SCALABLE PREDICTIVE CODING BY NESTED QUANTIZATION WITH LAYERED SIDE INFORMATION

Huisheng Wang and Antonio Ortega

Signal and Image Processing Institute
Integrated Media Systems Center and Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564
E-mail: {huishenw, ortega}@sipi.usc.edu

ABSTRACT

An efficient scalable predictive coding method is proposed for the Wyner-Ziv problem, using nested lattice quantization followed by multi-layer Slepian-Wolf coders (SWC) with layered side information. The proposed coder can support embedded representation and high coding efficiency by exploiting the high quality version of the previous frame in the enhancement-layer coding of the current frame. Specifically, the decoder generates the enhancement-layer side information with an estimation approach to take into account all the available information to the enhancement layer. On the other hand, a practical switching algorithm is applied at the encoder to simplify the correlation estimation on the channel code design by assuming either the current reconstructed base-layer frame or prior enhancement-layer reconstruction as side information. Experiments based on a DPCM model show great benefits to the enhancement layer reconstruction. The paper also discusses the possible adaptation of this approach to practical video compression.

1. INTRODUCTION

Scalable coding has become an active research area in recent years with the growing popularity of network visual communication. Predictive coding, in which reconstructed previous samples are used as a predictor for the current sample, is an important technique to remove temporal redundancy in multimedia signals. Efficient scalable coding becomes more difficult if predictive techniques are used because scalability leads to multiple possible reconstructions of each source sample. In this situation either a single prediction is used, which leads to either drift or coding inefficiency, or a different prediction is obtained for each reconstructed version, which leads to added complexity.

MPEG-2 SNR scalability and MPEG-4 FGS exemplify the first approach. MPEG-2 SNR uses the enhancement-layer information in the motion-compensated prediction (MCP) loop for both base and enhancement layers, which leads to drift if the enhancement layer is not received. On the other hand, MPEG-4 FGS completely ignores the enhancement layer information of the previous frames in the MCP loop. The enhancement layer is represented by coding residual error with respect to the current base-layer reconstruction, which results in very low coding efficiency. Rose and Regunathan [1] proposed a multiple-MCP-loop approach, in which the enhancement-layer predictor is optimally estimated by considering all the available information from both base and enhancement layers. However, closed-loop prediction has the disadvantage of requiring the encoder to generate all possible decoded versions for each frame, so that each of them can be used to generate a predictor residue. Thus, the complexity is high at the encoder especially for multi-layer coding scenarios. Moreover, it also suffers the inherent limitation that the reconstruction of the predictor symbol at the decoder must be same as that used at the encoder to avoid drift.

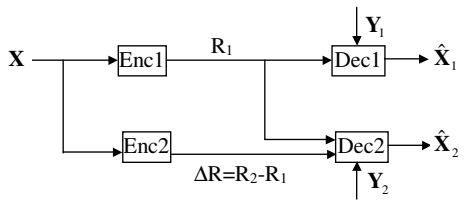

Fig. 1. Two-stage successive refinement with different SI Y_1 and Y_2 at the decoders, where Y_2 has better quality than Y_1, i.e. $X \rightarrow Y_2 \rightarrow Y_1$.

Based on the Wyner-Ziv framework [2], several side information (SI) based video codecs have been proposed in the recent literature [3, 4]. These can be thought of as an intermediate step between closing the prediction loop and coding independently. The closed-loop prediction (CLP) approach requires the exact value of the predictor to create the residue, whereas Wyner-Ziv coding only requires the correlation structure between the current signal and the predictor, so there is no need to generate the decoded signal at the encoder as long as we can find the correlation structure. Some of the recent work addresses the problem of scalable coding in this setting. Sehgal *et al* [5] provided a theoretical approach by constructing several redundant Wyner-Ziv descriptions targeted at different fidelities. Based on the encoder's knowledge of the reconstruction status of previous samples at the decoder, a decision is made about which of those descriptions to send. Xu and Xiong [6] proposed an MPEG-4 FGS-like scheme by treating a standard coded video as a base layer, and building the bit-plane enhancement layers using Wyner-Ziv coding with current base and lower layers as SI. Steinberg and Merhav [7] formulated the theoretical problem of successive refinement of information, originally proposed by Equitz and Cover [8], in a Wyner-Ziv setting, as shown in Fig. 1. The achievable region is given, and the necessary and sufficient conditions are also provided for successive refinability in the sense that both stages can asymptotically achieve the Wyner-Ziv R-D function simultaneously.

Here we propose a practical Wyner-Ziv scalable (WZS) coder for the setting illustrated by Fig. 1. Our approach supports embedded representation and high coding efficiency by exploiting the

0-7803-8554-3/04/$20.00 ©2004 IEEE.

high quality version of the previous frame in the enhancement-layer coding of the current frame. Specifically, the decoder generates the enhancement-layer SI with an estimation approach, same as that proposed in [1] for scalable predictive coding, to take into account all the available information to the enhancement layer. On the other hand, a practical switching algorithm is applied at the encoder to simplify the correlation estimation on the channel code design by assuming either the current reconstructed base-layer frame or prior enhancement-layer reconstruction as SI. The proposed coder, first developed on a DPCM source model, uses nested lattice quantization followed by multi-layer Slepian-Wolf coders (SWC) with layered side information. The switching algorithm can be potentially adapted to the standard DCT-based video coding. The paper is organized as follows: We describe the coding algorithm in Section 2, followed by the discussion on video application in Section 3. Section 4 presents simulation results on DPCM samples and shows substantial improvement on the PSNR of the enhancement layer reconstruction. Finally, conclusions are made in Section 5.

2. SCALABLE CODER DESIGN

2.1. Preliminaries

The proposed coding algorithm uses nested lattice quantization for scalable source coding. For a set of n basis vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$ in $\mathbf{R}^n$, an n-dimensional lattice Λ is composed of all integer combinations of the basis vectors, i.e., $\Lambda = \{\lambda = \mathbf{G}\mathbf{i} : \mathbf{i} \in \mathbf{Z}^n\}$ with $n \times n$ generator matrix $\mathbf{G} = [\mathbf{g}_1|\mathbf{g}_2|\ldots|\mathbf{g}_n]$ [9]. The nearest neighbor quantizer $Q(\cdot)$ associated with Λ is defined by $Q(\mathbf{x}) = \text{argmin}_{\lambda\in\Lambda} \| \mathbf{x} - \lambda \|$, where $\mathbf{x} \in \mathbf{R}^n$. The mod-$\Lambda$ operation is defined as $\mathbf{x} \bmod \Lambda = \mathbf{x} - Q(\mathbf{x})$, which is the quantization error of $\mathbf{x}$ with respect to Λ. The basic Voronoi cell is $\mathcal{V} = \{\mathbf{x} : Q(\mathbf{x}) = \mathbf{0}\}$ and its volume $V = \int_{\mathcal{V}} d\mathbf{x}$.

A pair of n-dimensional lattices (Λ_2, Λ_1) is nested, $\Lambda_1 \subset \Lambda_2$, i.e. $\lambda \in \Lambda_1 \Rightarrow \lambda \in \Lambda_2$, if there exists corresponding generator matrices $\mathbf{G}_1 = \mathbf{G}_2 \cdot \mathbf{J}$, where $\mathbf{J}$ is an $n \times n$ integer matrix and $\det(\mathbf{J}) \geq 1$. Λ_2 and Λ_1 are called fine and coarse lattice respectively. The nesting ratio $r = \sqrt[n]{V_1/V_2}$, where V_i is the volume of the Voronoi cells of Λ_i, $i = 1, 2$. Thus Λ_1 induces a partition (Λ_2/Λ_1) of Λ_2 into $|\Lambda_2/\Lambda_1| = V_1/V_2$ cosets of Λ_1. A partition chain $\Lambda_n/\Lambda_{n-1}/\ldots/\Lambda_1$ is a sequence of lattices such that $\Lambda_n \supset \Lambda_{n-1} \supset \ldots \supset \Lambda_1$.

2.2. Wyner-Ziv scalable (WZS) coding algorithm

Consider a zero-mean first-order Markov source $x_k = \rho x_{k-1} + z_k$, $x_{k-1} \perp z_k$, where $E[x_k^2] = \sigma_x^2$, $E[z_k^2] = (1-\rho^2)\sigma_x^2$. Let x_k, $\hat{x}_k^b$ and $\hat{x}_k^e$ be the current sample, its base and enhancement layer reconstruction, respectively. y_k^b and y_k^e denote the base-layer and enhancement-layer SI. We assume that to decode x_k the decoder has access to $\hat{x}_{k-1}^b$ as SI at the base layer, and both $\hat{x}_{k-1}^b$ and $\hat{x}_{k-1}^e$ at the enhancement layer. Fig. 2 depicts the block diagram of our proposed two-layer DPCM coder.

Encoding: The encoder in Fig. 2 consists of a pair of nested lattices (Λ_e, Λ_b) to support two-layer source quantization, and multi-layer SWC encoders to exploit the correlation between the quantized sample and the base and enhancement layer reconstruction of the previous sample. An ideal SWC can code the quantization output $Q(X)$ with vanishing probability of error at the expected rate equal to the conditional entropy $H(Q(X)|Y)$ given the SI Y. At

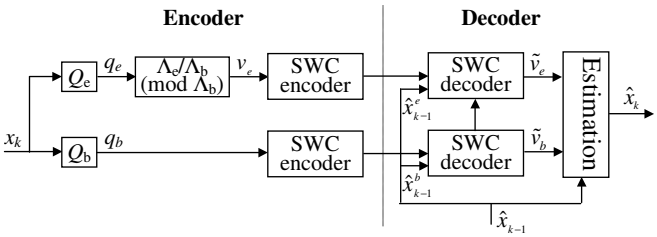

Fig. 2. Two-layer DPCM coder by nested quantization with layered SI.

the *base layer*, first quantize x_k to the nearest point in Λ_b, resulting in $q_b = Q_b(x_k)$, then code it by the base-layer SWC encoder with the SI $y_k^b = \hat{x}_{k-1}^b$ available at the decoder. Similarly, at the *enhancement layer*, quantize x_k to $q_e = Q_e(x_k)$, then create an index which identifies $v_e = q_e \bmod \Lambda_b$, the leader of the unique relative coset containing q_e. In addition to the base-layer information $x_k \in \mathcal{V}_b$, where $\mathcal{V}_b$ is the corresponding Voronoi cell of Λ_b, the enhancement-layer decoder has access to $\hat{x}_{k-1}^e$ as well. Taking into account both of them, the enhancement-layer SI is given by

$$y_k^e = E[x_k|\hat{x}_{k-1}^e, x_k \in \mathcal{V}_b] \tag{1}$$

The index stream is then coded by the enhancement-layer SWC with SI y_k^e.

Decoding: Decoding of x_k proceeds by using the reconstruction of the previous sample $\hat{x}_{k-1}$ to form SI, and the parity bits received as the coset information of SWC. The limitation here is that the coset bits received at the enhancement layer may not be decoded correctly if the corresponding SI $\hat{x}_{k-1}^e$ is not available. Let $\tilde{v}_b$ and $\tilde{v}_e$ be the decoded quantization indices from the base and enhancement-layer SWC decoders, respectively. $\tilde{v}_b$ indicates that $x_k \in \mathcal{V}_b$, where $\mathcal{V}_b$ is a Voronoi cell of the base layer Λ_b with index $\tilde{v}_b$. If $\tilde{v}_e$ is also obtained, the extra information provided by $\tilde{v}_e$ can refine the quantization cell of x_k to $\mathcal{V}_e$ of the fine lattice Λ_e at the enhancement layer. Assuming the finest cell of x_k decoded is $\mathcal{V}$, the final reconstruction of x_k is computed as the conditional centroid $E[x_k|\hat{x}_{k-1}, x_k \in \mathcal{V}]$. The decoder performs *sequential* decoding since the enhancement-layer decoding requires the information from base layer.

The proposed WZS approach can be extended to the multi-layer coding scenario in a straightforward way with a partition chain as the multi-layer nested quantization. We could also combine the proposed enhancement-layer coding with the traditional CLP base-layer coder to achieve improved enhancement-layer coding performance.

The key advantage of this approach is that for coding the enhancement layer (EL) of the current sample x_k it exploits the information from both its base layer (BL) and the prior EL reconstruction of x_{k-1} efficiently, as shown in Fig. 3a. We conclude this subsection by showing two special cases of this general framework: (1) **WZ-FGS**[1]: The EL of x_k is coded only with the information from its BL by nested quantization (NQ) without the SWC at EL. When the nesting ratio is 2, it becomes a Wyner-Ziv FGS-like bit-plane coding, similar to that proposed in [6]; (2) **WZ-Simulcast**: The EL of x_k is coded directly with SI $\hat{x}_{k-1}^e$ discarding its BL information, i.e., each layer is coded separately or by simulcast.

[1]Here, the term "FGS" does not really indicate fine granularity scalability since we only discuss about two layers. We use the term FGS here because the EL of a frame is coded only with the information from its BL, as in MPEG-4 FGS.

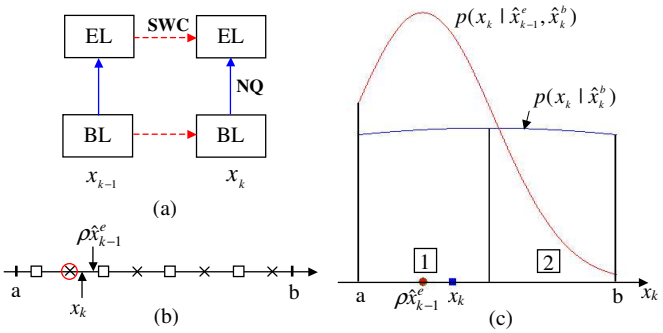

Fig. 3. (a) Coding dependence. (b) EL coded by scalar quantization and memoryless coset construction with SI $\hat{x}^e_{k-1}$. (c) an example of conditional pdf of x_k with respect to different SI at EL with nesting ratio 2, where x_k is inside bin 1.

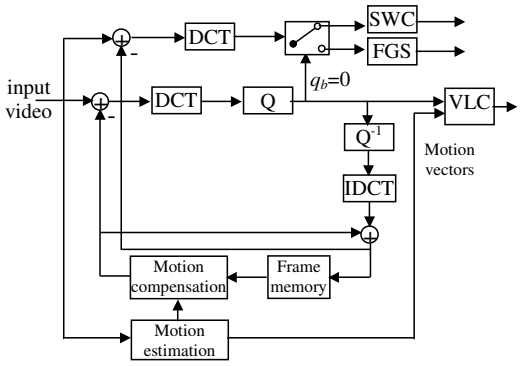

Fig. 4. Block diagram of the proposed video coder.

2.3. Practical WZS switching algorithm

Though the decoder has great flexibility to generate the enhancement layer SI y^e_k as (1), it is hard to design channel codes with appropriate rates to match the general dependence model between x_k and y^e_k for each base-layer quantization cell $\mathcal{V}_b$. For simplicity, let us consider the scalar quantization with $\mathcal{V}_b = (a, b)$. (1) is approximated as $y^e_k \approx \rho\hat{x}^e_{k-1} + E[z_k | z_k \in (a - \rho\hat{x}^e_{k-1}, b - \rho\hat{x}^e_{k-1})]$. It follows that the conditional density $p(x_k|y^e_k)$ has the mean y^e_k and the same shape as $p(z_k)$ truncated and normalized to the interval $I = (a - \rho\hat{x}^e_{k-1}, b - \rho\hat{x}^e_{k-1})$, [2]

$$p(x_k|y^e_k) = \begin{cases} \frac{p(z_k)}{\int_I p(z_k)dz_k}, & \forall z_k \in I, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The above statistics may vary significantly depending on the position of base-layer interval. Therefore we propose a simplified practical switching algorithm to code the enhancement layer: If $x_{k-1} \in (a, b)$ which implies $\hat{x}^e_{k-1} \in (a, b)$, use $\hat{x}^e_{k-1}$ as enhancement-layer SI and select a channel code at rate close to $H(Q_e(x_k)|\hat{x}^e_{k-1})$ to match the approximated additive Gaussian model. Else, code the enhancement layer directly from the base layer as that of WZ-FGS.

To illustrate the basic concept, we examine a simple example with scalar quantization and memoryless coset construction, shown in Fig. 3b. Let Δ_b and Δ_e be the quantization step size for the base and enhancement layer, respectively. We find a channel code that is matched to the correlation noise between x_k and $\hat{x}^e_{k-1}$ by partitioning the quantized codeword space into N cosets, with the minimum distance $d_{min} = N\Delta_e$. The enhancement-layer rate $R_e = \log N$. Two kinds of distortion exist in the enhancement layer, one due to quantization, and another introduced by probability of decoding error. When $\hat{x}^e_{k-1}$ is inside the base-layer interval (a, b), we can derive the total distortion D_e bounded by the sum of the two distortions

$$D_e \leq \frac{\Delta_e^2}{12} + \frac{\Delta_b^2 \cdot \text{erfc}(\frac{N\Delta_e}{\sqrt{8}\sigma_v}))}{Pr(x_k \in (a, b)|\hat{x}^e_{k-1})}, \quad (3)$$

where $v = Q_e(x_k) - \rho\hat{x}^e_{k-1} \approx z_k$ indicates the correlation noise between x_k and x_{k-1}. When the number of cosets $N = 2^{R_e}$ is large and σ_z^2 is small, the erfc function drops exponentially with Δ_e such that the distortion introduced by the decoding error is relatively small. However, compared to the WZ-FGS coding which achieves $D_e \approx \frac{\Delta_e^2}{12}$ at rate $R_e = \log \frac{\Delta_b}{\Delta_e}$, the rate of this method is reduced because $N < \frac{\Delta_b}{\Delta_e}$. Fig. 3c shows an example of conditional pdf of x_k with respect to different SI when $\hat{x}^e_{k-1} \in (a, b)$. Coding enhancement layer by conditioning on $\hat{x}^b_k$ as in WZ-FGS requires about one bit per sample since the two refinement bins have approximately equal probabilities. However, if conditioning with $\hat{x}^e_{k-1}$, as proposed by our approach, the refinement probabilities are biased leading to lower entropy. If $\hat{x}^e_{k-1}$ is far away from (a, b), the two conditional pdfs are close, and therefore we switch to WZ-FGS for low complexity. The coding advantage of WZS over WZ-FGS and WZ-Simulcast depends on the relative rate of base and enhancement layer and the correlation coefficient ρ of the source model.

3. SCALABLE VIDEO CODING

This section adapts the proposed WZS algorithm to scalable video coding, focusing on the combination with a standard CLP base-layer video coder like MPEG-4/H.26L. Fig. 4 shows a block diagram that uses the WZS switching algorithm at the enhancement layer. The temporal evolution of DCT coefficients can be usually modelled by a first-order Markov process $x_k = \rho x_{k-1} + z_k$, where x_k is a DCT coefficient in the current frame and x_{k-1} is the corresponding DCT coefficient after motion compensation in the previous frame. The motion vectors are transmitted separately at the standard base layer and the same motion vector is shared for each macroblock among all layers. Slight modifications are made in the algorithm to work with the prediction-based base layer. The switching condition of the first enhancement layer (if there are multiple layers) changes to whether the quantized base-layer residual is zero or not. If it is zero, apply the SWC (such as Turbo and LDPC codes [4, 6]) to the enhancement-layer DCT residue $(x_k - \hat{x}^b_k)$ with the SI $(\hat{x}^e_{k-1} - \hat{x}^b_k)$ available at the decoder. Else, use the traditional FGS bit-plane coding. It should be noted that the decoder could do the switching based on the reconstruction of the previous frame without any extra side information from the encoder. The model parameters ρ and σ_z^2 may be estimated from a set of training data. The correlation structure depends on those parameters and the sizes of quantization intervals, and would be easily estimated once the source model is known. The parameters can also be sent to the decoder for optimal reconstruction.

[2] We follow the analysis similar to that in [1], where this fact is exploited by conditional entropy coding.

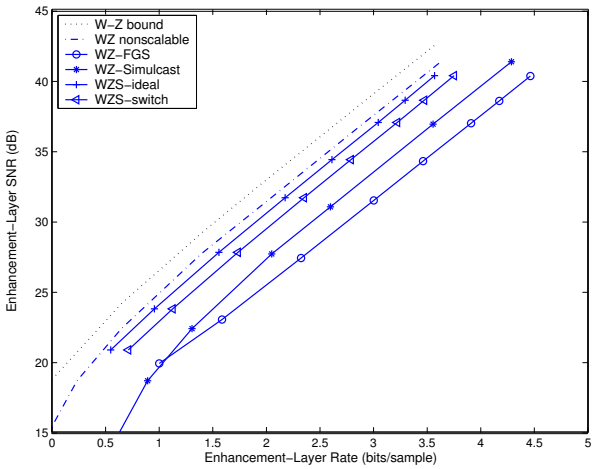

Fig. 5. Comparison between the proposed two-layer WZS coders, WZ-FGS and WZ-Simulcast for Gauss-Markov source with $\rho = 0.99$. Base-layer rate is 0.66 bits/sample.

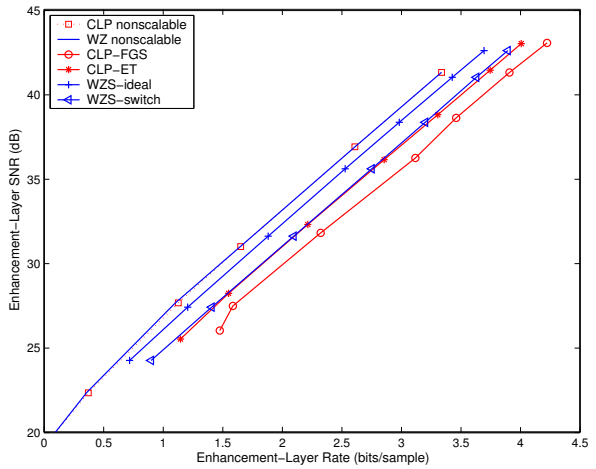

Fig. 6. Comparison between the proposed two-layer WZS coders, CLP-FGS and CLP-ET for Gauss-Markov source with $\rho = 0.99$. Base-layer rate is 0.95 bits/sample.

4. EXPERIMENTAL RESULTS

To demonstrate the performance of the proposed approach, we consider two-layer coding of first order Gauss-Markov sources of length 10^5 samples. The results are calculated as an average of these samples. We use nested scalar quantization and ideal SWC (i.e., the conditional entropy is calculated to approximate the rate after SWC) in our approach.

The following enhancement-layer coding methods are compared: (1) **WZ-FGS**; (2) **WZ-Simulcast**; (3) Proposed WZS coder which uses the SI y_k^e computed as (1) (**WZS-ideal**); (4) Simplified WZS switching coder (**WZS-switch**); (5) Closed-loop prediction using base-layer reconstruction only (**CLP-FGS**); (6) The multiple-MCP-loop approach proposed in [1] which estimates the optimal enhancement-layer predictor by exploiting both current base layer and previous enhancement layer information (**CLP-ET**). The base layer of first four methods are identical using Wyner-Ziv coding as shown in Fig. 2. An optimal entropy-constrained uniform threshold quantizer (UTQ) is used in the base-layer coding of the CLP-based methods (5) and (6), in which the rate is calculated as the first-order entropy of the quantizer indices. Also provided for reference is the Wyner-Ziv bound and the performance of both Wyner-Ziv and CLP nonscalable coder at the same total rate.

Fig. 5 shows the performance comparison between the proposed WZS coders, WZ-FGS and WZ-Simulcast for various enhancement layer rates with $\rho = 0.99$. Though there is a gap between the ideal WZS coder and the simplified WZS-switch coder, the WZS-switch coder provides significant gains over both WZ-FGS and WZ-Simulcast. Fig. 6 depicts the simulation results comparing with the CLP-based methods. The WZS-switch coder consistently outperforms the CLP-FGS coder, and performs similar to the CLP-ET method which also accounts for the enhancement-layer information in the prediction. It is also noted that the non-scalable CLP and Wyner-Ziv coders perform quite closely in the middle and high rate range.

5. CONCLUSIONS

This paper presents a new approach on scalable predictive coding in the Wyner-Ziv setting, using nested lattice quantization followed by multi-layer Slepian-Wolf coders. The enhancement layer of the current sample is thus coded by exploiting all the available information from its base layer and the enhancement layer reconstruction of the previous sample. Simulation results show a consistent gain on the PSNR of the enhancement layer reconstruction over a large rate range.

6. REFERENCES

[1] K. Rose and S.L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Image Processing*, vol. 10, pp. 965–976, Jul. 2001.

[2] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[3] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2002.

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conf. Signals and Systems*, Nov. 2002.

[5] A. Sehgal, A. Jagmohan, and N. Ahuja, "Scalable predictive coding as the Wyner-Ziv problem," in *The 8th Int. Conf. Commun. Systems*, Nov. 2002, vol. 1, pp. 101–106.

[6] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," in *Proc. VCIP'04*, Jan. 2004.

[7] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *Submitted to IEEE Trans. Inform. Theory*, Mar. 2003.

[8] W.H. Equitz and T.M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[9] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250 – 1276, Jun. 2002.