값을, 시각민감도가 작은 영역에 대해서는 큰 QP 값을 양자화 과정에 적용함으로써, 시각 민감도가 작은 영역에 대해서는 사용 비트양을 절약하고, 절약된 비트를 상대적으로 시각 민감도가 큰 영역을 위해 사용함으로써 비디오의 주관적 화질을 향상시킬 수 있었다. 뿐만 아니라 이를 하드웨어에 적용 가능하게 하기 위해 HM 12.0 기반 하드웨어 구현을 위한 소프트웨어 플랫폼에 구현하여 실험한 결과, R-λ 모델 울 제어 알고리즘으로 울 제어 하여 부호화 한 경우 Y-PSPNR(peak signal to perceptual noise ratio)에 대한 BD-rate 는 평균 9.4%의 이득이 있었음을 확인하였다.

1. 서론

최근 HEVC (high efficiency video coding) [1] 표준화가 마무리되면서 HEVC 기반 비디오 인코더를 하드웨어로 구현하기 위한 많은 시도들이 이루어지고 있다. HEVC 표준 이전에도 MPEG-2[2], H.264/MPEG-4 AVC (advanced video coding) [3] 등의 비디오 압축 기술 표준들이 개발되어왔다. 이러한 전통적인 압축 기술들은 통계적인 중복을 줄이거나 신호처리적인 관점에서만 압축을 수행해왔다. 그러나 비디오 매체가 최종적으로 도달하여 소비되는 곳은 인간의 시각 인지 체계임을 생각하면, 비디오를 압축할 때 인간의 시각 인지 특성을 고려하지 않은 채 MSE (mean squared error)나 PSNR (peak signal to noise ratio) 등의 객관적인 지표와 비트율의 관계를 최적화 하는 기존의 전통적인 비디오 압축 기술에 인간의 시각 인지 특성을 고려한 기술을 접목하면 더 높은 부호화 효율을 얻을 수 있을 것으로 기대할 수 있다. 이같이 시각 인지 특성을 비디오 압축 과정에 반영하는 것을 인지 비디오 부호화라고 하며 최근 많은 연구가 진행되고 있다. 본 논문에서도 전통적인 비디오 부호화 방식의 한계점을 극복하기 위해 시각 인지적인 특성을 이용하여 비디오 부호화의 효율을 향상시키기 위한 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2 장에서 화소 영역의 최소가지차에 대해 살펴본 후 3 장에서는 HM 12.0 참조 소프트웨어에서 이용되는 R-λ 모델 기반의 울 제어 기술에

대해 설명한다. 4 장에서는 본 논문에서 제안하는 기법을 설명하고, 5 장에서는 제안한 알고리즘의 성능을 실험을 통해서 검증한다. 마지막으로 6 장에서는 본 논문에 대한 결론을 맺는다.

2. 화소 영역의 최소가지차

비디오 매체가 최종적으로 도달하는 지점은 인간의 시각 인지 체계이기 때문에 인간의 시각 인지 체계를 이해하는 것은 매우 중요하다. 다행히도 비디오 압축 기술에 적용할 만한 몇 가지 시각 인지 특성들이 존재한다. 예를 들어, 대조 민감성(contrast sensitivity)이라는 특성이 있는데 이는 대조 값이 특정 임계치를 넘지 못하는 영상 신호에 대해서는 시각 인지 체계가 인식하지 못하는 특성을 나타낸다. 이 임계치는 배경 그림의 밝기나 복잡도에 따라 달라진다. 최소가지차 (JND, just noticeable distortion)란 대조 민감성과 유사하게 시각 인지 체계의 특징을 정량적으로 모델화한 것이다. 이로부터 알 수 있듯이 JND 란 시각 인지 체계가 인지할 수 있는 최소한의 왜곡을 의미한다. 즉, JND 값 이하의 왜곡에 대해서는 시각 인지 체계가 인지할 수 없는 것이다. JND 값을 모델화 하는 과정을 크게 두 가지로 나눌 수 있는데, 하나는 subband JND 이고 나머지는 화소 영역의 JND 이다[4]. 본 논문에서는 시각 인지 특성의 모델을 위해 화소 영역에서의 Yang 의 JND 모델 [5]을 이용하였다. Yang 의 JND 모델에서는 배경

밝기로부터 유발되는 JND 값과 배경의 복잡도로부터 유발되는 JND의 합을 이용하여 전체 JND 값을 다음과 같이 계산한다.

$$JND_S(x, y) = JND_{LM}(x, y) + JND_{CM}(x, y) - C \times \min(JND_{LM}(x, y), JND_{CM}(x, y)) \quad (1)$$

여기서 JND_{LM} 은 배경의 밝기에 의한 JND 값이고 다음과 같이 계산 된다.

$$JND_{LM}(x, y) = \begin{cases} 17 \left(1 - \sqrt{\frac{\bar{I}_Y(x, y)}{127}} \right) + 3 & \text{if } \bar{I}_Y(x, y) \leq 127 \\ \frac{3}{128} (\bar{I}_Y(x, y) - 127) + 3 & \text{otherwise} \end{cases} \quad (2)$$

JND_{CM} 은 배경의 복잡도에 의해 계산되는 값으로 다음과 같이 계산된다.

$$JND_{CM}(x, y) = \beta G(x, y) W(x, y) \quad (3)$$

여기서 G 는 현재 화소에서 네 방향으로의 변화도 (gradient) 중 최대값을 나타내고, W 는 Canny edge detector로부터 얻은 에지 정보이다. 따라서 변화도가 클수록 JND 값이 커지는 동시에 영상의 에지 부분에서는 JND 값을 낮게 해준 것이다.

3. R-λ 모델 기반 HEVC 율 제어

R-λ 모델 기반 율 제어 방법은 HM 12.0 참조 소프트웨어에 포함되어 있는 율 제어 알고리즘이다 [6]. R-λ 모델 기반의 율 제어 알고리즘은 크게 두 부분으로 나뉘는데 첫 번째는 비트 할당이고, 두 번째는 할당된 비트를 달성하기 위해 양자화 파라미터 값을 구하는 과정이다. 비트 할당은 주어진 목표 비트율이 주어지면 GOP (group of pictures) 단위로 필요 비트를 할당하고, 다음으로 프레임 레벨, 그리고 CTU (coding tree unit) 레벨로 필요 비트를 할당한다. 목표 비트가 할당된 후에는 아래의 식을 이용해서 λ와 QP 값을 계산한다.

$$\lambda = \alpha \cdot bpp^\beta \quad (4)$$

$$QP = 4.2005 \ln \lambda + 13.7122 \quad (5)$$

여기서 α와 β는 비디오의 특성에 따라서 부호화 하며 계속해서 갱신해나가는 모델 매개변수 값이고, bpp는 CTU 레벨로 할당된 비트에 대해 계산된 화소당 필요 비트 (bit per pixel) 값이다. 따라서 여기에서 CTU 마다 할당되는 bpp 값을 CTU의 시각 민감도에 따라서 변화시켜 주면 그에 따라 QP가 변화되어 시각적으로 민감한 영역에는 낮은 QP 값이, 시각적으로 둔감한 영역에는 높은 QP 값이 할당되는 것을 예상할 수 있다.

4. 제안 방법

앞에서 언급한 R-λ 모델 기반 비트율을 할당하고 QP 값을 예측함에 있어, 기존 처럼 QP 값을 구하기 위해 CTU 단위로 할당된 목표 비트 값을 그대로 이용하지 않고 시각 민감도에 따라서 시각적으로 둔감한 CTU에 대해서는 상대적으로 적은 양의 비트를 할당하고 시각적으로 민감한 CTU에 대해서는

상대적으로 많은 양의 비트를 할당하면 프레임 내에서 작은 왜곡에도 시각적으로 둔감하게 반응하는 영역에 대해서는 낮은 비트율로 부호화 하여도 발생하는 왜곡이 상대적으로 덜 눈에 띄므로 주관적 화질을 유지하면서도 비트를 절약할 수 있다. 반대로 여기서 절약한 비트를 이용하여 시각적으로 민감한 영역에 더 많은 비트를 할애하여 부호화 해주면 시각적으로 민감한 영역에서는 화질을 높일 수 있으므로 전체적으로 동일한 비트를 가지고 더 좋은 주관적 화질을 얻을 수 있다. 이를 위해 다음과 같은 식으로 CTU 단위로 할당되는 비트 수를 변화시켜 주었다.

$$bpp = bpp \cdot (JND_{Frame} / JND_{CTU}) \quad (6)$$

여기서 우변의 bpp는 율 제어 알고리즘에 의해 CTU에 할당된 목표 비트를 CTU의 화소 수로 나뉜 값이고 JND_{Frame} 은 (1)식으로부터 구한 JND를 한 프레임에 대해 평균을 구한 값이고, JND_{CTU} 는 (1)식으로부터 구한 JND 값을 현재 처리하고 있는 CTU에 대해 평균을 구한 값이다. JND_{CTU} 의 역수 값을 해당 CTU의 시각 민감도로 정의하여 계산해준 것이다. JND 값에 따라 bpp 값을 변화시켜 줌으로써 λ 값이 변화되고 그에 따라 QP 값이 CTU 마다 서로 다르게 설정되는 것이다. 식 (6)을 식 (4)와 (5)에 대입해보면 실제로 할당되는 QP 값은 아래와 같이 변화됨을 알 수 있다.

$$QP = QP + 4.2005 \beta \ln(JND_{Frame} / JND_{CTU}) \quad (7)$$

5. 실험 결과

4장에서 제안한 알고리즘을 검증하기 위해 소프트웨어에 구현하여 실험을 통해 검증하였다. 구현에 이용된 소프트웨어는 HM 12.0을 하드웨어로 구현하기 위한 간략화된 참조 소프트웨어 플랫폼이다. GOP 구조는 Low delay P 구조가 이용되었고, 크기는 4이다. HEVC의 Main 프로파일을 이용했다. 실험 영상은 1920×1080 크기의 영상인 *Kimono*, *BasketballDrive*, *BQTerrace*가 이용되었다. 이 3개의 영상에 대해 R-λ 모델 기반의 율 제어 알고리즘을 적용하여 각각 주어진 목표 비트율로 부호화하였다. 목표 비트율은 3Mbps, 5Mbps, 10Mbps, 20Mbps 4 가지이며, 각각의 목표 비트율로 부호화된 영상은 PSPNR (peak signal to perceptual noise ratio)을 이용하여 화질을 측정하였다. PSPNR은 다음과 같이 계산된다.

$$PSPNR = 10 \cdot \log_{10} \frac{255 \times 255}{\frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (I(x, y) - \hat{I}(x, y) - JND(x, y))^2 \delta(x, y)} \quad (8)$$

여기서 $\delta(x, y)$ 는 다음과 같다.

$$\delta(x, y) = \begin{cases} 1, & \text{if } |I(x, y) - \hat{I}(x, y)| \geq JND(x, y) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

즉 PSPNR은 PSNR과 거의 유사해 보이지만 JND에 의한 왜곡을 보상해 줌으로써 마치 그 오차가 0인 것처럼 간주한다는 의미이다. 표 1은 제안 알고리즘을 구현하여 부호화한 결과를 비교하여 나타낸다. 표 1의 실험 결과로부터 알 수 있듯이 비트율 제어되는 상황에서 제안 알고리즘이 평균적으로 우수한 주관적 화질을 나타내는 것을 확인할 수 있었다. 동일한 목표 비트율로 부호화한 결과를 보면, 목표 비트율로부터 오차가 발생하여 부호화된 실제 비트율이 서로

다르긴 하지만 모든 경우에서 제안 알고리즘이 더 높은 PSPNR 값을 나타내는 것을 볼 수 있다. 또한 PPSNR 값에 대한 BD-rate 값도 Y 성분에서 평균 9.4% 이득이 있음을 확인할 수 있다.

표 1. 기존 부호화기로 부호화한 결과와 제안 알고리즘으로 부호화한 결과 비교

실험영상	기존 부호화기				제안 알고리즘				BD-rates (%)		
	kbps	Y-PSPNR	U-PSPNR	V-PSPNR	kbps	Y-PSPNR	U-PSPNR	V-PSPNR	Y	U	V
Kimono	20032.65	65.98	71.16	74.71	20023.56	66.26	71.86	75.57	-6.5%	-8.6%	-10.1%
	10006.43	61.50	66.73	68.40	10004.52	61.99	67.52	69.58			
	5001.85	56.30	61.50	61.67	5001.57	56.88	62.00	62.47			
	3001.03	51.59	57.40	57.32	3000.87	52.17	57.91	58.13			
BasketballDrive	20010.85	47.65	57.43	57.24	20006.95	47.72	57.53	57.36	-9.5%	-13.6%	-11.5%
	10002.98	44.09	53.52	52.23	10001.84	44.59	54.24	53.00			
	5001.12	40.40	50.09	48.16	5000.76	41.09	51.00	49.06			
	3000.12	37.59	48.11	45.85	2999.87	38.18	48.86	46.47			
BQTerrace	20006.42	43.82	55.36	60.23	20005.73	44.15	55.94	61.18	-12.1%	-19.7%	-29.8%
	10002.37	42.00	53.89	58.71	10004.36	42.35	54.38	59.56			
	4999.68	40.20	52.34	57.24	4999.65	40.56	52.91	58.07			
	3000.70	38.66	50.73	55.78	3000.63	39.01	51.19	56.44			
								평균	-9.4%	-13.9%	-17.1%

그림 1 은 기존 부호화기와 제안 알고리즘을 포함한 부호화기로 부호화한 비디오의 일부분을 나타낸다. 그림 1 에서 볼 수 있듯이 제안 알고리즘으로 부호화한 영상이 주관적으로 더 좋은 화질을 나타냄을 확인할 수 있다.



(a) 기존 부호화기로 부호화한 결과



(b) 제안 부호화기로 부호화한 결과

그림 1. 기존 부호화기와 제안 부호화기의 부호화 결과 비교 BasketballDrive, 3000kbps, 500th frame

6. 결론

본 논문에서는 인간의 시각 인지 특성을 고려하여 HEVC 부호화기 성능 향상 방법을 제안하였다. 시각 인지 특성을 나타내는 지표로는 JND 값을 이용하였고, 프레임 전체에서의 JND 값의 평균과 현재 부호화 하는 CTU 에서의 JND 값의 평균을 비교하여 시각 민감도에 따라 CTU 별로 비트 수를 다르게 할당해 주었다. R-λ 모델 기반의 윌 제어 알고리즘을 이용하여 윌 제어를 하는 경우, 제안하는 알고리즘을 구현하여 실험한 결과 PPSNR 을 기준으로 측정된 BD-rate 에서 평균 9.4%의 성능 향상을 나타냈다.

참고문헌

- [1] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, "High Efficiency Video Coding (HEVC) Text Specification Draft 9," JCTVC-K1003, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Oct. 2012.
- [2] Information Technology: Generic Coding of Moving Pictures and Associated Audio, part 2: visual, ISO/IEC 13818-2:1994, 1994.
- [3] Draft ITU-T Recommendation and Final Draft International standard of Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC), JVT-G050, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG 2003.
- [4] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile from grey scale image/video in DCT domain", IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 3, pp. 337-346, Mar. 2009.
- [5] X. K. Yang, W. Lin, Z. K. Lu, E. P. Ong and S. S. Yao, "Just noticeable distortion model and its applications in video coding", Signal Process.: Image Commun., vol. 20, no. 7, pp. 662-680, Aug. 2005.
- [6] B. Li, H. Li, L. Li, and J. Zhang, "Rate Control by R-lambda Model for HEVC," JCTVC-K0103, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Oct. 2012.