

계층적 부호화 구조의 H.264/AVC에서 효율적인 비트율 제어를 위한 향상된 비트율-양자화 모델링

*서찬원 *한중기

*세종대학교

*sachownon@gmail.com, *hjk@sejong.ac.kr

Improved R-Q model for rate control in hierarchical coding structure of H.264/AVC

*Seo, Chan-Won *Han, Jong-Ki

*Sejong University

요약

계층적 부호화 구조는 H.264/AVC의 부호화 효율을 최대화하고 시간적 스케일러빌리티를 지원하는 등 다양한 장점을 제공한다. 이러한 구조에서 비트 할당 및 비트율 제어 기술은 비디오 코덱의 성능을 향상시킬 수 있는 중요한 요소들이다. 기존의 비트율 제어 기술에서는 계층적 구조의 특성을 고려하지 못한 비트율-양자화 모델을 사용하여 비디오 코덱의 성능을 최적화하기에 어려움이 있었다. 따라서, 본 논문에서는 계층적 구조에서도 효율적으로 비트 할당 및 비트율 제어를 할 수 있도록, 기존보다 향상된 비트율-양자화 모델을 제안한다. 그리고 실험을 통하여 제안하는 비트율-양자화 모델이 기존 기술에 비해 정확함을 보인다.

1. 서론

ISO/IEC MPEG와 ITU-T은 JVT(Joint Video Team)을 통해 H.264/AVC을 표준화 하였다[1]. 계층적 부호화 구조는 H.264/AVC의 부호화 효율을 최대화하고 시간적 스케일러빌리티를 지원하는 등 다양한 장점을 제공한다.

계층적 구조에서 비트 할당 및 비트율 제어 기술은 비디오 코덱의 성능을 향상시킬 수 있는 중요한 요소이다. 이에 따라 많은 연구에서 효율적인 비트 할당과 비트율 제어기술들을 제안하였다[2][3]. [2]에서는 MAD(Mean Absolute Difference)을 이용하는 비트율-양자화 모델을 이용하여 비트 할당 및 비트율 제어기술을 제안하였다. 그러나 이 기술은 IPPP 구조에만 적합하도록 고안되어, [3]에서 계층적 부호화 구조에서도 적절하게 동작할 수 있도록 수정되었다. 기존 기술들에서 사용하는 비트율 양자화 모델은 (1)와 같다.

$$\frac{V_i}{MAD} = \frac{X_1}{Q} + \frac{X_2}{Q^2} \quad (1)$$

위의 수식에서, V_i 은 할당된 비트량, Q 은 양자화 크기, X_1 와 X_2 은 모델 파라미터를 나타낸다. MAD 은 현재 프레임과 참조 프레임의 차이신호에 절대값을 취하여 평균한 값이다.

기존 기술들에서 사용하는 비트율-양자화 모델은 MAD값에만 의존하여 비트율을 예측하고, 그 결과에 따라 양자화 파라미터를 결정한다. [3]에서는 P 프레임의 양자화 파라미터를 (1)을 이용하여 할당하고 I와 B 프레임의 양자화 파라미터들은 P 프레임들의 부호화 결과와 현재 프레임의 계층을 기준으로 결정함으로써 I와 B 프레임의 비트율 제어를 정확하게 수행하지 못한다. 이 때문에 기존의 기술을 계층적 부호화 구조에 적용하는 경우, 효율적인 비트 할당 및 정확한 비트율 제어를 하기 어렵다. 따라서 본 논문에서는 계층적 구조에서도 효율적으로 비트 할당 및 비트율 제어를 할 수 있도록 향상된 비트율-양자화 모델을 제안한다.

본 논문의 제 2장에서는 향상된 비트율-양자화 모델을 소개하고, 제 3장에서는 제안한 비트율-양자화 모델의 파라미터를 계산하는 방법을 설명한다. 제 4장에서는 제안한 비트율-양자화 모델을 사용한 비트율 제어 결과를 보이고, 마지막 제 5장에서 결론을 서술한다.

본 연구는 지식경제부 및 정보통신산업진흥원의 IT융합 고급인력과정 지원사업의 연구결과로 수행되었음 (NIPA-2010-C6150-1001-0013)

2. 향상된 비트율-양자화 모델

기존 기술에서 사용하는 (1)은 MAD 기반의 비트율-양자화 모델이다. H.264/AVC 부호화기에서 발생하는 비트량 중 많은 부분을 현재 프레임과 참조 프레임의 차이신호를 부호화 하는데 사용하고 있다. 차이신호는 크게 DC(저주파)신호와 AC(고주파)신호로 나누어 볼 수 있다. 그 중 MAD는 차이신호의 평균 절대값을 사용하여 DC신호가 발생시키는 비트량을 상대적으로 나타내고 있다. 본 논문에서는 DC신호뿐만 아니라, AC신호가 발생시키는 비트량을 상대적으로 반영하기 위하여 VOD(Variance Of Difference)값을 적용하여 새로운 비트율-양자화 모델을 제안한다.

$$Z_{VOD} = \sigma \cdot \frac{\sqrt{VOD}}{Q} + \frac{\sqrt{MAD}}{Q} \quad (2)$$

σ 은 제안하는 비트율-양자화 모델의 파라미터이다. H.264/AVC에서는 4x4 또는 8x8 블록단위의 주파수 변환을 사용하기 때문에, VOD값은 각 4x4 블록 단위로 계산하여 평균값을 계산한다.

실험을 통해 Z_{VOD} 와 프레임 단위로 발생한 비트량의 관계를 확인하면, 2차 함수 형태의 비선형 관계임을 확인할 수 있다. 비트 할당과 비트율 제어기술에서 보다 간단한 연산을 이용하여 비트 발생량을 예측할 수 있도록 제곱근을 사용하여 (2)를 선형으로 변환한다.

$$\sqrt{Z_{VOD}} = \sqrt{\sigma \cdot \frac{\sqrt{VOD}}{Q} + \frac{\sqrt{MAD}}{Q}} \quad (3)$$

(3)의 $\sqrt{Z_{VOD}}$ 와 프레임단위로 발생한 비트의 관계는 그림 1과 같이 선형임을 확인할 수 있다. 발생비트 S와 $\sqrt{Z_{VOD}}$ 의 선형 관계를 이용하여 발생 비트량 S값은 다음과 같이 예측 할 수 있다.

$$S' = \xi_1 \times \sqrt{Z_{VOD}} + \xi_2 \quad (4)$$

S' 은 발생 비트량 S의 예측값을 나타내며, ξ_1 와 ξ_2 은 선형 모델의 파라미터이다.

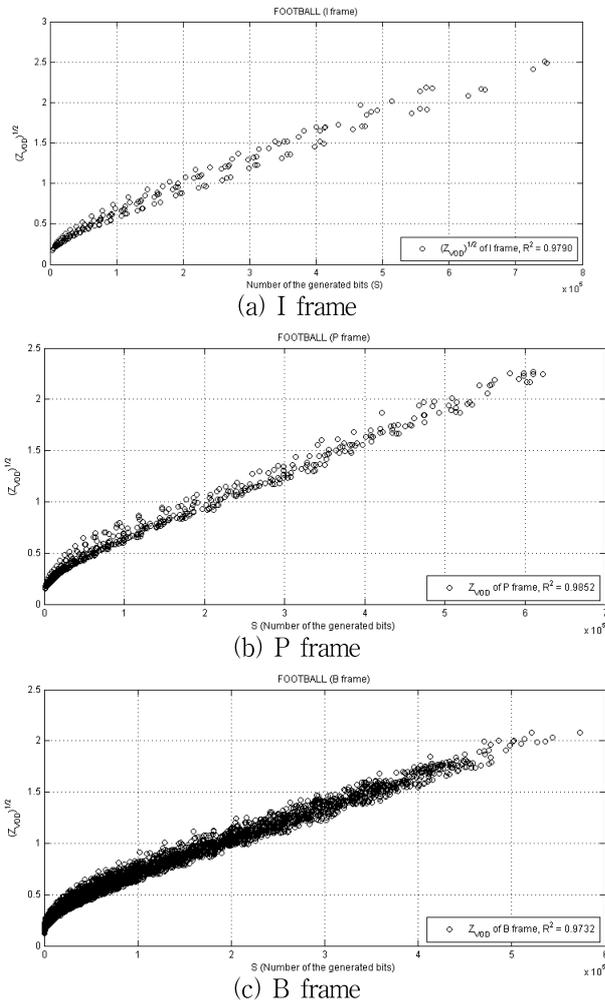


그림 1. \sqrt{VOD} 와 발생비트 S의 관계

3. 향상된 비트율-양자화 모델의 적용 방법

제 2장에서 제안한 비트율-양자화 모델을 이용하여 양자화 파라미터를 결정하기 위해서는 실제 발생비트 S와 예측한 발생 비트량 S'의 오차인 수식 (4)을 최소화하는 파라미터 σ 와 ξ_1 , ξ_2 를 각각 계산해야 한다.

$$E_1 = \sum_{i=0}^{N-1} \left[S_i - \left\{ \xi_1 \times \sqrt{\sigma \cdot \frac{\sqrt{VOD}_i}{Q_i} + \frac{\sqrt{MAD}_i}{Q_i}} + \xi_2 \right\} \right]^2 \quad (5)$$

예측 오차를 최소화하기 위해서는 E_1 을 각 파라미터 σ 와 ξ_1 , ξ_2 에 대하여 편미분을 수행하여야 한다. 그러나 이중 제곱근 때문에 편미분이 쉽지 않다. 따라서 총 3단계에 걸쳐서 각 파라미터를 최적화한다.

첫 번째 단계에서는 σ 의 값을 고정된 상태에서 ξ_1 와 ξ_2 의 최적화된 값을 계산하고, 두 번째 단계에서는 ξ_1 와 ξ_2 의 계산된 값을 이용하여 σ 를 계산한다. 이 때, 이중 제곱근 내부에 존재하는 σ 를 계산하기 위하여 오차 수식을 다음과 같이 수정하여 계산한다.

$$E_2 = \sum_{i=0}^{N-1} \left[\left(\frac{S_i - \xi_2}{\xi_1} \right)^2 - \left(\sigma \cdot \frac{\sqrt{VOD}_i}{Q_i} + \frac{\sqrt{MAD}_i}{Q_i} \right) \right]^2 \quad (6)$$

마지막으로 ξ_1 와 ξ_2 은 두 번째 단계에서 계산된 σ 를 이용하여 최적화를 수행한다.

각 파라미터가 최적화된 후, (4)에서 S'에 목표 비트량을 대입하여

Q를 계산한다. 그리고 계산된 Q에 대응하는 양자화 파라미터를 결정하여 현재 프레임에 할당한다.

4. 실험 결과

본 실험에서는 H.264/AVC의 참조 소프트웨어인 JM16.1[4]을 사용하였다. 테스트 시퀀스는 CIF(352x288) 해상도의 Soccer, City, Ice, Crew를 사용하였으며, 기존 방법은 [2]에서 제안한 기술과, [3]에서 제안한 RC_MODE_2와 RC_MODE_3을 적용하여 제안하는 비트율-양자화 모델의 효율성을 확인한다. GOP(Group Of Picture) 크기는 4와 8로 설정하여 다양한 계층의 부호화 구조를 실험하였다. 비트율 제어는 매크로 블록 단위로 양자화 파라미터를 결정하는 방법을 사용하였으며, 그 방법은 프레임 단위 계산방법과 같은 방법을 사용하였다.

각 기술의 성능을 평가하기 위하여 (7)와 같이 프레임의 목표 비트량과 실제 발생한 비트량의 차이를 측정하였다. 각 프레임의 목표 비트량은 I와 P, B1, B2, B3에 대하여 각 10와 7, 5, 3, 2 kbits을 할당하였다.

$$e_r = \sum_{i=1}^{N_{data}} |T_i - S_i| / \sum_{i=1}^{N_{data}} T_i \quad (7)$$

T은 프레임의 목표 비트량을 나타낸다.

(7)의 오차율을 근거로한 실험결과는 표 1와 같다.

표 1. 각 기술의 비트율 제어 정확도 측정 결과

Seq.	HGOP Size	[2]	[3]		Proposed scheme
			RC2	RC3	
Soccer	4	0.3599	0.3959	0.3115	0.1593
	8	0.5656	0.6637	0.8988	0.2118
City	4	0.6349	0.7265	0.7040	0.1965
	8	0.8708	0.8430	1.0950	0.2066
Ice	4	0.3567	0.3651	0.7378	0.1586
	8	0.7377	0.6747	1.4927	0.2413
Crew	4	0.5966	0.6388	0.6288	0.2430
	8	0.8852	1.0581	0.9334	0.3080

표 1은 (7)로 측정된 프레임단위 비트율 제어 정확도를 보여준다. 모든 실험결과에서 제안하는 방법의 결과가 기존 기술의 결과 보다 정확도가 높음을 알 수 있다. 제안하는 방법은 기존 [2]와 [3]의 방법과 차별화되어, P 프레임 뿐만 아니라, I와 B 프레임에 대한 비트율-양자화 모델을 각각 정립하여 정교하게 양자화 파라미터를 결정한다. 또한 MAD와 함께 VOD를 고려하여 발생 비트량을 예측함에 따라 기존 기술들에 비해 보다 정확한 예측이 가능하게 되었다.

5. 결론

본 논문에서는 H.264/AVC의 계층적 부호화 구조에서도 효율적으로 비트 발생량을 예측할 수 있는 비트율-양자화 모델을 제안하였다. 또한 실험을 통해 제안하는 비트율-양자화 모델이 비트율 제어의 정확도를 높이는 것을 보였다.

6. 참고 문헌

- [1] ITU-T Rec. H.264 and ISO/IEC 14496-10, "Advanced Video Coding for Generic Audiovisual Services", Nov. 2007.
- [2] Z. Li, F. Pan, K. Pang, "Adaptive Basic Unit Layer Rate Control for JVT", Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-G012, Mar. 2003.
- [3] A. Leontaris and A. M. Tourapis, "Rate Control reorganization in the Joint Model(JM) reference software", Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-W042. Apr. 2007.
- [4] K. Suehring, H.264/AVC reference software (JM16.1) [Online]. Available : <http://iphome.hhi.de/suehring/tml/download/>