VISUALLY OPTIMIZED MULTIPLE DESCRIPTION IMAGE CODING

Chao Tian[*]

School of Computer and Communication Sciences
Swiss Federal Institute of Technology
Lausanne, Switzerland, CH 1015

Sheila S. Hemami

School of Electrical and Computer Engineering
Cornell Univerisity
Ithaca, NY 14853

ABSTRACT

We consider the problem of constructing visually optimized balanced multiple descriptions of images, instead of being optimized with the conventional measure of mean squared error (MSE). The recently proposed Modified Multiple Description Scalar Quantizer (MMDSQ) is used because its central and side quantizers all have convex and uniform cells. Since the majority of research results on the visual distortion caused by quantization noise assumes convex and uniform quantizer cells, they can be conveniently applied to the system based on MMDSQ. The technique provides substantially higher side reconstruction perceived quality when compared at the same central reconstruction quality, which is verified through perceptual tests.

1. INTRODUCTION

The multiple description (MD) problem considers encoding a source into two descriptions, either of which can be used to reconstruct the source, while the two descriptions can jointly provide a better quality reconstruction. This coding approach is useful in situations when two unreliable channels are present between the transmitter and the receiver, as well as in packet networks with loss (see [1] for an excellent review). From an information theoretic point of view, we consider encoding a source X into two descriptions with rate constraints R_1 and R_2, respectively, and the reconstructions by using the individual description and both descriptions induce the side distortions D_1, D_2, and the central distortion D_0, respectively [2]. The most often considered case is when the descriptions are balanced, where $R_1 = R_2 = R$ and $D_1 = D_2$, which we assume in this work. In this rate-distortion setup, the mean squared error (MSE) is most commonly taken as the distortion measure, especially when the source X is continuous [3].

Though Peak Signal to Noise Ratio (PSNR), which is inversely proportional to MSE, is widely used as a measure of the reconstructed image quality in image coding literature, it is indeed the quality of the image perceived by the human visual system (HVS), i.e., the *visual quality*, that matters in the applications when the end users are human observers. PSNR can be misleading: it was shown [4,5] that one (reconstructed) image can be ranked as better than another in terms of visual quality, but ranked worse in terms of PSNR .

MD wavelet image coding has previously been addressed in [6, 7], however with the goal of maximizing PSNR. In this work, we consider constructing a visually optimized MD system by incorporating existing results on HVS, particularly, the Dynamic Contrast-Based Quantization (DCQ) algorithm [5]. DCQ is selected over other algorithms such as the visual optimization tools in JPEG-2000 because it offers a single solution for all rates. The visually optimized MD system is practically implemented using the Modified Multiple Description Scalar Quantizer (MMDSQ) [7]. MMDSQ is simple yet efficient, furthermore, it has one desirable feature that the side quantizer and central quantizer all have convex and uniform quantization cells, and it is under this condition the psycho-visual tests are usually conducted [5, 8]; in contrast, the previous design of the Multiple Description Scalar Quantizers (MDSQ) [9] does not promise such convexity. We show that the DCQ algorithm can be conveniently applied to an MMDSQ-based system, and the resulting wavelet codec provides a significant visual quality improvement comparing to the MSE-based system in [7]: the technique provides substantially higher side reconstruction perceived quality when compared at the same central reconstruction quality. This work highlights the unique advantage of MMDSQ over previously proposed quantization techniques in both its simplicity and efficiency.

2. MMDSQ AND THE DCQ ALGORITHM

2.1. MMDSQ and its properties

When quantization is used to form multiple descriptions of a source, the goal is to create two coarse side quantizers which produce acceptable side distortions when used alone, while combining them together can produce a finer central quantizer, which provides lower distortion than the side quantizers. The side quantizers can be unconventional in the sense that the quantization cells might not be convex. A general design of MDSQ was pioneered by Vaishampayan [9].

Different from the framework proposed in [9], MMDSQ has two stages: in the first stage, the two side quantizers are two uniform quantizers with their bins staggered by half of

[*]The author performed the work while at Cornell University.

the step size (Fig. 1), and the joint quantizer formed by the two side quantizers has a half-sized quantization bin; in the second stage, each bin in the joint quantizer is refined further into a fixed number of smaller bins. The two descriptions are formed as follows: the quantization indices from each side quantizer are entropy-coded independently into the bitstream of each description (base layer); then the indices from the second stage are again entropy-coded, but split evenly between each description (refinement layer). When the decoder receives only one description, it discards the refinement layer information, and only decodes the base layer information in this description; with both descriptions available, the decoder can reconstruct to a better quality by using both the base layer and the refinement layer information.

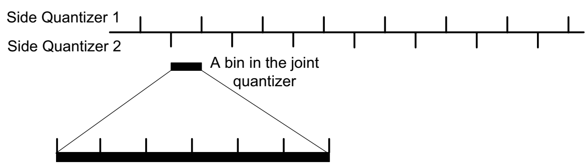

Fig. 1. The structure of MMDSQ.

Surprisingly, this extremely simple structure offers very competitive performance. Its asymptotic performance is the same as MDSQ with a uniform central quantizer in terms of MSE (see [7] for details). Furthermore, notice in Fig. 1 that both side quantizers have convex and uniform central quantizer cells; in contrast, the side quantizers in MDSQ [9] usually have non-convex side quantizer cells.

The MMDSQ framework offers two parameters to control the overall quantization. One parameter is the first stage quantization step size Δ, which determines the base layer rates; the other is the number of finer quantization cells in each joint quantizer cell. Denote this number as N, which implies the central quantizer in fact is a uniform quantizer with step size $\frac{\Delta}{2N}$. When the individual description is constrained by a given one-description rate R, the choice of the parameters (Δ, N) has only one degree of freedom, which in fact determines the tradeoff between the central and side distortions; in other words, under a fixed rate constraint, a lower central distortion necessitates a higher side distortion and vice versa.

2.2. The DCQ algorithm

Lossy image compression is widely used in practice (e.g. in both JPEG and JPEG2000 standards) to achieve higher compression ratio than lossless compression. An image codec should also take into account the properties of HVS, such that it can offer the most visually pleasing reconstruction given a fixed bit rate budget. Particularly, different step sizes can be strategically selected to quantize coefficients in wavelet subbands to achieve this goal. The Dynamic Contrast-based Quantization (DCQ) algorithm is one of such algorithms.

For the purpose of this work, it suffices to understand the DCQ algorithm as follows. Let G_{VD} be a control parameter defined on the interval of $(0, 1]$ that reflects the visual distortion. Given a target G_{VD}, DCQ computes a set of appropriate step sizes $[\Delta(0), \Delta(1), ...]'$, where $\Delta(s)$ is used to uniformly quantize subband s (see [5] for more details).

3. A VISUALLY OPTIMIZED MD IMAGE CODEC

In [7], an MD image codec using MMDSQ was proposed. It is based on a wavelet bit-plane image coder, namely the Tarp filter image coder with classification for embedding (TCE) [10]. The TCE coder is simple, yet its performance is competitive (comparable to JPEG2000). In this section, the DCQ algorithm is further incorporated into the system in [7] to optimize the visual quality, instead of minimizing the MSE.

To encode an image into a single description, the DCQ algorithm provides a set of visually optimized quantization step sizes $\Delta_{SD}(s)$, $s = 0, 1, ..., m-1$, where m is the number of subbands after performing DWT. It is clear that in an MMDSQ-based system, there are two quality layers, i.e., the base layer and the refinement layer. For the base layer, the DCQ algorithm is again able to provide a set of visually optimized quantization step sizes $\Delta_b(s)$, $s = 0, 1, ..., m-1$ for the first stage side quantizers. For the refinement layer, a set of values of $N(s)$ can be used to approximate another set of optimal quantization step sizes with a lower visual distortion as $\Delta_r(s) = \frac{\Delta_b(s)}{2N(s)}$. Notice $N(s) = 1$ implies that there is no refinement layer in subband s. Since $N(s)$ can take only integer values, a rounding approximation is in fact made on the true optimal step sizes $\Delta'_r(s)$ such that $\Delta'_r(s) \approx \Delta_r(s)$.

If bitplane image codecs, such as TCE [10], are used, $N(s)$ has to be in the form of 2^n, where n is a non-negative integer, and thus further approximations on the refinement layer step sizes have to be made. More precisely, instead of approximating the true step sizes $\Delta'_r(s)$ by an even integer fraction of $\Delta_b(s)$, we in fact approximate them by a dyadic fraction of $\Delta_b(s)$. The visual quality degradation by using this approximation is minor: a similar approximation is made when the DCQ algorithm is applied to JPEG2000 in [5].

For higher visual distortion G_{VD} (or a low bit budget), the DCQ algorithm requires certain frequency subbands be discarded altogether, which corresponds to a quantization step size of $\Delta(s) = \infty$. However, in the refinement layer, the visual distortion is reduced, and some of these discarded subbands will be quantized by a quantizer with a finite step size and subsequently encoded. In this case, it is meaningless to require that $\Delta_r(s) = \frac{\Delta_b(s)}{2N(s)}$ still holds. Additional header information can be used to signal such event.

It is also desirable if the encoder can offer precise rate-control. In the proposed system, rate control can be performed as follows. For the base layer, the rate control algorithm in [5] can be used such that a reasonable range of G_{VD} value is chosen for the base layer. For the refinement layer, the remaining bit budget can be used to encode the subbands in an embedding manner, and encoder can choose to stop the encoding when the remaining bit budget is depleted. Denote the step sizes corresponding to a visual distortion G_{VD} as $\Delta^{G_{VD}}(s)$, and also define the indicator function

$$I(\Delta, \Delta') = \begin{cases} 1 & \Delta = \infty, \Delta' < \infty; \\ 1 & \Delta < \infty, 2\Delta' < \Delta; \\ 0 & otherwise. \end{cases}$$

Note $\Delta(s) < \infty$ means that the subband s will be quantized rather than discarded. As G_{VD} decreases, the step size of quantization for each subband monotonically decreases. Thus we have the following Algorithm 1.

Algorithm 1 Embeddedly encoding the refinement layer

Given:
$\Delta_r(s) = \Delta_b(s)/2$, $s = 0, 1, ..., m-1$: quantization step sizes; $G_{VD}(0) = G_{VD_b}$: the base layer visual distortion parameter; R': the remaining bit budget per description; $n = 1$.
while $R' > 0$ **do**
  Search for a $G_{VD}(n) < G_{VD}(n-1)$ such that $\sum_s I(\Delta_r(s), \Delta^{G_{VD}(n)}(s)) = 1$;
  Let $s' = \arg_s [I(\Delta_r(s), \Delta^{G_{VD}(n)}(s)) = 1]$;
  if $\Delta_r(s') = \infty$ **then**
    $\Delta_r(s') = \Delta^{G_{VD}(n)}(s')$;
  else
    $\Delta_r(s') = \Delta_r(s')/2$.
  end if
  Encode the subband s', with step size $\Delta_r(s')$; denote the increment rate as ΔR;
  Let $R' = R' - \Delta R/2$; $n = n + 1$;
end while

Note that reducing the step size by a factor of two is equivalent to encoding a lower bit plane. In searching for the appropriate value of $G_{VD}(n)$, there is no need to perform the actual entropy coding, thus it can be carried reasonably fast; this step can be made even faster by adopting the model-based rate-control approach in [5]. When the algorithm is terminated with $G_{VD}(n)$, the resulting step sizes Δ's are as if the visual distortion parameter was specified as $G_{VD}(n)$ before encoding, with the additional constraint that each step size has to be a dyadic fraction of $\Delta_b(s)/2$.

4. RESULTS

A five-level wavelet decomposition is used with the Daubechies 9/7 filters, and all the subbands are uniformly quantized by quantizers whose step sizes are determined by the DCQ algorithm, as given in the last section. The test images are resolution 512-by-512 gray-scale images.

A perceptual test was conducted in order to compare the performance of the proposed system (denoted as MMDSQ-TCE-VO) with that of the reference system [7], which is essentially the same system but without the visual optimization (denoted as MMDSQ-TCE). It is desirable to compare the visual qualities of the side (central) reconstructions when the two systems operate at the same central (side) visual quality. However, the visual quality is not readily quantified with such accuracy, which makes this task difficult. To simplify the overall perceptual test, the following approach is taken. Preliminary experiments are used to determine operating points for a given image at the same rate, where the central reconstruction quality of MMDSQ-TCE-VO is slightly better than that of MMDSQ-TCE; then the corresponding side reconstructions for the two systems can be compared. In doing this, the better visual quality of the central reconstruction of MMDSQ-TCE-VO is in fact taken to be equal to that of the reference system, and the main comparison is made on the side reconstructions, thus in a sense we discriminate against the proposed system.

Table 1. Results of the perceptual test for eight test subjects, each of whom were asked to compare 12 pairs of side reconstruction images and 12 of central reconstructions.

Category of comparison	# of choices			Total
	Proposed	Can not rank	MMDSQ-TCE	
Side reconstructions	96	N/A	0	96
Central reconstructions	84	11	1	96

Three typical test images were used, each compressed at 0.5 bpp/description and 0.25 bpp/description; at each rate, two operating points were chosen, one with relatively high redundancy between the two descriptions, the other one with relatively low redundancy. This amounts to 12 comparisons between the two systems. The side reconstruction image pairs were shown to the test subjects, and the eight test subjects were asked to select the better of the two images; the original images were shown to the subjects before the test started. To confirm that the central reconstruction images chosen by the preliminary experiments are indeed discriminating against the proposed system, they were also shown to the test subjects, but with the ternary choices among "better", "worse" or "can not rank". Overall, 24 pairs of images were included in the test: 12 of them are for the central reconstructions, and the other 12 are for the side reconstructions. With eight test subjects, we have a total of 96 comparisons for the central reconstructions, and the same number for the side reconstructions. Results of the test are summarized in Table 1, where the results for different images and operating points are summed up together instead of listed separately.

From Table 1, we can see that the proposed system is in fact operating at a better central reconstruction quality than the reference system, while at the same time its side reconstruction quality is significantly better: the test subjects overwhelmingly preferred the proposed system in the test on the side reconstructions, without even once preferring the reference codec. Recall that our aim was to choose the operating points such that the perceived quality of the central reconstructions of the two systems are approximately the same, and to compare the side reconstructions. The results imply that the side reconstruction perceived quality of the proposed sys-

Fig. 2. Comparison of side reconstructions at 0.5 bpp/description, with approximately the same central perceived quality: (a) without and (b) with visual optimization. These images are meant to be viewed from approximately 2-3 picture heights.

tem can be further improved, when the central reconstruction is properly adjusted such that the two systems operate at the exact same central reconstruction quality.

In conventional single description image coding, it is a fact that visual optimization can provide a significant improvement in terms of visual quality in the relatively low bit rate region. In MD image coding, though the overall bit rate is usually reasonably high (in the test at 0.5 bpp and 1.0 bpp in total), the effective bit rate is actually lower, which is due to the amount of redundancy introduced by the MD coding. In this sense, visual optimization is more important in MD image coding than in single description coding.

A set of comparison is given in Fig. 2. The two systems are compared at approximately the same central reconstruction visual quality (MMDSQ-TCE-VO gives a slight better visual quality). The difference is particularly visible in the water ripple and crane edge area. More test images can be found at http://foulard.ece.cornell.edu/tch/MDTCEVO.htm.

5. CONCLUSION

We considered the problem of constructing visually optimized balanced multiple descriptions of images. By combining the recently proposed MMDSQ and existing psycho-visual results, such a system is formed. Furthermore the embedded feature of TCE make the rate-control of the proposed system straightforward. Tests on natural images show a significant improvement in terms of visual quality, comparing to its counterpart without considering HVS properties.

6. REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 47, no. 6, pp. 2199 – 2224, Sep. 2001.

[2] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Information Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[3] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Information Theory*, vol. 45, no. 7, pp. 2629–2636, Nov. 1999.

[4] J. Chen and T.N. Pappas, "Perceptual metrics and perceptual coders," in *Proc. Human Vision and Electronic Imaging*, San Jose, CA, Jan. 2001.

[5] D.M. Chandler and S.S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE Trans. Image Processing*, vol. 14, no. 4, pp. 397– 410, Apr. 2005.

[6] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 813–826, May. 2000.

[7] C. Tian and S. S. Hemami, "A new class of multiple description scalar quantizers and its application to image coding," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 329–332, Apr. 2005.

[8] A. B. Watson, G. Y. Tangand, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.

[9] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Information Theory*, vol. 39, no. 3, pp. 821–834, May. 1993.

[10] C. Tian and S. S. Hemami, "An embedded image coding system based on tarp filter with classification," in *Proceedings of 2004 International Conference on Acoustics, Speech and Signal processing*, Montreal, QC, Canada, May 2004.