

적응 불규칙 양자화와 반복 복호를 이용한 비디오 코딩 방식에의 응용

정회원 최 강 선*

Adaptive Irregular Binning and Its Application to Video Coding Scheme Using Iterative Decoding

Kang-Sun Choi* *Regular Member*

요 약

센서 네트워크와 같이 한정된 리소스만을 사용할 수 있는 차세대 소형 무선 비디오 시스템을 위해 움직임 추정(motion estimation)이 필요 없는 저 복잡도의 비디오 인코딩 방식과 이를 효과적으로 디코딩하는 새로운 비디오 코딩 방식을 제안한다. 비디오 프레임에 주기적으로 인트라 코딩(Intra coding)하고 인트라 코딩된 프레임 사이에 있는 프레임에 대해선 제안하는 불규칙 양자화(irregular binning) 방식을 이용하여 효과적으로 코딩한다. 불규칙 양자화 방식은 코딩할 데이터의 Local statistics에 따라 각 구간의 길이를 조절함으로써 적은 수의 불규칙적인 구간으로도 효과적으로 데이터를 표현할 수 있다. 이와 함께, 연속된 프레임 간에 존재하는 높은 상관 관계(correlation)를 활용하기 위해 디코딩 시 불규칙 양자화 정보로 복원된 영상을 기초로 하여, 주변 프레임으로부터 움직임이 보상된 선명한 영상 부분(motion compensated image components)을 가져와 합성함으로써 복원된 프레임의 화질을 향상시키는 방법을 제안한다. 제안된 비디오 코딩 방식은 실험 결과에서 최신의 비디오 압축 방식인 H.264/AVC의 IB 코딩 모드와 유사한 성능을 보여주고 있다. 따라서 제안된 비디오 코딩 방식은 저 복잡도의 인코더가 필요한 응용 분야에 H.264/AVC를 대신하여 사용될 수 있다.

Key Words : low complexity encoder, iterative decoding, adaptive quantizer, irregular binning

ABSTRACT

We propose a novel low complexity video encoder, at the expense of a complex decoder, where video frames are intra-coded periodically and frames in between successive intra-coded frames are coded efficiently using a proposed irregular binning technique. We investigate a method of forming an irregular binning which is capable of quantizing any value effectively with only small number of bins, by exploiting the correlation between successive frames. This correlation is additionally exploited at the decoder, where the quality of reconstructed frames is enhanced gradually by applying POCS(projection on the convex sets). After an image frame is reconstructed with the irregular binning information at the proposed decoder, we can further improve the resulting quality by modifying the reconstructed image with motion-compensated image components from the neighboring frames which are considered to contain image details. In the proposed decoder, several iterations of these modification and re-projection steps can be invoked. Experimental results show that the performance of the proposed coding scheme is comparable to that of H.264/AVC

※ 이 논문은 한국과학재단의 해외 Post-doc. 연수지원에 의하여 연구되었음.

* 삼성전자 정보통신총괄 무선사업부 S/W Lab

논문번호 : KICS2005-04-142, 접수일자 : 2005년 4월 4일, 최종논문접수일자 : 2006년 3월 16일

coding in IB mode. Since the proposed video coding does not require motion estimation at the encoder, it can be considered as an alternative for versions of H.264/AVC in applications requiring a simple encoder.

I. 서론

MPEG-X와 H.26X 같은 비디오 압축 표준에서는 연속된 프레임 간에 존재하는 중복성(redundancy)을 효과적으로 제거하기 위해 인코더에서 움직임 추정(ME: motion estimation)을 통한 프레임 간 예측 코딩(inter-frame predictive coding)을 이용한다. 그러나 움직임 추정은 많은 계산량을 필요로 하기 때문에 인코더가 디코더보다 5~10배 정도 더 높은 복잡도를 지니게 된다¹⁾. 복잡도 상의 이러한 비대칭 형태는 고성능 처리 단말기(processing unit)를 통해 비디오가 한 번만 인코딩되고, 여러 단말기에서 디코딩되는 방송이나 스트리밍 VoD(video-on-demand)에 적합하다. 그러나 제한된 처리 능력과 에너지로 운용되어야 하는 센서 네트워크나 실시간 감시 시스템과 같은 차세대 비디오 시스템에서는 저 복잡도의 신호 처리 방식이 더욱 적합하다.

최근에 디코더의 복잡도는 증대되지만 인코더에 ME가 없는 저 복잡도의 인코딩 방식을 이용하는 비디오 코딩 알고리즘이 제안되었다. Side information을 이용하여 효과적으로 비손실/손실 압축을 할 수 있음은 오래 전에 이론적으로 증명되었지만²⁾, 이를 이용한 실용적인 구현 방식에 대해서는 최근 들어서 나타나고 있다^{3,4)}. Pradhan과 Ramchandran이 처음으로 Syndrome에 기초한 실용적인 Wyner-Ziv 코딩 방식을 제안한 이후⁴⁾, Aaron이 Turbo codes를 이용한 Distributed video compression을 제안하였다⁵⁾. 그러나 Aaron의 방식은 표준 MPEG-4 방식과 비교했을 때 R-D(rate-distortion) 면에서 상당한 차이를 보였다. Sehgal은 Coset을 응용한 Wyner-Ziv 비디오 코딩을 제안하고 이 방식이 비디오 전송 시 발생하는 에러로 인한 예측 불일치(prediction mismatch) 문제를 완화시킴을 보여줬다⁶⁾. Yaman은 Shape-adaptive SPIHT(set partitioning in hierarchical trees) 인코더로 디코더에서 예측하기 어려운 부분을 가려내 고화질로 인코딩하는 웨이블릿(wavelet) 비디오 코딩 방식을 제안했다⁷⁾.

본 논문에서는 기존의 비디오 표준과는 반대되는 위와 같은 복잡도의 비대칭 코딩 특성을 갖는 새로운 비디오 코딩 방식을 제안한다. 제안하는 방식에서는 키프레임(key frame)이 주기적으로 기존의 Intra-frame 인코더를 통해 인트라 모드로 코딩되어 전송

되며, 키프레임 사이의 중간 프레임들에 대해서는 제안하는 불규칙 양자화 방식을 이용하여 효율적으로 코딩된다. 불규칙 양자화 방식은 중간 프레임의 데이터를 코딩하기 위해 상관도가 높은 주변 키프레임으로부터 코딩할 데이터와 관련된 Local statistics을 근사화하여(approximated) 구하고 이에 적합한 네 개의 불규칙한 양자화 구간을 형성한다. 디코더에서는 대응되는 Intraframe 디코더를 이용하여 인코더에서 이용한 것과 동일한 키프레임을 복원하고, 복원된 키프레임과 간단한 보간법(interpolation)을 통해 중간 프레임에 대한 Estimate을 얻는다. 또한 인코더와 디코더 모두 동일한 키프레임을 갖고 있기 때문에 디코더에서도 불규칙 양자화 방식을 통해 동일한 불규칙 양자화 구간을 형성할 수 있다. 형성된 불규칙 양자화 구간과 전송된 양자화 정보/인덱스(quantization index)를 이용하여 중간 프레임을 위한 제한 조건 집합인 QCS(quantization constraint set)를 구성한다. 중간 프레임에 대해 얻어진 Estimate을 QCS로 투사(projection)하는 POCS(projection on the convex sets) 방식을 사용하여 영상의 화질을 개선시킨다⁸⁻¹⁰⁾. QCS에 투사된 Estimate은 선명하지 않지만 중간 프레임의 전체적인 모습을 지니고 때문에, 이 모습에 맞춰 키프레임에 있는 선명한 윤곽선(edges)과 재질감(textures) 등을 움직임 보상하여 Estimate에 융합시킴으로써 Estimate의 화질을 더욱 개선시킬 수 있다. 제안하는 디코딩 방식에서는 이와 같이 주변의 키프레임으로부터 선명한 영상 부분을 취해 움직임 보상 후 결과에 융합하고 QCS에 투사하는 과정을 반복함으로써 점진적으로 비디오의 질을 향상시키게 된다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 제안하는 불규칙 양자화 방식에 대해 설명하고, 3장과 4장에서 각각 인코더와 디코더에 대해 세부적으로 설명한다. 5장에서는 다양한 실험 결과를 통해 최신의 Interframe 코더인 H.264/AVC와 제안하는 비디오 코딩 방식을 비교하며 마지막으로 6장에서 결론을 맺는다.

II. 순서통계에 기반을 둔 불규칙 양자화

Temporal redundancy를 효과적으로 줄이기 위해 기존의 Interframe 인코더에서는 ME를 수행하게 되

는데, 이는 인코더가 디코더보다 복잡도 면에서 5~10 배 정도 더 복잡하게 되는 요인이 되고 있다. 제안하는 비디오 코딩 방식에서는 이런 문제를 피하기 위해 ME 과정이 없는 Intraframe 인코더가 사용된다. 가능한 많이 비트율을 줄이기 위해 비디오의 프레임 중 일부분만을 키프레임으로 할당하여 일반 Intraframe 인코더를 통해 원하고자 하는 화질로 코딩을 하여 디코더로 전송, 복원하고 키프레임의 사이에 있는 프레임들은 복원된 키프레임을 보간하여(interpolation) 얻어낸다. 물론 시간축 상으로 보간되어 얻어지는 프레임(interpolated frame)은 원 영상과 많이 차이가 날 수 있다. 이런 문제를 해결하기 위해서 제안하는 인코더에서는 매우 성긴 양자화기(coarse quantizer)를 이용해 중간 프레임에 대한 아주 적은 양의 정보를 추가로 보내어 보간된 영상의 화질을 향상시킨다. 가장 쉬운 방법으로 Uniform scalar quantizer를 고려할 수 있지만, 데이터 범위를 아주 적은 수의 균일한 구간으로 나누게 되면 각 구간이 차지하는 범위가 상당히 넓어져 중간 프레임에 대해 추가로 보내는 정보의 효용성이 떨어지게 된다.

본 논문에서는 코딩할 데이터의 Local statistics에 따라 각 구간의 길이를 조절함으로써 적은 수의 불규칙적인 구간으로도 효과적으로 데이터를 표현할 수 있는 적응 양자화기(adaptive quantizer)를 제안한다. 제안하는 불규칙 양자화 방식에서는 인코더와 디코더 모두 사용이 가능한 키프레임으로 간단히 평균 보간하여 영상을 얻고 이로부터 코딩하고자 하는 데이터에 대해 동일한 Local statistics를 추정해 낸다.

f_n 과 \hat{f}_n , F_n , \hat{F}_n 을 각각 코딩하고자 하는 비디오의 원 영상과 그에 대한 Estimate, 그리고 두 영상의 변환된(transformed) 데이터라 할 때, f_n 과 \hat{f}_n 의 같은 위치에 있는 값들 사이에는 상관도가 매우 높음을 볼 수 있다. 하지만, 웨이블릿 변환 같이 Shift variant 특성의 transform을 거치게 되면, 윤곽선의 작은 움직임으로도 변환된 F_n 과 \hat{F}_n 에서는 부호(sign)나 크기 면에서 매우 다르게 나타나게 된다.

표 1은 "Foreman" 시퀀스에서 동일한 위치에 있는 3x3 웨이블릿 계수(wavelet coefficient)를 보여준다. Estimate인 \hat{F}_n 은 $(F_{n-1} + F_{n+1})/2$ 로 얻어진다. 움직임이 작은 물체와 정지된 배경을 포함하고 있는 1~3번째 프레임에는, F_n 과 \hat{F}_n 이 서로 매우 유사하여, 표 1(a)의 가운데에 있는 F_n 의 계수도

표 1. "Foreman" 시퀀스에서 동일 위치에 있는 F_n 과 \hat{F}_n 의 웨이블릿 계수 간의 비교 예

Table 1. Examples of wavelet coefficients of F_n and \hat{F}_n at the same location in "Foreman" sequence

F_n	544.01	352.01	665.90
	420.62	573.91	360.09
	240.17	442.94	538.23
\hat{F}_n	551.45	348.26	662.81
	424.15	574.86	359.80
	239.35	445.02	536.49

(a) 2번째 프레임의 LL3 부밴드

F_n	-0.55	0.09	-56.11
	-39.85	-0.70	23.57
	0.03	-0.49	-23.97
\hat{F}_n	12.12	-0.78	-56.57
	-31.83	0.35	31.99
	-0.24	12.27	-20.12

(b) 2번째 프레임의 HL2 부밴드

F_n	547.51	521.02	536.12
	394.40	681.78	561.10
	621.64	522.57	642.30
\hat{F}_n	503.20	540.33	467.37
	691.93	492.41	614.33
	440.89	608.78	555.00

(c) 186번째 프레임의 LL3 부밴드

F_n	225.34	-2.20	46.80
	-70.89	-37.67	-20.14
	36.71	114.92	-130.71
\hat{F}_n	-79.02	48.04	-38.11
	48.13	19.76	-16.25
	-6.36	-83.76	90.08

(d) 186번째 프레임의 HL3 부밴드

\hat{F}_n 의 동일 위치상에 있는 계수로 예측이 가능하다. 반대로 큰 카메라 움직임을 보이고 있는 186번째 프레임에서는 위에 설명한 변환 특성으로 인해 표 1(c)와 (d)의 F_n 과 \hat{F}_n 사이에는 상관도가 매우 낮음을 볼 수 있다. 그러나 \hat{F}_n 내의 주변에 있는 계수들을 살펴보면 F_n 의 중앙 계수와 유사한 값을 찾을 수 있다. 이런 현상은 근접해 있는 계수들이 유사한 지역적 특성(local characteristics)을 갖는 값들이면서 위치적으로는 다양한 변화(shift)를 표현하고 있기 때문이다. 따라서 F_n 내 임의의 위치에 있는 계수를 예측하기 위해 \hat{F}_n 의 해당 위치와 그 주변에 있는 계수들 중에서 유용한 후보값(candidate)을 골라낼 수 있다.

제안하는 불규칙 양자화 방법은 이와 같은 특성을 이용하여 우선 표 1과 같이 \hat{F}_n 에서 해당 위치를 중심으로 3x3 크기 윈도우 내의 9개 계수들을 구하고, 그 중에서 중앙에 위치한 중앙값과 함께 최소값(minimum), 최대값(maximum), 중간값(median), 총 4개의 후보값을 결정한다. 3개의 순서통계값(order-statistics value)을 포함하고 있는 이 후보값들은 Local statistics를 효과적으로 나타낼 수 있기 때문에 각 후보값을 양자화기를 구성하는 구간의 대표값(codeword)으로 사용하여 웨이브렛 계수의 실효 범위(dynamic range)를 4개의 구간(interval)으로 나눌 수 있다.

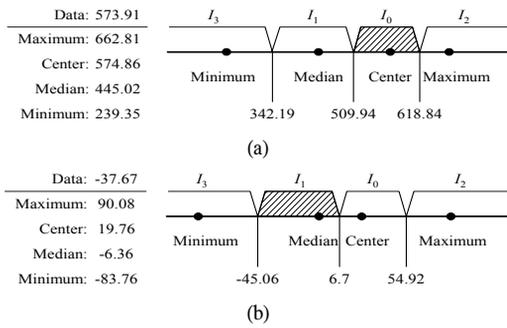


그림 1. OS 기반 불규칙 양자화. (a) 표 1(a) 데이터에 대한 불규칙 양자화 모습. (b) 표 1(d) 데이터에 대한 불규칙 양자화 모습
 Figure 1. Examples of the OS-based irregular binning. (a) Irregular binning with data of Table 1(a). (b) Irregular binning with data of Table 1(d)

그림 1은 표 1에 있는 값들을 이용해 순서통계 기반의 불규칙 양자화를 수행하는 과정을 보여준다. 표 1(a)에서 코딩하고자 하는 데이터인 573.91에 대해 Estimate \hat{F}_n 의 3x3 크기 윈도우 내 계수들로부터 위에 설명한 4개의 후보값, 즉 동일 위치에 해당하는 값인 574.86, 최대값 662.81, 최소값 239.35, 중간값 445.02를 찾는다. 이를 크기순으로 나열한 후 인접한 후보값들의 평균을 구간의 경계로 하여 그림 1(a)에서와 같이 4개의 구간을 얻는다.

$$\begin{aligned}
 I_0 &= \{x|509.94 \leq x < 618.84\} \\
 I_1 &= \{x|342.19 \leq x < 509.94\} \\
 I_2 &= \{x|618.84 \leq x\} \\
 I_3 &= \{x|x < 342.19\}
 \end{aligned} \tag{1}$$

코딩할 데이터 573.91은 I_0 에 속하기 때문에 심볼 0으로 표현되며 이때 얻어진 구간 I_0 의 크기가

8.9로 매우 세밀하다는 사실을 주목할 필요가 있다. 이는 제안된 방법이 \hat{F}_n 로부터 얻어지는 근사화된 (approximated) Local statistics에 따라 F_n 의 각 웨이브렛 계수마다 그에 맞는 4개의 후보값들이 조정되기 때문에 4개의 심볼만으로도 효과적으로 계수값을 표현할 수 있기 때문이다. 본 논문에서는 이산 웨이브렛 변환과 함께 인코더와 디코더에서 사용될 Estimate을 얻기 위해 평균 보간법(average interpolation)을 이용하기로 한다.

III. 인코더 구조

그림 2에서는 제안하는 인코더의 구조를 보여주고 있다. 두 프레임 간격으로 키프레임이 할당되며 키프레임은 JPEG2000 방식으로 인코딩되며^[11], 키프레임 사이에 낀 프레임은 제안된 불규칙 양자화 로 코딩된다.

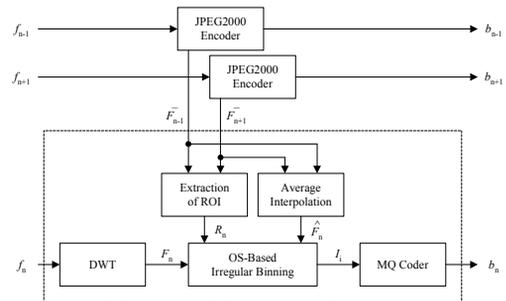


그림 2. 제안하는 인코더의 블록도
 Figure 2. Block diagram of the proposed encoder

\bar{F} 를 JPEG2000 디코더에서 복원되는 영상의 웨이브렛 계수라 할 때, 이는 JPEG2000 인코딩 시 각 코드블록(code-block)마다 몇 개의 Coding pass가 전송될 지 결정되면 구할 수 있다. 이 영상을 이용하여 평균 보간을 통해 다음과 같이 Estimate을 얻을 수 있다.

$$\hat{F}_n = \frac{\bar{F}_{n-1} + \bar{F}_{n+1}}{2} \tag{2}$$

$\mathbf{p}_0 = [k_0, l_0]$ 위치에 있는 웨이브렛 계수 $F_n(\mathbf{p}_0)$ 을 코딩하기 위해서 일단 \hat{F}_n 에서 9개의 관련 계수들 $\{\hat{F}_n(\mathbf{p}) | \mathbf{p} \in \mathbf{P}\}$, $\mathbf{P} = \{[k, l] | k_0 - 1 \leq k \leq k_0 + 1, l_0 - 1 \leq l \leq l_0 + 1\}$ 을 뽑아낸다. 이 9개의 값으로부터 다음과 같이 4개의 후보값을 선택한다.

$$\begin{aligned}
 c_0 &= \widehat{F}_n(\mathbf{p}_0) \\
 c_1 &= \text{med}(\{\widehat{F}_n(\mathbf{p}) | \mathbf{p} \in \mathbf{P}\}) \\
 c_2 &= \min(\{\widehat{F}_n(\mathbf{p}) | \mathbf{p} \in \mathbf{P}\}) \\
 c_3 &= \max(\{\widehat{F}_n(\mathbf{p}) | \mathbf{p} \in \mathbf{P}\})
 \end{aligned} \tag{3}$$

제안하는 불규칙 양자화 방식은 인접한 두 후보 값의 중앙을 구간의 경계로 하여 데이터 범위를 4 개의 구간으로 나눈다. 인코더는 실제 코딩될 값인 $F_n(\mathbf{p}_0)$ 가 구해진 4개의 구간 중 어디에 속하는지 찾아 그 구간의 인덱스를 엔트로피 부호화하여 디코더에 보낸다. 엔트로피 코더로는 JPEG2000 인코더에서 사용되고 있는 MQ 코더를 재사용한다.

이 때 후보값 중 하나인 $\widehat{F}_n(\mathbf{p}_0)$ 가 다른 순서통계 후보값과 같은 값을 갖게 되는 경우가 발생할 수 있다. 이러한 경우는 실제적으로 3개 구간을 지닌 양자화기가 되기 때문에 효율이 떨어지게 된다. 따라서 이런 경우 식(4)을 이용해 가상 후보값 c_v 를 구하여 4개의 구간으로 나눈다.

$$c_v = \begin{cases} (c_0 + c_2)/2, & \text{if } c_0 = c_1 \text{ and } |c_0 - c_2| > |c_0 - c_3|, \\ (c_0 + c_3)/2, & \text{if } c_0 = c_1 \text{ and } |c_0 - c_2| \leq |c_0 - c_3|, \\ 2c_0 - c_1, & \text{otherwise.} \end{cases} \tag{4}$$

인코딩 시 코딩 효율을 더욱 향상시키기 위해 중간 프레임의 웨이브렛 계수 중 Estimate의 웨이브렛 계수와 차이가 많이 나는 부분(ROI: Region of Interest)인 R_n 을 예측하여 그 부분에 대해서만 선별적으로 적응 불규칙 양자화를 이용하도록 한다. 인코더의 복잡도를 높이지 않기 위해 간단한 프레임 차(FD: frame difference) 연산자를 사용하였다.

$$R_n(\mathbf{p}) = \begin{cases} 1, & |\overline{F}_{n-1}(\mathbf{p}) - \overline{F}_{n+1}(\mathbf{p})| > T, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

FD는 복잡도 측면에서 간단할 뿐 아니라, 대부분의 경우 실제 ROI와 유사한 결과를 제공해준다. 또한 FD에 사용되는 문턱값(threshold) T 를 조절하여 전송할 웨이브렛 계수의 수를 결정하여 R-D를 조절할 수 있다.

IV. 디코더 구조

제안하는 디코더에서는 그림 3과 같이 먼저 인코더에서 키프레임으로 전송된 비트스트림 b_{n-1} 과 b_{n+1} 을 JPEG2000 디코더에 입력하여 키프레임 \overline{F}_{n-1} 과 \overline{F}_{n+1} 을 복원한다. 인코더에서 사용된 것과

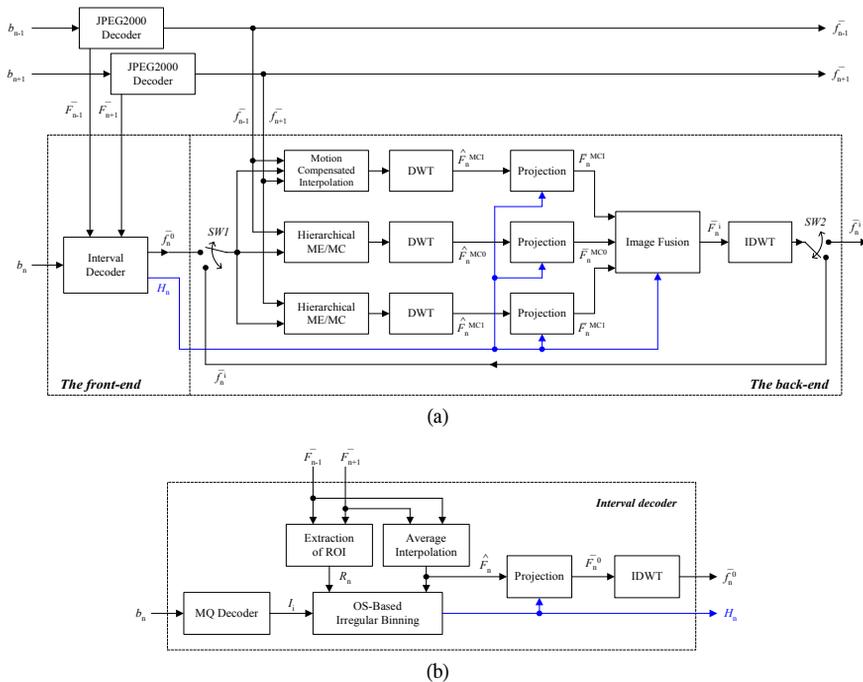


그림 3. (a) 제안하는 디코더의 블록도. (b) 구간 디코더의 세부도
Figure 3. (a) Block diagram of the proposed decoder. (b) Detail of the interval decoder

동일한 키프레임이 복원이 되면, 그림 3(b)의 구간 디코더에서 인코더와 동일한 R_n 과 \hat{F}_n 을 구하게 된다. 또한, 중간 프레임에 대해 부호화된 비트스트림 b_n 은 엔트로피 디코더인 MQ 디코더를 통해 구간 인덱스 정보인 I_i 로 변환된다.

$R_n(\mathbf{p}) = 1$ 인 위치 \mathbf{p} 에 대해 불규칙 양자화 방식을 이용해 \hat{F}_n 에서 4개의 후보값을 추출한 후 I_i 에 해당하는 구간을 구하고 이 구간을 복원되는 계수 $\overline{F}_n(\mathbf{p})$ 를 위한 구간 제한 조건(interval constraint) [$\overline{F}_n^L(\mathbf{p}), \overline{F}_n^U(\mathbf{p})$]으로 이용한다. $R_n(\mathbf{p}) = 1$ 인 모든 위치에 대해 얻어진 구간 제한 조건 집합 H_n 에 평균 보간 방식으로 얻어진 \hat{F}_n 를 투사하고 그 결과를 공간 영역으로 역변환하여 첫 추정 영상(initial estimate) \overline{f}_n^0 을 구한다. 구간 디코더는 위와 같이 \overline{f}_n^0 과 구간 제한 집합 H_n 을 구하여 디코더의 후반부(back-end)에 제공한다.

디코더 후반부는 \overline{f}_n^0 과 H_n 을 이용하여 POCS를 수행함으로써 점진적으로 화질을 향상시키게 된다. 비디오의 연속된 프레임들 간에는 높은 상관관계가 있다는 사실에 기반하여, 주변 프레임으로부터 현재

을 사용한 FSBMA의 결과가 계층적 Neighborhood relaxation 결과보다 PSNR 면에서 0.5dB 정도 낮은 문제가 있다.

디코더의 후반부에는 또 하나의 경로가 존재하는데 이는 블록 기반 움직임 보상 보간법(block-based motion compensated interpolation)으로 \overline{f}_n^0 과 $\overline{F}_{n-1}, \overline{F}_{n+1}$ 을 받아들여 보간된 영상을 추출한다.

최종적으로 중간 프레임을 위해 2개의 계층적 ME/MC 결과물과 하나의 MCI 결과물이 얻어지며 각 결과물의 웨이브렛 변환인 \hat{F}_n^{MCI} 과 $\hat{F}_n^{MC_0}$, $\hat{F}_n^{MC_1}$ 은 H_n 에 투사된다. 투사된 세 영상 \overline{F}_n^{MCI} 과 $\overline{F}_n^{MC_0}$, $\overline{F}_n^{MC_1}$ 는 하단에 주어진 식(6)과 같은 영상 융합(image fusion) 과정을 통해 하나의 영상으로 통합된다. 여기서 Δ_{min} 은 세 계수 $\overline{F}_n^{MCI}(\mathbf{p})$ 과 $\overline{F}_n^{MC_0}(\mathbf{p})$, $\overline{F}_n^{MC_1}(\mathbf{p})$ 중 투사 과정에서 웨이브렛 계수가 가장 적게 영향을 받은 계수를 나타낸다. 만약 $|\overline{F}_n^{MCI}(\mathbf{p}) - \overline{F}_n^{MC_0}(\mathbf{p})|$ 가 $|\overline{F}_n^{MCI}(\mathbf{p}) - \overline{F}_n^{MC_1}(\mathbf{p})|$, $|\overline{F}_n^{MC_0}(\mathbf{p}) - \overline{F}_n^{MC_1}(\mathbf{p})|$ 중 최소이면 $|\overline{F}_n^{MCI}(\mathbf{p}) - \overline{F}_n^{MC_0}(\mathbf{p})|$ 이 Δ_{min} 으로 선택된다.

$$\overline{F}_n^i(\mathbf{p}) = \begin{cases} \text{mean}(\overline{F}_n^{MCI}(\mathbf{p}), \overline{F}_n^{MC_0}(\mathbf{p}), \overline{F}_n^{MC_1}(\mathbf{p})), & \text{if } R_n(\mathbf{p}) = 0 \\ \Delta_{min}, & \text{otherwise.} \end{cases} \quad (6)$$

프레임에 도움이 될 영상 요소를 움직임 보상을 덧입히고 다시 주어진 구간 제한 집합으로 투사함으로써 화질을 향상시킬 수 있다. 디코더의 후반부에서는 이와 같은 변경과 투사 과정을 반복적으로 수행한다.

\overline{f}_n^0 가 후반부에 들어오게 되면, 이를 Reference로 삼아 복원된 키프레임 \overline{F}_{n-1} 과 \overline{F}_{n+1} 로부터 움직임 추정/움직임 보상(ME/MC)을 통해 두 개의 영상을 얻어낼 수 있다. 키프레임에 있는 영상 조각의 정확한 움직임을 찾기 위해 본 논문에서는 Neighborhood relaxation 방식을 계층적으로 사용하기로 한다^[12]. Full search block-matching algorithm (FSBMA)이 구현하기 쉽고 미디언 필터링을 통해 튀는 움직임(outlier)을 상당량 제거할 수 있는 장점이 있지만, 제안하는 디코딩 방식에 적용했을 때, 수렴하기 위해서 더 많은 반복 수행을 거쳐야 하는 단점이 있다. 또한 수렴한 이후에도 미디언 필터링

이렇게 얻어진 결과는 Feed-back되어 다시 디코더 후반부의 입력으로 사용될 수 있는데, 계층적 ME와 MCI에 구간 디코더에서 제공되는 \overline{f}_n^0 보다 향상된 화질의 \overline{f}_n^i 가 Reference로 제공되기 때문에 계층적 ME에서 더욱 정확한 움직임 정보를 얻게 된다. 따라서 현재 프레임에 덧입힐 영상 요소를 가져올 때 더욱 정확한 부분을 가져와 융합함으로써 비디오의 화질을 점진적으로 향상시킬 수 있다. 반복적으로 ME/MC를 수행함으로써 디코딩에 필요한 계산량이 증가되지만, 실제로는 $(i-1)$ 번째 반복에서 얻어지는 움직임 정보(motion information)가 i 번째 처리에서 얻어지는 움직임과 유사하기 때문에, 이전 단계에서 얻어진 움직임을 기준으로 작은 탐색 구간(search range)을 갖는 Motion refinement를 통해서도 정확한 움직임을 찾아낼 수 있어 추가되는 계산량을 효과적으로 줄일 수 있다.

V. 실험 결과

제안된 방식의 성능을 H.264/AVC Interframe 코더와 비교 평가하기 위해 두 프레임 간격으로 키프레임을 할당하여 JPEG2000 코덱으로 인트라 코딩을 하며, 키프레임 사이에 있는 프레임은 제안된 불규칙 양자화 방식으로 코딩된다. H.264/AVC 코덱은 최근 참조 소프트웨어인 Joint Model 9.2를 이용하여 30Hz QCIF 크기의 프로그램시브 비디오를 다음과 같은 환경에서 코딩하였다. (두 프레임 간격으로 I 프레임을, 중간의 프레임에 대해서 B 프레임을 코딩, 엔트로피 코딩으로 CAVLC를 사용, R-D 최적화 사용)

양자화 파라미터 QP는 28, 32, 36, 40으로 설정했으며 수평 수직 방향으로 ± 32 의 탐색 구간에 대해 1/4-pel ME/MC를 수행하였다. Motion refinement을 위한 탐색 구간으로는 ± 2 를 설정하고 계층적 ME 수행 시에는 3 계층 레벨에 첫 블록 크기를 32x32 픽셀로 정하여 최종 계층 ME 시 8x8 픽셀의 블록 크기를 이룬다.

그림 4와 5에서는 정적인 비디오와 동적인 비디오에 대해 제안된 디코더가 화질을 향상시키는 과정을 보여주고 있다. 그림 4의 장면에선 정지된 배

경과 함께 사람의 머리가 조금 움직이고 있는데, 그림 4(d)에 나타난 Initial estimate에는 작은 움직임으로 인해 얼굴이 흐려져 있다. 구간 디코더에서 얻어진 H_2 로 투사하여 영상이 개선되긴 했지만, 아직 얼굴의 윤곽이 선명하지 못함을 볼 수 있다. 하지만 디코더 후반부를 한번 통과하여 주변 키프레임으로부터 알맞은 영상 요소를 융합시킨 결과인 그림 4(f)는 디코더에 사용된 주변 키프레임의 화질과 유사함을 볼 수 있다. 그림 5는 빠른 카메라 패닝(panning)으로 이루어진 장면을 보여주고 있는데, 예상대로 평균 보간된 Initial estimate \hat{f}_{186} 은 매우 열화되어 있다. 그러나 구간 디코더의 투사 과정을 거치면서 f_{186} 의 전체적인 모습이 되살아나게 되며, 이를 Reference로 시작해 디코더 후반부의 POCS 과정을 네 번 반복함으로써 원 영상과 매우 유사한 영상을 얻었다.

그림 6에서는 두 비교 알고리즘의 R-D 곡선을 보여주고 있다. 제안된 코딩 방식은 키프레임의 비트율과 중간 프레임의 ROI를 결정하는 T , 두 개의 파라미터에 의해 R-D가 조절될 수 있다. "Foreman" 시퀀스에서 제안된 방식이 $T=30$ 인 경우 H.264/AVC와 0.1~0.25 dB의 차이를 보여주고 있고,

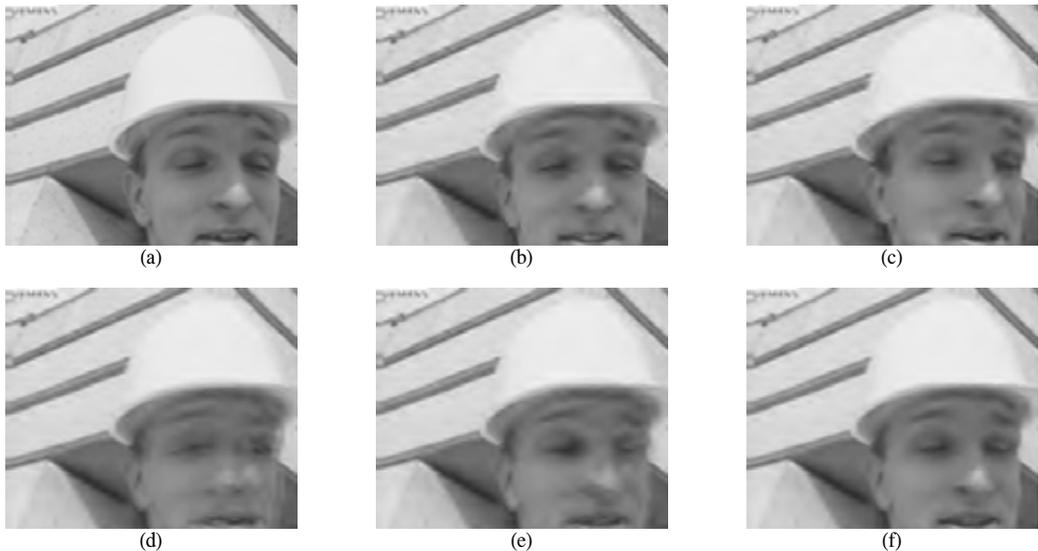


그림 4. 제안된 디코더에서 정적인 비디오를 개선시키는 과정. (a) "Foreman" 시퀀스의 2번째 프레임. (b) 0.7 bpp (픽셀당 비트 수)로 인코딩된 1번째 키프레임. (c) 0.7 bpp로 인코딩된 3번째 키프레임. (d) (c)와 (d)를 평균 보간하여 얻은 (a)의 Initial estimate \hat{f}_2 . (e) 구간 디코더에서 나온 \bar{f}_2^0 . (f) 디코더 후반부에서 나온 첫 결과물 \bar{f}_2^1
 Figure 4. Process of enhancing a static video in the proposed decoder. (a) 2nd frame of "Foreman" sequence. (b) 1st key frame encoded at 0.7 bpp (bit per pixel). (c) 3rd key frame encoded at 0.7 bpp. (d) Initial estimate of (a), \hat{f}_2 , average of (b) and (c). (e) Output of interval decoder \bar{f}_2^0 . (f) Result at first iteration \bar{f}_2^1

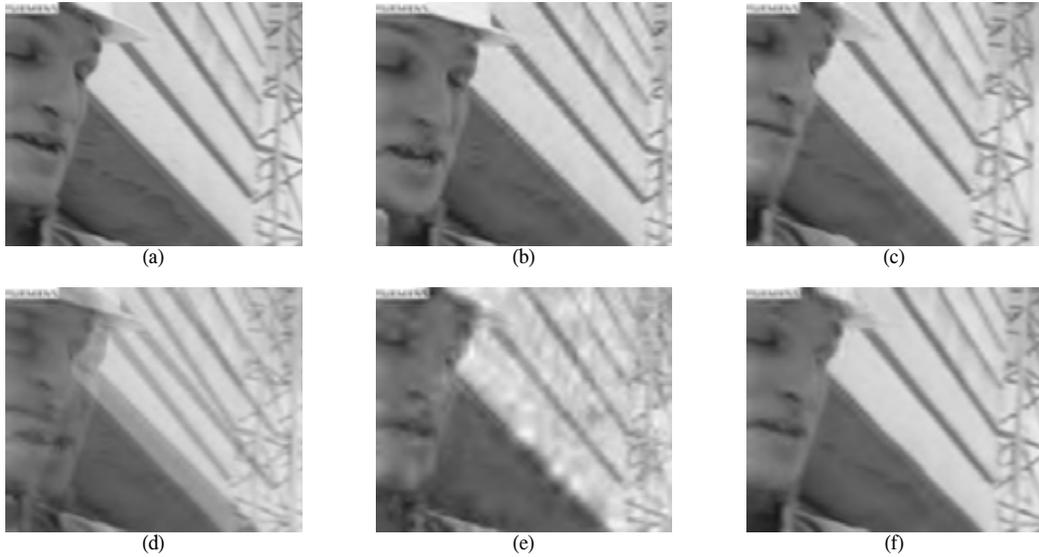


그림 5. 제안된 디코더에서 동적인 비디오를 개선시키는 과정. (a) “Foreman” 시퀀스의 186번째 프레임. (b) 0.7 bpp (픽셀당 비트수)로 인코딩된 185번째 키프레임. (c) 0.7 bpp로 인코딩된 187번째 키프레임. (d) (c)와 (b)를 평균 보간하여 얻은 (a)의 Initial estimate \hat{f}_{186}^0 . (e) 구간 디코더에서 나온 \hat{f}_{186}^0 . (f) 디코더 후반부에서 네 번 반복 후 나온 결과물 \hat{f}_{186}^4

Figure 5. Process of enhancing a static video in the proposed decoder. (a) 186th frame of “Foreman” sequence. (b) 185th key frame encoded at 0.7 bpp (bit per pixel). (c) 187th key frame encoded at 0.7 bpp. (d) Initial estimate of (a), \hat{f}_{186}^0 , average of (b) and (c). (e) Output of interval decoder \hat{f}_{186}^0 . (f) Result at first iteration \hat{f}_{186}^4

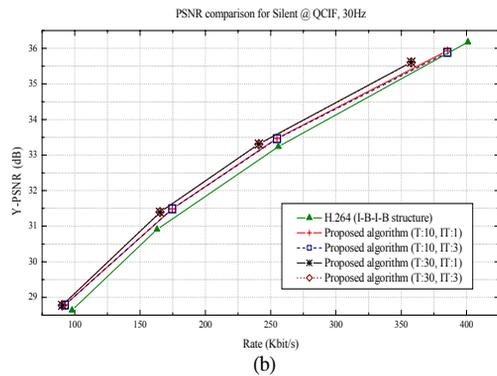
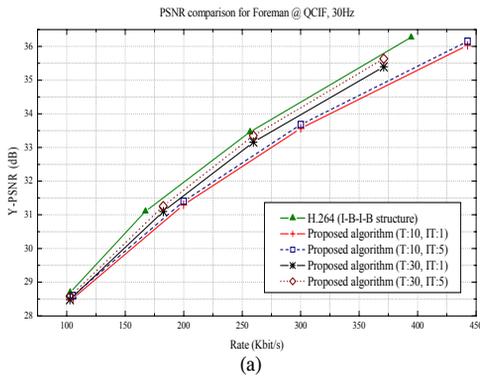


그림 6. R-D 곡선. (a) “Foreman” 시퀀스, (b) “Silent” 시퀀스
Figure 6. Rate-distortion curves (a) for “Foreman” sequence, (b) for “Silent” sequence

$T=10$ 인 경우에는 0.5~1.0 dB의 차이를 보여주고 있다. 정적인 배경이 주로 있는 “Silent” 시퀀스에서는 평균 보간법으로도 효과적으로 Initial estimate을 생성할 수 있기 때문에 제안된 방식이 오히려 H.264/AVC보다 0.25~0.5 dB 더 좋은 화질의 결과를 보여줬다.

VI. 결론

디지털 방송이나 스트리밍 VoD 응용분야에서 성

공적으로 사용되고 있는 MPEG-X나 H.26X와 같은 비디오 압축 표준은 인코더에서 움직임 추정에 의한 프레임 간 예측을 수행하기 때문에 디코더보다 5~10배 정도 더 높은 복잡도를 갖고 있는 특성이 있다. 그러나 센서 네트워크나 실시간 감시 시스템과 같은 차세대 비디오 시스템에서는 제한된 처리 능력과 에너지로 운용되어야 하는 시스템 특성 상 저 복잡도로 운영될 수 있는 인코딩 시스템이 절실히 요구되어지고 있다.

본 논문에서는 OS를 기반으로 한 간단한 적응

불규칙 양자화 방식과 반복적인 영상 복원 기술을 제안하여 저 복잡도의 인코더를 이용하면서도 효과적으로 동영상 코딩하여 기존의 비디오 표준과 유사한 성능을 지닌 비디오 압축 방식을 제안하였다. 제안한 방식에서는 비디오 중 일부를 키프레임으로 할당해 JPEG2000 코덱으로 인트라 코딩/디코딩하고 키프레임 사이의 중간 프레임에 대해서는 제안한 불규칙 양자화를 이용해 효과적으로 코딩한다. 본 논문에서는 제안한 불규칙 양자화 방식이 코딩할 데이터와 상관도가 높은 주변 정보로부터 코딩할 데이터의 Local statistics를 근사화하여 얻어 내 양자화기에 사용되는 구간의 경계를 조정함으로써 단지 네 개의 적은 심볼로도 효과적으로 계수를 코딩할 수 있음을 보여줬다. 디코더에서는 우선 키프레임을 복원한 이후, 전송 받은 불규칙 양자화 심볼 정보를 이용해 중간 프레임에 대한 영상을 복원한다. 이때 얻어진 중간 프레임에 대한 Estimate는 선명해진 않지만, 중간 프레임 영상을 구성하는 여러 가지 객체들의 전체적인 모습 및 위치를 담고 있다. 비디오의 연속된 프레임 간에 상관도가 높음을 이용하여, 주변의 키프레임으로부터 선명한 윤곽선 및 재질감과 같이 복원한 영상 프레임에 도움이 될 영상 요소들을 Estimate에 맞게 움직임 보상이며 반복적으로 융합함으로써 점진적으로 화질을 향상시키게 된다.

실험 결과에서는 최신의 H.264/AVC 코덱을 IB 모드로 이용하여 제안한 방식의 압축 성능을 비교하였다. 사용되는 시퀀스의 동적 특성에 따라 근소한 압축 성능의 차이를 보였는데, 일반적인 시퀀스에 대해서는 0.1~1 dB의 미세한 화질 저하를 보여줬으며 정적인 시퀀스에 대해서는 오히려 0.25~0.5 dB 정도 더 좋은 화질을 보여줬다. 따라서 많은 계산량을 필요로 하는 움직임 보상이 없이 간단히 인코딩될 수 있는 장점을 감안해 볼 때 제안된 방식은 저 복잡도의 인코더가 필요한 실시간 감시 시스템이나 센서 네트워크 같은 응용 분야에 H.264/AVC를 대신할 수 있는 비디오 코딩 방식으로 평가된다.

참 고 문 헌

[1] S. Yaman, G. AlRegib, "A low-complexity video encoder with decoder motion estimator," *IEEE Int. Conf. ASSP*, 3, pp. 157-160, 2004.
 [2] D. Slepian, J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, IT-19, pp. 471-480, July 1973.

[3] A. D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, IT-22, pp. 1-10, May 1976.
 [4] S. Pradhan, K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. on Information Theory*, 49, pp. 626-643, Mar. 2003.
 [5] A. Aaron, S. Rane, R. Zhang, B. Girod, "Wyner-Ziv coding for video: applications to compression and error resilience," *IEEE Data Compression Conference*, pp. 93-102, Mar. 2003.
 [6] A. Sehgal, A. Jagmohan, N. Ahuja, "A state-free causal video encoding paradigm," *IEEE Int. Conf. Image Processing*, 1, pp. 605-608, 2003.
 [7] R. Puri, K. Ramchandran, "PRISM: a reversed multimedia coding paradigm," *IEEE Int. Conf. Image Processing*, 1, pp. 617-620, 2003.
 [8] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. on Circuits Syst. Video Technol.*, 2, pp. 91-95, 1992.
 [9] 김윤, 정재한, 김재원, 고성제, "POCS를 이용한 효과적인 블록 현상 제거 기법", *신호처리합동학술대회 논문집*, 14(1), pp. 755-758, Sep. 2001.
 [10] K.-S. Choi, J.-K. Cho, M.-C. Hwang, S.-J. Ko, "POCS-based enhancement of de-interlaced video," *Lecture Notes in Computer Science*, 2869, pp. 667-674, Nov. 2003.
 [11] D. S. Taubman, M. W. Marcellin, *JPEG 2000: Image compression fundamentals, standards and practice*, Kluwer Academic Publishers, 2002.
 [12] Y.-K. Chen, Y.-T. Lin, S. Y. Kung, "A feature tracking algorithm using neighborhood relaxation with multi-candidate pre-screening," *IEEE Int. Conf. Image Processing*, 2, pp. 513-516, 1996.

최 강 선 (Kang-Sun Choi)

정회원



1997년 고려대학교 전자공학과 학사
 1999년 고려대학교 전자공학과 공학석사
 2003년 고려대학교 전자공학과 공학박사
 2003년-2005년 University of Southern California Post-doc.

연구원

2005년~현재 삼성전자 정보통신총괄 S/W Lab. 책임 연구원
 <관심분야> 멀티미디어 통신, 센서 네트워크, 신호처리, 영상처리