

변환영역에서의 Wyner-Ziv 코딩을 위한 개선된 상관 잡음 모델

*김병희 **고봉혁 ***전병우

성균관대학교

*huhu0930@skku.edu **o3eda@skku.edu ***bjeon@skku.edu

Improved Correlation Noise Modeling for Transform-Domain Wyner-Ziv Coding

*Kim, ByungHee **Ko, BongHyuck ***Jeon, Byeungwoo

Sungkyunkwan University

요약

최근 센서네트워크와 같은 에너지 제한 환경을 위한 경량화 부호화 기술의 필요성이 대두됨에 따라 분산 소스 부호화 기술(Distributed Source Coding)의 응용기술로 비디오 부호화 복잡도의 대부분을 차지하는 움직임 예측/보상과정을 부호화기가 아닌 복호화기에서 수행하는 분산 비디오 부호화 기술(Distributed Video Coding)에 대한 연구가 활발히 이루어져 왔다. 이에 가장 대표적인 기술인 Wyner-Ziv 코딩 기술은 채널 코드를 이용하여 원본 프레임과 이에 대한 복호화기의 예측영상인 보조정보 사이의 잡음을 제거하여 영상을 복원한다. 일반적으로 보조정보는 원본영상에 유사한 키 프레임간의 프레임 보간을 통하여 생성되며 채널 코드는 Shannon limit에 근접한 성능을 보이는 Turbo 코드나 LDPC 코드가 사용된다. 이와 같은 채널 코드의 복호화는 채널 잡음 모델에 기반하여 수행되어지며 Wyner-Ziv 코딩 기술에서는 이 채널 잡음 모델을 '상관 잡음 모델'(Correlation Noise Modeling)이라 하고 일반적으로 Laplacian이나 Gaussian으로 모델화 한다. 하지만 복호화기에는 원본 영상에 대한 정보가 없기 때문에 정확한 상관 잡음 모델을 알 수 없으며 잡음 모델에 대한 예측의 부정확성은 잡음 제거를 위한 패리티 비트의 증가를 야기해 부호화 기술의 압축 성능 저하를 가져온다. 이에 본 논문은 원본 프레임과 보조정보 사이의 잡음을 정확하게 예측하여 잡음을 정정할 수 있는 향상된 상관 잡음 모델을 제안한다. 제안 방법은 잘못된 잡음 예측에 의해 Laplacian 계수가 너무 커지는 것을 방지하면서 영상내의 잡음의 유무에 별다른 영향을 받지 않는 새로운 문턱값을 사용한다. 다양한 영상에 대한 제안 방법의 실험 결과는 평균적으로 약 0.35dB에 해당하는 율-왜곡 성능 향상을 보여주었다.

1. 서론

H.264/AVC와 같은 기존의 비디오 코딩 기술에서는 부호화기의 복잡도와 에너지소비의 대부분을 움직임 예측/보상과정이 차지하였다. 하지만 센서네트워크와 같은 에너지 제한 환경을 위한 경량화 부호화 기술의 필요성이 대두됨에 따라 분산 소스 부호화 기술(DSC: Distributed Source Coding)의 응용기술로 부호화기가 아닌 복호화기에서 움직임 예측/보상과정을 수행하는 분산 비디오 부호화 기술(DVC: Distributed Video Coding)에 대한 연구가 활발히 이루어져 왔다[4, 5]. 이에 가장 대표적인 기술인 Wyner-Ziv 코딩 기술은 채널 코드를 이용하여 원본 프레임과 이에 대한 복호화기의 예측영상인 보조정보 사이의 잡음을 제거하여 영상을 복원한다[4]. 따라서 이러한 구조에서는 부호화기가 프레임간의 유사성을 이용하는 어떠한 과정도 수행하지 않고 단순히 현재 프레임에 대한 인트라코딩만을 수행하기 때문에 매우 경량화된 구조를 가질 수 있다.

Wyner-Ziv 코딩 기술에서 프레임은 키 프레임과 Wyner-Ziv 프레임(WZ 프레임)으로 구분되며 이 중 키 프레임은 통상 종래의 비디오 부호화 기술에 기반하여 인트라 부호화 되고 복호화기로 전송된다. 복호화기는 원본 영상에 관한 정보가 유효하지 않기 때문에 전송된 키 프레임과 WZ 프레임간의 유사도에 기반하여 보조정보를 생성하며 이는 일반적으로 프레임 보간법에 기반한다[4]. 생성된 보조정보와 원본 영상 간에 발생하는 차이로 인해 보조정보는 잡음이 섞여있는 원본 영

상으로 간주되며 Wyner-Ziv 복호화기는 채널 코드를 이용하여 보조정보 상의 잡음을 제거하여 영상을 복원한다. 채널코드는 Shannon limit에 근접한 성능을 보이는 Turbo 코드나 LDPC 코드가 일반적으로 사용된다. 이러한 채널코드의 복호화는 채널잡음 모델에 기반하여 수행되어지며 Wyner-Ziv 코딩 기술에서는 이 채널 잡음 모델을 '상관 잡음 모델'(Correlation Noise Modeling)이라 하고 일반적으로 Laplacian이나 Gaussian으로 모델화 한다. 하지만 복호화기에는 원본 영상에 대한 정보가 없기 때문에 정확한 상관 잡음 모델을 알 수 없으며 잡음 모델에 대한 예측의 부정확성은 잡음 제거를 위한 패리티 비트의 증가를 야기해 부호화 기술의 압축 성능 저하를 가져온다.

본 논문은 상관 잡음 모델에서 잘못된 잡음 예측에 의해 잡음이 없거나 작다고 가정하는 경우에 Laplacian 계수가 실제와는 달리 너무 큰 값을 가지는 것을 방지하기 위한 새로운 문턱값을 이용하여 영상내의 잡음의 유무에 별다른 영향을 받지 않고 성능 이득을 보이는 상관 잡음 모델을 제안한다.

본 논문의 구성은 다음과 같다. 우선 2장에서는 변환영역에서의 Wyner-Ziv 비디오 부호화의 전반적인 구조(TDWZ: Transform Domain Wyner-Ziv Coding)에 대해 제시한다. 3장에서는 상관 잡음 모델의 필요성과 제안하는 상관 잡음 모델을 제시한다. 4장에서는 실험 및 그 결과에 대해서 논의한다. 마지막으로 5장에서는 결론과 향후 연구 과제를 도출한다.

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

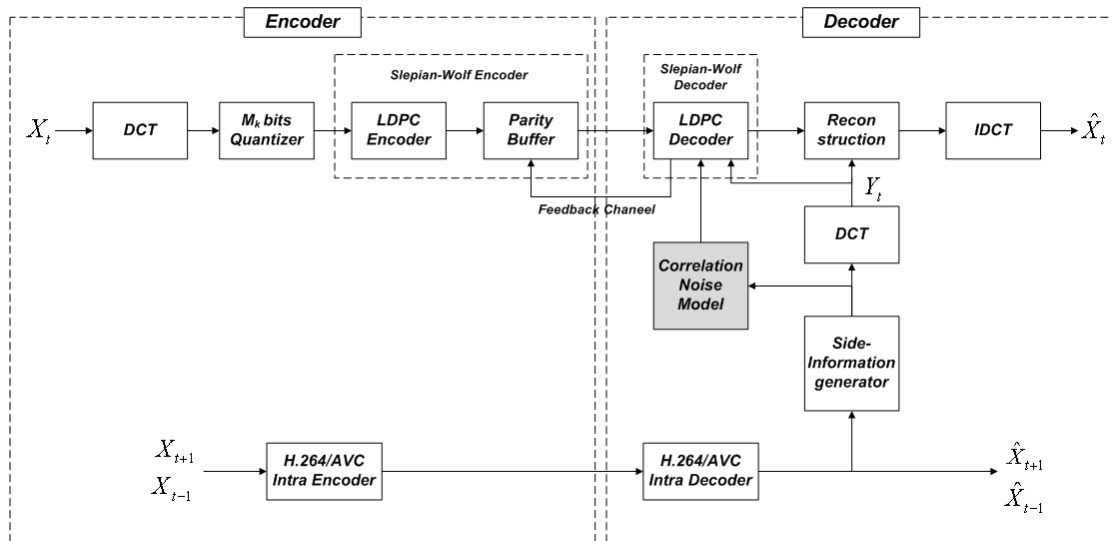


그림 1. 제안하는 TDWZ 코딩 구조

2. 변환영역 Wyner-Ziv 비디오 부호화

그림 1은 제안하는 TDWZ의 부호화 구조이다. 앞서 설명하였듯이 부호화되는 프레임은 종래 부호화 기술로 인트라 부호화 되는 키 프레임과 제안하는 Wyner-Ziv 부호화 방법으로 부호화 되는 WZ 프레임으로 구분된다. 본 논문에서는 매 짝수 번째 프레임이 키 프레임에 해당하며 매 홀수 번째 프레임은 WZ 프레임에 해당한다.

가. 부호화 및 복호화 과정

부호화기에서는 우선 현재 부호화하는 WZ 프레임에 상응하는 키 프레임을 H.264/AVC intra 부호화하여 복호화기로 전송한다. 복호화기에서는 부호화기에서 전송되어온 비트스트림으로부터 키 프레임을 복원한 후, 이를 이용하여 WZ 프레임에 해당하는 보조정보를 프레임 보간법에 기반하여 생성한다[7]. 부호화기는 WZ 프레임을 부호화하기 위해 우선 이 프레임을 변환한다. 변환영역에서의 부호화를 통하여 영상내 공간적 중복성 (spatial redundancy)을 줄일 수 있기 때문에 향상된 압축성능을 기대할 수 있다. 제안 TDWZ에서는 이를 위해 WZ 프레임을 4x4 integer DCT(Discrete Cosine Transform)를 취하여 변환한다. WZ 프레임이 변환되면 부호화기는 동일한 주파수 성분의 변환 계수들을 모아 각각의 주파수 밴드를 형성한다. 형성된 주파수 밴드에 대해서는 각각의 밴드마다 서로 다른 양자화 레벨을 적용하여 양자화한다[8]. 이때 DC 밴드에 대해서는 uniform scalar 양자화를 수행하며 나머지 AC 밴드에 대해서는 dead-zone 양자화를 수행한다[9]. 양자화 과정을 통해 생성된 각 밴드의 양자화 심볼스트림은 MSB에서 LSB에 이르는 비트플레인으로 재구성되며 각각의 비트플레인은 이 후 독립적으로 LDPC 부호화기에 입력된다. 그 결과 각 비트플레인에 대해 생성되는 패리티 비트는 버퍼에 저장하였다가 복호화기의 요청에 따라 그 중 일부를 복호화기에 전송한다. 복호화기의 LDPC 복호화기는 사전에 정의한 임계값 이하로 에러율이 내려가지 않는 경우 부호화기에 추가적인 패리티를 요청하여 복호화를 반복한다. 몇 차례의 이러한 패리티 요청과 전송을 통하여 에러율이 임계값 이하로 내려가면 LDPC 복호화기는 패리티 요청을 종료하고 다음 비트플레인에 대한 복호화를 수행한다. 한 주파수 밴드에 대한 모든 비트플레인이 복호되면 양자화 심볼 스트림을 복원하며 복원된 스트림은 영상복원기로 보내진다.

영상복원기는 LDPC복호화기에서 받은 양자화 심볼 스트림과 보조정보 및 잡음 모델을 이용하여 WZ 프레임의 각 변환계수 값을 복원한다 [6]. 복호화기는 마지막으로 모든 변환계수에 대한 복원을 마치면 이를 역변환하여 WZ 프레임을 복원한다.

3. 제안하는 상관 잡음 모델

가. 상관 잡음 모델의 필요성

Wyner-Ziv 코딩의 복호화기에서는 보조정보와 WZ 프레임간의 잡음을 Laplacian으로 모델화하여 LDPC 복호화기에 입력값으로 들어가는 연판정(soft input)값을 계산하여 복호화 과정이 진행된다. 이 때, Laplacian으로 모델화하는 과정을 상관 잡음 모델에서 수행한다. 둘 사이의 잡음이 적을수록 양자화된 심볼이 나타내는 구간에서 보조 정보에 가까운 값이 복원되고, 잡음이 많을수록 Laplacian 모델은 균일분포에 가까워지므로 구간의 중간값으로 복원이 된다. 따라서 둘 사이의 잡음을 어떻게 계산하고 Laplacian 계수를 정하느냐에 따라 Wyner-Ziv 코딩의 성능이 달라진다. 하지만 Wyner-Ziv 코딩의 복호화기에는 WZ 프레임에 대한 아무런 정보도 없기 때문에 상관 잡음 모델은 부호화기에서 전송된 키 프레임 사이의 잡음을 이용하여 Laplacian으로 모델화한다. 키 프레임 사이의 영상에 움직임이 많지 않다면 키 프레임 사이에 비슷한 블록들은 많이 있다. 따라서 둘 사이의 잡음은 많이 발생하지 않고, 상관 잡음 모델도 잘 예측된다. 하지만 만약 키 프레임 사이의 영상에 움직임이 많다면 한쪽 키 프레임에만 존재하는 블록이 다수 존재할 가능성이 커진다. 그렇다면 이런 경우에는 키 프레임 사이의 잡음이 커지고 상관 잡음 모델도 제대로 예측하기 어려워진다. 그렇기 때문에 키 프레임을 이용하여 상관 잡음 모델을 제대로 정의해주는 것은 어렵지만 Wyner-Ziv 코딩의 성능 향상을 위해선 중요하다. 따라서 본 논문에서는 종래의 방법[3]보다 개선된 상관 잡음 모델을 제안한다.

나. 제안하는 방법

제안하는 방법에서는 보조정보를 생성할 때 사용된 움직임 보상을 키 프레임의 블록간의 차이를 식(1)과 같이 구한다.

$$R(x, y) = X_F(x + dx_f, y + dy_f) - X_B(x + dx_b, y + dy_b) \quad (1)$$

여기서 X_F 는 순방향의 키 프레임의 의미이고, X_B 는 역방향의 키 프레임의 의미이다. 그리고 dx_f 와 dy_f 는 순방향의 움직임 벡터, dx_b 와 dy_b 는 역방향의 움직임 벡터를 의미한다.

그 후에 4x4 블록 DCT를 적용한다 ($T_b(u, v)$). 그리고 여기에 절대값($|T_b(u, v)|$)을 적용한 후에 DCT 밴드의 분산($\hat{\sigma}_b^2$)을 구한다. 실제 잡음을 이용하여 각각의 $T_b(u, v)$ 에 대해 $\alpha_b(u, v)$ 를 구할 때에는 식(2)와 같다.

$$\alpha_b(u, v) = \sqrt{\frac{2}{[T_b(u, v)]^2}} \quad (2)$$

하지만 위에서도 언급했듯이 복호화기에는 WZ 프레임에 대한 아무런 정보가 없기 때문에 복호화기에서 전송된 키 프레임간의 차이를 잡음으로 가정한다. 따라서 식(2)의 신뢰도가 떨어져 사용하지 않고 제안하는 방법에서는 식(3)을 이용하여 구한다.

$$\hat{\alpha}_b(u, v) = \begin{cases} \sqrt{\frac{2}{[|T_b(u, v)|]^2}}, & [\hat{\sigma}_b^2]^\lambda \leq |T_b(u, v)| \\ \frac{\sqrt{2}}{[\hat{\sigma}_b^2]^\lambda}, & [\hat{\sigma}_b^2]^\lambda > |T_b(u, v)| \end{cases} \quad (3)$$

식(3)은 식(2)를 기반으로 만들었는데 $[\hat{\sigma}_b^2]^\lambda$ 와 $|T_b(u, v)|$ 을 비교하는 것은 상관 잡음 모델에서 키 프레임간의 차이가 실제 잡음과 같을 것이라는 것에 대한 신뢰도가 떨어지기 때문이다. 따라서 식(3)을 이용하되 Laplacian 계수가 너무 커지게 되는 것을 방지할 수 있는 문턱값도 필요하다. 이에 $\hat{\sigma}_b^2$ 는 DCT 계수들이 DCT 밴드의 평균값 주위로 어느 정도로 분포되어 있는지 측정할 수 있는 수단이라는 것을 근거로 해서 $[\hat{\sigma}_b^2]^\lambda$ 보다 $|T_b(u, v)|$ 이 작은 값을 가지게 되면 실제 잡음보다 너무 작은 값으로 잡음을 가정한다고 보고 $\frac{\sqrt{2}}{[\hat{\sigma}_b^2]^\lambda}$ 를 문턱값으로 사용한다. λ 는 실험을 통해 최적의 값을 얻는다.

종래의 방법도 제안하는 방법과 마찬가지로 Laplacian 계수가 너무 커지게 되는 것을 방지할 수 있는 문턱값을 이용한다. 제안하는 방법과 다른 점은 문턱값을 선택하게 되는 조건이 다르고 이에 따라 선택하게 되는 문턱값도 다르다. 종래의 방법에선 각 DCT 계수의 Laplacian 계수와 DCT 밴드의 평균값의 차를 이용해서 문턱값을 선택하는 조건을 만든다[3].

4. 실험 방법 및 결과

가. 실험 방법

본 논문에서 제안하는 개선된 상관 잡음 모델을 평가하기 위하여 15Hz, 150 프레임의 Foreman(지멘스 로고 있는 것), Soccer, Stefan, Coastguard 영상에 대하여 실험을 하였다.

실험은 다음 두 가지 Wyner-Ziv 코딩 방법과 H.264/AVC Intra 코딩 방법의 비교로 이루어졌다.

- Previous : 종래의 상관 잡음 모델을 사용한 TDWZ
- Proposed : 개선된 제안 상관 잡음 모델을 사용한 TDWZ
- H.264/AVC Intra

구체적인 양자화 파라미터 설정은 표 1과 같다.

코딩 구조는 I-WZ-I-WZ ... 이고 보조정보 생성을 위한 블록의 크기는 8x8이고 탐색 영역은 ± 32 이다. 전송률은 키 프레임과 WZ 프레임에 대한 값을 더해서 계산했으며, PSNR은 원본 영상과 복원된 영상 간에 측정된 값이다.

나. 실험 결과

그림 2는 화질평가를 위한 Foreman 영상과 Soccer 영상과 Stefan 영상과 Coastguard 영상의 R-D 곡선이다. 그래프를 통해 알 수 있듯이 전체적으로 종래의 방법보다 향상된 성능을 보인다. 특히, Soccer 영상에서 BDPNSR 측면에서 최대 0.52dB, BDBR 측면에서 최대 8.41%의 성능 이득을 보인다. 이는 제안하는 방법이 보조정보에 잡음이 많은 Soccer와 같은 영상에서 문턱값을 적절히 선택하여 Laplacian 계수를 잘 생성한다는 것을 보인다. 하지만 Coastguard 영상과 같이 선형적인 움직임 외에 별다른 움직임이 없는 영상 같은 경우에는 상대적으로 성능 이득이 적다. 높은 bit rate에서의 BDPNSR 측면에서 최대 0.21dB, BDBR 측면에서 최대 4.01%의 성능 이득이 있지만 낮은 bit rate에서의 BDBR 측면에서 최대 10.32%의 성능 손실을 보인다. 이는 잡음이 적은 보조정보의 경우에 문턱값이 불필요한 부분에서 제안하는 방법이 제대로 예측하지 못하고 문턱값을 선택하여 손실이 발생하는 경우가 발생하기 때문이다. Foreman 영상과 Stefan 영상에서도 BDPSNR과 BDBR 측면에서 각각 최대 0.39dB, 5.65%와 0.24dB, 4.03%의 성능 이득을 보인다.

5. 결론

본 논문에서는 보조정보와 WZ 프레임간의 잡음을 제거하기 위하여 이를 Laplacian으로 모델화하는 상관 잡음 모델을 개선하여 제안하였다. 제안하는 방법은 이 잡음이 실제와는 달리 너무 작게 예측되어 Laplacian 계수가 너무 커지는 것을 방지하고 이를 대신할 수 있는 적

표 1. 양자화 파라미터 설정

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
	Soccer	44	43	41	36	36	34	31	25
	Stefan	40	39	38	34	34	32	29	25
	Coastguard	38	37	37	34	33	31	30	26

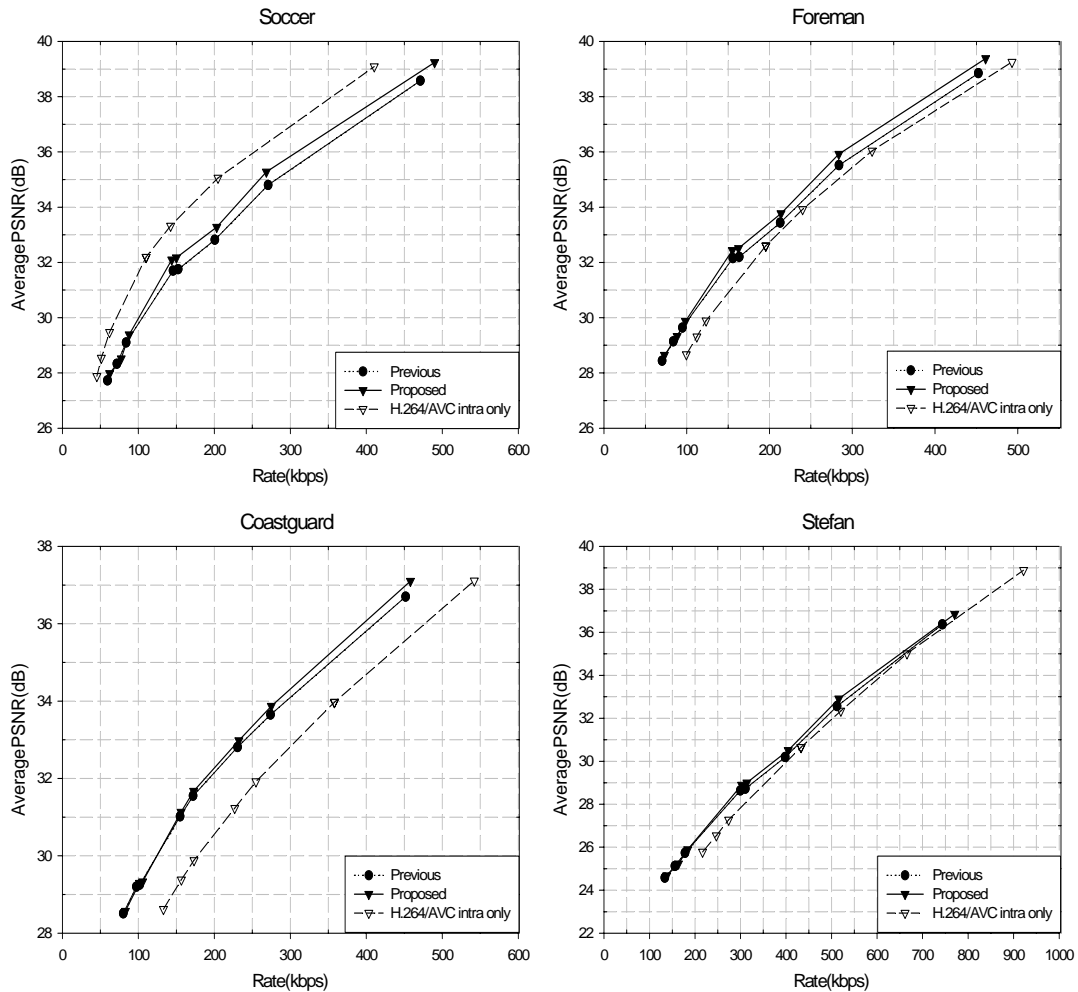


그림 2. R-D 성능 비교

절한 문턱값을 찾는다. 따라서 종래의 방법보다 제안하는 방법은 R-D 성능에 있어서 대부분의 영상에서 성능 이득을 보인다. 하지만 Coastguard 실험 결과를 보면 낮은 bit rate에서의 BDBR 측면에서 최대 10.32%의 손실을 보인다. 이것은 잡음이 적은 보조정보의 경우에서 문턱값이 불필요한 부분이 발생했을 때 제안하는 방법이 제대로 예측하지 못하고 문턱값을 선택하여 손실이 발생했기 때문이다. 그렇기 때문에 향후에는 이러한 문제를 해결하기 위해 낮은 bit rate나 잡음이 적은 보조정보에서도 종래의 방법보다 더 정교한 문턱값과 이를 선택하는 조건을 통해 Laplacian 계수를 제대로 생성하고 나아가 보조정보에 잡음이 많고 높은 bit rate에서도 현재보다 더 잘 선택할 수 있는 정교한 상관 잡음 모델에 대한 연구가 필요하다.

참고문헌

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. IT-19, pp. 471 - 480, July 1973.
 [2] D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Inform. Theory*, vol. IT-22, pp. 1 - 10, Jan. 1976.
 [3] 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, Sep. 2008.

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, Nov. 2002.
 [5] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, pp. 71-83, Jan. 2005.
 [6] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video", *Proc. SPIE Conference on Visual Communication and Image Processing*, Jan. 2004.
 [7] B. Ko, H. J. Shim, and B. Jeon, "Wyner-Ziv coding with spatio-temporal refinement based on successive turbo decoding", *IEEE Intern. Conf. on Multimedia and Expo*, Hannover, Germany, Jun. 2008.
 [8] C. Brites, J. Ascenso, F. Pereira, "Improving Transform domain Wyner-Ziv video coding performance", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006.
 [9] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Oualet, "The DISCOVER codec: architecture, techniques and evaluation", *Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007