In Proc. IEEE Data Compression Conference, March 2001

Enhancing analog image transmission systems using digital side information: a new wavelet-based image coding paradigm †

S. Sandeep Pradhan and Kannan Ramchandran
EECS Dept. Univ. of California, Berkeley

Abstract

In this work we address digital transmission for enhancing, in a backward compatible way, the quality of analog image transmission systems. We propose a practical algorithm that treats the problem as one of wavelet image compression with side information (available in the form of a noisy analog version of the image) present at the decoder. We propose a rate allocation technique to efficiently allocate the rate among the wavelet coefficients of the image. In typical instances of the problem, we get gains up to 2.5 dB over conventional methods that ignore the side information. Surprisingly, this is typically achieved by modifying a very small fraction of the wavelet coefficients (typically around 10-20 %) of the conventional source coder. Extensions of our proposed image transmission framework to that of video transmission finds application in the upgrade of current analog television broadcast systems to digital TV.

1 Introduction

Recent advances in communication and signal processing technologies have spurred the need for upgrading and rehauling of analog communication systems with digital transmission schemes. Simulcast of digital television with analog television (e.g. upgrading of NTSC television to digital TV, or the PAL+ digital upgrade of the PAL system in Europe) are examples of digital side channels of existing analog transmission systems. All currently known digital side-channels of legacy analog transmission systems independently re-encode the information source digitally, totally ignoring the existing base analog transmission system. This paper is motivated by the question of how one can systematically and practically leverage the analog side information available at the receiver. This pertains to the concept of digital transmission with full backward compatibility [1] with legacy analog transmission systems.

It might thus be of interest to design a digital transmission scheme that upgrades the existing analog transmission mode instead of "blindly" replacing it. Developing a practical image transmission scheme that is motivated by this idea forms the crux of this work. Quantifying the gains in upgrading rather than totally replacing the existing noisy analog version that is present in the form of side-information at the receiver is of particular interest. It is important to note that the digital encoder is not privy to the noisy analog version of the signal available at the only at the decoder. The encoder is assumed to be informed only of the statistics of the noise (assumed to be AWGN in this paper).

The concept of such upgrading with an additional digital channel was first studied in information theory [2] under the name of source coding with side information at the decoder. Here the encoder needs to compress a source of information, X, using the knowledge

†This work was supported in part by gift award from Microsoft Research, NSF MIP 97-03181(CAREER) and ONR under Grant No. N00014-97-1-0864.

that the decoder has an information source, Y, which is related to the source X. Practical codes inspired by this theory were first explored in [3] and referred to as Distributed Source Coding Using Syndromes (DISCUS). An encoder/decoder pair is built for such a system revealing the potential for significant performance gains over systems that ignore the side information. The signal model assumed is a somewhat idealized independent identically distributed (i.i.d.) Gaussian model.

In this paper we address the application of this framework to the transmission of "real world" images using a digital side-channel, decoding with the help of a noisy version of the image at the decoder. To the best of our knowledge, there has been no prior work at building actual digital upgrade systems based on these elegant concepts. We hasten to add however that this work represents but the start of a potentially promising class of digital simulcast schemes, and much work lies ahead before a practical, viable system can be built. Of particular interest is the extension of our image transmission scheme to video transmission with applications to digital television simulcast, which is beyond the scope of this work.

One of the contributions of this paper is to extend the concept of coding with side information by bridging the gap between the simplistic i.i.d. Gaussian signal model assumption of [3] to the more complicated case of real-world images. In doing so, we recognize that most of the recent gains obtained in the (now mature) image coding community come from more accurate statistical image models. While there is a large body of work on image compression based on statistical models, in this paper, we leverage the wavelet image coding work in [4] which uses a simple but efficient Gaussian mixture model for the wavelet image representation, resulting in the so-called Estimation-Quantization (EQ) coder, whose performance is state-of-the-art. A similar Gaussian mixture model has been used successfully in [5] for the denoising of images, again with state-of-the-art performance. The most interesting aspect of this model is that it helps to build a framework that serves as a bridge between the world of information theory (where elegant concepts have been formalized but for rather simplistic signal models – e.g., i.i.d. Gaussian), and the world of image processing, where real-world images are modeled far more accurately, but where connections to information theory are tenuous at best.

Motivated by the promising wavelet image models studied in the literature, in this work we consider a wavelet-based digital upgrade system where the digitally encoded image is decoded with the help of a noisy version of that image at the decoder. We formulate a new practical image transmission system which involves an intricate yet computationally efficient combination of source coding, channel coding [6] and estimation theory [7]. The paper is organized as follows. In Section 2, we give the intuition and illustration of the key concepts. In Section 3, we formulate the system design, and in Section 4 we give simulation results based on real images. Section 5 concludes the paper.

2 Intuition and Illustration

In this paper we use some of the techniques developed in [3] and assume familiarity with them. We briefly mention the key ideas for completeness. Consider the encoding of an *i.i.d.* zero mean, unit variance Gaussian source X, with only the decoder having a noisy version of X which is given by $Y = X + N$ where N is *i.i.d.* zero mean Gaussian which

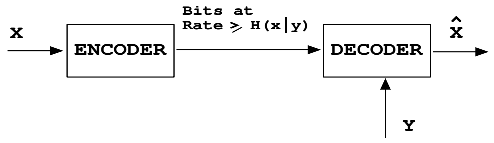

Figure 1: Communication System: Only the decoder has access to the side information Y (which is correlated to X). The Slepian-Wolf theorem shows that X can still be described with $H(X|Y)$ bits/sample.

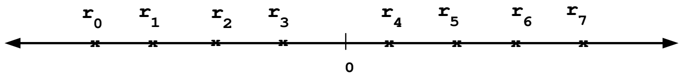

Figure 2: Reconstruction levels of an 8-level scalar quantizer

is independent of X, and of variance σ_n^2 as shown in Fig. 1. The source needs to be encoded using the knowledge that the decoder has access to Y. The method based on a scalar quantizer is illustrated in the following. Let us consider an 8-level fixed-length scalar quantizer as shown in Fig. 2. We quantize X using an 8 level fixed-length scalar quantizer. We have to spend 3 bits/sample for this encoding. The goal of the encoder is to reduce the rate of transmission by exploiting the correlation between X and Y while still encoding the source using an 8-level quantizer. The reconstruction levels are partitioned into cosets containing the codevectors such that the minimum distance between any two codevectors in a coset is kept as large as possible while maintaining the symmetry among the cosets. For a rate of transmission of 1 bit per source sample, the codevectors, r_0, r_2, r_4 and r_6 are grouped into coset-0 and r_1, r_3, r_5 and r_7 are grouped into coset-1. The encoder sends the index of the coset containing the codevector (referred to as an active codeword), which X is quantized to. The decoder with the help of Y (side information) obtains the most likely codeword in the coset whose index is received from the encoder. Then the decoder gets the optimal estimate of X, given the output of the side information and the fact that X belongs to the given voronoi region. Due to the probabilistic nature of the noise N, there is a finite probability of making an error in the decoder. Thus for the present example, 8 codevectors are partitioned into 2 cosets, and the encoder sends 1 bit of information for every sample of X. Thus the key concept is to first choose a source code to represent the source X, and partition this source codeword space into cosets of "channel codes". The information source, X, is first quantized using the source code. The encoder sends the index of the "channel code" containing the active codeword. At the decoder, we need to find a codeword in the coset of the channel code whose index is received from the encoder, closest (in some metric) to the side information Y. This can be interpreted as follows: each coset is acting as a channel code between the fictitious channel between the source coded outcome of X and the side information Y.

The method of coset construction described above is based on a scalar quantizer in a memoryless manner. Using the same scalar quantizer as a source code, we can construct cosets having memory (built on product scalar quantizer reconstruction codeword space) which will reduce the probability of decoding error. The performance of such codes for various root scalar quantizers can be obtained [3]. Using trellis based coset construction we get a coding gain of 3-6 dB (with 4-128 states in the trellis) in terms of tolerable signal to noise ratio. In other words, sophisticated codes reduce the probability of decoding error. The key concept employed here is to first design a source code for the optimal representation of X, and then reduce the rate if transmission by exploiting the correlation between X and Y while minimizing the probability of decoding error.

3 System Design

3.1 Mixture Gaussian model for the images

In this paper, inspired by the EQ coder, we restrict ourselves to simple models (without complex data structures) for typical natural images. We model the wavelet coefficients of the image to come from a realization of independent Gaussian distribution with a slowly varying variance field. This model has been shown to give good performance in compression and conventional denoising of images. There is a strong relationship between compression and denoising which has been explored [8, 9] in the literature. The variance field is estimated using a 3×3 centered square-shaped window around each pixel. This model captures the the level of spatial energy (to model edges, textures) at a coefficient level (micro details) and also captures slowly varying image characteristics (macro details) using the variance field which are captured by the coefficients of higher and lower scales respectively.

3.2 Rate allocation

Since the dynamic range of the variances of the orthonormal wavelet coefficients of real images is typically very high, a system where all the coefficients are allocated an equal number of bits is not optimal. We assume that the image pixels are corrupted with additive white Gaussian noise with variance given by σ_n^2. We use unitary wavelet transforms which preserve the energy and maintains independence of noise coefficients. Thus it can be assumed that the wavelet coefficients are corrupted with AWGN with variance σ_n^2. We define the correlation-SNR (c-SNR) as the ratio of the signal variance to the noise variance. In conventional source coding, the optimum rate allocation techniques have been studied under the name of "inverse-water pouring" algorithms. In [10], an information-theoretic rate allocation algorithm for source coding with side information of Gaussian sources with memory has been studied. We state the results of this algorithm in the following.

Let $\mathbf{X} = [X_1, X_2, \ldots, X_k]$ be a random vector with the i^{th} random variable having an independent Gaussian distribution, $N(0, \sigma_{xi}^2)$ and $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ where $\mathbf{N} = [N_1, N_2, \ldots, N_k]$ where *i.i.d.*, $N_i \sim N(0, \sigma_{ni}^2)$ and independent of X_i. The decoder has access to $\mathbf{Y}$ and the encoder wishes to compress $\mathbf{X}$ with a total rate of R bits/source vector. The rate-distortion with side information for such a random vector is given by

$$R(D) = \sum_{i=1}^{k} \frac{1}{2} \log \frac{\sigma_{xi}^2 \sigma_{ni}^2}{(\sigma_{xi}^2 + \sigma_{ni}^2) D_i} \tag{1}$$

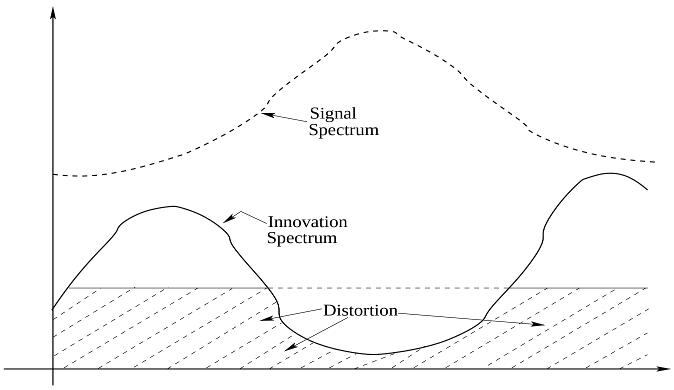

Figure 3: Optimum rate allocation: Rate-distortion function with side information involves "inverse water-pouring" on the innovation spectrum of X given Y. All the components whose innovation spectral density is below a threshold are discarded even though they have large signal power.

$$D_i = \begin{cases} \lambda, & \text{if} \quad \lambda < \frac{\sigma_{xi}^2 \sigma_{ni}^2}{\sigma_{xi}^2 + \sigma_{ni}^2} \\ \frac{\sigma_{xi}^2 \sigma_{ni}^2}{\sigma_{xi}^2 + \sigma_{ni}^2}, & \text{if} \quad \lambda \geq \frac{\sigma_{xi}^2 \sigma_{ni}^2}{\sigma_{xi}^2 + \sigma_{ni}^2} \end{cases} \tag{2}$$

where λ is such that $\sum_{i=1}^{k} D_i = D$. Thus the optimum encoding involves "inverse-water pouring" on the innovation spectrum [7] of X given Y. This is illustrated in Fig. 3 with a sample spectrum. It shows that the algorithm allocates the same level of distortion for those components whose innovation spectral density is higher than some constant λ, and discards the rest. Thus for the i^{th} component random variable X_i (assuming $\lambda < \frac{\sigma_{xi}^2 \sigma_{ni}^2}{\sigma_{xi}^2 + \sigma_{ni}^2}$),

$$R_i = \frac{1}{2} \log \frac{\sigma_{xi}^2 \sigma_{ni}^2}{(\sigma_{xi}^2 + \sigma_{ni}^2)\lambda}. \tag{3}$$

To encode this component optimally in the rate-distortion sense, we use a quantizer with $2^{LR_{si}}$ codewords where,

$$R_{si} = \frac{1}{2} \log \frac{\sigma_{xi}^2 (\sigma_{ni}^2 - \lambda)}{\sigma_{ni}^2 \lambda}, \tag{4}$$

(L is the block-length in encoding), and partition it into 2^{LR_i} cosets each having $2^{LR_{ci}}$ codewords, where

$$R_{ci} = \frac{1}{2} \log \frac{(\sigma_{ni}^2 - \lambda)(\sigma_{xi}^2 + \sigma_{ni}^2)}{\sigma_{ni}^4}. \tag{5}$$

Note that although the signal power of a given component is high, if its innovation spectral power is low, it will be discarded.

For real images, we need to develop a practical version of this information theoretic algorithm. Now we approximate this allocation technique using a system based on fixed-length scalar quantizer and trellis based coset construction. It was pointed out in [3] that the performance of the root quantizer depends on the c-SNR for a given rate of transmission and for a given probability of decoding error. For example, for a memoryless coset construction with an 8 level root scalar quantizer and for a probability of decoding error $< 10^{-4}$, with 1 bit/source sample, the c-SNR must be more than 21.3 dB. The distortion performance of the system is guaranteed during decoding success. This distortion performance improves as we increase the number of levels in the root quantizer. Thus if the c-SNR is between 15.6 dB and 21.3 dB, then only the 4-level root quantizer can be used. Any given root quantizer can only operate in certain c-SNR ranges. With a trellis based coset construction on the scalar root quantizer, we get a gain of $3-6$ dB over that of memoryless construction. Similar performance plots can be obtained for other root quantizers and other rates of transmission. We restrict our attention to integer rates of transmission and use root scalar quantizers with 2^n levels with $n = 1, 2, \ldots, 7$. When the c-SNR is low, then the reduction in the rate of transmission using the above-mentioned methods may not be possible. In such cases the rate of transmission is equal to the rate of the root quantizer. For the same example of $R = 1$ bit/sample, when $c - SNR < 15.6$ dB, we use 2-level scalar quantizer. For the present work, we consider a trellis based coset construction using 16-state trellis as given by Ungerboeck [11]. For every system based on a given scalar quantizer, we use the same trellis structure to have uniform coding gain.

Thus we have 7 choices for the root quantizer and the available choices for the rate of transmission, R, for a root quantizer with 2^n levels is given by $1 \leq R \leq n$. We allocate an "optimal" root quantizer with an "optimal" rate of transmission to each wavelet coefficient. We formulate a practical version of the "inverse-water pouring" on the innovation spectrum algorithm of information theory: we fix some $\lambda > 0$, and allocate a distortion to each component (whose innovation power $> \lambda$) characterized by a wavelet coefficient, close to λ while minimizing the rate of transmission and discard the rest of the components. The discarded coefficients are estimated at the decoder using traditional denoising methods like Wiener Estimation [5, 7]. Thus we can classify the coefficients into three groups. The group-1 consists of those coefficients whose c-SNR is sufficiently high and there is a considerable reduction in the rate of transmission by exploiting the correlation. The group-2 consists of coefficients which have low to moderate c-SNR, implying that they need to be sent using traditional source coding techniques (no reduction in rate). The group-3 consists of coefficients which have very low c-SNR, implying that they are discarded at the encoder and these coefficients are estimated using Wiener Estimation. The framework of DISCUS captures the task of source coding, channel coding and estimation theory in an intricate way.

4 Simulation Results

In this section we present simulation results on real images. We use Daubechies 8-length wavelet transform with 3 layers of decomposition. The wavelet coefficients are modeled as a mixture of independent Gaussian random variables. We estimate the variance of the coefficients using a 3×3 square shaped, centered window. We assume that the signals are zero mean except in the first band. For those coefficients with higher c-SNR and use DISCUS concepts, we estimate the variance and mean field based on noisy data and for the rest of the coefficients we estimate in a backward adaptive fashion based on quantized coefficients. As the wavelet transforms are orthonormal, we assume that the wavelet coefficients are corrupted with the same noise characteristics. This noisy image is the side information available at the decoder.

Table 1: Rates and the root quantizer minimum performance criterion in terms of necessary c-SNR for a probability of decoding error $< 10^{-4}$ using 16-state trellis. R and L specify the rate of transmission in bits/sample and the number of levels in the root scalar quantizer respectively.

R	L	c-SNR	R	L	c-SNR
1	128	$>$ 43 dB	2	128	$>$32 dB
1	64	$>$ 34 dB	2	64	$>$26 dB
1	32	$>$ 28.5 dB	2	32	$>$20 dB
1	16	$>$ 23 dB	2	16	$>$18 dB
1	8	$>$ 17 dB	2	8	$>$14 dB
1	4	$>$ 11.5 dB	4	128	$>$ 19 dB
3	128	$>$ 25 dB	4	64	$>$ 12.5 dB
3	64	$>$ 19 dB	4	32	$>$ 6 dB
3	32	$>$ 13 dB	5	128	$>$ 12.5 dB
3	16	$>$ 6.5 dB	5	64	$>$ 5.5 dB

As we discussed in section 3, the DISCUS framework is applicable for some c-SNR ranges. The probability of error plots, as given in [3], is only for 4- and 8-level quantizer for 1 bit/sample of rate of transmission and using a trellis based coset construction. We obtain similar performance curves for the scalar quantizers with 2^n levels for $n = 2, 3, \ldots, 7$, for rate of transmission R, such that $1 \leq R \leq (n-1)$ using the 16-state trellis. The distortion performance of these root quantizers can also be obtained. These are not shown in this paper due to lack of space. The essential components are given in Table.1. The table can be interpreted as follows. R specifies the rate of transmission, L specifies the number of levels in the root quantizer and c-SNR specifies the minimum c-SNR required to ensure a probability of decoding error $< 10^{-4}$. We fix the allocated distortion level and allocate those rates and root quantizers with a distortion as close to this level as possible. The DISCUS framework is applicable to only those coefficients which have c-SNR greater than 5.5 dB. These belong to the group-1 as discussed in Section 3. For the next set of coefficients which have low to moderate c-SNR, we use conventional source coding techniques by choosing a scalar quantizer with 2^n levels, and $n = 1, 2, \ldots, 7$, and use full information to transmit them. These coefficients belong to group-2 and they have c-SNR > -6.9 dB. The rest of the coefficients are not allocated any rate.

This algorithm can be implemented in a computationally efficient way. For example, for $\lambda = 10.0$ dB, the rate allocation is given in Table. 2. Similar rate allocation tables can be obtained for any λ. These tables can be precomputed and stored for a fixed grid of λ. For a given λ, the rate of transmission and the quantizer for a given c-SNR can be read out from the corresponding table and the corresponding system can be implemented.

Table 2: The rate allocation for a $\lambda = 10.0$. R and L specify the rate of transmission and the number levels in the quantizer used for the given c-SNR.

R	L	c-SNR	R	L	c-SNR
1	128	$>$ 43 dB	2	128	32-43 dB
3	128	25-32 dB	3	64	19-25 dB
3	32	13-19 dB	3	16	6.5-13 dB
3	8	2-6.5 dB	2	4	(-3) - (2) dB
1	2	(-7)-(-3) dB	0	1	$<$ (-7) dB

At the decoder the side information is used to get the active codewords, in the coset whose index is received from the encoder, for those coefficients which are in group-1. Now for all the coefficients, with the help of side information and the information received from the encoder, we obtain a minimum mean squared error estimate. The probability of decoding error is kept at a level lower than 10^{-4}. The minimum mean squared error estimate is given by

$$\hat{\mathbf{X}} = \begin{array}{c} argmin \\ \mathbf{a} \in \mathbf{R}^n \end{array} \left[E \left\{ \rho(\mathbf{X}, \mathbf{a}) \middle| \begin{array}{c} \mathbf{Y} = \mathbf{y} \\ \mathbf{X} \in \mathbf{\Gamma} \end{array} \right\} \right] \tag{6}$$

Bold faced letters denote vectors and $\mathbf{\Gamma}$ is the sequence of intervals associated with each element $\mathbf{X}$ and $\mathbf{R}$ is the real line and $\rho(\mathbf{x}, \mathbf{a}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - a_i)^2$ and x_i and a_i are the i^{th} components of $\mathbf{x}$ and $\mathbf{a}$ respectively.

To evaluate the gains obtained by the DISCUS framework, we also use conventional source coding of the image in its wavelet domain using the same model and same set of quantizers. For this case we optimally allocate the quantizer to each coefficient to minimize the average mean squared error. For the wavelet coefficients, we use the same mixture model as discussed in Section 3. This conventional source coder does not use the noisy image available at the decoder. We have simulation results on 5 different images with different characteristics. They are lena, baboon, pepper, F-16, and lake images.

Figure 4: (a) Upgraded Lena image using fixed-length scalar quantizers and trellis based coset construction: PSNR=39.87 dB, rate of transmission = 0.6233 bits/pixel, decoded with the availability of a noisy image with PSNR of 28.1 dB. (b) Source Coded Lena image: PSNR=38.63 dB, rate of transmission= 0.6233 bits/pixel. This is encoded using conventional source coding techniques.

Figure 5: Four images used in the simulations: (a) Upgraded Baboon image , PSNR=38.91 dB (b) Upgraded Pepper image, PSNR=39.10 dB (c) Upgraded F-16 image, PSNR=40.74 dB (d) Upgraded Lake image, PSNR= 39.22 dB.

The reconstructed Lena images are as shown in Fig. 4. The image encoded using the DISCUS framework has a peak signal to noise ratio (PSNR) of 39.87 dB for a rate of transmission of 0.62 bits/sample and the conventional image coder for the same average bit rate gets PSNR of 38.60 dB. The noisy image is at a PSNR of 28.1 dB. Thus we can obtain an improvement of 1.28 dB using the noisy image. The performance of the proposed system on other images is shown in Table. 3. The rates are allocated using $\lambda = 10.0$ and this gives a distortion performance close to 38.1 dB of PSNR. The gains are higher for the Baboon, F-16 and Lake images as the signal power in these images are higher. The upgraded images are shown in Fig. 5. Similar performance results are shown in Table. 4 for a noise variance of 144.0 and similar gains are obtained.

Table 3: Performance of upgraded version and conventional source coded images. Noise variance= 20 dB

Image	Rate	Denoised	Quantized	Gain
Lena	0.6233	39.87 dB	38.63 dB	1.24 dB
Baboon	2.0447	38.91 dB	36.37 dB	2.54 dB
Pepper	0.6883	39.10 dB	38.00 dB	1.1 dB
Lake	1.2803	39.22 dB	37.61 dB	1.61 dB
F-16	0.6892	40.74 dB	38.95 dB	1.79 dB

Table 4: Performance of upgraded version and conventional source coded images. Noise variance= 21.6 dB

Image	Rate	Denoised	Quantized	Gain
Lena	0.5248	38.92 dB	37.78 dB	1.14 dB
Baboon	1.8474	37.32 dB	35.29 dB	2.03 dB
Pepper	.5392	38.01 dB	36.97 dB	1.04 dB
Lake	1.067	37.69 dB	35.80 dB	1.89 dB
F-16	0.5993	39.65 dB	38.00 dB	1.65 dB

Even though for the Lena image, 85 % and for the Baboon image, 43% of the wavelet coefficients are “buried” in noise (c-SNR $<$ 0.0 dB), there is still lot to be gained using the DISCUS framework. Similar performance gains are obtained for other rates of transmission which are not presented in this paper. The performance gain is a monotonically decreasing function of the noise variance. When the noise variance is large, the DISCUS framework reduces to that of conventional source coding and conventional denoising techniques. Thus this framework gracefully bridges the gap between source coding with side information, conventional source coding and denoising.

5 Conclusion

We have proposed a method for upgrading with side information using distributed source coding principles. In a typical application setting, we obtain a gain of 1.24 dB and 2.54 dB gain in PSNR for the Lena and Baboon images respectively. We used a very simple mixture model for the wavelet coefficients without using any complex data structures, while using fixed-length scalar quantizers with integer rates. There is a lot of scope for improvement which will be studied in future work. The first goal is to have a continuum of rates for allocation and rates of the root quantizer. This will require a better understanding of the codes that need to be built on such quantizers. The second goal is to use better models using sophisticated data structures to capture all the inter-coefficient correlations as proposed in [4, 5]. In this paper we have used a heuristic rate allocation algorithm inspired by the "inverse water-pouring" algorithm of rate-distortion with side information. An optimum practical version of this algorithm will give increased gains. Incorporation of entropy coding is another issue that needs further study. Another goal would be the incorporation of such ideas into the upgrading of video sequences.

References

[1] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, March 1998.

[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. IT-22, pp. 1–10, January 1976.

[3] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS) : Design and construction," *Proc. IEEE Data Compression Conference, Snowbird, UT*, March 1999.

[4] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," *Proc. IEEE Data Compression Conference, Snowbird, UT*, March 1997.

[5] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, pp. 300–303, Dec 1999.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[7] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 146–181, March 1974.

[8] B. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. Signal Processing*, vol. 43, pp. 2595–2605, Nov 1995.

[9] H. Krim, D. Tucker, and S. Mallat, "On denoising and best signal representation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2225–2238, Nov 1999.

[10] S. S. Pradhan, "On rate-distortion function of Gaussian sources with memory in the presence of side information at the decoder," *Project Report, Course ECE480, Univ. of Illinois*, Dec 1998.

[11] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.