ADAPTIVE RATE-DISTORTION-BASED THRESHOLDING: APPLICATION IN JPEG COMPRESSION OF MIXED IMAGES FOR PRINTING

Marcia Ramos[1] *Ricardo L. de Queiroz*[2]

[1] School of Electrical Eng., Cornell University, Ithaca, NY, 14850
[2]Xerox Corporation, 128-27E, Webster, NY, 14580
mramos@ee.cornell.edu, queiroz@wrc.xerox.com

ABSTRACT

In this paper, we propose a new technique for transform coding based on rate-distortion (RD) optimized *thresholding* (i.e. discarding) of wasteful coefficients. The novelty in this proposed algorithm is that the distortion measure is made adaptive. We apply the method to the compression of mixed documents (containing text, natural images, and graphics) using JPEG for printing. Although human visual system's response to compression artifacts varies depending on the region, JPEG applies the same coding algorithm throughout the mixed document. This paper takes advantage of perceptual classification to improve the performance of the standard JPEG implementation via adaptive thresholding, while being compatible with the baseline standard. A computationally efficient classification algorithm is presented, and the improved performance of the classified JPEG coder is verified. Tests demonstrate the method's efficiency compared to regular JPEG and to JPEG using non-adaptive thresholding. The non-stationary nature of distortion perception is true for most signal classes and the same concept can be used elsewhere.

1. INTRODUCTION

Efficient signal coding systems often rely on some degree of optimization based on a rate-distortion (RD) trade-off. For most signals, the perception of distortion is non-stationary and heavily depends on the signal contents. This accounts for masking properties and for the perceptual information necessary for critical decisions, e.g. the role of edges in the location and identification of objects. In fact, psychophysical studies have suggested that there are three perceptually significant regions in an image: *edge*, *smooth* and *detailed* (*textured*) regions [1]–[4]. The *edge* regions include text areas and other strong edges of higher perceptual significance. The *smooth* and *detailed* regions have less importance in this order. Edges play a major role in the human ability to recognize objects [4],[5], while compression artifacts in a smooth region are easily perceived. Detailed regions, however, can mask more errors, and therefore, can tolerate higher compression. The different sensitivity of the human visual system (HVS) to these regions can be used in our advantage in image compression schemes. Therefore, it is desirable to conceive distortion measures that are adapted to different regions of the signal.

With the growing popularity of digital technologies, documents are no longer limited to simple sheets of paper with text, but can be manipulated in the digital realm and contain a mixture of several data types, including text, still images, and graphics. The most popular image compression scheme is the JPEG baseline system [6], or JPEG for short. JPEG is based on the DCT and requires that a single quantization matrix be used for all blocks in an image component [6]. The number of bits generated per block is therefore *adaptive* on a block-by-block basis but the distortion masking properties are not. In a mixed document context, the compression algorithm should be able to trade lower compression in the text regions for higher compression in the smooth and detailed regions. In fact, oversmoothing unimportant areas is not as important as damaging edges of text.

Blockwise distortion adaptivity in JPEG has been conceived through the use of multiple quantizer tables [7]–[10]. In these techniques, quantizer tables are switched on a block-by-block basis but coders are often non-compliant with the JPEG baseline standard. Apart from adaptive quantization, the thresholding technique [11] can achieve blockwise adaptation within JPEG. In that, less relevant coefficients in an RD sense are simply discarded (thresholded) from a block, and the RD analysis can be made globally [11] or locally [12]. Also, in a non-RD-based framework, perceptual models can be used for discarding DCT coefficients which are visually less important [13]. In [14] an image segmentation algorithm is used to classify blocks into a few classes which are associated with predefined quantizer tables used to guide the discarding of DCT coefficients.

This paper presents a modified JPEG algorithm that takes advantage of perceptual classification and classified thresholding to improve the JPEG performance for mixed documents while being fully compliant with the JPEG baseline standard.

2. ADAPTIVE THRESHOLDING

The proposed algorithm is based on the *thresholding* technique which in this context sets to zero some AC DCT coefficients in JPEG based on RD characteristics [11].

The framework in [11] is general enough to be applied to any kind of data processing combined with compression. In general the signal is divided into units x_i, each unit at a particular instantiation will contribute to the bit-rate by R_i bits and distortion is some function of the quantization error $\hat{x}_i - x_i$, where $\hat{x}_i$ is the reconstructed unit. The global rate and distortion are given by

$$R = \sum_i R_i \quad D_i = f(\{\hat{x}_i - x_i, \forall i\}) \quad \left(\text{e.g.} = \sum_i (\hat{x}_i - x_i)^2\right) \tag{1}$$

By using a well behaved distortion function such as MSE any processing can be accounted in the RD balance by minimizing a cost funtion J which combines rate and distortion through a Lagrangian multiplier [11]:

$$J = R + \lambda D. \tag{2}$$

We do not challenge the optimization principle, rather we accommodate adaptation by defining s pace varying meaning for distortion as opposed to adapting the algorithm. For that, f is modified and distortion is

$$D = \sum_i f_i(\hat{x}_i - x_i) \quad \left(\text{e.g.} = \sum_i (\hat{x}_i - x_i)^2 w_i\right) \tag{3}$$

where w_i is a distortion weighting factor specific for the i-th unit. This algorithm clearly demands a priori classification of the signal in order to identify units and assign proper weights.

3. INCORPORATION INTO JPEG

In this paper, we incorporate perceptual classification into the thresholding decision in JPEG. In JPEG the units are image blocks and we use perceptually-weighted mean-squared-error (MSE) as a distortion measure. The HVS-based weights change from block to block according to the block classification. The approach we use is a variant of the one in [12] by incorporating adaptive HVS weights[1] . The goal of this perceptual weighting is to guarantee that most of the high frequency coefficients to be thresholded are from smooth and detailed blocks, while preserving the edge regions of the image. As in [12], for each non-zero AC coefficient in a given block, a cost-benefit ratio is computed where the cost is defined as the number of bits required to encode the coefficient and the benefit is defined as the decrease in distortion achieved when the coefficient is kept. The cost-benefit ratio is then compared to a threshold to decide whether or not the coefficient will be discarded.

In JPEG, quantized DCT coefficients of a block are mapped into a vector $zz(n)$ by scanning the block in a zigzag path. For a non-zero quantized coefficient $zz(n)$, assume the next non-zero quantized coefficient in the vector order is $zz(l)$ at index l. Without getting into the details of the JPEG variable length coding algorithm [6], it suffices to say that a non-zero sample $zz(i)$ produces two bit quantities: (a) b bits are used to encode how many zero samples are there before $zz(n)$ (and after the last non-zero coefficient in the path) and to provide information on the magnitude of $zz(i)$; (b) $SSSS$ bits are used to encode its sign and part of the information relative to the magnitude. Let $b = b_1$ and $SSSS = SSSS_1$ for $zz(n)$ and $b = b_2$ and $SSSS = SSSS_2$ for $zz(l)$. If $zz(n)$ is discarded, the run of zeroes before $zz(l)$ increases and it then spends $b_3 + SSSS_2$ bits to be encoded. The cost $R(n)$ of keeping the coefficient is then the cost of sending $zz(n)$ and $zz(l)$ minus the cost of sending $zz(l)$ if $zz(n) = 0$, that is:

$$R(n|zz(n) \neq 0) = b_1 + b_2 - b_3 + SSSS_1 \tag{4}$$

The benefit achieved with keeping the coefficient is then the decrease in distortion given by the information conveyed in $zz(n)$. Let the original non-quantized coefficient be $d(n)$, the quantizer step size be $q(n)$, and the perceptual weight associated with that coefficient be $w_K(n)$, where K is the class associated with the block. The distortion resulting from quantizing and keeping the coefficient is then

[1] In [12], thresholding was simplified for speed purposes by analyzing each non-zero quantized AC coefficient independently.

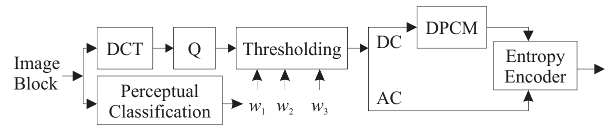


Figure 1. Block diagram of classified thresholding applied to a JPEG encoder.

$|d(n) - zz(n)q(n)|^2 w_K(n)$, while the distortion resulting from thresholding the coefficient is simply $|d(n)|^2 w_K(n)$. The decrease in distortion given by keeping $zz(n)$ is then:

$$D(n|zz(n) \neq 0) = w_K(n) zz(n) q(n)\ (2d(n) - zz(n)q(n)) \tag{5}$$

Finally, the cost-benefit, i.e., RD ratio is given by:

$$\nu(n) = \frac{b_1 + b_2 - b_3 + SSSS_1}{w_K(n) zz(n) q(n)\ (2d(n) - zz(n)q(n))} \tag{6}$$

$\nu(n)$ is compared to a threshold τ and $zz(n)$ is set to zero whenever $\nu(n) > \tau$. A block diagram of the encoder is shown in Fig. 1.

4. PERCEPTUAL CLASSIFICATION

The classification algorithm introduced here is a computationally efficient technique which classifies blocks through a simple analysis of the luminance difference values inside a block. Let $x(i,j)$ $(0 \leq (i,j) \leq 7)$ be the pixels in an 8×8 block and let $x'(i,j)$ be the pixels in a 4×4 block found by subsampling the 8×8 block through averaging of 2×2 neighbour pixels. The activity measures computed are the maximum differences among neighbour pixels in a block:

$$\mu_1 = \max\{|x(i,j) - x(i-1,j)|, |x(i,j) - x(i,j-1)|\} \quad \text{for } 1 \leq (i,j) \leq 7 \tag{7}$$

$$\mu_2 = \max\{|x'(i,j) - x'(i-1,j)|, |x'(i,j) - x'(i,j-1)|\} \quad \text{for } 1 \leq (i,j) \leq 3\}. \tag{8}$$

Two thresholds T_{lo} and T_{hi} $(T_{hi} > T_{lo})$ are required for classification:

Condition:	Class:
$\mu_1 > T_{hi}$	edge
$\mu_1 < T_{lo}$ and $\mu_2 < T_{lo}$	smooth
else	detailed

Note that computation can be greatly simplified if the algorithm is bypassed when a pixel difference is found larger than T_{hi} while computing μ_1. Also, if $\mu_1 \geq T_{lo}$, it is not necessary to compute $x'(i,j)$ or μ_2.

The proposed algorithm was tested against several other methods (commonly much more complex). Despite its simplicity it has been shown to be more efficient than its competitors in two experiments: (i) comparing classifier output map to manual classification; (ii) RD plots for different images[2].

5. COMPRESSION RESULTS FOR MIXED IMAGES IN PRINTING

Several experiments were performed on different mixed documents and the results shown here are just a sample for the *wine* image shown in Fig. 2.

We suggest using the following perceptual weights (which can be changed for different applications):

[2] Unweighted MSE for overall distortion, adaptive weighted MSE for thresholding decision

Table 1. HVS-based weights for the 8×8 DCT coefficients chosen for the 3 perceptual classes.

Smooth areas							
246	854	1000	935	791	631	486	364
854	952	997	915	771	616	475	356
1000	997	955	852	715	573	443	334
935	915	852	752	631	509	397	302
791	771	715	631	533	433	341	262
631	616	573	509	433	356	284	221
486	475	443	397	341	284	229	180
364	356	334	302	262	221	180	143
Detailed areas							
246	1000	791	486	267	139	69	33
1000	955	715	443	247	130	65	32
791	715	533	341	197	106	55	27
486	443	341	229	139	78	41	21
267	247	197	139	88	52	29	15
139	130	106	78	52	32	18	10
69	65	55	41	29	18	11	6
33	32	27	21	15	10	6	4
Edge areas							
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000
1000	1000	1000	1000	1000	1000	1000	1000

Those matrices were found by calculating DCT-domain energy of a linear HVS transfer function assuming maximum viewing frequency of 56 and 28 cycles/degree for smooth and detailed regions, respectively. For edges, we use uniform weighting.

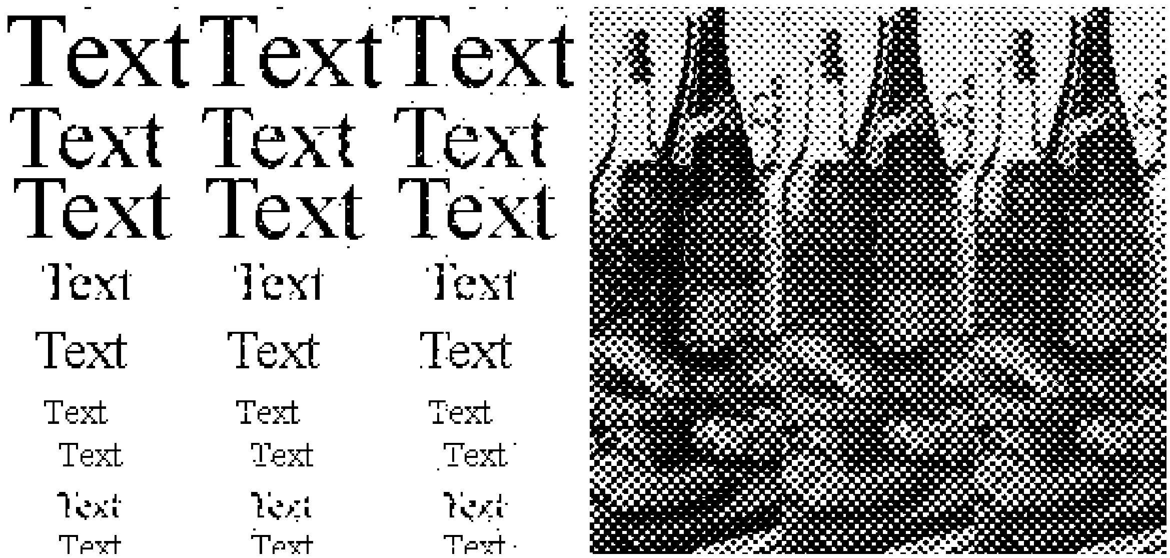
In one example, using the proposed approach we obtain the wine image at 1.11 bpp and 31.62 dB PSNR, while JPEG attains a PSNR of 35.69 dB for the same bit rate. A comparison of the classified thresholding coded image versus the baseline JPEG coded image reveals that the visual quality around the edges of the document (most noticeably the text areas) for the proposed JPEG coder is superior to the standard JPEG coder. However, regular JPEG has better quality on the picture region. When both images are halftoned for printing, the errors in the picture region are masked while ringing around text edges is shown through the halftone. In other words, after halftoning, the image compressed using the proposed coder is fairly superior in all aspects. This is so because the adaptive distortion measure (weights) intentionally took into account masking properties. So, overblurring smooth or detailed regions was not as penalized as blurring text areas. Enlarged portions of the halftoned images are presented in Fig. 3.

6. CONCLUSIONS

Adaptive thresholding has the ability of tailoring distortion measures to specific signal regions, therefore adding a new flexibility in compression systems without large computational penalty. We applied the concept to JPEG in the context of image compression for printing, in which perceptual differences accross regions are very significant. Nevertheless the concept may be applied to transform-based coding of virtually all clases of signals.

The processing described in this paper yields a JPEG-compliant coder with higher subjective performance than the regular JPEG (with or without non-adaptive thresholding). The price paid is a small computational complexity

Figure 2. (a) Enlarged Sample Mixed Image *wine*, (b) its halftone and (c) correspondent classification map (with $T_{lo} = 30$ and $T_{hi} = 120$). (white=smooth, gray=detailed, black=edge).



(a) (b)

Figure 3. Zoom of halftones of original and reconstructed images: (a) edge regions; (b) mixed regions. In each, the left image is the halftone of the orignal image, the center corresponds to the proposed algorithm (1.11 bpp, 31.62 dB PSNR). and the right one corresponds to JPEG (1.11 bpp, 35.69 dB).

penalty for the classification algorithm and for the thresholding algorithm. The perceptual classification was performed using a pixel-based classification algorithm which was shown to be visually accurate. A more detailed comparison of several classification algorithms will be published soon. The goal in embedding perceptual classification into JPEG coders is to allow the adaptive-thresholding JPEG compressor to preserve the visual quality of the most perceptually significant regions in a mixed document.

REFERENCES

[1] D. Marr, *Vision*, W.H. Freeman and Company, 1982.

[2] X. Ran and N. Farvardin, "A perceptually motivated three-component image model - part I: description of the model, " *IEEE Trans. Image Processing*, Vol. 4, no. 4, pp.401-415, April 1995.

[3] X. Ran and N. Farvardin, "A percepually motivated three-component image model - part II: application to image coding," *IEEE Trans. Image Processing*, Vol. 4, no. 4, pp.430-447, April 1995.

[4] M. G. Ramos, S. S. Hemami, and M. A. Tamburro, "Psychovisually-based multiresolution image segmentation," *Proc. IEEE Intl. Conf. on Image Proc.*, Santa Barbara, CA, Vol. 3, pp. 66–69, 1997.

[5] M. G. Ramos and S. S. Hemami, "Perceptually-based scalable image coding for packet networks," *Journal of Electronic Imaging*, Vol. 7, no. 3, pp. 453-463, July 1998.

[6] W.B. Pennebaker and J.L. Mitchell, *JPEG - Still Image Data Compression Standard*, Van Nostrand Reinhold, 1993.

[7] N. Chaddha, A. Agrawall, A. Gupta, and T.H.-Y. Meng, "Variable compression using JPEG," *Proc. Intl. Conf. on Multimedia Computing and Systems*, Boston, MA, pp. 562-569, May 1994.

[8] R. Rosenholtz and A. B. Watson, "Perceptual adaptive JPEG coding," *Proc. IEEE Intl. Conf. on Image Processing*, Lausanne, Switzerland, Vol. 1, pp. 901-904, Sept. 1996

[9] M. G. Ramos and S. S. Hemami, "Edge-adaptive JPEG image compression," *Proc. SPIE Visual Communications and Image Processing*, Orlando, FL, Vol. 2727, pp. 1082-1093, March 1996.

[10] ISO/IEC CD 10918-3, Info. Technology - Digital compression and coding of continuous tone still images - Part 3: Extensions, 13 Nov. 1994.

[11] K. Ramachandran and M. Vetterli, "Rate-distortion fast thresholding with complete JPEG-MPEG decoder compatibility," *IEEE Trans. on Image Processing*, vol. 3, no. 7, pp. 700-704, Sept. 1994.

[12] R.L. de Queiroz, "Processing JPEG-compressed images and documents," *Proc. IEEE Intl. Conf. on Image Processing*, Santa Barbara, CA, Vol. 2, pp. 334-337, 1998.

[13] R. Safranek, "A comparison of the coding efficiency of perceptual models," *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV*, San Jose, CA, Vol. 2411, pp. 83-91, February 1995.

[14] A. Nafarieh, "Segmentation and adaptive quantization scheme within the baseline JPEG sequential mode," Xerox Corp. internal report.