

특집논문-09-14-2-05

주파수 적응 채널 잡음 모델링에 기반한 변환영역 Wyner-Ziv 부호화 방법

김 병 희^{a)}, 고 봉 혁^{a)}, 전 병 우^{a)‡}

Transform domain Wyner-Ziv Coding based on the frequency-adaptive channel noise modeling

Byunghee Kim^{a)}, Bonghyuck Ko^{a)}, and Byeungwoo Jeon^{a)‡}

요 약

최근, 사용자 제작 콘텐츠(UCC: User Created Contents) 또는 다시점 비디오(Multiview Video) 등의 응용을 위한 경량화 부호화 기술의 필요성이 대두됨에 따라 비디오 부호화 복잡도의 대부분을 차지하는 움직임 예측/보상 과정을 부호화기가 아닌 복호화기 측에서 수행하는 분산 비디오 부호화 기술(Distributed Video Coding)에 대한 연구가 활발히 이루어지고 있다. Wyner-Ziv 부호화 기술은 채널 코딩을 이용하여 원본 영상에 대한 복호화기 측의 예측영상인 보조정보에 포함된 잡음을 제거함으로써 영상을 복원하는 구조를 가진다. 일반적인 Wyner-Ziv 부호화 기술은 키 프레임 간의 움직임 예측/보상 과정에 기반한 프레임 보간법을 통해 보조정보를 생성하며, Shannon limit에 근접한 성능을 보이는 Turbo 코드나 LDPC 코드를 통해 잡음을 제거한다. Wyner-Ziv 부호화 기술은 채널 코드의 복호화를 위해 보조정보에 포함된 잡음의 정도를 예측하는데, 이를 ‘가상 채널 잡음(Virtual Channel Noise)’이라 하며 일반적으로 Laplacian이나 Gaussian으로 모델화 한다. 본 논문은 변환영역에서의 주파수 단위에 적응적인 채널 잡음 모델링에 기반한 Wyner-Ziv 부호화 방법을 제안한다. 다양한 영상에 대한 제안 방법의 실험 결과는 기존 방법과 비교하여 최대 약 0.52dB에 해당하는 율-왜곡 성능의 향상을 보여준다.

Abstract

Recently, as the necessity of a light-weighted video encoding technique has been rising for applications such as UCC(User Created Contents) or Multiview Video, Distributed Video Coding(DVC) where a decoder, not an encoder, performs the motion estimation/compensation taking most of computational complexity has been vigorously investigated. Wyner-Ziv coding reconstructs an image by eliminating the noise on side information which is decoder-side prediction of original image using channel code. Generally the side information of Wyner-Ziv coding is generated by using frame interpolation between key frames. The channel code such as Turbo code or LDPC code which shows a performance close to the Shannon's limit is employed. The noise model of Wyner-Ziv coding for channel decoding is called Virtual Channel Noise and is generally modeled by Laplacian or Gaussian distribution. In this paper, we propose a Wyner-Ziv coding method based on the frequency-adaptive channel noise modeling in transform domain. The experimental results with various sequences prove that the proposed method makes the channel noise model more accurate compared to the conventional scheme, resulting in improvement of the rate-distortion performance by up to 0.52dB.

Keywords : DSC, DVC, Wyner-Ziv coding, Channel noise modeling

a) 성균관대학교 정보통신공학부

School of Information and Communication Engineering, Sungkyunkwan University

‡ 교신저자 : 전병우(bjeon@skku.edu)

* 이 논문은 과학기술부의 재원으로 한국과학재단 국가지정 연구실사업으로 수행된 연구임(R0A-2006-10286-0(2008))

· 접수일(2009년1월30일), 수정일(2009년3월19일), 게재확정일(2009년3월23일)

1. 서론

최근 소비자들의 미디어에 대한 능동적 활용과 미디어에 관련된 각종 콘텐츠에 대한 참여도가 높아짐에 따라 디지털 콘텐츠 시장의 규모가 급격히 성장하였다. 그리고 이와 같은 시장의 흐름에 발맞추어 사용자 제작 콘텐츠로 불리는 개인 제작 비디오 콘텐츠 및 관련 시장 규모가 급증하였다. 이러한 개인 제작 비디오 콘텐츠를 제작하기 위하여는 고가의 전문적인 기기대신 저가의 모바일 캠코더가 더욱 선호되고 있는데 이는 개인 제작 비디오 콘텐츠의 특성상 콘텐츠 제작에 많은 비용이 들이지 않고 전문성이 없어도 제작이 용이하기 때문이다. 하지만 이러한 저가의 모바일 캠코더와 같은 사용 가능한 에너지와 컴퓨팅 자원이 한정적일 수 밖에 없는 모바일 기기를 사용해 고화질 콘텐츠를 압축하기 위하여는 경량화 부호화 기술이 절실히 요구된다. 종래의 MPEG-1/2/4나 H.26x와 같은 표준 압축 기술은 영상에 존재하는 중복성(Redundancy)을 제거하는 일련의 처리과정을 부호화기에서 수행하기 때문에 부호화기 자원이 한정된 환경에서는 압축 성능을 떨어뜨리는 방법으로 복잡도를 낮추어 부호화기를 경량화시킬 수 밖에 없고, 따라서 현재 요구되고 있는 경량화 부호화에는 적합하지 못하다는 문제점이 지적되어 왔다.

이러한 연유로 분산 소스 부호화 기술(DSC : Distributed Source Coding)은 경량화 부호화 문제의 주요 솔루션으로써 관심을 받아왔다. Slepian-Wolf이론과 Wyner-Ziv이론에 기반한 이 기술은 상관도가 있는 여러 소스들을 소스 데이터 사이의 중복성을 제거하는 일련의 처리없이 독립적으로 부호화하는 문제를 다룬다. 그리고 이론적으로 분산 소스 부호화 기술이 소스간 유사도를 이용하는 기존의 부호화 기술의 율-왜곡(R-D: rate-distortion) 성능과 동일할 정도의 성능을 얻을 수 있음을 증명하였다^{1,2}.

위의 이론들은 기존의 비디오 압축 기술에서는 부호화기가 프레임 간 중복성을 줄이기 위한 모든 처리절차를 행하였으나, 부호화기에서 행하였던 모든 처리절차를 부호화이득에 별다른 손실 없이 복호화기로 옮길 수 있음을 의미한다. 분산 소스 부호화 기술에 기반하여 각 소스들을 서로 독립적으로 부호화하며 복호화시 소스간 유사도를 이용하

여 연계 복호화하는 구조의 비디오 부호화 기술에 대한 많은 연구가 이루어졌으며 통칭하여 분산 비디오 부호화 기술(DVC : Distributed Video Coding)로 불린다.

이러한 분산 비디오 부호화 기술 중 현재 가장 대표적인 기술은 Wyner-Ziv 부호화 기술이다. 이 기술은 현재 부호화하는 프레임의 예측정보인 보조정보를 복호화기에서 현재 부호화하는 프레임과 유사한 키 프레임 간의 프레임 보간을 통해 생성한다. 이때 부호화기는 보조정보를 알 수 없으므로 현재 프레임과 보조정보간의 차이를 일종의 잡음으로 간주한다³. 이 잡음은 전송채널에서 발생하는 것이 아니라 복호화기에서 현재 부호화되는 프레임에 유사한 보조정보를 생성할 때 그 정확도가 떨어져 발생하는 것이므로 이를 ‘가상의 채널 잡음(Virtual Channel Noise)’이라 한다. 이 가상의 채널 잡음을 제거하기 위해 Wyner-Ziv 부호화 기술에서는 오류정정능력이 Shannon's limit에 근접하다고 알려진 Turbo 코드나 LDPC 코드를 이용하여 현재 프레임에 대한 패리티 비트를 생성하여 이를 복호화기에 전송한다. 복호화기는 부호화기로부터 패리티 정보를 전송받아 보조정보내의 잡음을 제거하여 영상을 복원하게 된다. 채널 코드의 복호화는 채널 잡음 모델에 기반하여 수행되어지며 일반적으로 Wyner-Ziv 부호화 기술에서는 이를 Laplacian 이나 Gaussian으로 모델화 하여 채널 코드의 복호화기에 입력값으로 들어가는 연판정-입력(Soft-input)을 계산하여 복호화과정이 진행된다. 현재 부호화되는 프레임과 보조정보 사이의 잡음이 적을수록 양자화된 심볼이 나타내는 구간에서 보조정보에 가까운 값으로 복원되고 잡음이 많을수록 채널 잡음 모델은 균일분포에 가까워지므로 구간의 중간값으로 복원이 된다.

따라서 Wyner-Ziv 부호화 기술에서는 채널 잡음 모델을 얼마나 정확히 모델화 하느냐에 따라 전체적인 시스템의 부호화 효율이 큰 영향을 받게 된다. 하지만 복호화기에는 현재 부호화되는 프레임에 대한 정보가 없고 부호화기에는 보조정보가 없기 때문에 부/복호화기 모두 정확한 채널 잡음 모델을 알 수 없고, 따라서 잡음을 제거하기 위해 필요한 패리티 비트의 양 또한 알 수 없다. 그렇기 때문에 Wyner-Ziv 부호화 기술에서는 역방향 채널을 두고, 잡음이 제거될 때까지 복호화기는 부호화기로부터 패리티 비트를 점진적

으로 전송받는다. 하지만 부정확한 채널 잡음 모델에 따른 채널 코드의 연관성 입력의 신뢰도 저하는 잡음 제거를 위한 패리티 비트의 증가를 야기해 부호화 효율의 저하를 가져오게 된다. 또한 변환영역에서의 주파수 밴드별로 보조 정보 상의 잡음의 차이 때문에 주파수 밴드별 채널 잡음의 차이를 가져오게 되어 채널 잡음 모델의 부정확성을 야기한다. 따라서 이를 해결하기 위해 원본 영상에 대한 정보를 전송받아 주파수 밴드별로 다른 채널 잡음 모델을 예측하는 방식의 접근이 가능하다. 기존에는 보조정보를 생성할 때 사용하는 키 프레임 간의 시간적 유사성을 이용해 프레임 단위로 채널 잡음 모델을 예측하는 방식의 접근이 있었다^[4]. 하지만 복호화기에 현재 부호화되는 프레임에 대한 정보가 없는 상태에서 키 프레임 간의 시간적 유사성만을 이용해서 채널 잡음 모델을 정확하게 예측한다는 것은 쉽지 않으며 특히 프레임 간 움직임이 복잡하거나 occlusion이 있는 경우엔 키 프레임 간 시간적 유사성이 떨어지기 때문에 더욱 어려워진다. 따라서 키 프레임 간의 시간적 유사성뿐 아니라 프레임내의 공간적 유사성도 함께 이용하여 채널 잡음 모델을 예측하는 방식을 통해 이 문제를 해결하려는 접근이 나왔다^[5]. 이 방법은 블록 단위 혹은 화소 단위로 채널 잡음 모델을 예측하여 키 프레임 간의 시간적 유사성뿐 아니라 프레임내의 공간적 유사성도 이용하여 채널 잡음 모델의 성능을 향상시켰다. 하지만 단순히 키 프레임 간의 차를 이용해 시간적 유사성을 계산하여 채널 잡음을 예측하는 접근 방법은 위에서 언급한 것과 같이 프레임 간 움직임이 복잡하거나 occlusion이 있는 경우 프레임 간의 시간적 유사성이 떨어지는 구간에서 제대로 된 채널 잡음 모델을 예측하기 어렵다. 따라서 기존의 방법^[6]은 변환영역에서 시·공간적 유사성을 이용하여 채널 잡음을 계산할 때 단순히 DCT(Discrete Cosine Transform) 변환을 통해 구한 키 프레임의 DCT 변환계수만을 이용하는 것이 아니라 각 DCT 밴드의 평균도 이용하여 거리(Distance)라는 개념을 정의하고 이를 통해 현재 프레임과 보조정보 사이의 실제 잡음을 예측하는 방법을 사용하였다. 또한, Laplacian 계수가 너무 커지는 것을 방지하기 위해 문턱값을 사용하였는데 이는 실제로는 잡음이 존재하나 채널 잡음 모델이 제대로 예측하지 못하는 경우, 즉 보조정보 내에 잡음이 없

거나 거의 없다고 판단하게 되는 경우엔 Laplacian 계수가 무한대의 값을 가지거나 혹은 실제값보다 무척 큰 값을 가지게 되는 문제가 발생하기 때문이다. 그렇기 때문에 이전 방법^[7]에서는 각 밴드의 DCT 변환계수들의 분산을 DCT 변환계수들이 그 밴드의 평균 주위로 어느 정도로 분포되어 있는지 측정할 수 있는 수단이라는 것에 근거하여 Laplacian 계수에 대한 잘못된 예측을 방지하기 위한 문턱값으로 사용하였다.

하지만 일반적으로 고주파로 갈수록 원본 영상과 보조정보의 값이 작아지기 때문에 잡음의 양 또한 줄어든다. 따라서 이전 방법^[7]처럼 단순하게 모든 주파수 밴드에 대해 고정된 값을 문턱값으로 사용하게 되면 실제 잡음의 값이 작아서 Laplacian 계수가 커져야 하는 경우에 고정된 문턱값이 이를 방해한다. 이를 해결하기 위해 각 밴드별로 서로 다른 문턱값을 사용하는 방법이 가능하다. 이에 본 논문은 Template matching^[8]을 이용해 더욱 정교하게 예측된 잡음과 부호화기에서 전송된 원본영상에 대한 정보를 기반으로 실제 잡음의 분산을 예측하여 주파수에 적응적인 개선된 채널 잡음 모델링 방법을 사용하는 변환영역 Wyner-Ziv 부호화 기술(TDWZ: Transform-domain Wyner-Ziv coding)을 제안한다.

II. 제안하는 변환영역 Wyner-Ziv 부호화 방법

1. 부호화기의 구조

본 논문에서 사용하는 Wyner-Ziv 이론 기반의 경량화 비디오 부호화 구조는 그림 1과 같다. Wyner-Ziv 부호화 기술에서 프레임은 키 프레임과 Wyner-Ziv 프레임으로 구분되며 이 중 키 프레임은 통상 종래의 비디오 부호화 기술에 기반하여 인트라 부호화 되어 복호화기로 전송된다.

본 논문에서는 매 홀수 번째 프레임이 이에 해당하며 부호화에는 H.264/AVC intra codec이 사용되었다. 키 프레임 부호화 후에는 DCT 변환, DCT 변환계수들에 대한 주파수 밴드 형성, 양자화, 비트플레인 생성, LDPC 부호화, Wyner-Ziv 프레임에 대한 패리티 비트의 생성을 순차적으

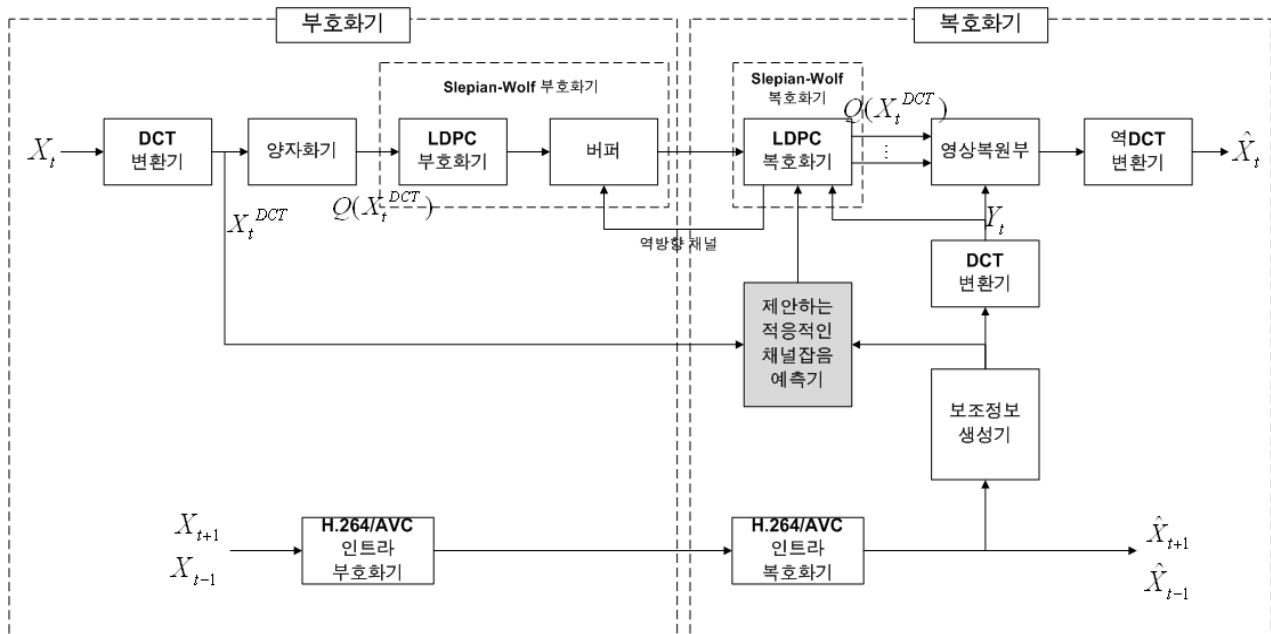


그림 1. 제안하는 변환영역 Wyner-Ziv 부호화 구조
 Fig. 1. Proposed Transform-domain Wyner-Ziv coding structure

로 진행한다.

인트라 부호화된 키 프레임이 복호화기로 전송되고 나면 부호화기는 Wyner-Ziv 프레임을 식 (1)과 같이 4x4 정수 DCT를 취하여 변환한다. 왜냐하면 DCT 변환을 통해 영상 내 공간적 중복성을 줄일 수 있어 향상된 압축 성능을 기대할 수 있기 때문이다. 이때 변환과정의 스케일링은 역변환 과정에 포함된다.

$$\begin{aligned}
 F &= ICT \times f \times ICT^T \\
 f' &= ICT^T \times (F \otimes S \otimes S') \times ICT
 \end{aligned}
 \tag{1}$$

식 (1)에서 F는 변환영역의 4x4블록, ICT는 4x4 integer cosine transform matrix, f는 화소영역의 4x4블록, ICT^T는 ICT의 transposed matrix, f'는 역 변환된 화소영역의 4x4블록, ⊗는 두 행렬의 동일 위치 요소들 간 곱 연산을 뜻하며 S는 4x4 스케일링 행렬, S'은 4x4 역 스케일링 행렬을 의미한다.

부호화기는 DCT 변환 후 프레임 내 동일 DCT 변환계수 성분들을 모아 주파수 밴드를 형성한다. 형성된 각 주파

수 밴드에 대해서는 각각의 밴드마다 서로 다른 양자화 레벨을 적용하여 양자화 한다^[9]. 그림 2는 저전송률(M1)에서 고전송률(M8)에 상응하는 양자화 표^[9]를 나타내며 표 안의 숫자는 각 밴드에 해당하는 양자화 레벨의 수를 나타낸다. 예를 들어 그림 2의 M8을 보면 DC 밴드(1행 1열)에는 128개의 양자화 레벨이 주어지고 첫 번째(1행 2열), 두 번째(2행 1열) AC 밴드에는 64개의 양자화 레벨이 주어진다. 따라서 DC 밴드에 대해서는 총 7개의 비트플레인(2⁷ = 128)이 생성된다. 같은 방법으로 첫 번째와 두 번째 AC 밴드에는 각각 6개의 비트플레인이 생성된다. 이 때 DC 밴드에 대해서는 uniform scalar 양자화를 수행하며 dynamic range는 [0, 2¹²)로 고정한다. 원래는 [0, 2¹⁰)이나 스케일링을 역변환 과정에 포함하였기 때문에 DC 밴드의 스케일링 요소인 0.25의 역수만큼 dynamic range가 확장되었다. 그리고 나머지 AC 밴드에 대해서는 dead-zone 양자화를 수행하는데 DC 밴드와는 달리 각 밴드의 dynamic range를 고정하지 않고 밴드별로 각 계수의 절대값을 측정하여 [-Max Val_b, Max Val_b)로 사용한다. 여기서 Max Val_b는 b 번째 밴드의 계수 중 절대값의 크기가 가장 큰 값으로 부호

16	8	0	0	32	8	0	0	32	8	4	0	32	16	8	4
8	0	0	0	8	0	0	0	8	4	0	0	16	8	4	0
0	0	0	0	0	0	0	0	4	0	0	0	8	4	0	0
0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
M1				M2				M3				M4			
32	16	8	4	64	16	8	8	64	32	16	8	128	64	32	16
16	8	4	4	16	8	8	4	32	16	8	4	64	32	16	8
8	4	4	0	8	8	4	4	16	8	4	4	32	16	8	4
4	4	0	0	8	4	4	0	8	4	4	0	16	8	4	0
M5				M6				M7				M8			

그림 2. 윌 제어를 위한 양자화 표(저전송률(M1) ~ 고전송률(M8))^[9]
 Fig. 2. Quantization matrices for rate control(low bit-rate(M1) ~ high bit-rate(M8))^[9]

화되는 모든 AC 밴드에 대해 복호화기로 전송된다.
 양자화 과정을 통해 생성된 각 밴드의 양자화 심볼스트림은 MSB에서 LSB에 이르는 비트플레인으로 재구성되며 각각의 비트플레인은 이후 독립적으로 LDPC 부호화기에 입력된다. 그림 1의 LDPC 부호화기^[10]는 각 밴드의 양자화 심볼스트림의 각 비트플레인을 입력으로 받아 syndrome 비트를 생성하며 생성된 syndrome 비트는 버퍼에 저장하였다가 복호화기의 요청에 따라 그 중 일부를 복호화기에 전송한다.

2. 복호화기의 구조

Wyner-Ziv 부호화 기술의 복호화기에는 현재 부호화되는 프레임에 대한 아무런 정보가 없기 때문에 일반적으로 부호화기에서 전송된 키 프레임들 사이의 선형적인 움직임을 가정한 프레임 보간법을 이용하여 보조정보를 생성한다^[11]. 이에 대한 예로써 생성된 Foreman 영상의 보조정보를 그림 3^[12]에서 나타낸다. 그림 3(a)와 같이 프레임 간에 움직임이 많지 않은 경우에 보조정보를 생성하면 객관적 그리고 주관적으로 뛰어난 화질의 보조정보가 생성된다. 반면에 프레임 간 움직임이 복잡한 경우나 occlusion이 존재할 경우에는 선형적인 움직임을 가정한 프레임 보간법^[11]을 이용하기 때문에 그림 3(b)와 같이 움직임 예측을 잘못할 가능성이 높아진다. 따라서 그림 3(b)의 생성된 보조정보

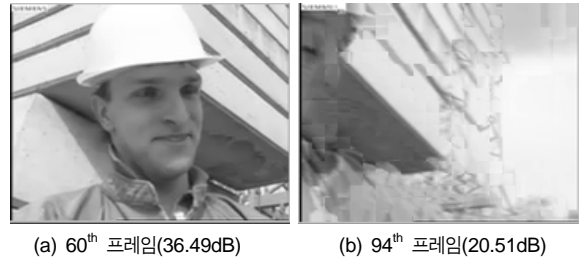


그림 3. 보조정보 비교(QP = 25, 포만 시퀀스, 지멘스 로고 있는 것)^[12]
 Fig. 3. Side information comparison(QP = 25, Foreman sequence with Siemens logo)^[12]

는 그림 3(a)의 보조정보에 비해 상대적으로 많은 오류를 포함하게 되며 이는 전체 시스템의 성능 저하의 원인이 된다. 보조정보 생성 후에는 채널 잡음 모델 예측, 채널 복호화, 영상복원, 역 DCT 변환이 진행된다.

일반적으로 Wyner-Ziv 부호화 기술에서 현재 부호화되는 프레임과 보조정보 간 채널 잡음은 Laplacian이나 Gaussian 분포로 모델화 되는데 본 논문에서는 Laplacian 분포를 사용한다. 채널 잡음 모델은 Laplacian 계수값만으로 정의가 가능하지만 그림 3에서 나타나듯이 프레임 간 움직임이 많고 적응에 따라 보조정보 내에 발생하는 채널 잡음의 양이 주로 결정되므로 한 프레임 내에서 non-stationary한 특성을 띤다. 따라서 채널 잡음을 모델링하기 위해 모든 변환계수에 각각 Laplacian 계수값을 구하여 적용한다^[5]. 그리고 현재 DVC 기술의 선두 연구그룹 중 하나인 DISCOVER에서 제안하는

방법^[6]과 달리 본 논문에서는 채널 잡음을 예측할 때 단순히 키 프레임 간의 차이만을 이용하는 것이 아니라 그림 1에서 보는 바와 같이 Template matching^[8]을 통해 잘못된 움직임 예측에 의해 보조정보가 생성될 시 추가적인 잡음을 더하여 더욱 정교하게 잡음을 예측한다. 또한, 일반적으로 고주파로 갈수록 잡음의 양이 줄어들기 때문에 각 주파수 밴드별로 서로 다른 문턱값을 적용하는데 부호화기로부터 원본 영상에 대한 정보를 전송받아 실제 잡음의 분산을 예측하여 이를 문턱값에 사용한다.

본 논문의 LDPC 복호화기^[10]는 연관정-입력을 계산하고 여기에 부호화기로부터 전송되는 syndrome 비트를 더해 복호화 과정을 수행하게 되는데 이를 위해 보조정보와 채널 잡음 모델로부터 받은 Laplacian 계수를 이용한다. 따라서 연관정-입력이 주는 정보의 양과 그 신뢰도에 따라 윌-왜곡 성능이 달라진다. 그리고 이전에 복호된 비트플레인의 결과를 현재 비트플레인의 복호화 과정에 이용함으로써 채널 잡음 모델이 정교하게 되고 이로 인해 연관정-입력의 엔트로피를 더욱 낮출 수 있다. 즉, LDPC 복호화기는 연관정-입력과 syndrome 비트를 이용해 복호화 과정을 수행하고 그 결과로 연관정-출력(Soft- output)을 산출한다. 그리고 그 결과로부터 에러율을 산출하여 사전에 정의한 임계값(예: 10^{-4}) 이하로 에러율이 내려가지 않는 경우 부호화에 추가적인 패리티 요청을 하고 복호화를 반복한다. 몇 차례의 이러한 패리티 요청과 전송을 통하여 에러율이 임계값 이하로 내려가면 LDPC 복호화기는 패리티 요청을 종료하고 다음 비트플레인에 대한 복호화를 수행한다. 하지만 에러율이 임계 에러율 이하로 내려갔는지의 여부를 정확하게 판단하기에는 이 값의 신뢰도가 너무 낮기 때문에 실제로 이를 판단하기 위해 부호화기로부터 CRC 비트를 전송받아 이를 판별한다^[13]. CRC 체크를 통해 에러율이 임계값 이하로 내려간 것을 확인하면 복호화를 끝마치고 그렇지 않을 경우에는 피드백 채널을 통하여 부호화기에 추가적인 패리티 요청을 한다.

한 주파수 밴드에 대한 모든 비트플레인이 복호화되면 복호화기는 원본 변환계수의 양자화 심볼스트림을 복원한 후 영상복원기로 보낸다. 영상복원기는 보조정보와 복원된 양자화 심볼 그리고 채널 잡음 모델에 기반하여 원본정보

의 기대값을 구하는 과정을 통해 양자화 과정에서 버려진 이 심볼스트림의 나머지 하위 비트들을 보상한다.

III. 제안하는 주파수 적응 채널 잡음 모델링 방법

1. 가상의 채널 잡음(Virtual Channel Noise) 예측 방법 및 과정

Wyner-Ziv 부호화 기술에서는 원본 영상과 보조정보 사이의 잡음을 가상의 채널 잡음으로 간주하고 이를 채널 코드를 이용해 복호화한다. 하지만 복호화기에는 원본 영상에 대한 정보가 없고 부호화기에는 보조정보에 대한 정보가 유효하지 않기 때문에 Wyner-Ziv 부호화 기술에서 가상의 채널 잡음을 예측하는 것은 어렵다. 따라서 실제적으로는 복호화기에서 키 프레임을 이용하여 가상의 채널 잡음을 예측한다^[4,5,6,7]. 하지만 Wyner-Ziv 부호화 기술에서는 프레임 간 선형적인 움직임을 가정한 프레임 보간법^[11]을 사용하여 보조정보를 생성하기 때문에 프레임 간 움직임이 많거나 occlusion이 발생할 경우 잘못된 움직임 예측이 발생할 가능성이 높아진다. 즉, 이런 경우에는 단순히 키 프레임만을 가지고 예측한 잡음은 실제 잡음과 더욱 많은 차이를 보일 수 있다. 따라서 본 논문에서는 더욱 정교하게 가상의 채널 잡음을 예측하기 위해 Template matching^[8]을 이용한다. Template matching이란 현재 부호화되는 블록의 주변 블록을 비교하여 움직임 예측을 하는 방법인데 본 논문에서는 이를 이용하여 보조정보의 각 블록의 주변 블록과 이에 대응되는 키 프레임의 주변 블록을 비교하여 잘못된 움직임 예측이 발생한 곳을 찾는다. 그 후, 실험을 통해 찾은 문턱값보다 에러가 많이 발생하면 움직임 예측이 잘 이루어지지 않았다고 판단하고 주변 블록에서 발생한 잡음을 기존에 키 프레임을 이용하여 예측한 가상의 채널 잡음에 더하게 된다.

2. 주파수 적응 채널 잡음 모델링 방법

일반적으로 고주파로 갈수록 원본 영상과 보조정보의 값

이 작아지기 때문에 잡음의 양 또한 작아지게 된다. 따라서 주파수 밴드별로 잡음의 양이 다른데 이전 방법^[7]에서는 모든 주파수 밴드에 대해 고정된 문턱값을 사용하였다. 따라서 본 논문에서는 일반적인 잡음의 경향을 고려하여 각 주파수 밴드별로 다른 문턱값을 적용하였다. 그리고 문턱값을 결정하기 위해서는 실제 잡음의 분산이 필요한데 Wynzer-Ziv 부호화 기술에서는 복호화기에 원본 영상에 대한 정보가 없기 때문에 이를 구할 수 없다. 따라서 본 논문에서는 부호화기로부터 원본 영상의 정보를 전송받아 실제 잡음의 분산을 예측한다.

$$\begin{aligned} \text{Var}(X-Y) &= E[(X-Y)^2] - (E[X-Y])^2 \\ &= E[X^2 - 2XY + Y^2] \quad (\because X-Y=N, (E[X-Y])^2=0) \\ &= E[X^2] - 2E[XY] + E[Y^2] \end{aligned} \quad (2)$$

식 (2)는 실제 잡음의 분산을 구하는 과정으로 여기서 Var 은 분산, X 와 Y 는 각각 원본 영상과 보조정보, E 는 평균을 의미한다. N 은 원본 영상과 보조정보 사이의 실제 잡음을 나타내는데 random noise 이므로 식 (2)와 같이 N 의 평균을 0으로 볼 수 있다. 식 (2)의 $E[XY]$ 는 식 (3)과 같이 표현될 수 있다.

$$\begin{aligned} E[XY] &= E[X]E[Y] + \text{Cov}(X, Y) \\ &= E[X]E[Y] + E[(X-E[X])(Y-E[Y])] \\ &= E[X]E[Y] + E[(X-E[X])(X-N-E[X-N])] \\ & \quad (\because Y=X-N) \\ &= E[X]E[Y] + E[(X-E[X])(X-N-E[X])] \\ & \quad (\because E[N]=0) \\ &= E[X]E[Y] + E[(X-E[X])^2] - N(X-E[X]) \\ &= E[X]E[Y] + E[(X-E[X])^2] - \text{Cov}(N, X) \quad (\because E[N]=0) \end{aligned} \quad (3)$$

여기서 Cov 는 공분산을 뜻하며 N 은 영상에 의존적이지 않고 영상내의 움직임에 의존적이므로 N 과 X 의 관계가 서로 독립적이라고 가정하면 $\text{Cov}(N, X) = 0$ 이 되어 식 (3)은 식 (4)와 같이 정리된다.

$$\begin{aligned} E[XY] &= E[X]E[Y] + E[(X-E[X])^2] \\ &= E[X]E[Y] + \sigma_x^2 \end{aligned} \quad (4)$$

식 (4)에서 σ_x^2 은 X 의 분산을 나타낸다. 최종적으로 $\text{Var}[X-Y]$ 는 식 (5)와 같이 나타낼 수 있다.

$$\begin{aligned} \text{Var}(X-Y) &= E[X^2] - 2\{E[X]E[Y] + \sigma_x^2\} + E[Y^2] \\ &= E[X^2] - 2E[X]E[Y] - 2E[X^2] + 2E[X]^2 + E[Y^2] \\ & \quad (\because \sigma_x^2 = E[X^2] - E[X]^2) \\ &= E[Y^2] - E[X^2] - 2\{E[X]E[Y] - E[X]^2\} \end{aligned} \quad (5)$$

식 (5)를 통해 실제 잡음의 분산을 계산하기 위해서는 복호화기에 $E[X]$ 와 $E[X^2]$ 의 값이 필요하다는 것을 확인하였다. 하지만 $E[N] = 0$ 이란 가정은 다수의 샘플값이 존재한다는 가정 하에 성립될 수 있기 때문에 $E[N]$ 이 0이 아닌 값을 가지게 되면 식 (5)에 의한 계산이 틀리게 되어 예측한 실제 잡음의 분산이 음수가 나오는 경우가 발생한다. 따라서 본 논문에서는 식 (5)를 그대로 시스템에 적용시키지 않고 예측한 실제 잡음의 분산이 음의 값을 가지는 것을 방지하기 위해 식 (5)의 결과에 절대값을 취하여 실제 잡음의 분산으로 예측한다.

부호화기에서 복호화기로 보내주는 원본 영상에 대한 정보는 시스템의 윌 성능을 저하시키는 원인이 되기 때문에 본 논문은 전체 시스템의 윌-왜곡 성능에 크게 영향을 미치는 일정량의 저주파 밴드에 대해서는 원본 영상의 정보를 이용하여 예측한 분산을 문턱값에 적용하고 나머지 고주파 밴드에는 예측된 가상의 채널 잡음의 분산을 통해 문턱값을 구한다. 즉, 부호화기에서 전송된 $E[X]$ 와 $E[X^2]$ 의 값을 이용하여 실제 잡음의 분산을 예측하고 이를 통해 각 주파수 밴드별로 다른 채널 잡음 모델의 문턱값을 결정한다.

IV. 실험 및 결과

1. 실험 방법

본 논문에서 제안하는 변환영역의 Wyner-Ziv 부호화 방법의 성능을 평가하기 위해 다음 세 가지 부호화 방법의 윌-왜곡 성능을 비교하였다.

- DISCOVER: DISCOVER 그룹의 TDWZ^[14,15]
- Proposed method: 주파수 적응 채널 잡음 모델링을 하는 제안 TDWZ
- H.264/AVC intra only: JM 12.2

구체적 실험 조건은 다음과 같다.

- 영상: 15Hz 150프레임의 QCIF 사이즈 Foreman(지멘스로고 있는 것), Hall monitor, Soccer, Stefan 영상
- 양자화: M1 ~ M8의 8개의 양자화 표에 대하여 상응하는 (표 1의)키 프레임의 QP는 표 1과 같다.
- 보조정보 생성기: 양 키 프레임 간 움직임 예측을 위한 블록 크기는 8x8이며 움직임 예측을 위한 탐색영역은 ± 32 .

표 1. 양자화 파라미터

Table 1. Quantization setting

Wyner-Ziv 프레임 (양자화 표 값)		1	2	3	4	5	6	7	8
키 프레임 (QP for H.264/AVC Intra)	Foreman	40	39	38	34	34	32	29	25
	Hall monitor	37	36	36	33	33	31	29	24
	Soccer	44	43	41	36	36	34	31	25
	Stefan	40	39	38	34	34	32	29	25

전송률과 PSNR은 Wyner-Ziv 프레임과 키 프레임을 포함하는 모든 프레임에 대한 값이며 PSNR은 원본 영상과 복원된 영상 간에 측정되었다.

2. 실험 결과

그림 4는 실험에 사용한 영상들에 대한 율-왜곡 성능을 나타낸다. 제안하는 주파수 적응 채널 잡음 모델의 성능을

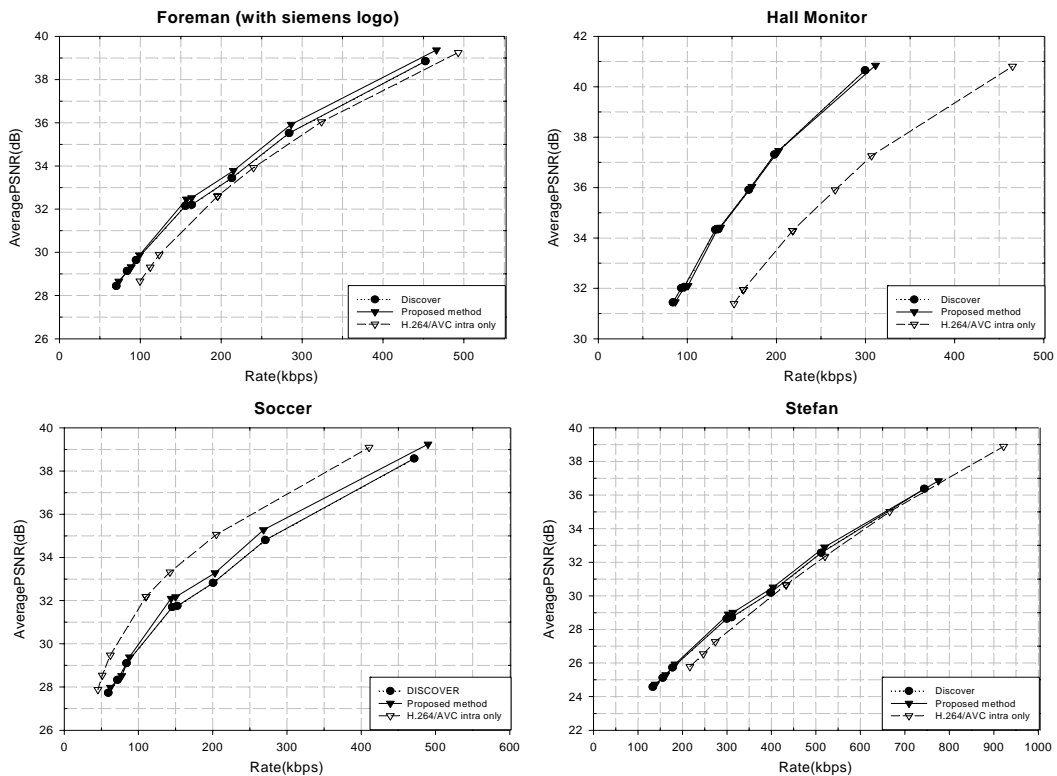


그림 4. R-D 성능 비교
Fig. 4. R-D performance comparison

평가하기 위하여 제안 방법을 적용한 Proposed method와 DISCOVER의 실험결과를 비교하면 우선 Foreman, Soccer, Stefan 영상과 같이 비교적 움직임이 복잡한 영상에서는 제안 방법을 통하여 채널 잡음 모델이 개선되었음을 실험결과를 통해 볼 수 있는데 DISCOVER 방법과 비교했을 때 최대 0.52dB에 해당하는 PSNR 이득을 보인다. 이 결과는 채널 잡음이 많이 발생하는 영상에서 제안하는 방법이 DISCOVER 방법에 비해 때 더 정확하게 실제 잡음을 예측하고 적절한 문턱값을 정하였다는 것을 나타낸다. 그렇지만 저전송률로 갈수록 이득을 얻지 못하는데 이는 저전송률로 갈수록 키 프레임 상에 존재하는 양자화 에러가 많아지기 때문이다. 원래 움직임이 복잡하여 잡음이 많이 발생하는데 여기에 양자화 에러까지 더해짐으로 실제 잡음이 더 커져 제안하는 방법이 정확하게 예측하지 못하게 된다. 그리고 Stefan 영상과 같이 움직임이 복잡하면서 배경이 복잡한 영상은 잡음을 예측하기 어려워 고전송률에서도 다른 영상에 비해 많은 성능 개선을 보이지는 못한다. Hall monitor 영상과 같이 움직임이 적은 영상에서는 제안 방법을 통한 성능 향상이 거의 없음을 알 수 있다. 이는 영상 내 움직임이 적거나 단순한 경우 보조정보 생성 시 이미 상당한 화질의 보조정보를 생성할 수 있어 보조정보 내의 잡음의 양이 줄어들게 된다. 따라서 채널 코드를 이용해 보조정보 내의 잡음을 제거하여 얻을 수 있는 이득 또한 줄어들어 채널 잡음 모델이 성능에 큰 영향을 미치지 못하게 되고 결과적으로 제안 방법을 통한 채널 잡음 모델 성능 개선의 여지가 줄어든다. 하지만 결과로부터 고전송률일 때 보다 저전송률일 때 성능 향상이 발생하는 것을 알 수 있는데 이는 저전송률일수록 두 키 프레임 상에 존재하는 양자화 에러도 더 많아서 채널 잡음 모델로 인해 개선될 수 있는 부분이 더 많아지기 때문이다.

IV. 결론 및 향후 연구과제

본 논문에서는 Template matching^[8]을 이용해 더욱 정교하게 예측된 잡음과 부호화기에서 전송된 원본영상에 대한 정보를 기반으로 실제 잡음의 분산을 예측하여 주파수에

적응적인 개선된 채널 잡음 모델링 방법을 사용하는 변환영역 Wyner-Ziv 부호화 기술을 제안하였다. 움직임이 많아 실제 잡음이 많이 발생하는 실험 영상의 모의실험에서는 제안 방법을 통하여 더 정확하게 실제 잡음을 예측하고 적절한 문턱값을 찾아서 채널 부호화기가 요구하는 패리티 비트양이 줄어들고 영상의 화질도 좋아져서 개선된 윌-왜곡 성능을 얻을 수 있음을 확인할 수 있었다. 이후의 연구에서는 채널 잡음 모델을 더욱 발전시켜 더 정확하게 실제 잡음을 예측하고 저전송률에서도 전체 시스템의 성능을 개선할 수 있는 방안에 대해서 연구할 예정이다.

참고 문헌

- [1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. IT-19, pp. 471 - 480, July 1973.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, vol. 22, pp. 1-10, July 1976.
- [3] A. Aaron, S. Rane, R. Zhang, and B. Girod, "Wyner-Ziv coding for video: Applications to compression and error resilience," Proc. of IEEE Data Compression Conference, Snowbird, UT, pp. 93-102, March 2003.
- [4] C. Brites, J. Ascenso, and F. Pereira, "Modeling Correlation Noise Statistics at Decoder for Pixel Based Wyner-Ziv Video Coding," Picture Coding Symposium, Beijing, China, April 2006.
- [5] C. Brites, J. Ascenso and F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," IEEE International Conference on Image Processing, Atlanta, USA, pp. 273-276, October 2006.
- [6] C. Brites and F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner-Ziv Video Coding," IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 9, pp. 1177-1190, September 2008.
- [7] 김병희, 고봉혁, 전병우, "변환영역에서의 Wyner-Ziv 코딩을 위한 개선된 상관 잡음 모델," 방송공학회 학술대회, pp. 81-84, 2008년 11월.
- [8] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and Choong Seng Boon, "Inter frame coding with template matching spatiotemporal prediction," Proc. IEEE Int. Conf. Image Processing, vol. 1, pp. 465-468, October, 2004.
- [9] C. Brites, J. Ascenso and F. Pereira, "Improving Transform domain Wyner-Ziv video coding performance," IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, vol. 2, pp. 525-528, May 2006.
- [10] D. Varodayan, A. Aaron and B. Girod, "Rate-adaptive distributed source coding using Low-Density Parity-Check codes," Proc.

Asilomar Conference on Signals and Systems, Pacific Grove, CA, pp. 1203-1207, November 2005.

- [11] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Slovak Republic, July 2005.
- [12] 고봉혁, 심혁재, 전병우, "복호 신뢰도에 기반하여 점진적으로 보조정보를 향상시키는 변환영역 Wyner-Ziv 부호화 방법," 방송공학회논문지, 제13권, 제6호, pp. 892-904, 2008년 11월.
- [13] D. Kubasov, K. Lajnef and C. Guillemot, "A hybrid encoder/decoder rate control for a Wyner-Ziv video codec with a feedback channel," IEEE MultiMedia Signal Processing Workshop, Chania,

Crete, pp. 251-254, October 2007.

- [14] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," Picture Coding Symposium, Lisbon, Portugal, November 2007.
- [15] <http://www.discoverdvc.org/>
- [16] 고봉혁, 심혁재, 전병우, "공간적 유사성과 심볼단위 오류정정 채널 코드를 이용한 경량화 비디오 부호화 방법," 방송공학회논문지, 제13권, 제2호, pp. 188-199, 2008년 3월.
- [17] B. Ko, H. J. Shim, and B. Jeon, "Wyner-Ziv coding with spatio-temporal refinement based on successive turbo decoding," IEEE International Conference on Multimedia and Expo, Hannover, Germany, pp. 785-788, June 2008.

저 자 소 개



김 병 희

- 2007년 : 성균관대학교 정보통신공학부 졸업(학사)
- 2008년 ~ 현재 : 성균관대학교 정보통신공학부 석사과정
- 주관심분야 : 멀티미디어 신호처리, 영상압축, 신호처리



고 봉 혁

- 2006년 : 아주대학교 전자공학부 졸업(학사)
- 2008년 : 성균관대학교 정보통신공학부 졸업(석사)
- 주관심분야 : 멀티미디어 신호처리, 영상압축, 신호처리



전 병 우

- 1985년 : 서울대학교 전자공학과 졸업(학사)
- 1987년 : 서울대학교 전자공학과 졸업(석사)
- 1992년 : Purdue Univ, School of Elec. 졸업(공학박사)
- 1993년 ~ 1997년 : 삼성전자 신호처리연구소 수석연구원
- 1997년 ~ 현재 : 성균관대학교 정보통신공학부 교수
- 주관심분야 : 멀티미디어 신호처리, 영상압축, 영상인식, 신호처리