A JOINT SOURCE-CHANNEL DISTORTION MODEL FOR JPEG COMPRESSED IMAGES

M. Farooq Sabir, Hamid R. Sheikh, Robert W. Heath and Alan C. Bovik

Department of Electrical and Computer Engineering,
The University of Texas at Austin, Austin, TX 78712-1084, USA.
Email: {mfsabir, sheikh, rheath, bovik}@ece.utexas.edu

ABSTRACT

The need for efficient joint source-channel coding is growing as new multimedia services are introduced in commercial wireless communication systems. An important component of practical joint source-channel coding schemes is a distortion model to measure the quality of compressed digital multimedia such as images and videos. Unfortunately, models for estimating the distortion due to quantization and channel bit errors in a combined fashion do not appear to be available for practical image or video coding standards. This paper presents a statistical model for estimating the distortion introduced in progressive JPEG compressed images due to both quantization and channel bit errors. Important compression techniques such as Huffman coding, DPCM coding, and run-length coding are included in the model. Examples show that the distortion in terms of peak signal to noise ratio can be predicted within a 2 dB maximum error.

1. INTRODUCTION

With the introduction of high data rates in second and third generation wireless communication systems, real-time imaging and multimedia data transmission are becoming relevant applications in wireless cellular communication systems. Real-time imaging and multimedia data communication is especially difficult in wireless systems, however, due to the scarcity of bandwidth available. Techniques that reduce bandwidth consumption by making tradeoffs between distortion and compression rate, such as joint source-channel coding (JSCC) and unequal error protection (UEP), are consequently of great interest [1–3].

In JSCC, the goal is to minimize the distortion introduced in the received data by optimizing the distribution of available bits amongst source and channel coding. This can be done either on a 'per image' basis using simulations, or for an ensemble of images by developing statistical distortion models. In [2], a distortion model is formulated for the discrete wavelet transform (DWT) compressed images. Bit sensitivities for different classes of bits are derived, and this model is used for efficient joint allocation of source and channel bits. In [4], the authors have first derived an expression for the expected value of distortion for a general class of images. This model is then applied to different classes of source and channel coders. UEP is a pragmatic approach to JSCC in which different levels of error protection are provided to different parts of the data to minimize distortion at the receiver. In [3], an optimization problem based on an 'incremental award' with each correctly decoded source packet is formulated. This optimization is then carried out for different rate channel codes. Different source packets are then unequally protected using these channel codes.

Most of the JSCC and UEP schemes in literature are based on the minimization of distortion on a 'per image' basis rather than developing expressions for evaluating average distortion as a function of source and channel parameters for a large set of images. The 'per image' based approach is not an ideal solution for real-time image communication, as it increases the complexity of the transmission system, in addition to introducing delay. Therefore, it is essential to have distortion models that can predict the average distortion over a set of images as a function of source coding rate and channel bit error probability. Efficient JSCC and UEP schemes can be designed based on these models. Unfortunately, the models employed in the existing JSCC and UEP techniques do not account for practical compression techniques such as Huffman coding, DPCM coding, and run-length coding that are present in standard source coders.

In this paper, we develop a statistical distortion model for progressive JPEG compressed images that jointly estimates distortion due to both quantization and channel errors. Our approach is to divide the discrete cosine transform (DCT) coded JPEG images into different layers based on different subbands. The expected value of the mean squared error (MSE) is then found as a function of source coding parameters and the channel bit error probability for each layer. The total distortion is then the sum of these distortion terms due to individual layers. In contrast with prior work, our model takes Huffman coding, DPCM coding and run-length coding into account. We compute the parameters of our model from a 'training' database of images. In our simulations, our model performs within 2 dB of peak signal to noise ratio (PSNR) when tested on a test database of images. While we derive the expressions explicitly for the JPEG standard using Huffman coding, our model can be extended to other coefficient-by-coefficient coding schemes employing any kind of entropy coding.

2. SYSTEM MODEL

Source coding often makes the compressed bitstream highly sensitive to channel errors due to the presence of entropy coding. A single bit error has the potential to corrupt an entire image or video sequence. Therefore, error-resilient features are required for constructing feasible multimedia communication systems. In this paper, we use the JPEG standard with certain error-resilient features. Specifically, we use RST (reset) markers in progressive DCT based mode, and assume that the headers are not corrupted by any bit error. These features are important for constructing any practical image communication system based on the JPEG algorithm.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

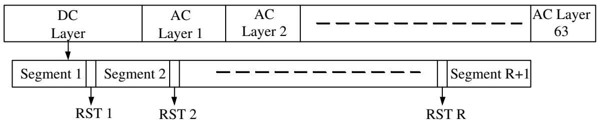


Fig. 1. JPEG Layers and Segments. The headers for each layer are not shown for simplicity.

2.1. The Source Coding Model

In our JPEG encoding, we use the progressive DCT based mode of operation with spectral-selection [5]. In the progressive DCT mode, the data is arranged in different quality layers. In the spectral-selection method, the DCT coefficients are divided into subbands that are encoded in separate passes. After quantization, the DC coefficients are DPCM coded in the first pass, whereas the AC coefficients are run-length coded in subsequent passes. Both the DC and the AC coefficients are then entropy coded. We use Huffman coding for the model presented in this paper, however, this model would also work for arithmetic coding.

In the model discussed in this paper, the 64 different subbands of the DCT coefficients are organized in separate layers, thus giving a total of 64 layers, the first one being the DC layer followed by 63 AC layers. In this way the resolution and the quality of the decoded image is improved as more layers are decoded. RST (reset) markers are inserted in each layer regularly. We call the portion of each layer within two consecutive RST markers a 'segment'. Decoding is reinitialized whenever a RST marker is encountered [5], and a bit error occurring in the bitstream only corrupts the image until the next RST marker. In case bit errors occur, we assume that the decoder detects the first bit error (due to loss in synchronization of Huffman decoding) and decodes all the coefficients in the rest of the segment as zero. The structure of the JPEG compressed stream is shown in Fig. 1.

2.2. The Channel Model

We consider the channel to be a binary symmetric channel (BSC) and derive our distortion model for a given bit error probability p_e. Both the AWGN and the Rayleigh fading channels can be represented as a BSC, given the bit error probabilities for these channels and the fact that the probability of making an error from 0 to 1 is the same as that of 1 to 0. Therefore, using this model we can find the distortion curves for any channel that can be represented as a BSC, if the source coding rate and the expression for bit error rate (BER) are known. This makes our distortion model independent of the modulation type and channel coding. Also, no packetization is considered, and it is assumed that the encoded bitstream is directly transmitted over the channel without any encapsulation or overhead by streaming protocols.

3. DISTORTION MODEL

In this section we derive the expressions representing the distortion in the JPEG image due to quantization and channel errors in a combined fashion. We use MSE as our distortion metric.

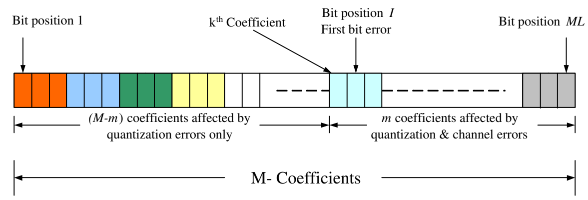


Fig. 2. Bit error in a segment. Different colors represent different coefficients.

3.1. Assumptions and Notation

Our goal is to relate image distortion in terms of MSE to the source coding rate and the bit error probability. This modelling is complicated due to the presence of Huffman coding, DPCM coding and run-length coding. A single bit error can cause the decoder to lose synchronization and corrupt the entire segment. Furthermore, it is not possible to precisely determine the coefficient position corresponding to a particular bit in error due to the different lengths of the entropy coded symbols. Due to these challenges such a modelling has not been attempted in the past.

In this paper, we make certain simplifying assumptions that allow us to derive the average MSE due to quantization and channel errors over a set of images. Specifically, we will model the DCT coefficients as random processes that are wide sense stationary and ergodic. Also, since by assumption, the first bit error in a segment corrupts the entire segment from the bit in error to the next RST marker, we only need to consider the position I of the first bit in error. Thus multiple errors in the same codeword, as well as all other errors in the same segment, can be ignored. We also assume that the bit errors are random and independent of each other. In our JPEG encoder implementation, a RST marker is inserted in every layer after every M coefficients, which is constant for all layers. The coefficient corresponding to the bit position I is denoted as k. All coefficients from k to the end of the segment are assumed to be decoded as zero, while all previous coefficients from the start of the segment to the $(k-1)^{st}$ coefficient are assumed to be decoded correctly.

If p_e is the probability of bit error, then the distribution for I, $p_I(i)$, is given by

$$p_I(i) = p_e(1-p_e)^{i-1}. \quad (1)$$

Hence an increase in the bit error rate will increase the probability of first bit error occurring at an earlier location in the bitstream. Consider a segment with M DCT coefficients indexed from 1 to M, as shown in Fig. 2. We assume that the average length of a coded coefficient is a random quantity denoted by L, whose distribution depends upon the source coding rate. Thus, on average $k \approx \lceil I/L \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Note that this relation holds only on average, and the distortion model based on it would only predict the average MSE over a set of images. The number of coefficients in a segment corrupted by a bit error is denoted by $m(L, I)$, or m for notational convenience. Thus $M - m$ coefficients are decoded correctly, while m coefficients are corrupted and decoded as zero. This is depicted in Fig. 2.

We will first derive an expression for MSE conditioned on the knowledge of L and I, and later average over L and I using their respective distributions. Since in JPEG encoding, DC is DPCM coded, while AC is coded directly, the models for MSE need to be derived separately for DC and AC layers.

3.2. Distortion Model for the DC Layer

We can write the expected value of MSE between the original and the erroneous coefficient at position k given L and I as

$$E\left[MSE^k \mid L, I\right]=\frac{1}{N}E\left[(X^u_{k(L,I)} - \widehat{X}^q_{k(L,I)})^2 \mid L, I\right] \quad (2)$$

where $X^u_{k(L,I)}$ is the unquantized DC coefficient, $\widehat{X}^q_{k(L,I)}$ is the erroneously decoded quantized DC coefficient and N is the total number of pixels in the entire image. Note that $X^u_{k(L,I)}$ and $\widehat{X}^q_{k(L,I)}$ are also random variables. We will write $k(L, I)$ as k to keep the notation simple. The DC coefficients in JPEG are DPCM coded, and only the DC 'prediction' is coded in the bitstream. Thus, $\widehat{X}^q_k = X^q_{k-1} + \widehat{P}^q_k$, where X^q_{k-1} is correctly decoded, while $\widehat{P}^q_k$ is the erroneous prediction value. By assumption, the decoder decodes $\widehat{P}^q_k$ as zero. Thus

$$\widehat{X}^q_k = X^q_{k-1}. \quad (3)$$

The quantized DC coefficient X^q_k can be expressed as the sum of the unquantized coefficient X_k and a quantization error ξ_k

$$X^q_k = X^u_k + \xi_k. \quad (4)$$

Let the variances of the unquantized DC coefficients, the quantized DC coefficients and the quantization error be σ^2_u, σ^2_q and σ^2_ξ respectively. Now, expanding (2) and using (3) and (4), the expected value of MSE can be written as

$$E(MSE^k|L, I)=\frac{1}{N}(\sigma^2_u + \sigma^2_q - 2E\left[X^u_k X^u_{k-1} \mid L, I\right] - 2E\left[X^u_k \xi_{k-1} \mid L, I\right]). \quad (5)$$

We can fairly assume that the quantization error is uncorrelated with the unquantized DC coefficients; i.e. $E\left[X^u_k \xi_{k-1} \mid L, I\right] = 0$. Hence the expected value of MSE becomes

$$E(MSE^k|L, I) = \frac{1}{N}(\sigma^2_u + \sigma^2_q - 2r(1)); \quad (6)$$

where $r(1)$ is the autocorrelation function of the DC coefficients at lag 1. Using a similar methodology, MSE at a distance of $j-1$ coefficients from the k^{th} coefficient can be written as

$$E(MSE^{k+j-1}|L, I) = \frac{1}{N}(\sigma^2_u + \sigma^2_q - 2r(j)) \quad (7)$$

Since a total of m coefficients are decoded erroneously, assuming additivity, the total MSE (MSE_m) due to these m coefficients is

$$E(MSE_m|L, I) = \frac{1}{N}\left[m \cdot (\sigma^2_u + \sigma^2_q) - 2\sum_{j=1}^{m} r(j)\right]. \quad (8)$$

Recall that $M-m$ coefficients are decoded correctly, and we only need to model quantization error for them. Adding this quantization distortion, and averaging over I and L, we get

$$E(MSE_M)=\frac{1}{N}\sum_l \left[\sum_{i=1}^{Ml}[(M-m)\sigma^2_\xi + m \cdot (\sigma^2_u + \sigma^2_q) - 2\sum_{j=1}^{m} a^j \sigma^2_u]p_I(i)\right] p_L(l). \quad (9)$$

where we have used a first order auto-regressive model for the DC coefficients ($r(j) = a^j \sigma^2_u$). The outer summation is taken over all possible lengths l of the coded DC coefficients. We also need to consider the event when no bit error occurs in the entire segment and the distortion is solely due to quantization. The probability of such an event given L is $p(\text{No error}|L = l) = 1 - \sum_{i=1}^{Ml} p_I(i)$. Including this in our expression for MSE and assuming the additivity of MSE due to all the $R + 1$ segments in the DC layer, the total expected MSE in the image due to the DC layer is:

$$E(MSE_{DC})=(R+1)\left[\sum_l \frac{1}{N}[1 - \sum_{i=1}^{Ml} p_I(i)]M\sigma^2_\xi p_L(l) + E(MSE_M)\right]. \quad (10)$$

3.3. Distortion Model for AC Layers

The 63 AC subbands in the JPEG compressed image constitute the next 63 quality layers in our model. Following similar steps as for the DC layer, the expected value of MSE due to the n^{th} AC layer can be shown to be by

$$E(MSE_{AC,n})=\frac{1}{N}\sum_l \left[\sum_{i=1}^{Ml}[(M-m)\sigma^2_{\xi,n} + m\sigma^2_{u,n}]p_I(i) + [1 - \sum_{i=1}^{Ml} p_I(i)]M\sigma^2_{\xi,n}\right] p_{L_n}(l) \quad (11)$$

where $\sigma^2_{u,n}$ and $\sigma^2_{\xi,n}$ are the variances of the unquantized AC coefficients and the associated quantization error respectively for the n^{th} AC layer, and L_n models the average length of coded coefficients in the n^{th} AC layer. Note that there is no correlation term in the MSE expression for the AC layers. This is because the AC layers are not DPCM coded. Since we use a different distribution for L_n for each layer, the effect of run-length coding in AC coefficients is incorporated automatically in our derivation.

3.4. Total Distortion

Assuming additivity of distortion due to individual layers (recall orthonormality of the DCT basis), the total expected value of the MSE due to quantization and channel errors in all the layers can be written as

$$E(MSE)=E[MSE_{DC}] + \sum_{n=1}^{63} E[MSE_{AC,i}] \quad (12)$$

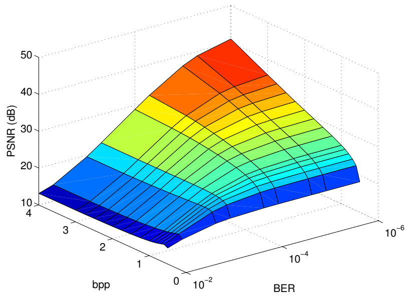
where $E(MSE_{AC,i})$ is the expected value of MSE for the i^{th} AC coefficient layer.

4. SIMULATIONS AND RESULTS

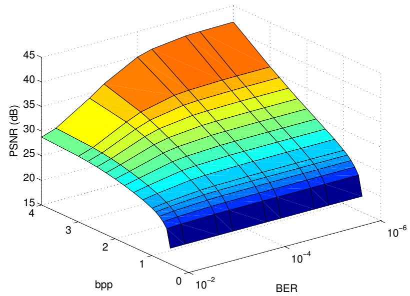
In this section we compare our model's prediction of average MSE against simulations. MSE is converted to PSNR using the simple relation $\text{PSNR} = 10\log_{10}\frac{255^2}{\text{MSE}}$, since PSNR is commonly used for image quality assessment.

4.1. Simulation Details

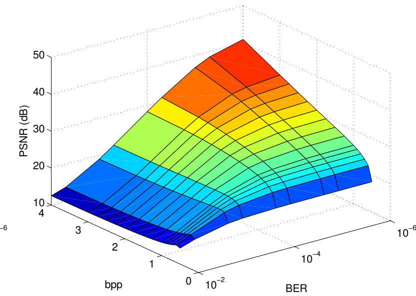
Two different databases of 200 512 $\times$ 512 natural grayscale images each are used in the simulations, one for training and the other for testing. For each layer, the variances σ^2_u, σ^2_q, and σ^2_ξ for DC, and $\sigma^2_{u,n}$, $\sigma^2_{q,n}$, and $\sigma^2_{\xi,n}$ for AC coefficients, and the empirical distributions for L and L_n are obtained from the training database. During simulations the headers and markers are separated and it is



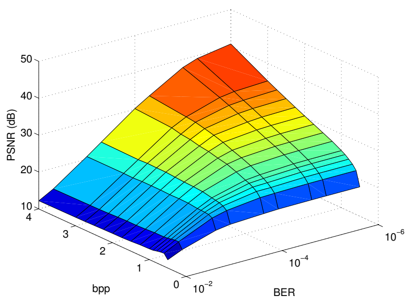
(a) DC model



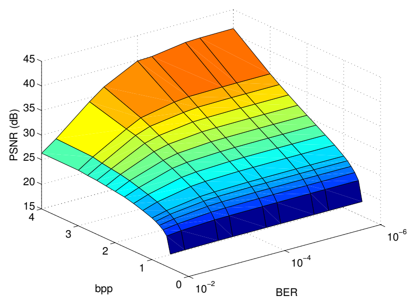
(b) AC model (first AC layer)



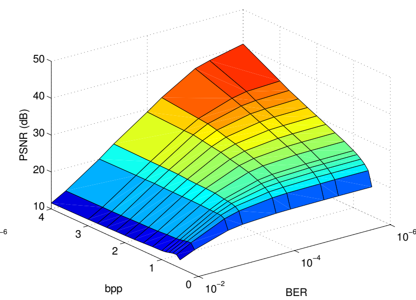
(c) Combined 64 layers model



(d) DC test simulation



(e) AC test simulation (first AC layer)



(f) Combined 64 layers test simulation

Fig. 3. PSNR vs BER and bpp curves for the DC layer, the first AC layer and all 64 layers combined.

assumed that they are transmitted without errors. This is a valid assumption since powerful channel codes can be used to transmit the headers and markers, which constitute a very small portion of the overall bitstream, without any bit errors. In our simulations RST markers are introduced after every 64 coefficients ($M = 64$). Random bit errors are introduced in the encoded bitstream without any channel coding. No error concealment is used at the decoder.

4.2. Results and Discussion

The model presented in this paper is evaluated at different source rates and channel bit error probabilities and plotted in Fig. 3 (top row). Using simulations, the average PSNR for the testing set of images using the system running under different source coding rates and error probabilities is plotted as a function of the source coding rate (in bits per pixels) and the bit error rate, as shown in Fig. 3 (bottom row). The model is compared against simulations for the DC layer, the first AC layer, and combined 64 layers.

As can be seen from Fig. 3, our models can accurately predict the distortion introduced in images due to source coding and channel errors. For the DC layer, the difference in PSNR obtained using the model and the simulations is within 2 dB at all the points. For the first AC layer this difference is within 2.5 dB at all the points, and for all the layers combined this difference is within 1.5 dB.

5. CONCLUSION

In this paper we presented a new model for estimating the distortion introduced in an image as a result of quantization and random bit errors when compressed by a JPEG encoder, and transmitted over a noisy/fading channel. This model is unique since it also incorporates modelling of DPCM and entropy coding for estimating distortion. To our knowledge such a model has not been presented previously. Simulation results show that the distortion predicted by the model estimates the true value obtained via simulations quite closely. Our model can be used to devise efficient JSCC and UEP schemes for transmission of JPEG compressed images over noisy and fading channels. We plan to formulate similar distortion models for MPEG-4 and H.263 video coding standards. In addition, we plan to devise joint source-channel coding and unequal error protection schemes based on these models for JPEG, MPEG-4 and H.263 standards.

6. REFERENCES

[1] J. W. Modestino, D. G. Daut, and A. L. Vickers, "Combined source-channel coding of images using the block cosine transmform," *IEEE Transactions on Communications*, vol. com-29, pp. 1261–1274, Sept. 1981.

[2] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Transactions on Image Processing*, vol. 8, pp. 305–320, Mar 1999.

[3] V. Chande and N. Farvardin, "Progressive transmission of images over memoryless noisy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 850 –860, June 2000.

[4] S. Appadwedula, D. L. Jones, K. Ramchandran, and L. Qian, "Joint source channel matching for wireless image transmission," *Proc. International Conference on Image Processing*, vol. 2, pp. 137–141, 1998.

[5] W. B. Pennebaker and J. L. Mitchell, "JPEG still image data compression standard." Van Nostrand Reinhold, 1993.