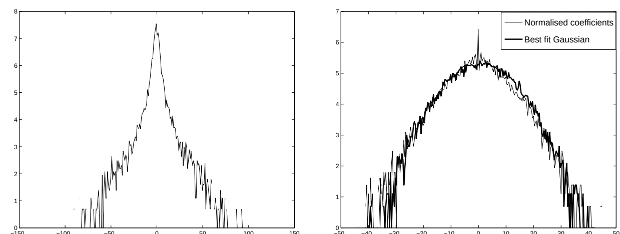
MULTIPLE DESCRIPTION IMAGE CODING USING NATURAL SCENE STATISTICS

Sumohana S. Channappayya, Robert W. Heath Jr., Alan C. Bovik

The University of Texas at Austin,
Department of Electrical and Computer Engineering,
1 University Station C0803 Austin, TX 78712-0240 USA

ABSTRACT

The statistics of natural scenes in the wavelet domain are accurately characterized by the Gaussian Scale Mixture (GSM) model. The model lends itself easily to analysis and many applications that use this model are emerging (for e.g., denoising, watermark detection). In this paper, we present an error-resilient image communications application that uses the GSM model and Multiple description coding (MDC) to provide error-resilience. We derive a rate-distortion bound for GSM random variables, derive the redundancy rate distortion function, and finally implement an MD image communication system.



(a) Raw coefficients (b) Normalized coefficients

Fig. 1. Log histograms of raw and normalized wavelet coefficients in a sub-band.

1. INTRODUCTION

Reliable image communication has always been an important research area and has recently received further attention due to the explosive growth in wireless personal communication systems. Packet based networks form a major chunk of all communication networks and these networks tend to be lossy. Lossy packet networks necessitate the design of error-resilient image communication systems.

Of the numerous techniques designed to combat packet loss such as FEC, retransmission, multiple description coding(MDC) [1] etc., MDC is particularly attractive for image sources. MDC schemes perform well over erasure channels due to the fact that all the descriptions are equally important and provide the equal amounts of redundancy. A key assumption made in most MD literature is that the source is Gaussian distributed. However, it has been decidedly shown that natural image sources in the transform domain (DCT or wavelet) are not Guassian[2, 3] . Therefore, it is not straightforward to apply standard MD coding results to image sources.

The fact that JPEG2000 – the latest image coding standard operates in the wavelet domain reflects the importance and popularity of image representation in the wavelet domain. It is therefore important to accurately model the statistics of natural images in the wavelet domain. Significant research effort has been expended towards this cause[3, 4]. It has been empirically shown in [3] that modelling wavelet coefficients as a scale mixture of Gaussian random variables captures the statistical properties of wavelet coefficients accurately and succinctly. The model lends itself to analysis since it provides a technique to “Gaussianize” the wavelet coefficients. This property has been used in developing applications such as image de-noising , image compression image quality assessment, watermark detection etc.

In this paper, we propose an error-resilient image communications scheme that exploits the properties of the GSM model and uses MDC techniques to achieve error resilience. We first derive a rate-distortion bound for GSM random variables and use the result in arriving at a redundancy rate-distortion function. Finally, we simulate an image communication system that uses an orthogonal pairwise correlating transform to form multiple descriptions and uses an erasure channel[1].

2. THE GAUSSIAN SCALE MIXTURE MODEL

In this section, we describe the model used to represent the statistics of natural images in the wavelet domain. It has been shown in [3] that the statistics of wavelet coefficients fit the GSM model very accurately. A random vector Y is a GSM if it satisfies the relation $Y = zU$ where z is a scalar random variable with $z \geq 0$ and $U \sim \mathcal{N}(0, Q)$ is a Gaussian random vector and z, U are independent [3]. The key feature of this model is that the conditional distribution $f_{Y|z}(y|z)$ is Gaussian. This feature lets us work with conditioned random vectors that are actually Gaussian distributed and therefore not make any assumptions about the Gaussianity of the source. Also, it has been shown in [3] that normalized wavelet coefficients $U(= Y/z)$ are jointly Gaussian.

Intuitively, z corresponds to local standard deviation of wavelet coefficients that scales the Gaussian random vector U. Figs.1(a) and 1(b) show log histograms of raw($f_Y(y)$) and normalized($f_{Y|z}(y|z)$) wavelet coefficients respectively. It is clear that normalizing the coefficients makes them closer to a Gaussian distribution. Further, it demonstrates the accuracy of the GSM model in representing wavelet coefficients. We use this model for wavelet coefficients in the sequel.

3. MULTIPLE DESCRIPTION CODING

Multiple Description Codes (MDC) are multiple representations of an information source. The descriptions are designed such that an acceptable quality of reconstruction is possible even from a subset of the descriptions. The source coder intelligently partitions the source information across descriptions. The communication channel is modelled as an erasure channel.

Several techniques have been proposed to generate multiple descriptions[1, 5, 6]. Of these options, we use the pairwise correlating transform (PCT) approach [1] in this work. Analytical amenability, provision of good error resilience, and ease of implementation motivated the use of the PCT technique.

The key idea of the PCT technique is the introduction of controlled correlation between uncorrelated random vectors (of dimension 2) via a correlating transform. Correlating the sources increases the rate required to encode them, which is called redundancy. Since we deal with erasure channels, a measure of the system's performance is the end-to-end distortion when one channel is available and the other is erased. This is called the one-channel distortion. The redundancy rate-distortion(RRD) function characterizes the one channel distortion as a function of redundancy.

The PCT technique assumes the source to be Gaussian distributed. From our discussion in Section [2] we see that normalized wavelet coefficients are Gaussian distributed, and hence lend themselves easily to the application of the PCT technique. In order to derive the redundancy rate-distortion function for GSM random vectors, we first derive a rate-distortion bound for them. This gives us a bound on the rate that needs to be spent to encode a GSM source at a given distortion.

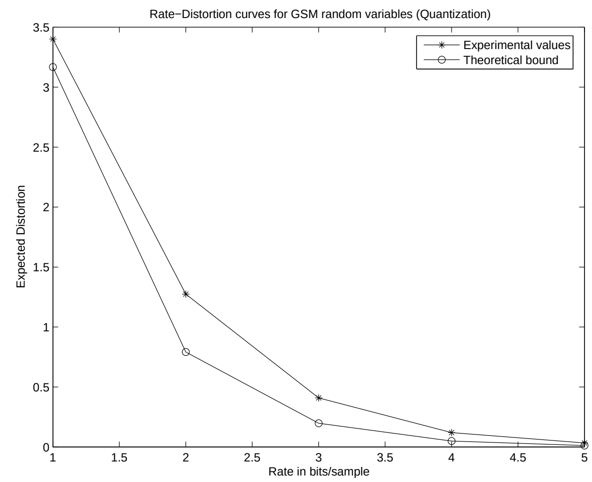


Fig. 2. Operational Rate-Distortion curves for a scalar GSM random variable - $f_Z(z) = \frac{ze^{-z}}{\Gamma(2)}, u \sim \mathcal{N}(0,1)$

4. RATE-DISTORTION ANALYSIS OF GSM RANDOM VARIABLES

We begin our analysis by deriving the rate-distortion bound for GSM random variables. This derivation is used for deriving the redundancy rate-distortion function for the case of multiple descriptions.

Let Y be a random vector that represents neighboring wavelet coefficients. $Y = zU$ where z is a scalar random variable and $U = [u_1, u_2]^T$ where u_1, u_2 are uncorrelated Gaussian random variables.

From the definition of the rate distortion function [7],

$$
\begin{aligned}
R(D) &= \min_{f(\hat{y}|y):E[Y-\hat{Y}]^2 \leq D} I(Y;\hat{Y}) && (1)\\
I(Y;\hat{Y}) &= h(Y) - h(Y|\hat{Y}) && (2)\\
&= h(z) + h(U) - h(Y-\hat{Y}|\hat{Y}) && (3)\\
&\geq h(z) + h(U) - h(Y-\hat{Y}) && (4)\\
\Rightarrow I(Y;\hat{Y}) &\geq h(z) + \frac{1}{2} log \frac{2\pi e \sigma_{u_1}^2 \sigma_{u_2}^2}{D} && (5)
\end{aligned}
$$

Eqn.(3) follows since subtracting a constant doesn't affect the entropy. Eqn. (4) results since conditioning reduces entropy and finally we get (5) since the Gaussian distribution is entropy maximizing. Furthermore, it is known that z is a positive random variable which can be approximated by a Gamma density defined as $f_Z(z) = \frac{\beta^\gamma z^{\gamma-1} e^{-z}}{\Gamma(\gamma)}$, where $\Gamma(\gamma)$ is the Gamma function ($\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx$). Evaluating the entropy of a Gamma density using the definition of differential entropy we get (letting $\beta = 1$)

$$
\begin{aligned}
h(z) &= -\int_0^\infty f_Z(z) log f_Z(z) dz \\
&= log[\Gamma(\gamma)] + \gamma - (\gamma-1)\Psi(\gamma)
\end{aligned} \quad (6)
$$

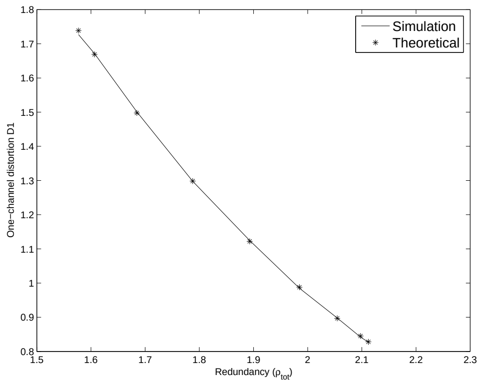


Fig. 3. Theoretical and simulated redundancy rate-distortion plot

where $\Psi(\gamma)$ is the digamma or psi function [8]. Substituting the result from (6) in (5) we get

$$R(D) \geq log[\Gamma(\gamma)] + \gamma - (\gamma-1)\Psi(\gamma) + \frac{1}{2}log\frac{2\pi e\sigma_{u_1}^2\sigma_{u_2}^2}{D}. \tag{7}$$

Fig.2 shows the operational rate-distortion curve for a scalar GSM random variable obtained from a generalized Lloyd's quantizer. We see that the simulation results are close to the bound defined in (7).

5. REDUNDANCY RATE-DISTORTION ANALYSIS

In this section we derive the redundancy rate-distortion function for the pair-wise correlated GSM source. We assume a two channel scenario with channel failure probabilities p_1 and p_2 respectively. We further assume that the rate-distortion bound in (7) is achieved with equality. Since Y is a GSM random variable, the pdf of the components of U i.e., u_1 and u_2 can be expressed as $f_{U_I|z}(u_I|z) \sim \mathcal{N}(0, z^2\sigma_{u_1}^2)$ where z is the scalar random variable (which is assumed to be known).

We derive the expression for redundancy ρ at a given two channel distortion D_0 needed to achieve a one-channel distortion D_1. An orthogonal correlating transform $\mathbf{T}$ is applied to the uncorrelated source pair $U = [u_1, u_2]^T$ to generate $V = [v_1, v_2]^T$ i.e, $V = \mathbf{T}U$. We evaluate $Cov(V) = E[VV^T]$ as

$$\begin{aligned} \mathbf{Cov(V)} &= E[VV^T] \\ &= \mathbf{T}E[UU^T]\mathbf{T}^T \text{ where,} \\ E[UU^T] &= \begin{bmatrix} z^2\sigma_{u_1}^2 & 0\;\theta \\ 0 & z^2\sigma_{u_2}^2 \end{bmatrix} \end{aligned} \tag{8}$$

The orthogonal transform $\mathbf{T}$ is defined as

$$\mathbf{T} = \begin{bmatrix} cos\,\theta & sin\,\theta \\ -sin\,\theta & cos\,\theta \end{bmatrix} \text{ substituting } \mathbf{T} \text{ in (8),}$$

$$\mathbf{Cov(V)} = z^2.\mathbf{W}, \text{where}$$

$\mathbf{W} =$

$$\begin{bmatrix} (\sigma_{u_1}^2\cos^2\theta \;+\; \sigma_{u_2}^2\sin^2\theta) & cos\theta\; sin\theta\,(\sigma_{u_2}^2 - \sigma_{u_1}^2) \\ cos\theta\; sin\theta\,(\sigma_{u_2}^2 - \sigma_{u_1}^2) & (\sigma_{u_1}^2\sin^2\theta \;+\; \sigma_{u_2}^2\cos^2\theta) \end{bmatrix}$$

We now compute the redundancy (or excess rate) required to encode $V = [v_1, v_2]^T$ with respect to the rate required to encode $U = [u_1, u_2]^T$ at two-channel distortion D_0. Let $R_{u_1,u_2|z}$ denote the encode rate for U, and $R_{v_1,v_2|z}$ denote the encode rate for V. Since U, V are conditionally Gaussian, we can use the rate-distortion function for Gaussian variables and arrive at

$$\begin{aligned} R_{u_1,u_2|z} &= \frac{1}{2}log\frac{z^4\sigma_{u_1}^2\;\sigma_{u_2}^2}{D_0} \;+\; K, \\ R_{v_1,v_2|z} &= \frac{1}{2}log\frac{z^4\sigma_{v_1}^2\;\sigma_{v_2}^2}{D_0} \;+\; K, \\ \rho = R_{v_1,v_2|z} - R_{u_1,u_2|z} &= \frac{1}{2}log\frac{z^4\sigma_{v_1}^2\;\sigma_{v_2}^2}{z^4\sigma_{u_1}^2\;\sigma_{u_2}^2}, \\ \Rightarrow \rho &= log\frac{\sigma_{v_1}\;\sigma_{v_2}}{\sigma_{u_1}\;\sigma_{u_2}} \end{aligned} \tag{9}$$

where K is a constant that accounts for entropy coding. We observe that since we are conditioning on z, z needs to be sent to the decoder without error for successful reconstruction. The rate needed to encode z is lower bounded by its entropy $h(z)$. Therefore the total excess rate needed is bounded by

$$\begin{aligned} \rho_{tot} &\geq h(z) + \rho \\ &= h(z) + log\frac{\sigma_{v_1}\;\sigma_{v_2}}{\sigma_{u_1}\;\sigma_{u_2}} \\ &= log[\Gamma(\gamma)] + \gamma - (\gamma-1)\Psi(\gamma) + log\frac{\sigma_{v_1}\;\sigma_{v_2}}{\sigma_{u_1}\;\sigma_{u_2}} \end{aligned} \tag{10}$$

The single channel distortion D_1 is defined as the average single channel distortion per random variable [1]. We now express the single channel distortion D_1 in terms of the excess rate in order to obtain the redundancy rate-distortion bound.

$$\begin{aligned} D_1 = E[z^2]\{&(1-p_1)p_2E[(u_1-\hat{u_1})^2 + (u_2-\hat{u_2})^2|ch.\,1] \\ &+p_1(1-p_2)E[(u_1-\hat{u_1})^2 + (u_2-\hat{u_2})^2|ch.\,2]\} \end{aligned}$$

The MMSE estimates $\hat{v}_1, \hat{v}_2$ are given by $E[v_1|v_2]v_2$ and $E[v_2|v_1]v_1$ respectively, since v_1, v_2 are Gaussian. The expressions for the estimates are given by $\hat{v_1} = \frac{E[v_1v_2]}{\sigma_{v_2}^2}v_2$ and $\hat{v_2} = \frac{E[v_1v_2]}{\sigma_{v_1}^2}v_1$. Finally $\hat{u_1}, \hat{u_2}$ are given by $\mathbf{T}'[\hat{v_1}\hat{v_2}]'$. Assuming $p_1 = p_2 = \frac{1}{2}$, neglecting effects of quantization and simplifying, we get,

$$D_1(\rho_{tot}) = E[z^2]\left(\frac{\sigma_{u_1}^2 + \sigma_{u_2}^2}{4}\right)2^{2(h(z)-\rho_{tot})} \tag{11}$$

Fig. 3 shows the plot of the redundancy rate-distortion function of the proposed multiple description coding system. The figure is plotted for $f_Z(z) = \frac{ze^{-z}}{\Gamma(2)}, u_1 \sim \mathcal{N}(0,1)$ and , $u_2 \sim \mathcal{N}(0, 0.16), u_1 \perp u_2$. We see that the theoretical curve and the simulation results match closely.

6. IMAGE COMMUNICATION SYSTEM

The basic premise of the proposed algorithm is the fact that normalizing wavelet coefficients by appropriate scale factors makes them Gaussian. Furthermore, it is assumed pairs of wavelet coefficients are uncorrelated and jointly Gaussian [3].

The procedure used to implement the system is described below. The image is sub-band decomposed to three levels using the Haar wavelet. The low-low sub-band is assumed to be transmitted without error. In order to normalize the wavelet coefficients, non-overlapping blocks of size 4x4 are formed and the covariance matrix Q and scaling factor estimate $\hat{z}$ are determined from these blocks. Once $\hat{z}$ is estimated, the coefficients are normalized by dividing them by $\hat{z}$. The pairwise correlating transform is then applied to the normalized coefficients. The resulting correlated coefficients are transmitted over a pair of erasure channels with erasure probability $p_1 = p_2 = \frac{1}{2}$. Since $\hat{z}$ has to be transmitted without error, it is sent over both channels. At the receiver, lost normalized coefficients are estimated using the MMSE estimator. The wavelet coefficient estimates are found by multiplying the normalized coefficient estimates with $\hat{z}$.

The main aim of the above experiment is to see if the GSM model performs better than the case where wavelet coefficients are assumed to be Gaussian. In order to compare the performance of the two systems, we used the exact same channel in the experimental setup for both cases. The key difference in the setup for the Gaussian case is that the wavelet coefficients are not normalized. The results of the above experiment on standard test images of size 512x512 are shown in Table 1. We clearly see from the table that using the GSM model results in an improvement in the reconstructed image quality when compared to the case where wavelet coefficients are assumed to be Gaussian.

7. CONCLUSION

We have shown that the GSM model for natural scene statistics lends itself very well to analysis in a multiple description coding context. This model does not require any assumptions about the Gaussianity of the source distribution. A rate-distortion bound was derived for GSM random variables. Further, the redundancy rate-distortion function was derived for GSM sources that are transformed using a PCT.

	GSM Model	Gaussian assumption
Barbara	22.32	21.59
Lena	24.48	23.79
Zelda	27.39	27.07

Table 1. Image reconstruction results for single description in PSNR (dB) at correlation angle $\theta = \pi/16$.

Finally, it was experimentally shown via a MD image coding system that the GSM model performs better than cases where the source is assumed to be Gaussian. We are currently looking at using other wavelet bases, using overlapping blocks, varying block sizes and non-orthogonal correlating transforms.

8. REFERENCES

[1] Yao Wang, Michael T. Orchard, Vinay Vaishampayan, and Amy R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 351–366, Mar. 2001.

[2] Edmund Y. Lam and Joseph W. Goodman, "A mathematical analysis of the dct coefficient distributions for images," vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[3] Martin J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural scenes," *Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861, 2000.

[4] Stephane G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[5] Vinay Anant Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, pp. 821–834, May 1993.

[6] Vivek K. Goyal and Jelena Kovačević, "Generalized multiple description coding with correlating transforms," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2199–2224, Sept. 2001.

[7] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, USA, 1991.

[8] Noah Gintis, Malcolm X. He, and Paolo E. Ricci, "Entropy of probability distributions and its applications in atm networks and dna sequences," in *International MultiConference in Computer Science and Computer Engineering*, 2003.