

# HEVC 쿼드트리 부호화를 위한 율-왜곡 모델

\*이범식 \*\*김문철

한국과학기술원

\*lbs@kaist.ac.kr, \*\*mkim@ee.kaist.ac.kr

## Rate-Distortion Model for HEVC Quadtree Coding

\*Lee, Bumshik \*\*Kim, Munchurl

Korea Advanced Institute of Science and Technology (KAIST)

### 요약

최근 ISO/IEC의 MPEG과 ITU-T의 VCEG이 JCT-VC (Joint Collaborative Team for Video Coding)를 구성하여 HEVC (High Efficiency Video Coding) 차세대 비디오 압축 표준 제정을 위한 작업을 진행 중이다. 과거 압축률이 가장 좋은 것으로 알려진 H.264/AVC 보다 최대 50%까지 부호화 효율 향상을 목표로 하고 있다. HEVC는 H.264/AVC와는 상이한 부호화 구조를 채택하고 있고 작은 크기의 영상뿐만 아니라 크기가 큰 영상까지도 효율적으로 부호화할 수 있도록 설계되고 있다. 예측 및 변환 부호화 과정이 계층적 쿼드트리 구조를 가지며, 특히 변환 부호화는 작은 크기의 변환 블록으로부터  $32 \times 32$  크기의 변환 블록까지 크게 확장되어 계층적 변환 구조를 이루며 부호화하도록 되어 있다.

본 논문에서는 기존 코덱과는 상이한 부호화 구조를 갖는 쿼드트리 부호화 기반 HEVC 코덱 표준을 위한 율-왜곡 (Rate-Distortion) 모델을 제안한다. 기존의 코덱에서는 부호화되는 기본 단위가  $16 \times 16$ 로 일정하고, 변환 및 양자화되는 블록의 크기 역시  $4 \times 4$  또는  $8 \times 8$  크기 단위로 그 블록의 크기가 작을 뿐만 아니라 고정된 크기를 사용한다. 따라서 단일 확률 모형을 사용하여 율-왜곡 모델을 만들었으며, 그 정확도 역시 비교적 정확한 결과를 얻었다. 그러나 HEVC에서는 계층적 가변 블록 크기를 갖는 기본 부호화, 예측 및 변환/양자화 기법을 사용하기 때문에 기존의 단일 모델로는 정확한 율-왜곡 모델을 만들어 내기 어렵다. 제안하는 방법은 HEVC의 기본 단위인 CU (Coding Unit)별로 독립적인 확률 모형을 사용하여 율-왜곡 모델을 사용하는 것으로 CU의 크기가 가변적이고 CU 내의 텍스처 역시 크기에 따라 매우 다른 특성을 가지고 있기 때문에 단일 모델을 사용하는 것보다 매우 효율적인 것을 실험을 통하여 확인하였다.

## 1. 서론 및 HEVC 부호화 구조

H.264/AVC [1] 표준 코덱은 현재 존재하는 비디오 압축 표준 중에서 압축 성능이 가장 뛰어난 것으로 알려져 있다. 2003년도 표준 제정 시 기존의 압축 표준 보다 50% 이상의 부호화 효율을 증대시키기 위하여 매우 많은 기술들이 채택되었다. 대표적으로 가변 예측 기술,  $1/4$  정확도의 부호소 움직임벡터 예측 기술, 정수변환 및 양자화 기술, 인루프 블로킹현상 제거 필터, 문맥적응형 엔트로피 부호화 기술 등이 채택되어 객관적 압축 성능 및 주관적 화질 향상을 가져왔다. 그러나 H.264/AVC는 CIF (Common Intermediate Format) 또는 QCIF (Quarter CIF) 등의 작은 영상의 압축에 효율적으로 부호화 되도록 설계되었다. 또한 가장 기본적인 부호화 처리 단위인 매크로블록 (Macroblock, MB) 역시  $16 \times 16$  크기로 고정되어 사용되고 있다. 변환 부호화 및 양자화의 경우 매크로블록 내에서 베이스라인 프로파일의 경우 16개의 고정 크기의  $4 \times 4$  블록을 사용하여 하이 프로 파일의 경우에는 16개의  $4 \times 4$  블록 또는 4개의  $8 \times 8$  블록을 사용한다. 이렇게 고정 크기의 매크로블록과 변환 블록의 사용은 입력 신호에 적응적인 변환 부호화를 수행하는데 한계가 있으며 변환 블록의 크기도  $8 \times 8$ 로 제한적이기 때문에 HD (High Definition)급 이상의 큰 공간 해상도를 갖는 영상을 높은 압축률로 부호화하기 어렵다.

이러한 H.264/AVC의 문제점을 해결하기 위하여 HEVC [2]에서

는 기본 부호화 처리 단위인 매크로블록을 확장하여 CU (Coding Unit)라는 개념을 도입하였으며 그 처리 단위가 가변 크기를 갖는 계층적 구조를 가지고 있다. 또한 변환 및 양자화 역시 CU의 크기를 넘지 않는 범위 내에서 계층적 가변 크기 구조를 가지고 있다. 변환 블록 단위 (Transform Unit, TU)의 크기 역시 기존의 최대  $8 \times 8$  크기에서 크게 확장하여 최대  $32 \times 32$  변환 크기 까지 사용할 수 있도록 설계되었다. 그림 1은 HEVC의 CU 및 TU에 대한 적용 예를 보여준다.

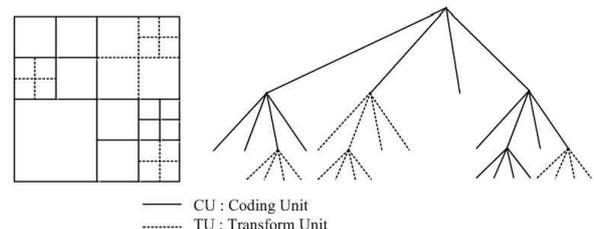


그림 1. HEVC에서 쿼드트리 구조를 갖는 CU 및 TU

그림 1에서 보는 것처럼 CU는 가변 블록 크기를 가지며 계층적으로 구성되며 그 형태가 쿼드트리 [3] 구조를 지니고 있다. CU는 H.264/AVC에서 MB와 유사한 기능을 하며 CU에서는 움직임 벡터 예측, 변환 및 양자화 수행, 엔트로피 부호화 등을 수행한다. CU의 최대

크기는 64×64이고 최소 크기는 8×8이다. 최대 크기와 최소 크기의 결정은 부호화시 파라미터 입력으로 결정되며 일단 최대 및 최소 크기의 CU가 결정되면 쿼드트리 계층의 깊이(depth 또는 level)가 결정되며 각 깊이에서 CU 블록의 크기를 알 수 있다. 그 구조가 계층적 쿼드트리 구조이기 때문에 입력 신호에 적응적으로 부호화할 수 있는 장점이 있다. TU (Transform Unit)는 변환 블록을 의미하며 CU블록 내에서 입력 잔차 신호에 대해 계층적 쿼드트리 구조를 이루며 변환 및 양자화를 수행한다. TU의 최대 크기는 32×32이고 최소 크기는 4×4가 되고 TU의 최대 및 최소 크기 역시 파라미터 입력으로 결정된다. CU 및 TU는 결합된 계층적 쿼드트리 구조를 가지며 각 레벨에서 CU의 TU의 크기는 윌-왜곡 최적화 (Rate-Distortion Optimization)을 통하여 결정된다.

이와 같이 HEVC는 H.264/AVC와는 매우 상이한 구조를 가지며 쿼드트리의 특성 때문에 입력 신호에 적응적으로 부호화되도록 설계되었다. 따라서 작은 크기의 영상부터 HD급 이상 영상의 크기가 매우 큰 영상까지 매우 효율적으로 부호화할 수 있는 장점이 있다. CU 및 TU의 크기 또는 레벨 (깊이)은 RDO에 의해 결정되기 때문에 그 크기가 크면 영상의 움직임 매우 느리거나 정적이며 텍스처도 영역 내에서 동일한 (homogeneous) 특성을 가질 때 유리하며, 반대인 경우는 영상의 움직임이 크거나 텍스처가 복잡할 때 쿼드트리 구조가 분할되기 때문에 더 작은 크기가 유리해진다.

본 제안 방법을 설명하기에 앞서 CU의 크기에 따른 신호의 특성을 대해 알아보기로 한다. 앞서 설명한 것처럼 CU의 레벨 (크기)에 따라 입력 신호의 특성이 달라지며 이를 입력 잔차 신호의 분산에 대해 분석하면 그림 2와 같다.

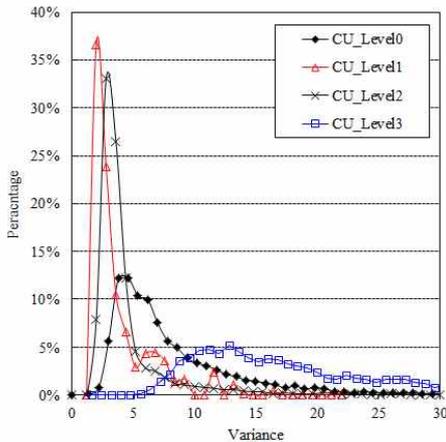


그림 2. CU 레벨에 따른 분산의 분포 (BQMall\_832x480, QP32)

그림 2에서 보는 것처럼 CU의 레벨(크기)에 따라 RDO 수행이후 입력 잔차 신호의 분산 분포가 상이하게 다름을 알 수 있다. 그림 2에 나타난 실험적 관찰을 통해 CU에 따라 각기 독립적 확률 모델을 이용한 윌-왜곡 모델을 제안한다.

## 2. 제안하는 HEVC를 위한 확률모델 기반 윌-왜곡 모델

그림 2에서 보는 것처럼 각 CU 레벨에서 RDO에 결정된 입력 잔차 신호의 특성이 매우 다르므로 기존의 단일 확률 모델을 이용해서는 정확한 윌-왜곡 모델을 만들기 어렵다. 본 논문에서는 라플라시안 확률 모델을 각 CU 레벨 별로 독립적으로 예측하여 프레임별로 윌-왜곡 모

델을 구성하는 방법에 대하여 제안한다. 따라서 제안하는 모델은 일종의 혼합 모델의 형태를 보이며 그 식은 (1)과 같다.

$$f_{LLM}(l) = \underbrace{w_0 \frac{\lambda_0}{2} e^{-\lambda_0|l|}}_{CU\ level0} + \underbrace{w_1 \frac{\lambda_1}{2} e^{-\lambda_1|l|}}_{CU\ level1} + \underbrace{w_2 \frac{\lambda_2}{2} e^{-\lambda_2|l|}}_{CU\ level2} + \underbrace{w_3 \frac{\lambda_3}{2} e^{-\lambda_3|l|}}_{CU\ level3} \quad (1)$$

(1)은 라플라시안 모델이 각 CU레벨에서 독립적으로 구성된 라플라시안 모델의 혼합모델로 구성된다. 여기서  $w_k$ 는 각 CU 레벨의 가중치,  $\lambda_k$ 는 CU 레벨  $k$ 에서의 모델 파라미터이며  $\lambda_k = \sqrt{2}/\sigma_k$ 로 계산된다.  $\sigma_k$ 는 CU 레벨  $k$ 에서 변환 계수의 표준편차를 나타낸다. 그림 2에서 보는 것처럼 CU 레벨에 따라 잔차 신호의 분산의 분포가 매우 다르므로 DCT 변환 영역에서 분산 역시 그 특성이 달라지기 때문에 본 제안 모델에서는 모델 파라미터  $\lambda_k$ 를 CU 레벨에 따라 독립적으로 예측하여 사용한다. 그림 3은 CU 레벨  $k$ 에서의 라플라시안 모델과 단일 모델을 보여준다.

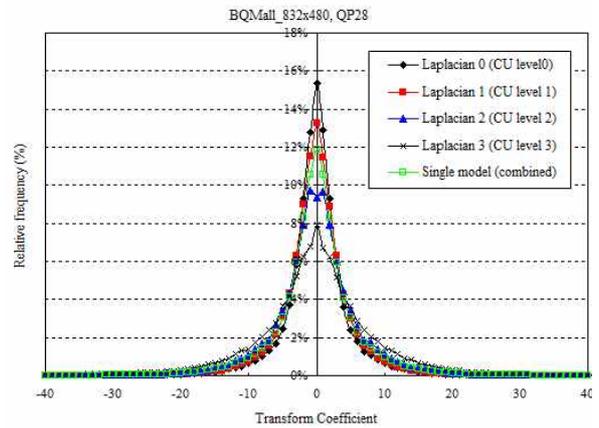


그림 3. CU레벨  $k$ 에서의 라플라시안 모델과 단일 모델

그림 3에서 보는 것처럼 단일 모델을 사용하였을 경우 CU 레벨  $k$ 와 오차를 보이며 이것은 윌-왜곡 모델을 적용 시 정확도 오차의 원인이 된다.

### 2-1. 제안하는 HEVC 왜곡 모델

제안하는 왜곡 모델은 (3)과 같이 SKIP 모드에 대한 왜곡 ( $D_{SKIP}$ )과 비SKIP 모드에 대한 왜곡 ( $D_{nonSKIP}$ )의 합으로 표현된다.

$$D_{Total} = D_{SKIP} + D_{nonSKIP} \quad (3)$$

(3)에서  $D_{SKIP}$ 은 스킵 모드에 대한 왜곡으로 이 전 프레임으로부터 예측을 하며  $D_{nonSKIP}$ 은 (1)로 표현되는 확률모델을 이용하여 각 CU에서 구해진 양자화오차를 평균자승에러 (MSE, Mean Square Error)를 이용하여 구한다. 따라서  $D_{nonSKIP}$ 은 (4)와 같이 구할 수 있다. 여기서 HEVC의 양자화 기법은 DZ + UTQ (Dead Zone + Uniform Threshold Quantizer) [4]이다.

$$D_{nonSKIP} = \int_{-(q-fq)}^{q-fq} l^2 f_{LLM}(l) dl + 2 \sum_{n=1}^{\infty} \int_{-(q-fq)}^{q-fq} l^2 f_{LLM}(l) dl \quad (4)$$

여기서  $q$ 는 양자화스텝 크기,  $t$ 는 라운딩 오프셋 파라미터를 나타낸다. (4)는 DZ+UTQ에서 MSE를 나타내며 (1)과 (4)를 이용하여 전체 비SKIP 모드에 대한 전체 왜곡 모델을 구하면 (5)와 같다.

$$D_{nonSKIP}(q) = \sum_{k=0}^{N-1} w_k \frac{(2 + q\lambda_k - 2fq\lambda_k) \cdot q\lambda_k e^{f\lambda_k q} + 2 - 2e^{-\lambda_k q}}{\lambda_k^2 (1 - e^{-\lambda_k q})} \quad (5)$$

여기서  $N$ 은 CU 레벨의 총 개수,  $w_k$ 와  $\lambda_k$ 는 CU레벨  $k$ 의 가중치와 모델 파라미터를 각각 의미한다. 두 파라미터가 모델의 정확도를 결정짓는 중요한 요소이다. 본 제안 방법에서는  $w_k$ 는 이전 프레임들에서 CU 레벨  $k$ 가 프레임에서의 차지하는 면적으로부터,  $\lambda_k$ 는 이전 프레임들에서 CU 레벨  $k$ 의 변환 계수로부터 예측하여 결정한다.

### 2-2. 제안하는 HEVC 울 모델

HEVC에서는 CU내에서는 서로 다른 변환 블록크기의 크기가 4종류가 존재할 수 있다. 예를 들어 CU의 크기가  $64 \times 64$ 인 경우 CU내에서 존재할 수 있는 변환 블록의 크기는  $32 \times 32$ ,  $16 \times 16$ ,  $8 \times 8$ ,  $4 \times 4$  4종류가 존재할 수 있다. CU내에서 여러 종류의 변환 블록의 존재한다고 하더라도 실제 비트와 변환양자화 계수와의 관계는 그림 4와 같다.

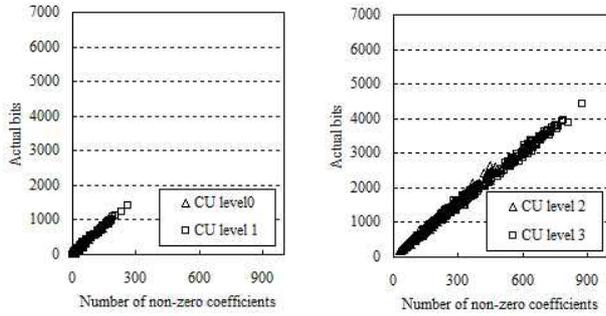


그림 4. 비영변환양자화 계수(non-zero coefficients)의 개수와 실제 비트와의 선형 관계

그림 4에서 보는 것처럼 영으로 양자화 되지 않는 변환 계수의 개수와 실제 비트와의 관계는 선형 관계를 가진다는 것을 알 수 있다. 이를 통해 CU 레벨  $k$ 에서 영변환 계수의 개수를 이용하여 울-모델을 만들 수가 있다. 확률 모델 (1)을 이용하여 영변환 계수를 계산하면 (6)과 같다.

$$N_{nz}^{(k)}(q) = 1 - 2 \int_0^{(1-f)q} f_L^{(k)} dl = e^{-(1-f)q\lambda_k} \quad (6)$$

영변환 계수와 실제비트와의 선형 관계를 이용하면 울 모델을 (7)과 같이 표현할 수 있다.

$$R(q) = \sum_{k=0}^{N-1} w_k N_{nz}^{(k)} \quad (7)$$

여기서  $N$ 은 CU 레벨의 총 개수,  $w_k$ 는 CU 레벨  $k$ 의 가중치이며 이전 프레임으로부터 CU 레벨  $k$ 의 면적을 예측함으로써 구할 수 있다.

### 3. 실험 결과

제안하는 모델을 검증하기 위하여 HEVC 모델 HM1.0을 사용하였다. 먼저 왜곡 모델을 검증하기 위하여 실제 왜곡과 모델을 이용한 왜곡을 이용하였을 때를 비교하였다. 먼저 비SKIP 모드에 대한 왜곡 모델을 검증한다. 이것은 모델 (4)가 얼마나 정확함을 검증하기 위한 것이다. BQSquare 영상을 QP 28을 이용하였을 때 자승에러를 측정하는 것이다. 그림 5에서 보는 것처럼 실제 왜곡과 모델로부터 예측한 왜곡은 거의 오차 없이 예측됨을 알 수 있다.

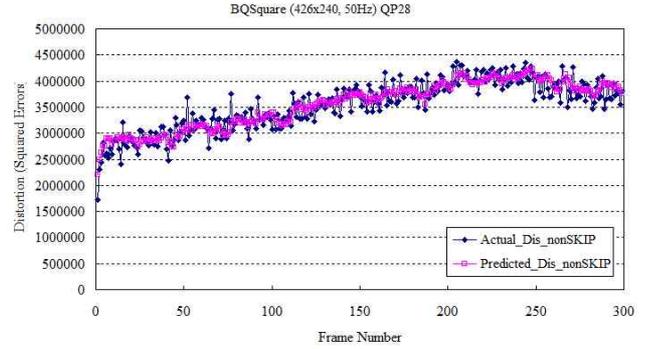


그림 5. 비SKIP 모드에 대한 왜곡 모델 검증  
그림 6은 SKIP 모드와 비SKIP모드를 모두 합산하여 얻은 결과이다.

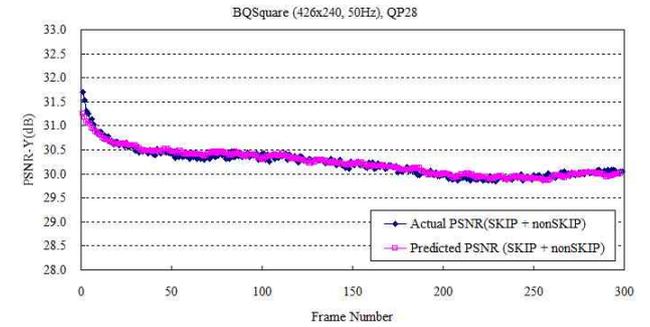


그림 6. 합산된 실제 PSNR과 모델을 통해 예측한 PSNR

합산된 PSNR의 경우도 역시 실제 PSNR과 예측된 PSNR의 오차가 거의 없음을 그래프를 통하여 알 수 있다.

울 모델에 대한 실험 결과를 그림 7에서 보여준다. 그림 7은 BQSquare 영상을 QP 28에서 부호화 하였을 때 실제 비트와 모델을 통한 예측 비트의 그래프를 나타낸다.

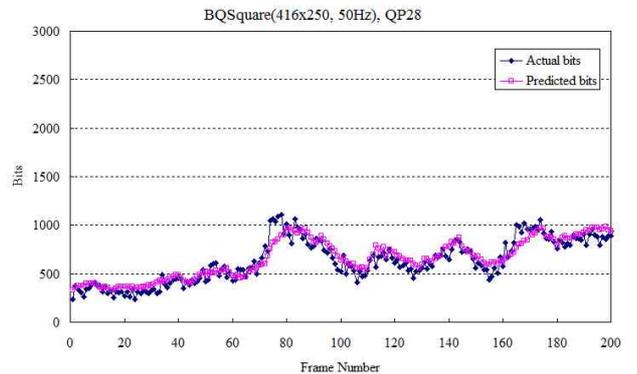


그림 7. 실제 비트와 제안된 모델을 통해 예측한 비트

그림 7에서 보는 것처럼 실제 비트와 모델을 통한 비트 예측이 아주 적은 오차로 정확히 예측되는 것을 알 수 있다.

### 3. 결론 및 향후 과제

본 논문에서는 HEVC에서 사용되는 있는 가변 크기를 갖는 부호화 단위 및 변환 블록에 적합한 울-왜곡 모델을 제안하였다. 제안된 모델은 하나의 프레임 단위 울-왜곡 모델 설정 시 하나의 모델 파라미터를 이용하지 않고 서로 다른 부호화 특성을 갖는 각 CU별로 독립된 라플라시안 모델을 구성한 후 가중치에 의해 혼합 라플라시안 모델 형태로 만들어 진다. 본 울-왜곡 모델은 실험결과에서 확인하였듯이 프레임 기반 실제 부호화를 통하여 얻은 울-왜곡 치를 적은 오차로 예측한다.

왜곡 모델은 MSE는 PSNR을 통한 결과 값을 이용하여 검증할 수 있으나 울 모델은 부호화되는 텍스처 외에 변환 블록의 크기나 움직임 벡터, CU의 분할 여부를 나타내는 정보 등을 나타내는 부가정보 (side information)등의 양이 현존하는 코덱에 비해 매우 많고 모델링하기 어렵기 때문에 본 논문에서 울 모델을 부호화되는 텍스처에 대한 모델만을 고려하였다. 향후 부가 정보에 대한 모델링을 통한 보다 정확한 비트율 예측이 필요하다고 할 수 있다.

### 참고 문헌

- [1] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10, Mar. 2005.
- [2] T. Wiegand, J.-R. Ohm, G. J. Sullivan, W.-J. Han, R. Joshi, T. K. Tan, K. Ugur, "Special Section On the Joint Call for Proposals on High Efficiency Video Coding (HEVC) Standardization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, Dec. 2010.
- [3] D. Marpe, H. Schwarz, S. Boße, B. Bross, P. Helle, T. Hinz, H. Kirchhoffer, H. Lakshman, T. Nguyen, S. Oudin, M. Siekmann, K. Sühring, M. Winken, and T. Wiegand: *Video Compression Using Nested Quadtree Structures, Leaf Merging and Improved Techniques for Motion Representation and Entropy Coding*, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1676-1687, Dec. 2010.
- [4] G. J. Sullivan and S. Sun, "On dead-zone plus uniform threshold scalar quantization," in *Proc. VCIP*, Beijing, China, Jul. 2005, pp. 10741-1052.