ANALYTICAL MODEL-BASED BIT ALLOCATION FOR WAVELET CODING WITH APPLICATIONS TO MULTIPLE DESCRIPTION CODING AND REGION OF INTEREST CODING

Phoom Sagetong and Antonio Ortega

Integrated Media Systems Center
Signal and Image Processing Institute
Department of Electrical Engineering-Systems
University of Southern California
3740 McClintock Ave., Los Angeles, CA 90089-2564
Email: sagetong@biron.usc.edu, ortega@sipi.usc.edu

ABSTRACT

We address the problem of allocating bits to the different regions in an image coded with a progressive wavelet coder such as SPIHT (Set Partitioning in Hierarchical Trees) [1]. This type of problem appears in applications such as Region of Interest (ROI) coding or Multiple Description Coding (MDC). The wavelet coefficients in both cases are divided by different factors before coding to enable different bit allocation to different regions, because the coefficients in each region are refined at different speeds. While this is a popular approach for ROI coding [2, 3], we propose using it for MDC as well. In this work, we introduce a **priority scaling factor** (psf) as a dividing factor. The main contribution of this work is to provide an analytical technique to determine what the psf should be, given criteria such as relative importance of the regions in ROI coding or degree of redundancy in an MDC. Our approach is based on an approximation to Mallat's model [4]. We show how our selection of psf is basically the same as that obtained by optimization of empirical data, with significantly less complexity.

1. INTRODUCTION

The problem of bit allocation has been studied in the literature [5, 6]. Two main approaches have been used: analytical approaches where Rate-Distortion (RD) data is based on a model (e.g., under the assumption of a Gaussian or Laplacian signal) and operational approaches where RD data is measured. In practice, however, there is no existing stochastic model that can be used to precisely incorporate the full diversity of image structures. In practical coding environments, a finite set of pre-selected admissible quantizers is used by the allocation algorithm to determine the best scheme to minimize the coding distortion subject to the given constraint in an operational RD framework. While this approach makes no assumption about input characteristics and can deal with arbitrary input images, it is obvious that the technique is limited in that the solution will have to be one of the discrete quantizer sets, so that optimality may suffer if a bad choice was made of those discrete quantizers. Another drawback of using empirical-based RD is that

collecting the RD data may be complex. The more admissible quantizers are used, the more time will be needed to compute the RD data. Therefore, it would be useful to design a bit allocation algorithm that can both efficiently work on arbitrary input images and enable fine granularity in the selection of arbitrary admissible quantizer sets and RD curve.

In this paper, we consider a new analytical model-based bit allocation across the regions in an image coded with a progressive wavelet coder. This technique extends Mallat's model [4] by incorporating a closed-form model and enabling it to be used in an ROI setting. A typical problem is to allocate bits optimally between the region of interest and the rests of the image, under a given constraint, by dividing the coefficients in each region by different factors at the encoder and multiplying these factors back to the corresponding coefficients at the decoder. While this bit allocation technique has been using in ROI coding [2, 3], we present a novel analysis for ROI coding. Furthermore, the proposed analytical bit allocation technique can be applicable to our MDC framework previously proposed in [7].

This paper is organized as follows: in Section 2, we introduce the analytical model-based bit allocation algorithm. We point out the accuracy of the proposed technique by comparing the analytical results with the empirical ones. We address the bit allocation problems in the MDC and ROI coding frameworks and show how this model can be applicable in Section 3. The conclusion of this work is discussed in Section 4.

2. ANALYTICAL MODEL

To start the analysis, we first introduce the empirical system we will analyze. Since SPIHT uses a successive approximation algorithm, if we divide the *whole* image by one constant, given the same bit rate, the reconstructed quality should be the same. However, if we take the wavelet coefficients and divide some of them, region 1, by a constant C_1 and others, region 2, by a constant C_2, then we can control their relative refinement levels at any level of resolutions. At the decoder we multiply these constants back to the corresponding decoded wavelet coefficients. This divide-and-multiply method is the same as the concept of quantization in typical ROI image codec [2, 3] in sense that, if $C_2 > C_1$, the quantization bin used in region 2 will be $\frac{C_2}{C_1}$ times larger than the

quantization bin used in region 1 at a given level of refinement. Therefore, we introduce higher distortion in region 2 than in region 1. In the case of SPIHT, if we divide region 2 by $\frac{C_2}{C_1}$ and leave region 1 unchanged, it generally yields the same performance, for a given total rate, as if we divide region 1 by C_1 and region 2 by C_2, because of the successive approximation algorithm.

Consider the Rate-Distortion function for SPIHT which has been proposed by Mallat [4]. Without loss of generality, we extend the original model by splitting an image into 2 regions. The distortion of each region can be represented as,

$$
\begin{aligned}
D_1(M_1) &= \frac{(1+\frac{1}{12})\sum_{i=M_1+1}^{N_1}\left|x_1(i)\right|^2}{N_1} = \frac{(1+\frac{1}{12})\alpha T_o^2 M_1 (1-(\frac{M_1}{N_1})^\beta)}{N_1} \\
D_2(M_2) &= \frac{(1+\frac{1}{12})\sum_{i=M_2+1}^{N_2}\left|x_2(i)\right|^2}{N_2} = \frac{(1+\frac{1}{12})\alpha T_o^2 M_2 (1-(\frac{M_2}{N_2})^\beta)}{N_2} \\
\text{where}\quad M_b &= (\tfrac{N}{5.5})R_b = M_1 + M_2 \\
\text{and}\quad N &= N_1 + N_2
\end{aligned}
\tag{1}
$$

where, for the jth region, $D_j(M_j)$ is the distortion, N_j is the total number of wavelet-transformed coefficients, $x_j(i)$ for $i = 1, \dots, N_j$ is the ith wavelet coefficient sorted by amplitude in monotonically decreasing order and M_j is the number of significant coefficients. Also, N is the total number of wavelet-transformed coefficients, M_b is the number of significant coefficients directly related to the desired average bit rate, R_b, α and β are fixed constants, and, finally, T_o is the quantization bin at final refinement. It is clear that the lower the bit rate R_b is, the smaller M_b is, and the larger the resulting distortion will be.

To analyze of the divide-and-multiply method, we divide the wavelet coefficients in region 2 by a psf which is a ratio of the dividing factor of region 2 and region 1 (i.e., $psf = \frac{C_2}{C_1}$), when $C_1 < C_2$. Therefore, the quantization bin of region 2 will be psf times larger than the quantization bin of region 1. It is worth noting that we use a normalized dividing factor so that the smallest analytical dividing factor is always equal to 1, e.g., $psf = 1$ at region 1. In Figure 1, we represent the sorted coefficients between region 1, X_1, and region 2, X_2, with the different quantization bins at $\hat{T}$ and $(psf)\hat{T}$, respectively. Since the coefficients in region 2 are divided while the coefficients in region 1 are left unchanged, at final refinement, the threshold of the significant coefficients is reduced (with respect to the threshold of the significant coefficients before dividing the coefficients in region 2 by the psf) from T_o to $\hat{T}$. The distortion equations are then modified as follows,

$$
\begin{aligned}
D_1(\hat{\mathbf{M}}_1) &= \frac{(1+\frac{1}{12})\alpha\hat{T}^2 \tilde{\mathbf{M}}_1 (1-(\frac{\tilde{\mathbf{M}}_1}{N_1})^\beta)}{N_1} = \frac{(1+\frac{1}{12})\sum_{i=\hat{\mathbf{M}}_1+1}^{N_1}\left|x_1(i)\right|^2}{N_1} \\
D_2(\hat{\mathbf{M}}_2) &= \frac{(1+\frac{1}{12})\alpha(\mathbf{psf}\hat{T})^2 \tilde{\mathbf{M}}_2 (1-(\frac{\tilde{\mathbf{M}}_2}{N_2})^\beta)}{N_2} \\
&= \mathbf{psf}^2 \left(\frac{(1+\frac{1}{12})\alpha\hat{T}^2 \tilde{\mathbf{M}}_2 (1-(\frac{\tilde{\mathbf{M}}_2}{N_2})^\beta)}{N_2}\right) \\
&= \mathbf{psf}^2 \left(\frac{(1+\frac{1}{12})\sum_{i=\hat{\mathbf{M}}_2+1}^{N_2}\left|\frac{x_2(i)}{\mathbf{psf}}\right|^2}{N_2}\right) \\
&= \frac{(1+\frac{1}{12})\sum_{\hat{\mathbf{M}}_2+1}^{N_2}\left|x_2(i)\right|^2}{N_2} \\
\text{where}\quad M_b &= \hat{M}_1 + \hat{M}_2
\end{aligned}
\tag{2}
$$

Clearly, one change is made with respect to the previous model of equation (1). That is, one of the two sets of coefficients is quan-
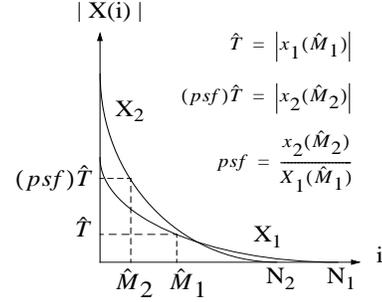

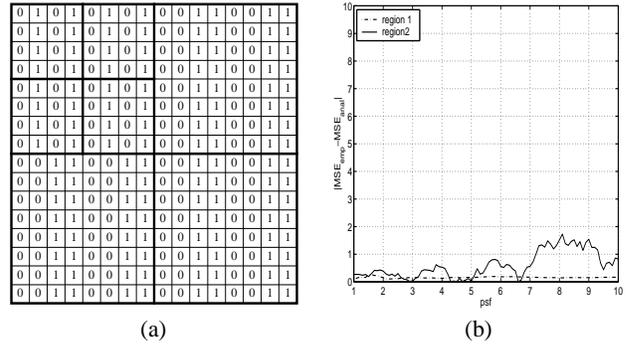Figure 1: Curves of two set of the sorted coefficients.

(a)   (b)

Figure 2: (a) Example of polyphase transform with a size of 16×16 to generate 2 polyphase components based on zerotree structure and (b) difference between the analytical and empirical MSE when using gray-level lena image of size 512×512 at rate 0.5 bps.

tized with a quantizer step size $(psf)\hat{T}$, instead of $\hat{T}$. In order to confirm the validity of the extended model, we identify the two regions generated using polyphase transform method by grouping the wavelet coefficients in a zerotree-based method as shown in Figure 2(a). We then divide the wavelet coefficients in the region 2 by the psf and vary its value to validate our analysis. The performance comparison obtained from the empirical experiment and analytical model-based method is illustrated in Figure 2(b). Note that based on the original Mallat's model, the accuracy of model will be highly achieved when the number of significant coefficients is large. Therefore, the larger the psf, the smaller the number of significant coefficients in region 2, and the larger the resulting mismatch between the empirical and analytical MSE will be.

As mentioned, the main contribution of this work is to find the appropriate psf such that the desired criteria will be satisfied (e.g., relative importance between regions in ROI coding framework or the level of redundancy in MDC framework). Having the distortion curve of each region from equation (2), given the explicit criteria such as the weighted distortion in each region, we can simply determine the parameter $\hat{M}_j$. The psf can be simply determined as shown in Figure 1 and used as the input parameter to the ROI codec afterwards. This modified model can be generalized to more than 2 regions which will be used in an MDC framework as shown next in Section 3.

3. APPLICATIONS AND EXPERIMENTAL RESULTS

Having established the intuition behind the proposed analytical bit allocation in 2 regions example, we now move to the practical applications: MDC in Section 3.1 and ROI coding in Section 3.2.

3.1. Multiple Description Coding (MDC)

We previously proposed a simple scheme for MDC using polyphase transform in [7]. In this paper, we improve our system by using psf to introduce the redundancy in each description. This is shown in Figure 3 for an MDC system with 2 descriptions where X_m stands for the mth polyphase component for $m = 0, 1$. In general, for the MDC system with S descriptions, the polyphase transform is used to generate the regions such that each region will have the same size, $N_j = \frac{N}{S}$ for $j = 0, \ldots, S-1$, and the similar property. The analytical psf technique has the advantage of being simple (a division operation) and flexible (the levels of redundancy can be easily changed).
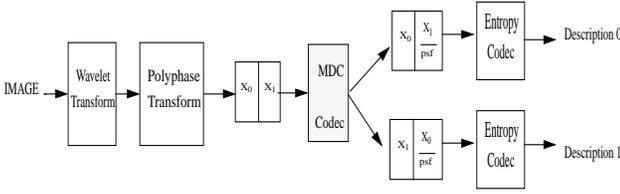
Figure 3: Block diagram of 2-description system using psf.

Without loss of generality, given that the total number of descriptions is 2 and the total bit rate R_b, we then code each description at the equally divided bit rate, $\frac{R_b}{2}$, to preserve our total bit rate budget. Let $D\left(\frac{X_{i\oplus j}}{psf_i} \middle| \frac{X_{i\oplus j\oplus 1}}{psf_{i\oplus 1}}\right)$ be the distortion associated with copy i of $i\oplus j$-polyphase component inside j-description (the modulo-2 addition is denoted by $\oplus$) for $i = 0, 1$ and $j = 0, 1$. As mentioned previously, $psf_0 = 1$ and $psf_1 = psf$. The argument in the condition part indicates the fact of the correlation between the main component and the other component inside a description. As derived in [7], the expected distortion can be computed as,

$$D_{avg} = \frac{1}{2}\left[\sum_{j=0}^{1}\sigma_j^2\right]P^2 + \frac{1-P}{2}\left[\sum_{i=0}^{1}P^i\sum_{j=0}^{1}D\left(\frac{X_{i\oplus j}}{psf_i}\middle|\frac{X_{i\oplus j\oplus 1}}{psf_{i\oplus 1}}\right)\right]. \quad (3)$$

Now our goal is to find the best psf, given the probability of packet loss P. Therefore, if we assume that we have access to the distortion curve $D\left(\frac{X_{i\oplus j}}{psf_i}\middle|\frac{X_{i\oplus j\oplus 1}}{psf_{i\oplus 1}}\right)$, searching all over the RD operating points should be the best way to determine the best psf. Using the models, we can search for the best values for psf without having to do extensive RD measurements or without restricting the psf to take discrete values.

In the 2-description case, the average distortion can be simplified from equation (3), by using the modeled distortion in equation (2), with two new parameters, M_0 and M_1, as

$$
\begin{aligned}
D_{avg}(M_0, M_1) &= \frac{1}{2}\left[\sum_{j=0}^{1}\sigma_{X_j}^2\right]P^2 + \frac{1-P}{2}\left[\sum_{i=0}^{1}P^i\sum_{j=0}^{1}D_j(M_i)\right] \\
&= \frac{1}{2}\left[\sum_{j=0}^{1}\sigma_{X_j}^2\right]P^2 + \frac{1-P}{2}\left[\sum_{i=0}^{1}P^i D(M_i)\right] \\
M &= M_0 + M_1 \quad\quad\quad\quad (4)
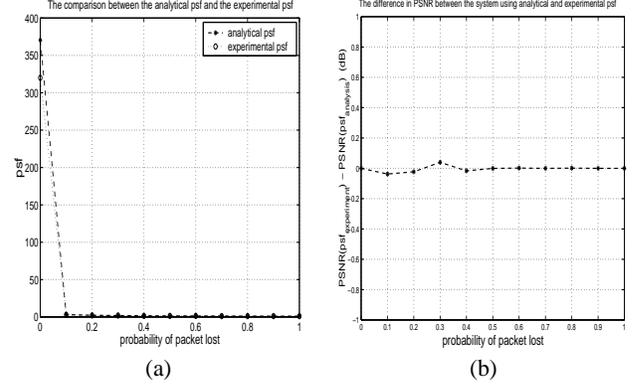\end{aligned}
$$

Figure 4: (a) Comparison between the characteristic distributions and (b) difference between the performances of the MDC system with 2 descriptions using the empirical psf and the analytical psf for gray-level lena image of size 512×512.

where $D_j(M_i)$ represents the distortion of the polyphase component that is divided by the psf_i in the jth description. Furthermore, since the polyphase transform has the property that the characteristics of each polyphase component can be approximately the same, they can approximately have the same curve of sorted coefficients. This implies that two curves of sorted wavelet coefficients for each polyphase component can be approximately represented by using the same parameters and it leads to the same distortion function for each description (i.e., $D_0(M_i) \approx D_1(M_i)$). Therefore, given the probability of packet loss, P, we can determine the optimal M_i^* by using Lagrangian optimization techniques [5]. With the optimal M_0^* and M_1^* which are determined from the optimization process, it is clear that the psf can be determined in the same way as shown in Figure 1. With varying packet loss rate, the comparison between the empirical-based and model-based psf as well as their performances are presented in Figure 4(a) and 4(b), respectively. The empirical-based psf is obtained from full search of all admissible psf of $\{1, 1.1, 1.2, \ldots, 400\}$ while the model-based psf is determined from our analysis as shown in Figure 1.

To confirm our validity of the proposed MDC algorithm, the gray-level lena image of size 512×512 is first wavelet transformed and 8 copies are generated. The polyphase components of each copy are divided by the appropriate psf values corresponding to the position inside each description. The set of psf used in the simulation is computed by our proposed algorithm based on the analytical model-based scheme. Given the total bit rate R_b, each uncompressed description from MDC codec then will be entropy-coded by SPIHT at the rate $\frac{R_b}{8}$ and the output bitstream will be transmitted as a description. At the receiver, the decoder will decide to use the available copy of each polyphase component that has the highest quality for reconstruction. Figure 5(a) shows the characteristics of bit allocation for the varying packet loss rates in terms of the psf assigned in each polyphase component inside a description. The psf_i represents the priority scaling factor for i^{th} component inside each description. As a comparison, when the packet loss rate increases (less number of receiving packets), the result is that the psf's tends to take similar values. Intuitively, this leads to the equal protection for each polyphase components across all descriptions, as expected. To evaluate the performance of our
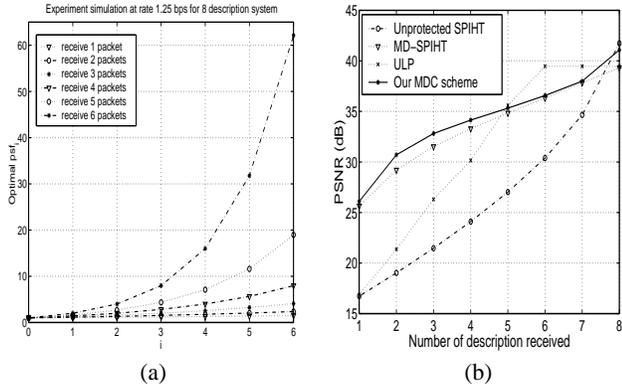
Figure 5: (a) Characteristic distribution of psf and (b) performance comparison between the proposed MDC, MD-SPIHT by Miguel *et al.* [8] and ULP by Mohr *et al.* [9] for the case of 8 descriptions with total bit rate at 1.25 bps.
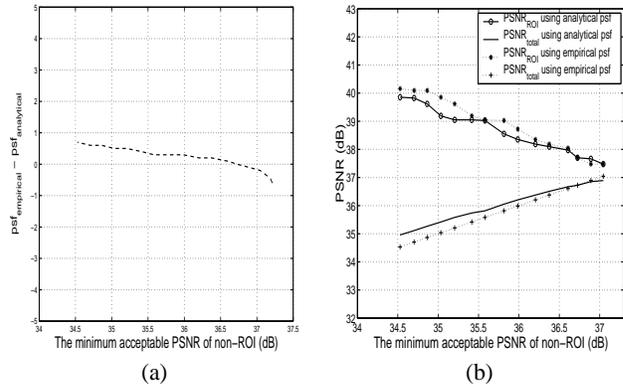

Figure 6: (a) Difference between the empirical psf and analytical psf and (b) performance comparison between ROI system using the empirical psf and analytical psf.

MDC scheme, we compare the results of our scheme with other MDC works [1] as shown in Figure 5(b). It is clear that our scheme can surpass the unprotected SPIHT and MD-SPIHT [8]. However, comparing with ULP [9] work, because of the advantage of using FEC, our approach is worse when several description have received, although the complexity of our scheme is lower. Note also that the version of SPIHT codec we use in the experiments is 7.05 while the one other MDC works use is 8.01.

3.2. Region of Interest Coding (ROI)

In this section, we apply our analytical model to a Region of Interest coding framework. We demonstrate the case of two region, ROI and non-ROI and choose the *minimum* acceptable quality of non-ROI as an example of optimization criteria. That is, given the lowest level of quality for non-ROI (e.g., the minimum PSNR that can be accepted) what the psf should be. Using the proposed models, we will be able to determine the psf without performing the RD data generation process. That is, assuming that region 1 is ROI and region 2 is non-ROI, we determine $\hat{M}_1$ and $\hat{M}_2$ based on the given distortion D_2 from equation (2), and then compute the psf in the same way as shown in Figure 1. The gray-level lena image of size 512×512 is used and the ROI is the rectangular block of size 16×16 in the middle of image. The whole image, after dividing the wavelet coefficients in the non-ROI by the psf, is coded by SPIHT at rate 0.5 bps. Figure 6(a) illustrates the difference between the empirical-based psf obtained from the full search of all admissible psf of $\{1, 1.1, 1.2, \ldots, 400\}$ and the model-based psf is determined from our proposed analysis. The performances of the whole image and ROI which are determined from ROI system using empirical and analytical psf are shown in Figure 6(b).

4. CONCLUSION

In this paper we propose the new analytical model-based bit allocation across the identified regions. We show that our analysis provides the close performance compared with the one obtained

from empirical experiments. We finally deploy our work in MDC and ROI coding framework to show how well our model performs in the practical applications.

5. REFERENCES

[1] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology, vol. 6, no. 4, pp. 243-250*, June 1996.

[2] E. Atsumi and N. Farvardin, "Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees," *Proc. IEEE International Conference on Image Processing (ICIP 98, pp. 87-91)*, Oct. 1998.

[3] D. Nister and C. Christopoulos, "Lossless region of interest with a naturally progressive still image coding algorithm," *Proc. IEEE International Conference on Image Processing (ICIP 98, pp. 856-860)*, Oct 1998.

[4] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. on Signal Processing, vol. 46, no. 4, pp. 1027-1042*, April 1998.

[5] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust. Speech Signal Process., ASSP-36(9), pp. 1445-1453*, Jan. 1988.

[6] J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated gaussian random variables," *IEEE Trans. Commun., vol. 11, pp. 289-296*, Sep. 1963.

[7] P. Sagetong and A. Ortega, "Optimal bit allocation for channel-adaptive multiple description coding," in *Proceedings of VCIP'2000*, San Jose, CA, Jan. 2000, pp. 53–63.

[8] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proc. of ICIP-99*, Kobe, Japan, Oct 1999.

[9] A. E. Mohr, E. A. Riskin, and R. Ladner, "Generalized multiple description coding through unequal forward error correction," in *Proc. of ICIP-99*, Kobe, Japan, Oct 1999.

---

[1] The authors would like to thank A. Miguel for providing the performance results in her website.