계층적 B-픽처 구조를 고려한 H.264/AVC 비트열의 PSNR 예측

서 정 동^{a)}, 손 광 훈^{a)*}

PSNR Estimation of H.264/AVC Bitstream for Hierarchical- B Picture Structure

Jungdong Seo^{a)} and Kwanghoon Sohn^{a)*}

요 약

부호화된 비트열의 PSNR 예측 알고리즘은 무 기준법에 속하는 화질 평가 방법으로 수신단에서 참조 영상 없이 수행할 수 있기 때문에 높은 효율성을 지닌다. 기존의 PSNR 예측 연구들은 주로 I-픽처나 일반적인 IBBP 예측 구조를 고려하여 이루어지는 반면에 본 논문에서는 계층적 B-픽처 구조를 고려한 PSNR 예측 기법을 제안한다. 제안된 알고리즘은 최하위 계층의 I-픽처를 위한 새로운 DCT 계수 모델링 방법과 상위 계층의 픽처들이 주로 선택되는 스킵 모드를 고려한 PSNR 예측 기법으로 구성되어 있다. 제안 알고리즘의 성능 평가를 위해 실험 영상을 H.264/AVC로 부호화 하고 생성된 비트열의 예측된 PSNR 값과 실제 PSNR 값을 비교하였다. 실험 결과를 통해 제안된 DCT 모델링 방법이 기존의 방법들에 비해 더 정확함을 확인하였으며 스킵 모드를 고려한 PSNR 예측 기법이 계층적 B-픽처 구조에 적합함을 확인하였다.

Abstract

This paper proposes a PSNR estimation algorithm of H.264/AVC bitstream for hierarchical B-picture structure. The proposed method consists of a modeling method for DCT coefficients for I-pictures and an error estimation method for blocks encoded by skip mode. The modeling method selects an appropriate model between Laplacian and Cauchy model, and the error of skip mode is estimated from MSE values of reference pictures. Experimental results show that the modeling method predicts more accurate PSNR values than Laplacian and Cauchy model and the error estimation method of skip mode enhances PSNR estimation of hierarchical B-picture structure.

Keyword : PSNR 예측, H.264/AVC, 계층적 B-픽처 구조, DCT 계수 모델링

1. 서 론

부호화된 동영상의 화질을 예측하는 연구는 여러 대학과

연구기관, 회사 등에서 진행되어 왔으며 표준화 기구인 Video Quality Expert Group (VQEG)을 중심으로 이와 관련된 표준화가 활발하게 진행되고 있다^[1]. 특히 네트워크 시스템과 재생 기기의 발전으로 실시간 동영상의 제공 환경이 다양해지면서 동영상 화질을 객관적으로 평가하는 기술은 더욱 주목을 받고있다^[2].

동영상의 화질평가 방법은 크게 주관적 평가와 객관적 평가로 나누어지는데 주관적 방법은 다수의 평가자가 영

a) 연세대학교 전기전자공학과

Department of Electrical and Electronic Engineering, Yonsei Univ.

* 교신저자 : 손광훈 (kshohn@yonsei.ac.kr)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 “대학IT연구센터 육성 지원사업”의 연구결과로 수행되었음 (NIPA-2011-C1090-1101-0006)

· 접수일(2011년6월25일), 수정일(2011년9월19일), 게재확정일(2011년11월8일)

상을 보고 느끼는 화질에 대한 관능적 평가를 토대로 결과를 획득한다. 인간이 느끼는 화질에 대한 보편적인 결과를 기대할 수 있으며 가장 정확한 평가 방법으로 알려져 있다. 반면에 객관적 평가는 수학적 모델링과 연산을 통해 주관적 평가 결과와 유사한 값을 이끌어내는 방법으로 주관적 평가와는 달리 자동화와 실시간 구현이 가능하다는 장점이 있다^[3]. 객관적 화질평가는 참조 영상(원 영상)의 사용 형태에 따라서 전 기준법(full reference), 감소 기준법(reduced reference), 무 기준법(no reference)으로 분류된다. 전 기준법은 평가 과정에서 원 영상을 사용할 수 있기 때문에 주관적 화질 평가와 높은 상관도를 보이지만 원 영상의 정보를 모두 사용하므로 활용성 측면에서 한계를 보인다. 이와 달리 감소 기준법과 무 기준법은 원 영상 정보를 선택적으로 사용하거나 전혀 사용하지 않기 때문에 객관적 평가 결과와 높은 상관도를 가지는 감소 기준법, 무 기준법 알고리즘은 화질 모니터링 시스템이나 전송 품질에 따른 과금 체계 같은 다양한 서비스에 적용될 수 있다.

비트열을 이용한 Peak Signal-to-Noise Ratio (PSNR) 예측 기법은 무 기준법에 속하는 객관적 화질평가 방법으로 부호화된 비트열로부터 원 영상의 정보를 추측하고 이를 기반으로 부호화된 영상의 PSNR 값을 예측한다. PSNR은 객관적 화질 평가 방법 중에서 가장 대표적인 Mean Square Error (MSE) 기반의 평가 척도인데 다양한 경우의 주관적 평가 결과를 민감하게 반영하지 못한다는 의견이 제시되지만 부호화된 영상에서는 평균적으로 80%이상의 상관도를 보이며 연산 과정이 간단하고 명확한 수치화가 가능하므로 화질 왜곡을 나타내는 척도로 널리 사용된다^[4]. Turaga는 전 기준법인 PSNR을 원 영상 없이 예측하는 알고리즘을 처음으로 제안하였다^[5]. MPEG-2로 부호화된 비디오 신호에서 Discrete Cosine Transform (DCT) 계수의 통계적 분포 특성을 추출하고 이를 기반으로 라플라시안 확률 밀도 함수 모델을 적용하여 원 영상과 부호화된 영상의 차이를 예측하였다. 하지만 Turaga의 제안 방식은 화면 내 부호화(intra coding)에만 적용된다는 한계가 있었다. Ichigaya는 가중치가 부여된 라플라시안 확률 밀도 함수를 사용하여 DCT 계수 분포 모델을 향상시켰으며

화면 간 부호화(inter picture coding)에도 적용할 수 있는 알고리즘을 제안하였다^[6]. 이 연구들이 모두 MPEG-2 부호화에 기반을 두고 진행된 것과 달리 Eden은 H.264/AVC 부호화를 위한 PSNR 예측 기법을 연구하였다. Eden의 제안 기법은 라플라시안 모델을 사용하였는데 원 영상의 계수 분포에 가까운 모델을 추정하기 위해 양자화에 의해 0이 된 DCT 계수의 비율만을 사용하였고 높은 정확성을 보여주었다^[7]. Shim은 Eden이 시도한 방법을 기반으로 코시(Cauchy) 확률 밀도 함수를 모델링에 사용하여 라플라시안 모델에 비해 향상된 결과를 보여주었다^[8].

본 논문에서는 기존의 PSNR 예측 기법들을 바탕으로 계층적 B-픽처 구조(hierarchical B-picture structure)에 적합한 PSNR 예측 알고리즘을 제안한다. 계층적 B-픽처 구조는 Scalable Video Codec (SVC)에서 적용되기 시작하여 H.264/AVC와 현재 표준화 중인 High Efficiency Video Coding (HEVC)에도 적용되는 부호화 구조로 계층적으로 양방향 예측 부호화를 수행하기 때문에 움직임 보상의 효율이 높으며 향상된 움직임 보상 성능 때문에 상위 계층에서는 대부분의 블록이 스킵 모드로 선택되는 특징을 가진다. 따라서 본 논문에서는 상위 계층 픽처의 참조 영상으로 사용되는 최하위 계층의 I-픽처를 위한 DCT 계수 분포 모델링 방법을 제안하고 스킵 모드를 고려한 PSNR 예측 방법을 제안하여 전체 PSNR 예측의 정확성을 높인다. 논문의 구성은 다음과 같다. 2장에서는 일반적인 PSNR 예측 과정의 구조를 설명하고 3장에서는 계층적 B-픽처 구조를 위한 PSNR 예측 기법을 제안한다. 4장에서는 실험 과정과 실험 결과를 보이고 5장에서는 결론과 향후 연구 방향을 제시한다.

II. PSNR 예측 알고리즘의 구조

PSNR은 원 영상과 부호화된 영상 간의 차이를 로그 규모(log scale)의 값으로 표현하는 방법으로 다음과 같은 수식을 통해 계산된다.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |s(i,j) - p(i,j)|^2$$

$$PSNR = 10 \log_{10} \left(\frac{(2^{depth} - 1)^2}{MSE} \right) \quad (1)$$

식 (1)에서 $s(i,j)$ 은 원 영상, $p(i,j)$ 은 부호화된 영상, m 과 n 은 영상의 크기를 의미하며 $depth$ 는 각 화소가 가지는 비트 깊이이다. 일반적으로 8비트 영상을 사용하므로 PSNR의 계산에서 분자에 들어가는 숫자는 255^2 이 된다. 식 (1)에서 MSE의 계산은 화소 도메인에서 이루어지는데 그 형태가 파워(power)의 계산과 동일하므로 Parseval의 원리(Parseval's theorem)에 의해 주파수 도메인에서 계산한 MSE는 화소 도메인에서 계산한 값과 동일하다. 많은 영상 부호화기에서 사용되는 DCT는 정규화된 직교 변환이므로 DCT 역시 Parseval의 원리가 적용가능하며 DCT 도메인에서 계산한 MSE는 화소 도메인에서 계산한 것과 같은 값을 가진다. DCT 도메인에서 PSNR을 계산하는 과정은 다음과 같다.

$$MSQE = \frac{1}{mn} \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} |S(u,v) - P(u,v)|^2$$

$$PSNR = 10 \log_{10} \left(\frac{(2^{depth} - 1)^2}{MSQE} \right) \quad (2)$$

식 (2)의 $S(u,v)$ 은 원 영상의 DCT 계수이고 $P(u,v)$ 은 부호화 된 영상의 DCT 계수를 의미한다. Mean square quantization error (MSQE)는 DCT 계수에 적용되는 양자화에 의한 오차를 나타내는데 Parseval의 원리에 의해 화소 도메인의 MSE와 동일한 물리량을 가지며 DCT 계수의 분포 특성을 고려하여 다음과 같이 계산할 수 있다.

$$MSQE = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \left(\sum_i \int_{a_i}^{b_i} (w - q_i)^2 p_{(u,v)}(w) dw \right) \quad (3)$$

N 은 블록의 크기, a_i 와 b_i 는 양자화 구간을 의미하며 q_i 는 양자화 구간의 대푯값이고 $p_{(u,v)}(w)$ 는 (u,v) 위치에 해당하는 원 영상의 DCT 계수에 대한 확률 밀도 함수이다. 식 (3)은 원 영상의 DCT 계수가 분포하는 특성인 확률 밀

도 함수를 적용하였기에 식 (2)와 달리 PSNR을 계산하기 위해 원 영상의 모든 정보가 필요하지 않으며 DCT 계수에 대한 확률적 분포의 정보만 획득하면 된다. 이러한 DCT 계수의 확률적 분포에 기반을 둔 PSNR 예측 기법의 계산 과정을 정리하면 다음과 같다.

- 1) 부호화된 비트열을 복호화하여 블록의 크기, 양자화 구간, 양자화 대푯값의 부호화 정보를 얻는다.
- 2) 양자화된 DCT 계수들을 복원하고 이를 이용하여 원 영상의 DCT 계수 분포를 예측한다.
- 3) 예측된 원 영상의 DCT 계수 분포로부터 MSQE를 계산한다.
- 4) MSQE 값으로 PSNR을 계산한다.

첫 번째 과정에서 블록의 크기, 양자화 구간, 양자화 대푯값을 얻는 과정은 부호화기의 종류에 따라 달라지는데 H.264/AVC 기반의 부호화기는 식 (4)처럼 정의된다^[10].

$$a_0 = 0$$

$$a_i = \text{sign}(i) \cdot (|i| - 1 + \alpha) QS \quad \text{for } |i| = 1 \cdots \infty$$

$$b_i = \text{sign}(i) \cdot (|i| + \alpha) QS \quad \text{for } |i| = 0 \cdots \infty \quad (4)$$

$$q_i = i \cdot QS$$

$$QS = c \cdot 2^{(QP \bmod 6)}$$

QS 는 양자화 단계의 크기로 양자화 파라미터에 따라 지수 규모로 증가하며 계수 c 는 미리 정의되어 있는 값이다. α 는 양자화 후에 DCT 계수들이 0으로 결정되는 구간을 조절하는 변수로 H.264/AVC 참조 소프트웨어에서는 화면 내 부호화에서 2/3, 화면 간 부호화에서 5/6으로 설정되어 있다^[4].

두 번째 과정에서 양자화된 DCT 계수를 통해 원 영상의 DCT 계수 분포를 예측하는데, 이 과정이 PSNR 예측 알고리즘에서 가장 핵심적인 부분이다. 일반적으로 각 DCT 계수들은 그림 1과 같이 변환 도메인 상의 위치에 따라 분포 특성이 다르기 때문에 각 DCT 계수 위치마다 분포를 모델링하며 모델링 함수는 주로 라플라시안 (Laplacian) 모델이나 코시 (Cauchy) 모델이 사용된다. 두 모델링 함수는 DCT 계수의 분포를 표현하기에 적합한 것으로 알려져 있으며^[11]

그림 1. 변환 도메인 상에서 위치에 따른 DCT 계수의 히스토그램
Fig. 1. Histogram of DCT coefficients in transform domain

적분된 함수 형태도 단순하여 수치 해석적인 접근이 용이하다. 두 모델의 수식은 다음과 같다.

$$p(w) = \frac{1}{2\sigma} e^{-\left|\frac{w}{\sigma}\right|} \quad (5)$$

$$p(w) = \frac{1}{\pi} \frac{\mu}{\mu^2 + w^2} \quad (6)$$

식 (5)는 라플라시안 모델의 수식이고 식 (6)은 코시 모델의 수식이다. 두 모델의 형태를 결정하는 변수 σ 와 μ 를 얻기 위해 몇 가지 방법들이 제안되었으나^{[6][7]} 0으로 양자화된 DCT 계수의 비율을 이용하는 방법이^[7] 복잡도가 낮으면서도 높은 정확도를 보이기 때문에 널리 이용된다. 그림 2와 같이 양자화 후에 0의 값을 가지는 DCT 계수들은 양자화가 적용되기 전 단계에서 α_{-1} 부터 α_1 의 구간에

존재한다. 따라서 이 구간 내의 확률 밀도 함수를 적분한 값은 전체 DCT 계수가운데 0으로 양자화 되는 계수의 비율과 동일한 값을 가진다. 이를 수식으로 표현하면 다음과 같다.

$$\frac{N_{zero}}{N} = \int_{\alpha_{-1}}^{\alpha_1} \frac{1}{2\sigma} e^{-\left|\frac{w}{\sigma}\right|} dw = \int_0^{\alpha_1} \frac{1}{\sigma} e^{-\frac{w}{\sigma}} dw$$

$$\sigma = -\frac{\alpha \cdot QS}{\ln\left(1 - \frac{N_{zero}}{N}\right)} \quad (7)$$

$$\frac{N_{zero}}{N} = \int_{\alpha_{-1}}^{\alpha_1} \frac{1}{\pi} \frac{\mu}{\mu^2 + w^2} dw = \frac{2}{\pi} \tan^{-1}\left(\frac{\alpha_1}{\mu}\right)$$

$$\mu = \frac{\alpha \cdot QS}{\tan\left(\frac{\pi}{2} \frac{N_{zero}}{N}\right)} \quad (8)$$

그림 2. 양자화되기 전 DCT 계수의 확률 밀도 함수
Fig. 2. probability density function of DCT coefficients before quantization

N_{zero} 와 N 은 0으로 양자화된 DCT 계수의 개수와 전체 DCT 계수의 개수를 의미하며 a_1 은 식 (4)에 따라 $\alpha \cdot QS$ 로 대체될 수 있다. 식 (7)은 식 (5)의 라플라시안 모델의 수식을 적용하여 0으로 양자화된 DCT 계수의 비율로부터 라플라시안 모델의 계수 σ 를 얻는 과정이고 식 (8)은 코시 모델을 적용하여 0으로 양자화된 DCT 계수의 비율을 통해 μ 를 구하는 과정이다. 두 수식 모두 N_{zero}/N 과 $\alpha \cdot QS$ 만을 변수로 가지는데 $\alpha \cdot QS$ 는 부호화 타입과 양자화 파라미터 값에 의해 결정되므로 N_{zero}/N 만 계산하면 얻고자하는 DCT 계수의 확률 밀도 함수를 예측할 수 있다. 이렇게 예측한 확률 밀도 함수를 식 (3)에 적용하여 MSQE를 계산하면 식 (2)에 의해 PSNR 값을 얻는 것이 가능하다.

III. 계층적 B-픽처 구조를 고려한 PSNR 예측 기법

계층적 B-픽처 구조는 프레임율 확장성을 위해 H.264/AVC의 추가 확장 버전인 SVC에서 처음 적용된 부호화 구조로 그림 3과 같이 계층적으로 화면 간 부호화를 수행한다. 부호화 구조상 최하위 계층을 제외한 픽처는 양방향 예측으로 부호화되며 일반적으로 IBBP 구조에 비해 높은 성능을 보이기 때문에 H.264/AVC와 multi-view video coding (MVC)에 적용되었고 현재 표준화 진행 중인 HEVC의 성능 실험에도 사용된다^{[12][13]}. 따라서 본 장에서는 PSNR 예측 알고리즘의 범용적인 활용을 위해 계층적 B-픽처 구조가 적용된 부호화 결과에 적용될 수 있는 기법을 제안한다.

계층적 B-픽처 구조는 대부분의 픽처가 양방향 예측으로 부호화된다. B-픽처는 두 방향의 참조 영상 리스트를 가지고 움직임 보상을 수행하기 때문에 차영상이 적게 발생되어 I-픽처나 P-픽처에 비해 적은 비트로 부호화 가능하다. 특히 상위 계층의 픽처나 저 비트율로 부호화된 픽처는 대부분의 블록이 스킵 모드로 선택되는데 스킵 모드의 경우 차영상에 대한 어떠한 정보도 부호화되지 않고 명시적인 방법으로 주변의 정보를 이용하여 움직임 벡터를 추측하고 움직임 보상을 통해 복호화된다. 이러한 특성을 가지는 계층적 B-픽처 구조를 위해 최하위 계층인 I-픽처의 DCT 계수 모델링 방법을 개선하고 상위 계층인 B-픽처에서 주로

그림 3. 계층적 양방향 예측 픽처 구조의 예
Fig. 3. Example of hierarchical B-picture structure

선택되는 스킵 모드를 고려하여 전체 PSNR 예측 알고리즘의 정확도를 높인다.

1. I-픽처를 위한 DCT 계수 모델링

DCT 계수 분포의 모델링을 위해 사용되는 라플라시안 모델과 코시 모델은 모두 표현하고자 하는 DCT 계수의 분포를 잘 나타내지만 AC계수의 표현에는 코시 모델이 더

적합한 것으로 알려져 있다^[14]. PSNR 예측에서도 코시 모델을 사용할 때 더 정확한 값을 예측하지만^[8] DCT 계수의 위치 혹은 영상의 특성에 따라서 두 모델의 성능이 바뀌는 경우가 발생하기도 한다. 그림 4는 식 (7)과 (8)을 book arrival 영상에 적용하여 예측한 값과 실제 DCT 계수의 분포의 차이를 보여준다. 그림 4(a)에는 QP 40일 때 양자화 되기 전 DC 값의 분포와 모델링 기법을 통해 예측한 결과가 나타나있는데 이 경우 라플라시안 모델이 코시 모델에 비

(a) QP40, 1번째 계수, 0으로 양자화된 계수의 비율: 0.934

(a) QP40, 1st coeff., Ratio of DCT coeff. quantized to zero: 0.934

(b) QP30, 2번째 계수, 0으로 양자화된 계수의 비율: 0.938

(b) QP30, 2nd coeff., Ratio of DCT coeff. quantized to zero: 0.938

그림 4. book arrival 영상에 적용된 DCT 계수 모델링의 예

Fig. 4. Example of modeling for DCT coefficients on 'book arrival' sequence

해 더 정확한 예측을 하고 있다는 사실을 확인할 수 있으며 그림 4 (b)는 QP30의 2번째 계수의 분포에 대한 결과를 보여주는데 코시 모델이 더 실제 DCT 계수 분포에 가깝게 예측했음을 확인할 수 있다. 따라서 두 모델링 방식 가운데 원 영상의 DCT 계수 분포에 가까운 방식을 선택할 수 있다면 좀 더 정확한 PSNR 예측이 가능하다.

먼저 각 모델링 기법의 특징을 파악하기 위해 두 방식을 비교하였다. 그림 5는 14개 실험영상에 대하여 20, 30, 40의 QP 값을 적용하여 얻은 N_{zero}/N 에 대해 두 모델링 방식 사이의 차이를 보여준다. N_{zero}/N 값이 0.6보다 작은 경우 두 모델링 기법에 의해 표현되는 값의 차이가 작지만 그 보다 큰 구간에서는 N_{zero}/N 이 증가함에 따라 지수함

그림 5. 0으로 양자화된 계수 비율에 따른 두 모델링 방식의 차이
Fig. 5. Difference between two models based on the ratio of DCT coeff. quantized to zero

표 2. 원 영상의 DCT 계수 분포와 두 모델링 방식의 오차
Table 2. Error between two modeling methods and original DCT coeff. distribution

0으로 양자화된 계수의 비율	원 영상의 DCT 계수 분포와의 오차	
	라플라시안 모델	코시 모델
0~0.6	0.371	0.348
0.6~0.7	0.544	0.496
0.7~0.8	0.492	0.472
0.8~0.9	0.454	0.457
0.9~0.95	0.535	0.516
0.95~1	0.680	0.506

수의 형태로 차이가 증가하며 1에 가까운 구간에서는 큰 차이를 보인다. 표 2는 두 모델링 방식의 차이를 수치적으로 보여주는데 N_{zero}/N 값에 따라 14개 영상이 가지는 DCT 분포 계수와 모델링에 의해 예측한 분포의 차이를 나타내었다. 대부분의 구간에서는 평균적으로 비슷한 오차를 보이지만 N_{zero}/N 값이 0.95에서 1 사이의 값을 가질 때에는 코시 모델이 더 적은 오차를 보인다. 이는 모델링 방식의 특징에 의한 현상으로 라플라시안 모델이 코시 모델에 비해 함수의 중심 부분에서 벗어날수록 감소하는 기울기가 가파르기 때문에^[14] 동일한 N_{zero}/N 값에서 유도될 경우 그림 4와 같이 코시 모델에 비해 낮은 높이의 모델이 획득된다. 따라서 라플라시안 모델은 높은 높이를 가지는 함수를 표현하기 어려우며 DCT 계수들이 대부분 0 근처에 집중되어 함수의 높이가 높은 0.95~1의 구간에서는 코시 모델을 선택하는 것이 효율적이다. 이것은 모델링 기법 선택의 첫 번째 기준으로 적용된다.

N_{zero}/N 값이 1에 근접하지 않은 영역에서는 표 2에서 볼 수 있듯이 두 모델링 방식이 서로 비슷한 성능을 보이므로 각 영상과 계수의 위치에 따라 원 영상의 DCT 계수 분포에 근접한 방식을 선택하기 위해 다른 조건을 사용하였다. 식 (9)와 같이 DCT 계수 w 의 제곱에 확률을 곱한 형태로 표현되는 DCT 계수의 분산은 양자화 된 후에도 양자화 전의 분산 값과 높은 상관도를 가지며 양자화 전과 후 분산

그림 6. 원 신호의 분산과 양자화 후의 분산 값의 비율 변화
Fig. 6. The ratio of variances of original and quantized signal

값의 차이는 양자화 스텝 사이즈에 비례하는 특성을 지닌다. 영상 특성과 계수의 순서 역시 양자화 전후의 분산 값의 차이에 영향을 미치지만 N_{zero}/N 값이 1에 근접하는 경우를 제외하면 두 분산 값의 비율은 그림 6과 같은 형태를 보이며 상수 혹은 변수를 사용하여 양자화 된 계수의 분산과 원 영상의 분산의 관계를 일반화 할 수 있다. 결국 양자화 된 DCT 계수의 분산 값을 통해 원 DCT 계수의 분산 값을 예측할 수 있으며 이 값과 모델링으로 얻은 함수의 분산을 비교하여 더 정확한 모델링 방법을 선택할 수 있다.

$$var = \int |w|^2 p(w) dw \quad (9)$$

그림 7은 제안 방법을 이용하여 화면 내 부호화 픽처에서

그림 7. 화면 내 부호화 픽처를 위한 MSQE 계산의 순서도
Fig. 7. Flowchart of the MSQE calculation method for I-picture

MSQE를 계산하는 과정을 나타낸 순서도이다. 먼저 비트열을 입력받으면 일반적인 동영상 복호화기에서 거치는 과정과 동일하게 엔트로피 부호화된 데이터로부터 H.264의 고준위 문법(high level syntax)을 따라 파싱을 수행하여 양자화 구간과 DCT의 크기, 양자화 대푯값 등의 데이터를 얻는다. 이후 복호화를 수행하여 DCT 계수들을 복원하고 각 계수의 위치에 맞게 0으로 양자화된 DCT 계수의 비율을 계산한다. 0으로 양자화된 DCT 계수의 비율이 미리 정의된 문턱 값보다 클 경우에는 코시 모델을 적용하는데 문턱 값은 1에 가까운 값을 가진다. 0으로 양자화된 DCT계수의 비율이 문턱 값보다 작은 경우에는 양자화된 계수의 분산을 통해 원영상의 분산을 예측하고 이 예측 값과 모델링을 통해 얻은 DCT계수 분포의 분산 값을 비교하여 적은 차이를 보이는 모델링 기법을 선택하여 MSQE를 계산한다. 모든 계수들에 대한 MSQE를 얻으면 전체 계수에 대한 평균 MSQE를 계산하고 식 (2)를 통해 PSNR을 얻는다.

2. 스킵 모드를 고려한 PSNR 예측

스킵 모드로 부호화된 블록에서는 차영상의 DCT 계수에 대한 정보가 전혀 존재하지 않으므로 DCT 계수 모델링에 의한 예측방법을 사용할 수 없다. 또한 스킵 모드는 화면 내 부호화가 아닌 움직임 벡터에 의한 움직임 보상으로 복호화가 이루어지므로^[15] 동일한 픽처 내에 존재하는 정보만으로는 스킵 모드로 부호화된 블록의 MSQE를 예측하기 어렵다. 따라서 제안 방식에서는 스킵 모드로 부호화된 블록의 MSQE를 다음과 같이 참조 영상의 MSQE를 사용하여 예측한다.

$$MSQE_{skip} = (\alpha MSQE_{ref0} + \beta MSQE_{ref1}) / (\alpha + \beta) \quad (10)$$

$MSQE_{ref0}$ 과 $MSQE_{ref1}$ 은 참조 리스트0과 참조 리스트1에서 현재 블록이 참조한 픽처의 MSQE 값을 의미하고 α 와 β 는 참조 리스트의 사용유무를 나타낸다. 해당되는 참조 리스트가 사용되면 1 그렇지 않으면 0의 값을 가진다. 식 (10)과 같이 참조 영상의 MSQE를 통해 예측하는 방

법은 두 가지 부분에서 잠재적인 문제의 소지가 있다. 첫째로 참조 영상 전체의 MSQE 값이 스킵 모드로 부호화된 어떤 블록의 MSQE를 대표할 수 있는가의 측면인데 제안된 방식 자체가 통계적 특성을 기반으로 이루어지므로 참조 영상 내 특정 위치의 MSQE를 얻기 어렵고 B-픽처 내에서 스킵 모드로 선택되는 블록의 비중이 표 3처럼 65%에서 90%에 이르기 때문에 스킵 모드로 부호화된 블록들의 평균 MSQE 값은 통계적으로 보았을 때 참조 영상의 MSQE 값을 따른다고 할 수 있다. 두 번째 문제의 소지는 하위 계층의 PSNR 예측 과정에서 오차가 발생하면 상위 계층으로 전파된다는 점인데 알고리즘의 특성상 이 부분은 피할 수 없으며 하위 계층에서 오차의 발생을 줄이기 위해 최하위 계층으로 적용되는 I-픽처에 사용되는 모델링 방법을 개선하여 전파되는 오차를 줄이고자 하였다.

표 3. QP값 변화에 따른 스킵 모드의 선택 비율 (kendo영상)
Table 3. The ratio of skip mode according to QP values (kendo seq.)

QP값	스킵 모드의 선택 비율 (%)
20	64.7
26	78.3
32	84.7
40	89.8

그림 8은 계층적 B-픽처 구조에서 B-픽처의 블록이 가지는 MSQE를 계산하는 과정을 보여준다. 비트열과 참조 영상의 MSQE가 입력되면 비트열을 파싱하여 양자화 구간과 변환 크기, 양자화 대푯값과 부호화 모드를 얻는다. B-픽처의 경우 블록이 화면 간 부호화 모드와 화면 내 부호화 모드, 스킵 모드로 부호화 가능하므로 각 모드에 따라서 예측 과정이 다르다. 스킵 모드가 아닌 경우에는 I-픽처의 PSNR 예측과 유사한 과정을 거치는데 B-픽처의 경우 효율적인 움직임 보상에 의해 차영상의 크기가 작아지므로 저 비트율에 적합한 코시 모델을 사용하여 MSQE를 계산한다. 블록이 스킵 모드로 부호화 되었다면 참조 리스트와 참조 영상을 추론하여 식 (9)를 통해 MSQE를 얻을 수 있다. 모든

블록에 대해 MSQE를 획득하면 I-픽처와 동일하게 전체 픽처에 대한 MSQE를 계산하고 이를 이용하여 PSNR을 예측한다. 그리고 현재 픽처가 최상위 계층이 아닌 경우에는 다른 픽처의 참조 영상으로 사용될 수 있으므로 현재 픽처의 MSQE 값을 저장한다.

그림 8. 스킵 모드를 고려한 양방향 예측 픽처의 MSQE 계산의 순서도
Fig. 8. Flowchart of the MSQE calculation method considering skip mode for B-picture

IV. 실험 결과

제안 알고리즘의 성능을 확인하기 위해 실험 영상에 여러 양자화 파라미터를 적용하여 부호화를 수행하고, 생성된 비트열로부터 예측된 PSNR 값과 실제 PSNR 값을 비교하였다. 실험에 사용된 영상은 14개로 표 4와 그림 9에서 확인할 수 있다^[16]. 실험 영상을 부호화하기 위해 Joint Multi-view Video Coding (JMVC) 8.3을 사용하였고 블록

그림 9. 실험 영상 (위: Balloons, Ballroom, Book arrival, Kendo, Newspaper, City, Crew, 아래: Foreman, Football, Harbour, Soccer, Ice, Parkrun, Stockholm)

Fig. 9. Video sequences for simulation (Top: Balloons, Ballroom, Book arrival, Kendo, Newspaper, City, Crew, Bottom: Foreman, Football, Harbour, Soccer, Ice, Parkrun, Stockholm)

표 4. 실험 영상의 정보

Table 4. Information of video sequence for simulation

영상	해상도	영상	해상도
Balloons	1024×768	Foreman	352×288
Ballroom	640×480	Football	352×288
Book arrival	1024×768	Harbour	704×576
Kendo	1024×768	Soccer	704×576
Newspaper	1024×768	Ice	704×576
City	704×576	Parkrun	1280×720
Crew	704×576	Stockholm	1280×720

표 5. 부호화 파라미터

Table 5. Coding parameter

항목	값	비고
GOP 길이	15	4개 계층이 존재
QP 값	20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40	
블록화 방지 필터	사용안함	
탐색 범위	64	
탐색 모드	고속 탐색	TZ 탐색

화 방지 필터는 적용시키지 않았다. 세부적인 부호화 파라미터는 표 5에 나타나있다.

I-픽처의 DCT 계수 모델링 알고리즘에 대한 성능 평가를

위해 각 실험 영상에 대해 화면 내 부호화를 수행하였고 생성된 비트열로부터 PSNR 값을 예측하여 실제 값과 비교하였다. 예측 과정에서 사용된 문턱 값은 0.97을 실험적으로 선택하였다. 그림 10은 예측한 PSNR과 실제 값의 분포도를 보여준다. 라플라시안 모델의 경우 대체적으로 실제 PSNR에 비해 낮게 예측하고 있으며 코시모델은 반대로 실제 값에 비해 높게 PSNR 값을 예측한다. 제안 알고리즘의 경우 특정 영상의 고 비트율에서 약간의 오차가 발생하는 것을 제외하면 전체적으로 실제 값에 가까운 예측이 이루어짐을 확인할 수 있다.

표 6은 각 모델링 방법에 따른 PSNR 예측 값과 실제 값의 상관도와 Root Mean Square Error (RMSE)를 나타낸다. 세 방식 모두 I-픽처의 PSNR 예측에 있어서 0.9 이상의 높은 상관도를 보이지만 RMSE 값은 차이를 보이며 제안 알고리즘이 가장 작은 RMSE 값을 가지는데 이는 제안 알고리즘이 실제 PSNR과 유사하게 예측한다는 것을 방증한다.

표 6. 화면 내 부호화 픽처의 예측된 PSNR과 실제 PSNR RMSE와 상관도
Table 6. RMSE and correlation between estimated PSNR and original PSNR for I pictures

	Laplacian	Cauchy	Proposed method
RMSE	1.97	1.56	1.13
상관도	0.96	0.97	0.97

그림 10. 화면 내 부호화 픽처에 적용된 제안 알고리즘의 PSNR 예측 분포도

Fig. 10. Scatter plot of the proposed algorithm for I-pictures (a) Laplacian model (b) Cauchy model (c) Proposed method

또 다른 제안 알고리즘인 스킵 모드를 고려한 PSNR 예측 기법의 성능을 확인하기 위해 실험 영상에 대한 Group Of Picture (GOP) 길이를 15로 설정하여 부호화 하고 PSNR 값을 예측하여 실제 값과 비교하였다. 그림 11은 스킵 모드를 고려한 PSNR 예측 값과 실제 값의 분포도를 나타내는데 전체적으로 예측 값과 실제 값이 유사한 경향성을 보임을 확인할 수 있다. 표 7은 각 영상에 대한 예측 된

PSNR과 실제 값 사이의 상관도와 RMSE를 보여주는데 RMSE 측면에서 관찰하면 전체적으로 비슷한 값을 보이지만 움직임이 빠른 kendo, soccer, ice 영상에서는 예측의 성능이 상대적으로 낮음을 확인할 수 있다. 또한 해상도가 작은 영상에서는 대체적으로 RMSE 값이 크게 나타남을 확인할 수 있다. 반면 화면 내 움직임이 적은 newspaper 영상에 대해서는 정확한 PSNR 예측이 이루어짐을 알 수 있다.

그림 11. 스킵 모드를 고려한 PSNR 예측 값과 실제 PSNR의 분포도

Fig. 11. Scatter plot of the proposed algorithm considering skip mode for B-picture

표 7. 스킵모드를 고려한 제안 알고리즘으로 예측한 PSNR과 실제 PSNR의 RMSE와 상관도

Table 7. RMSE and correlation between estimated PSNR by the proposed algorithm and original PSNR for hierarchical B-picture structure

영상	RMSE	상관도	영상	RMSE	상관도
Balloons	0.798	0.998	Foreman	3.271	0.930
Ballroom	0.672	0.982	Football	2.780	0.977
Book arrival	0.676	0.988	Harbour	0.615	0.990
Kendo	1.562	0.989	Soccer	1.226	0.966
Newspaper	0.337	0.995	Ice	3.107	0.940
City	0.795	0.986	Parkrun	0.975	0.976
Crew	0.775	0.974	Stockholm	1.526	0.963
전체	0.948	1.660			

V. 결 론

본 논문에서는 계층적 B-픽처 구조를 고려한 H.264/AVC 비트열 기반의 PSNR 예측 기법을 제안하였다. 제안된 알고리즘은 원 영상의 DCT 계수의 분포 특성을 모델링하여 PSNR을 계산하는 방법에 기반을 두고 새로운 DCT 계수의 모델링 방법과 스킵 모드를 고려한 PSNR 예측 방법으로 구성되어 있다. 제안된 DCT 계수 분포의 모델링 방법은 기존에 사용되는 모델을 상황에 맞게 복합적으로 사용하여 예측의 정확도를 향상 시켰으며 스킵 모드를 고려한 예측 방법은 참조 영상의 MSQE 값을 사용하여 스킵 모드로 부호화 된 블록의 MSQE를 예측하고 픽처 전체의 PSNR을 계산하였다. 제안 알고리즘의 성능을 확인하기 위하여 I-픽처로 부호화한 비트열과 계층적 B-픽처 구조를 적용하여 부호화한 비트열에 대해 PSNR을 예측하고 실제 값과 비교하였다. 실험 결과 제안된 DCT 계수 모델링 방법은 기존의 방법보다 높은 정확도를 보였으며 스킵 모드를 고려한 예측 기법은 DCT 계수가 존재하지 않는 상황에서도 비교적 높은 정확도의 결과를 보임을 확인하였다.

이 같은 연구 결과를 기반으로 하여 양자화된 DCT계수가 모두 0이 될 때 원 영상의 DCT 계수 분포를 정확하게 예측하는 방법에 대한 추가적인 연구와 계층적 B-픽처 구조를 사용하는 SVC와 Multi-view Video Coding (MVC)로의 확장이 이루어진다면 PSNR예측 알고리즘은 실용적으로 널리 사용될 수 있을 것으로 기대된다.

참 고 문 헌

- [1] <http://www.its.bldrdoc.gov/vqeg/>
- [2] B. Furht, O. Marques, Handbook of Video Databases: Design and Applications, CRC Press, 1041-1078, 2003
- [3] Z. Wang, A. Bovik, Modern Image Quality Assessment, Morgan & Claypool Publishers, 2006
- [4] ITU-T, Advanced video coding for generic audiovisual services, ITU-T Recommendation, 2005
- [5] D. S. Turaga, Y. Chen, J. Caviedes, No reference PSNR estimation for compressed pictures, Signal Process.: Image commun., 19, 2, 174-184, Feb., 2004
- [6] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nichida, E. Nakasu, A method of estimating Coding PSNR using quantized DCT coefficients, IEEE Trans. Circuits Syst. Video Technol., 16, 2, 251-259, Feb., 2006
- [7] A. Eden, No-reference estimation of the coding PSNR for H.264-coded sequences, IEEE Trans. Consumer Electronics, 53, 2, 667-674, May, 2007
- [8] S. Shim, J. Moon, J. Han, PSNR estimation scheme using coefficient distribution of frequency domain in H.264 decoder, Electronics Letters, 44, 2, 108 - 110, Jan., 2008
- [9] H. Schwarz, D. Marpe, T. Wiegand, Analysis of Hierarchical B Pictures and MCTF, IEEE International Conference on Multimedia and Expo, 2006, Toronto, 1929 - 1932, July, 2006
- [10] <http://www.vcodex.com/h264.html>
- [11] E. Lam, J. Goodman, A mathematical analysis of the DCT coefficient distributions for images, IEEE Trans. Image Process., 9, 10, 1661-1666, Oct., 2000.
- [12] P. Merkle, A. Smolic, K. Muller, T. Wiegand, Efficient Prediction Structures for Multiview Video Coding, IEEE Trans. Circuits Syst. Video Technol., 17, 11, 1461 - 1473, Nov., 2007
- [13] G. Sullivan, J. Ohm, Recent developments in standardization of high efficiency video coding (HEVC), SPIE Applications of Digital Image, 7798, 30-37, Aug., 2010
- [14] Y. Altunbasak, N. Kamaci, An analysis of the dct coefficient distribution with the H.264 video coder, IEEE International Conference on Acoustics, Speech, and Signal Processing 2004, vol.3, iii - 177-80, May 2004
- [15] Y. Peng, H. Tourapis, A. Tourapis, J. Boyce, Fast mode decision and motion estimation for JVT/H.264, International Conf. on Image Process. 2003, III-853-856 vol.2, Sept., 2003
- [16] MPEG, Call for Proposals on 3D Video Coding Technology, ISO/IEC JTC1 SC29 WG11, 2011

저 자 소 개

서 정 동

- 2004년 2월 : 연세대학교 전기전자공학부 학사
- 2006년 2월 : 연세대학교 전기전자공학과 석사
- 2006년 3월 ~ 현재 : 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 3차원 동영상 부호화, H.264/AVC, 영상 화질 평가

손 광 훈

- 1983년 : 연세대학교 전자공학과 공학사
- 1985년 : University of Minnesota, MSSE
- 1991년 : North Carolina State University, Ph.D
- 1993년 : 한국전자통신연구원 선임연구원
- 1994년 : Georgetown University, Post-doctor fellow
- 1995년 ~ 현재 : 연세대학교 전기전자공학부 교수
- 2002년 9월 ~ 2003년 8월 : Nanyang Technological University, Visiting Professor
- 주관심분야 : 3차원 영상처리, 영상 통신