Transform Image Coding with Global Thresholding: Application to Baseline JPEG

AZZA OULED ZAID, CHRISTIAN OLIVIER, OLIVIER ALATA, FRANCOIS MARMOITON

IRCOM-SIC-UMR 6615, SP2MI, Teleport 2, Boulevard M. et P.Curie, F-86962 Futuroscope Chasseneuil Cedex
{ouled, olivier, alata, marmoiton}@sic.sp2mi.univ-poitiers.fr

Abstract. Many image compression schemes perform the discrete cosine transform (DCT) to represent an image in frequency space. An analysis of a suite of images confirms that the luminance AC coefficients can be modeled by a Laplacian distribution. The distribution model can be used to drop the insignificant coefficients. In this paper we develop an image-adaptive JPEG encoding algorithm that incorporates global thresholding and near optimal quantization approach based on Lagrangian multiplier. Simulation results demonstrate that, with our thresholding technique, we can improve the reconstructed image quality compared to the one provided by other DCT image coding schemes without thresholding.

1 Introduction

The Discrete Cosine Transform (DCT) is used in many popular still image compression schemes. In particular JPEG baseline coding algorithm [1] uses the DCT to concentrate image information: JPEG partitions the image into 8x8 blocks, transforms each block using an 8x8 DCT, and scalar quantizes the resulting DCT coefficients using an 8x8 matrix of quantizer step sizes. The quantized coefficients are then entropy coded using zero run-length Huffman coding. The key to good compression (in the rate distortion sense) when using DCT lies in the quantizer selection. JPEG's quantizer step sizes largely determine the rate-distortion tradeoff in the compressed image. Many algorithms [2], [3], [4] have been proposed to optimize this tradeoff. However, even with image-adaptive quantizer selection, JPEG must apply the same quantizer to every image block. Thus, JPEG quantization lacks local adaptivity. To compensate this quantization inefficiency, intelligent adaptive thresholding of the coefficients can be incorporated in the DCT coding chain as shown in Figure 1. Thresholding strategy refers to adaptively dropping some nonzero-quantized coefficients to zero, in order to improve the rate-distortion tradeoff. This could refine quantizer scales for coefficients that are retained, without adding any complexity to the decoder. Several approaches have been tried in the literature in order to make *zeroing* decisions.

We mention the one found in [5], using a search strategy within a rate-distortion (R-D) framework, for the fixed scale quantization performed prior to thresholding.

Thus, starting from the "highest quality point" after quantization at a fixed scale (for JPEG), one can sweep the entire R-D curve over a continuous range of target bit

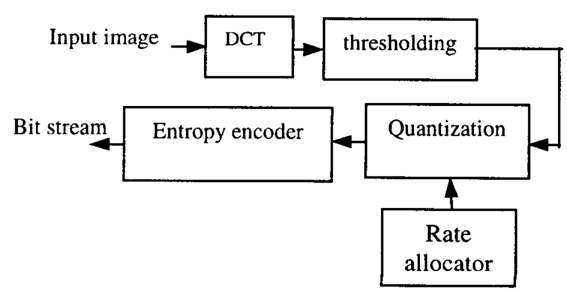


Figure 1: Schematic of the adapted image coding chain.

rates (or equivalently target coding qualities) by a pruning operation, in order to threshold insignificant coefficients in the image. Such strategy lies in its combination of R-D optimality and its complete compatibility with standard decoder. However, it is computationally expensive. Furthermore, the algorithm proposed in [5] needs to be rerun, every time a new rate of compression is desired.

Later, Ratnakar proposed to produce the global threshold tables [6] that decide the zeroing cutoffs levels for each DCT coefficient based on the rate/distortion estimations.

The basic goal of our research is to find a fast and simple thresholding strategy that speeds up quantizer scales for the retained coefficients and reduce the dequantization error. For simplicity, we only consider grayscale images, but the same method can be applied to color images as well.

In this paper, we propose two global thresholding techniques. The first one uses the information criteria

(IC) that consists in optimally classify the AC coefficient distribution. Thereby, we select the *dead-zone* that represents the less significant coefficients to be dropped to zero. The second consists in the selection of the thresholding table components based on theoretical coefficient's model distribution.

To determine a near optimal quantization table over a wide range of compression rates, we adopt RD-OPT (R-D optimality) algorithm because of its simplicity and performed results.

The organization of the remainder of this paper is as follows: Section 2 recalls definition about local and global thresholding and how they can be applied on DCT image coding. Section 3 describes the modeling of AC coefficients distribution. Section 4 gives the main notions to understand the origin and the interest of the IC in the modeling and parameter's estimation optimization. Then, explains our use of these criteria in the selection of the AC histograms bin numbers and the limit of zeroed class, and section 5 provides our theoretical thresholding technique based on the coefficient distribution modeling. Section 6 describes the RD-OPT algorithm [6], adopted in the quantization step of our coding chain. Finally, section 7 discusses experimental results and draws some conclusions about the thresholding techniques we have developed.

2 Thresholding scheme applied on DCT coefficients

We represent an NxN image I by a set of $\frac{N^2}{64}$ blocks each with 64 pixels, such that for $0 \le x, y \le 7$, I^b_{xy}, denotes the pixel at spatial location (x, y) in block b (see formula (1)), where the blocks are ordered in a raster scan order. By applying a 8x8 DCT transform to each block, we perform the block DCT:

$$c^b = DCT_{8X8}\left[I^b\right], b = 1,2,\cdots,\frac{N^2}{64} \qquad (1)$$

where c^b represents a vector of 64 DCT coefficients from block b.

To simplify the notation, we order the 8x8 set of spatial frequencies into a 1-D array of 64 coefficients indexed by n with using the zig-zag scan. Then, the block components through the transformed image are denoted by c^b_n, where b is the block indice and $n \in \{0,1,\cdots,63\}$ the coefficient indice location.

Let us recall that JPEG allows the encoder to optimize both the Huffman table and quantizer step sizes. However, these latter are not the only available degrees of freedom. An extra degree of freedom involves the coefficients themselves, since certain DCT coefficients may be dropped before the generation of the quantization table in an R-D sense. It has been shown that for JPEG with fixed quantizer step sizes Q, thresholding approximates the performance of finding the best quantization level for each coefficient [7].

We mention two thresholding types in the literature: The local and global thresholding.

The local thresholding process consists in making the zeroing decisions separately for each image block. It uses a set of binary thresholding values T^b_n that signal whether or not the transform coefficient c^b_n must be thresholded, i.e.,

$$T^b_n = \begin{cases} 1 \Rightarrow \hat{c}^b_n = Round\left(\frac{c^b_n}{Q_n}\right) \\ 0 \Rightarrow \hat{c}^b_n = 0. \end{cases} \qquad (2)$$

with $\hat{c}^b_n$ the thresholded-quantized coefficient to be transmitted through the entropy encoder.

The global thresholding is specified by a table of 64 non-negative real components called *the threshold table* T[64]. We define global thresholding as follows: if Q is the quantization table to be applied on each DCT block, then the combined result of quantization and thresholding is represented as follows,

$$\hat{c}^b_n = \begin{cases} Round\left(\frac{c^b_n}{Q_n}\right) \text{if } \left|c^b_n\right| \ge T_n \\ 0 \text{ otherwise}, \end{cases} \qquad (3)$$

where Round(x) is the closest integer to x. The table T does not need to be included with the compressed image, as the decoder does not need to know the thresholds. The decoder simply multiplies each thresholded-quantized coefficient by the quantization table entry, to calculate its approximation of the original DCT coefficient.

Through preliminary results, it has been confirmed that the benefit of using local thresholding for JPEG application is unprofitable in the complexity-performance tradeoff sense, since the zeroing decision is made through all the transform image coefficients. For this reason, we adopt the global thresholding in our work.

3. Modeling of DCT coefficient distribution

For DCT image coding schemes, there have been different assumptions concerning the distribution of the transform coefficients. Reininger [8] and Bellifemine [9] use the Kolmogorov-Smirnov goodness-of-fit tests in order to conclude that the AC transform coefficients can be better modeled as Laplacian than as Gaussian, Rayleigh, or Gamma distribution.

Other approaches propose a coefficient model based on the Cauchy law [10], or a mixture of Gaussian laws

[11]. Under these assumptions, Eude [11] concludes that a mixture of 1, 2 or 3 Gaussian laws is the best model of AC coefficients distribution.

Nevertheless, we retain a Laplacian model because of its reliability and simplicity. In fact, the distribution model can be readily matched by finding a single parameter (the Laplacian parameter α). In Figure 2, the plots depict the actual data against the model of coefficients in the space location (n = 30) of *Boat* image (see Figure 5.a).

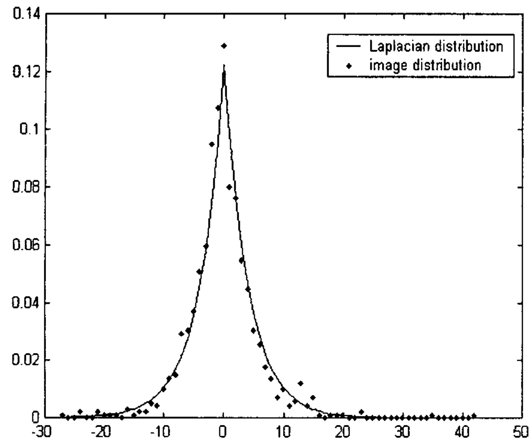


Figure 2: The model Laplacian distribution with α= 0.1244 (curve), and the actual distribution of the DCT coefficient (drash).

A sample Laplacian distribution is characterized by its probability density function (pdf) defined as follows:

$$p(x)=\frac{\alpha}{2}e^{-\alpha|x|} \tag{4}$$

Fitting the distribution to a particular coefficient requires finding Laplacian parameter α. The simplest way is via the standard deviation, σ.
Thus,

$$\alpha = \frac{\sqrt{2}}{\sigma}. \tag{5}$$

This calculation must be performed through the 63 AC coefficient frequency location because they have different distributions. The parameters for DCT coefficients of *Boat* image are shown in Table 1.

Next, we explore the DCT coefficient model to solve the thresholding problem.

DC	0.0065	0.0105	0.0146	0.0261	0.0401	0.0667	0.0787
0.0067	0.0129	0.0167	0.0252	0.0393	0.057	0.0866	0.1041
0.0143	0.0212	0.0289	0.0348	0.0477	0.0736	0.1006	0.1185
0.0249	0.0339	0.0437	0.0574	0.0767	0.0881	0.1244	0.143
0.0445	0.0557	0.0604	0.088	0.1141	0.127	0.1384	0.1396
0.0687	0.0912	0.104	0.1268	0.1655	0.1851	0.1853	0.1648
0.1242	0.1734	0.1759	0.1911	0.2162	0.2302	0.2079	0.1821
0.2062	0.2498	0.2702	0.2514	0.2742	0.2587	0.2215	0.1913

Table 1: Laplacian distribution parameters "α" for AC coefficients of the *Boat* image.

4 Information criteria and DCT coefficient thresholding

4.1 IC definition

Information criteria (IC) are some tools for the parameterized model estimation, as well as the model selection (clustering). Initially proposed to obtain the AR model order by Akaike [12], their usual form is:

$$IC(k)=-2\sum_{i=1}^{N}\log f\left(X_i\middle|\hat{\theta}_k\right)+C(N)\times k \tag{6}$$

where $X^N = \{X_1,\ldots,X_N\}$ is the sequence of observations that are supposed to be independent and:

- $\hat{\theta}_k$ is the maximum likelihood estimator of the theoretical model θ_k;of order k.
- $f(X_i|\hat{\theta}_k)$ is the likelihood term.
- $C(N)$ is the penalty term which depends on the observation number N. It differs according to the chosen criterion.

We regard a model with a smaller IC as the better one:

$$\hat{k}=\arg\min_k IC(k) \tag{7}$$

IC-type criteria take into account simultaneously both the goodness-of-fit of the model and the number of parameters used to achieve that fit. The different criteria differ by the penalty term, $C(N)$.

Initially, the AIC (Akaike Information Criterion) ($C(N) = 2$) results from the minimization of a cost, the Kullback-Leibler information between the original model and an approximated one. Theoretical research on this criterion gave matter to many works to palliate its inconsistency. Schwarz proposed a new criterion for an exponential family founded on a Bayesian justification. He suggested the BIC (Bayesian Information Criterion) [13]:

$$BIC(k)=-2\sum_{i=1}^{N}\log f\left(X_i\middle|\hat{\theta}_k\right)+k\log N \tag{8}$$

In a different way, Rissanen came up with an equivalent criterion MDL (Minimum Description Length)

[14] using a coding technique (minimizing the codelength in relation to the observations) for a parameterized density. This criterion is asymptotically convergent, i.e., it helps in finding the appropriate model when $N \rightarrow \infty$ (strong consistency). Note that the latter criterion penalizes more stringently the log-likelihood as the number of observations increases in comparison with AIC.

A third criterion φ was introduced by Hannan and Quinn [15] in the case of an autoregressive process. This criterion leads to convergence in probability of the order estimator (weak consistency) and stands as a compromise between AIC and BIC. Its penalty is fixed as follows:

$$\varphi(k) = -2\sum_{i=1}^{N} \log f\left(X_i / \hat{\theta}_k\right) + k \log(\log N) \tag{9}$$

Finally, we state El Matouat and Hallin's generalization [16] drawn on Rissanen's works on stochastic complexity [17] succeeded to φ_β criterion written as follows:

$$\varphi_\beta(k) = -2\sum_{i=1}^{N} \log f\left(X_i / \hat{\theta}_k\right) + kN^\beta \log(\log N) \tag{10}$$

where β is selected such as:

$$\frac{\log(\log N)}{\log N} \le \beta \le 1 - \frac{\log(\log N)}{\log N} \tag{11}$$

φ_β criterion is characterized by its strong consistency. The φ criterion can be seen as a borderline case of the φ_β criterion (i.e. $\beta = 0$).

The $C(N)$ penalties, indicated above correspond to the most traditional case (the parameterized model). In the very particular case of the histograms, those are slightly different (see section 4.2 for the retained φ_β criterion).

4.2 IC for optimal classification

We use the φ_β IC criterion to optimally classify AC DCT coefficients. We note that the objective aimed of is the selection of the less significant coefficient class.

The use of IC was extended to determine the number of equal or variable width bins (classes) of histogram, we refer to [18] for more theoretical details. We use this principle to optimally model the histogram built with N values of the AC coefficients (N is the number of blocks in the image sequence), in order to determine the limit (threshold value) of the class with low amplitude coefficients to be droped. Initially, the coefficients are merged in equal m classes. In [19], Sakamoto fixed the m value to:

$$m = O(2 \times \lfloor N - 1 \rfloor) \tag{12}$$

Let E be the random variable corresponding to the AC coefficient histogram and Ω the set of all the real values taken by E.

To find the best sub-partition (the optimal number c of classes), the sub-partition should optimally approximate the real probability law η of E to a priori law called μ (a Laplacian law in the case of AC coefficient distribution), in a cost sense. In the case of histogram, based on the works described in [20], the φ_β criterion will have the following expression:

$$\varphi_\beta(c) = c\left(1 + N^\beta \log(\log N)\right) - 2N\sum_{r=1}^{c} \theta_c(B_r) \log\left(\frac{\theta_c(B_r)}{\mu(B_r)}\right) \tag{13}$$

where θ_c is the maximum likelihood estimator of η. It is represented by the histogram with c classes $\{B_r\}$, $r=1,\ldots c$. On any sub-partition C of Ω, $\{\theta_c(B_r)\}_{r=1,\ldots,c}$ is equal to:

$$\theta_c(B_r) = \frac{\textit{Number of coefficients in Br}}{N} \tag{14}$$

We observe that the penalty term differs from the formula (10), since the objective is to find the optimal number of histogram classes and their borders not the number of model parameters and their corresponding values.

In the following section, we describe the merging algorithm used in our first thresholding technique.

Merging algorithm

Starting with m classes, a first histogram is built giving a partition M. The merge of classes is done according to an iterative process characterized by the IC variation. At a given iteration, let us suppose that we have a histogram with k classes B_r, $r=1,\ldots,j,\ldots,k$. If two adjacent classes B_j and B_{j+1} merge so that $B_r = B_j \cup B_{j+1}$, we obtain a new histogram with (k-1) classes. Among all possible merges with (k-1) classes we keep the one that minimizes $\varphi_\beta(k-1)$. We thus obtain the number of classes c and the associated optimal partition $\{(\theta_c(B_r)\}_{r=1\ldots c}$.

5 Theoretical thresholding

Referring to Eude's works [11] on distribution statistics modeling applied on baseline JPEG, we propose a second technique for global thresholding table generation. In this context, we take into account the probability density function of each AC coefficient through all the DCT blocks. Since the information worth differs from one coefficient to an other, the zeroing of coefficients in high frequencies space does not involve deterioration as much as the zeroing of coefficients in low frequencies. Thus, given the probability density function of the transform coefficients, we can define the limits over which they have low probability to appear. Thus, they should be retained to avoid considerable quality deterioration in the reconstructed image. This idea can be formulated as follows :

An AC coefficient c_n has a low probability if its absolute value is superior to a limit value S_n defined by :

$$\int_{-\infty}^{S_n} \mu_n(c_n)dc_n = 0.95 \qquad (15)$$

where μ_n is the probability density function of the distribution model which fits better to the real coefficient distribution (Laplacian law). Thus, through the 63 Laplacian models, we determine the thresholding table components T_n, $n \in \{0,...,63\}$, such as :

$$T_n = \begin{cases} \frac{F_e}{S_n} \text{if } n \neq 0 \\ 0 \text{ if } n = 0 \end{cases} \qquad (16)$$

with F_e a scaling factor (here F_e=50).

After computing the adequate S_n values for each AC coefficient, we perceive a low magnitude of S_n in high frequencies (see table 2) involving high threshold magnitudes (see Figure 3.a). So, only the high coefficients amplitude values are retained in high frequencies. However, in low frequencies the S_n magnitudes are high. Consequently, the corresponding threshold values are small (see Table 3), and most coefficients are conserved (see Figure 3.b).

DC	354.2	219.2	157.7	88.2	57.4	34.5	29.2
343.6	178.4	137.8	91.3	5.8	40.3	26.5	22.1
161	108.6	79.6	66.1	48.2	31.2	22.8	19.4
92.4	67.9	52.6	40.1	30	26.1	18.5	16.1
51.7	41.3	38.1	26.1	20.1	18.1	16.6	16.4
33.5	25.2	22.1	18.1	13.9	12.43	12.42	13.9
18.5	13.2	12.04	10.65	10	10	11	12.6
11.1	9.21	8.52	9.15	8.39	8.9	10.3	12

Table 2: the obtained S_n values for *Boat* image.

DC	0.140	0.227	0.317	0.567	0.870	1.449	1.709
0.144	0.279	0.362	0.546	0.854	1.238	1.879	2.260
0.309	0.460	0.626	0.754	1.035	1.598	2.183	2.573
0.541	0.735	0.949	1.247	1.665	1.912	2.702	3.104
0.966	1.208	1.312	1.910	2.477	2.758	3.005	3.031
1.492	1.981	2.259	2.753	3.594	4.019	4.022	3.578
2.696	3.766	3.819	4.150	4.694	4.998	4.514	3.955
4.477	5.424	5.867	5.459	5.953	5.618	4.810	4.153

Table 3: the obtained threshold values for *Boat* image.

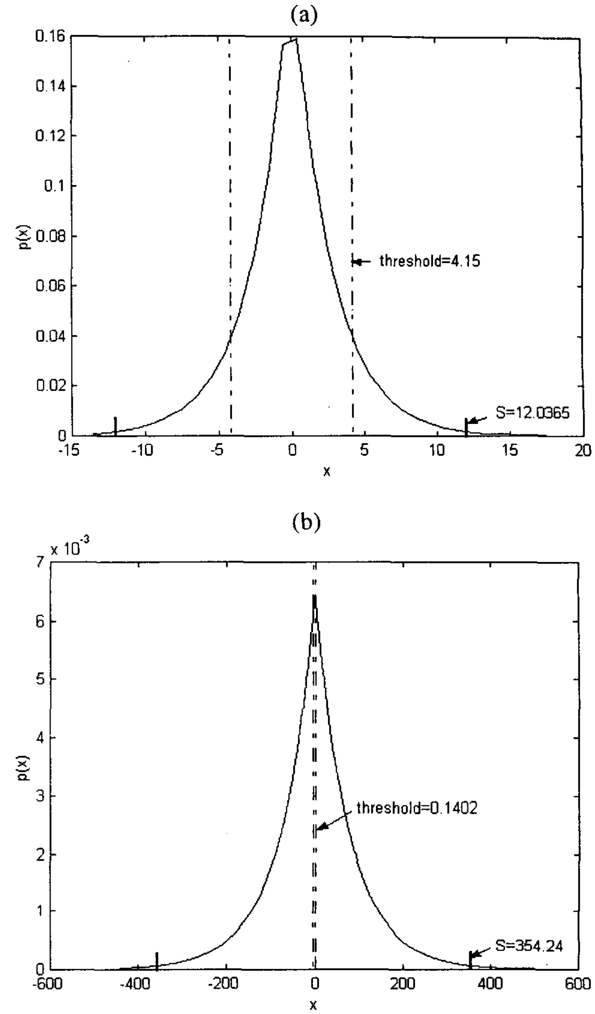


Figure 3: thresholding process according to coefficient location in the frequency space: (a) threshold and S_n values of DCT coefficient in the space location (n = 62); (b) threshold and S_n values of DCT coefficient in the space location (n = 1);

6 DCT quantization approach

Within the framework of our research, the coding chain is separated in two parts:

- Global thresholding table generation (by IC application and theoretical thresholding).
- Quantization tables generation by the RD-OPT algorithm.

In the following, we give an overview of the RD-OPT quantization algorithm.

As described in [21] the RD-OPT algorithm is an efficient algorithm for sub-optimal quantization table design. It produces quantization tables optimizing the trade-off between quality and compressed size. For each coefficient c_n and for each possible quantizer scale q, the contribution to total rate of the n^{th} coefficient c_n is calculated as $R[n][q]$, and the contribution to total distortion as $D[n][q]$. To estimate $R[n][q]$ and $D[n][q]$ for each possible q, a preliminary pass through the image is run to gather DCT statistics which are used to predict rate and distortion accurately.

Then, for any quantization table Q, the rate of compression is

$$R(Q)=\sum_{n=0}^{63} R[n][Q[n]], \tag{17}$$

and the distortion is

$$D(Q)=\sum_{n=0}^{63} D[n][Q[n]] \tag{18}$$

After calculating the contributions of individual coefficients to the total rate and total distortion, the RD-OPT runs the bit allocation algorithm by merging rate and distortion through the Lagrangian multiplier λ, in order to determine the quantization table components, which result in the minimum total Lagrangian cost.

For a more mathematical formulation, the RD-OPT principle can be stated as follows:

Given a target rate R_{budget}, find

$$D_{\min} = \min_{Q} D(Q) \text{subject to } R(Q) \leq R_{\text{budget}} \tag{19}$$

by introducing $J(\lambda)=[D(Q)+\lambda R(Q)]$ representing the Lagrangian cost of Q associated with the quality factor λ and solving the following equivalent unconstrained problem

$$J_{\min}(\lambda)=\min_{Q} J(\lambda)=\min_{Q}[D(Q)+\lambda R(Q)] \tag{20}$$

The desired optimal constant slope value λ^* is not known a priori and depends on the particular target budget or quality constraint. But, it is obtained using a fast convex search using the bisection algorithm (for more details we refer to the works presented in [22]).

$$J_{\min}(\lambda^*)=\min_{Q}[J_{\min}(\lambda)-\lambda R_{budget}] \tag{21}$$

Note that the algorithm described above is near optimal, since the minimum-Lagrangian cost operating point are not found independently for each block of the image sequence but only by considering a single block.

It is possible to improve the performance (i.e , the rate and distortion tradeoff), by selectively setting some coefficients to zero. This allows finer quantizer scales for the retained coefficients, and does not add any complexity to the decoder. In the following section, we present the results of our tests by using two global thresholding techniques (with IC approach and the coefficient distribution statistics), coupled with RD-OPT quantization algorithm.

7 Experimental results

The performance of our global thresholding techniques applied to JPEG baseline using compression results are given for two images. These are the well-known grayscale images, called Lena and *Boat* (see Figure 4. a and Figure 5. a). In Figures 4. b and 5. b are plotted the PSNR (Peak Signal to Noise Ratio) versus rate for these images compressed using our thresholding strategies coupled with RD-OPT quantization. Moreover, the PSNR-rate plots for these images compressed using the "default JPEG" quantization tables are depicted. In addition, PSNR-rate plots are also shown in the case of RD-OPT without thresholding, i.e., just quantization table optimization. The rates shown in these plots are the actual rates resulting from JPEG compression with Huffman coding, and not entropy estimates.

For each image, theoretical thresholding with quantization table in the rate-distortion sense results in PSNR gains of up to 2 dB (decibel), compared to RD-OPT algorithm without thresholding and 4 dB compared to JPEG with "default" quantization table. In the case of thresholding by IC, we find that our algorithm achieves nearly the same PSNR as unthresholded RD-OPT coding scheme in low bit rates. However, there is a point for both test images beyond which the performance starts to degrade. Intuitively, the use of the thresholding table generation by IC is inadvisable beyond a point.

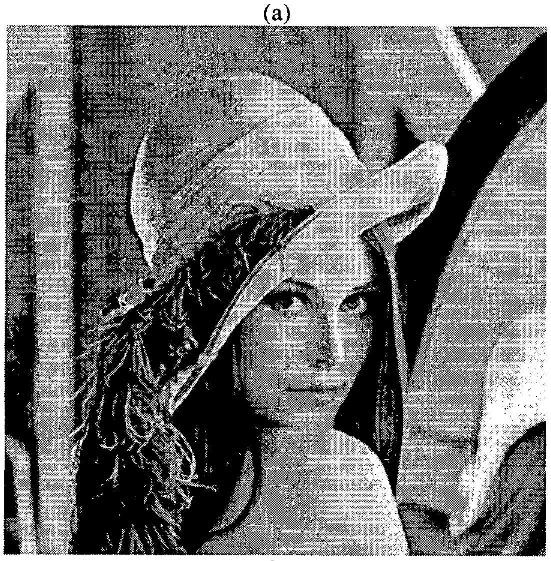
CONCLUSION

We have proposed an image coding technique based on the DCT coefficient statistic distribution. Our algorithm uses simply a global thresholding pass to approximate the zero set model and then completely quantizes the remaining coefficients in a second pass. Two thresholding strategies have been tested.

The first one is based on the IC application to optimally optimize the DCT coefficient histogram classes in maximum likelihood sense, in order to select the limit of zero coefficient class, this limit corresponds to the threshold value. The second strategy consists in the threshold table generation by using the theoretical model distribution (Laplacian). We note that unlike local thresholding, zeroing decisions are not made separately for each image block which simplifies the use of thresholding.

The coding results show that by incorporating the theoretical global thresholding in the coding chain, we can improve the performance (R-D tradeoff) of DCT-based image compression. However, the IC application for the thresholding table generation does not bring any improvements for the coding results in spite of its efficiency in image segmentation and optimal entropy coding.

(a)



(b)

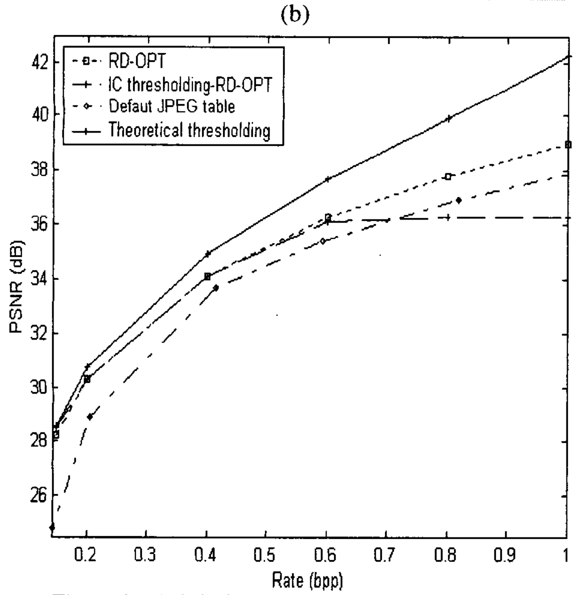
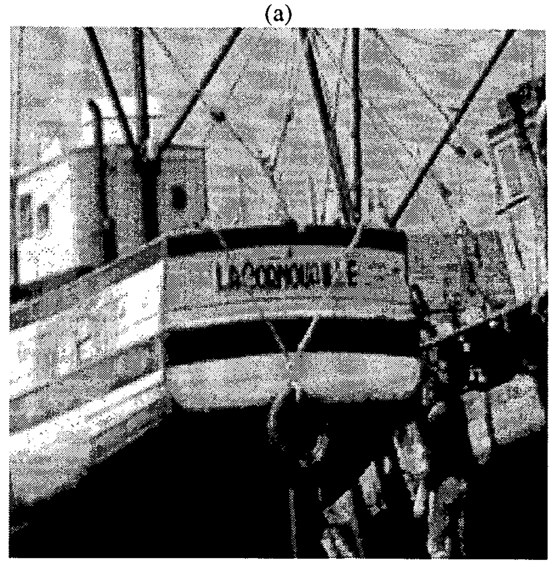


Figure 4: (a) 512x512 grayscale image of *Lena*; (b) Comparison in performance of global thresholding with φ_β versus global theoretical thresholding, unthresholded RD-OPT algorithm and classic JPEG for *Lena* image.

(a)



(b)

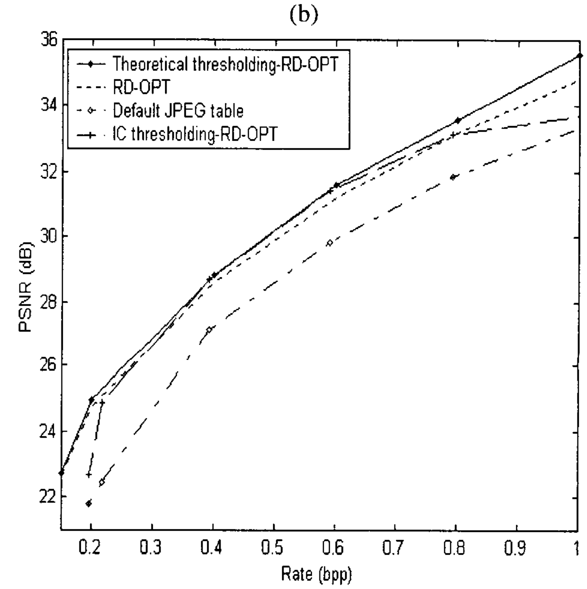


Figure 5: (a) 256x256 grayscale *Boat* image; (b) Comparison in performance of global thresholding with φ_β versus global theoretical thresholding, unthresholded RD-OPT algorithm and classic JPEG for *Boat* image.

REFERENCES

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

[2] S. Wu. and A. Gersho, Rate-constrained picture-adaptive quantization for JPEG baseline coders, In *Proc. Inter. Conf. Acoustics, Speech and Signal Processing*, vol. 5, pp. 389-392, April. 1993.

[3] D. M. Monoro and B. G. Sherlock, Optimum DCT Quantization, In *Proceding of Data Compression Conference,* pp. 188-194, 1993.
[4] V. Ratnakar. and M. Livny, RD-OPT: An efficient Algorithm For Optimizing DCT Quantization tables, *In Proceedings of Data Compression Conference*, pp. 332-341, 1995.
[5] M. Course. and K. Ramchandran, JPEG optimization using an entropy-constrained quantization framework, In *Proceedings of Data Compression Conference*, pp. 332-351, 1995.
[6] V. Ratnakar. and M. Livny, Extending RD-OPT with Global Thresholding for JPEG Optimization, *In Proceedings of Data Compression Conference*, pp. 379-386, 1996.
[7] M. Course and K. Ramchandran, Joint Thresholding and Quantizer Selection for Transform Image Coding: Entropy-Constrained Analysis and Applications to Baseline JPEG, IEEE Transaction on Image Processing, vol. 6, n°2, pp. 285-297, February 1997.
[8] R. C. Reininger and J. D. Gibson, Distributions of the Two-Dimentional DCT Coefficients for Images, *IEEE Trans. Communications*, vol. 31, n°6, pp. 835-839, June 1983.
[9] F. Bellifemine, A. Cappellino, A. Chimienti, R. Picco and R. Ponti, Statistical analysis of the 2D-DCT coefficients of the differential signal for images, Signal Processing: Image Communication, vol. 4, n°6, pp. 477-488, 1992.
[10] A. Baskurt, *Compression d'images numeriques par la transformation cosinus discrete,* Ph.D. dissertation, Univ. Lyon, France, June 1989.
[11] T. Eude, R. Grisel, H. Cherifi and R. Debrie, On the distribution of the DCT coefficients, ICASSP'94, vol. 5, pp. 365-368, South Australia, April, 1994.
[12] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, Second International symposium on Information Theory, pp. 267-281, Budapest, Akademiai Kaido, 1973.
[13] G. Schwarz, Estimating the Dimension of a model, The Annals of Statistics, vol. 6. pp. 461-464,1978.
[14] J. Rissanen, Modeling by Data Description, Automatica, vol. 14, pp. 465-471, 1978.
[15] E. J. Hannan and B. G. Quinn, The Determination of the Order of an Autoregression, *Journal of the Roy. Stat. B*, vol. 41, pp. 190-195, 1979.
[16] A.El Matouat and M. Hallin, Order Selection, Stochastic Complexity and Kulback-leibler Information, Time series Analysis, vol. 2, pp. 291-299. Springer Verlag. NY(USA), 1996.
[17] J. Rissanen, Stochastic Complexity and Modeling, The Annals of Statistics, vol. 14, pp. 1080-1100, 1986.
[18] C. Olivier, P. Courtellemont, O. Colot, D. de Brucq and A. El Matouat, Comparison of histograms: a tool of detection, *European Journal of Diagnosis and safety in Automation,* vol. 4, n°3, pp. 335-355, 1994.
[19] Y. Sakamoto, M. Ishiguro and G. Ktagawa, Akaike Information Criterion, *Statistics, Mathematics and its Applications*, KTK scientifics publishers, Japan, 1986.
[20] A. El Matouat and C. Olivier, Sélection du nombre de classes d'un histogramme et contraste de kullback, in *Proc. XXV*e *journées de statistique*, pp. 193-196, Vannes, France, May, 1993.
[21] V. Ratnakar and M. Livny, An Efficient Algorithm For Optimizing DCT Quantization, *IEEE Trans. Image Processing*, vol. 9, n°2, pp. 267-270, Februry, 2000.
[22] K. Ramchandran and M. Vetterli, Best Wavelet Packet Bases in a Rate-Distortion Sense, *IEEE Trans. Image Processing*, vol. 2, n°2, pp. 160-175, April, 1993.