Psychovisual and Statistical Optimization of Quantization Tables for DCT Compression Engines

S. Battiato, M. Mancuso, A. Bosco, M. Guarnera
STMicroelectronics – AST Catania Lab
Digital Still Camera & Multimedia Mobile Group

Abstract

The paper presents a new and statistical robust algorithm able to improve the performance of the standard DCT compression algorithm for both perceived quality and compression size. The approach proposed combines together an information theoretical/statistical approach with HVS (Human Visual System) response functions. The methodology applied permits to obtain suitable quantization table for specific classes of images and specific viewing conditions. The paper presents a case study where the right parameters are learned, after an extensive experimental phase, for three specific classes: Document, Landscape and Portrait. The results show both perceptive and measured (in term of PSNR) improvement. A further application shows how it is possible obtain significative improvement profiling the relative DCT error inside the pipeline of images acquired by typical digital sensors.

1. Introduction

The optimization of the quantization tables of DCT compression engine is connected to the possibility to improve the performance of the algorithm, gathering a better trade off between the final quality of the compressed image and the relative compression ratio. The standard JPEG [14] only suggests to use quantization tables that use perceptual criteria to quantize the corresponding DCT coefficients amplitudes that cause perceptible differences to a human observer. Even if only for useful suggestion the standard tables were obtained by Lohscheller [9] that conducted empirical studies measuring the sensitivity of the human visual system. This study shows that the observer is more sensitive to changes in the low frequency DCT components than in the high frequency DCT components. Later Pennebaker and Mitchell [14] found, however, that the tables yield noticeable artifacts when viewed on high-quality displays.

It is well known that optimum quantization tables are highly image dependent. For example, coarse quantization of the medium-frequency coefficients causes turbulence or speckling around text characters on a plain background. A complicated background may mask these artifacts, allowing coarse quantization to produce acceptable results. Indeed, it could be useful to flip between quantization tables, using one to compress textured areas and another to compress smoother areas or more in general to be able to generate quantization tables profiling the real content of the image both in term of theoretical and perceptive information. Unfortunately, the existing JPEG standard baseline does not allow changing the quantization table in the middle of compressing a component of the image. For low bit-rate compression side effects like *ringing* and the *blocking* are also present.

Several attempts have been conducted in order to generate *optimum* quantization tables, using different and various approaches.

The best results are however been obtained using iterative/adaptive procedure based on local, statistical or content-based, perceptive or not, analysis realized block-by-block [1], [6], [16], [20], [22], [23]. Other methods based on rate-distortion theory try to avoid and/or partially improve the well-known behavior of the scalar quantization with dead zone [11], [18]. In every case these methods aren't always compliant with the standard baseline JPEG, where the same table must be used for all blocks in the image; the complexity of the encoder, especially for real-time application, could be increase and become significantly high. Our approach, remaining fully compliant with the baseline standard, tends to bring both statistical and psycho visual techniques for specific classes of digital images. Experiments have shown as the specific statistical behavior of the DCT coefficients can be efficiently modeled, in a detailed way, considering specific classes of images. After an extensive experimental phase it has been possible derive an explicit model obtained combining together statistical analysis of DCT coefficients and heuristics related with Human Visual System.

The real implementation of DCT compression engine in consumer devices, requires specific well-driven optimization criterions. The study of typical pipelines, from real scene to final compressed image (e.g. in a digital CCD/CMOS camera) allowed us to derive useful information for further optimization. In particular we have obtained significative improvement evaluating numerically the absolute error in the

0-7695-1183-X/01 $10.00 © 2001 IEEE

DCT domain, according to the specific color interpolation algorithm inside the camera. Preliminary results showed us how invert, in some sense, such degradation process.

The next Section shows the details of the proposed model while the Section 3 will present some results obtained using such methodology. In Section 4 we briefly introduce, a new strategy of work useful to adapt the quantization table to the relative error due to the digital camera sensor. Finally a conclusion Section will give a brief overview on directions for future works and researches.

2. Q-Tables optimization

The next subsections are devoted to describe in details the main concepts and ideas followed to analyze, both perceptually and statistically, the images in each one of the different image database we used in the learning phase. In this way we have obtained a suitable model for the three classes under examination (Document, Portrait and Landscape) obtaining a sort of class-adaptive quantization table that sensibly improves the standard Q-tables.

2.1. Statistical analysis

Many others have tried modeling the global coefficient distribution to optimize the quantization step in the JPEG encoder [14], [19]. Earlier investigations were done by Pratt [13] and Reininger/Gibson [17]. Their works confirm that DCT coefficients can be modeled using Laplacian density functions; these studies are not so well applied to improve the JPEG algorithm because the coefficient modeling affects visual perception in a way that it is difficult to fully understand. We claim that only addressing this kind of analysis to specific classes of images it is possible to obtain effective improvement. The relative statistical regularities, if present, must be captured and properly used. We conducted our experiments using three large databases, respectively for *document*, *portrait* and *landscape* images, but the same kind of analysis could be conducted for every *effective* image class. It is our intention to be able to generalize this methodology obtaining an automatic way to discriminate, if possible, the key concept of *effectiveness*.

We have approximated the statistical distribution of the AC DCT coefficient, both luminance and chrominance components, of an 8x8 block, by a Laplacian distribution in the following way [17]:

$$p_{i,j}(x)=\lambda_{i,j}/2\ e^{-\lambda_{i,j}\,x} \qquad i,j = 0,1,2,\ldots,7;$$

where:

$\lambda_{i,j}= \mathrm{sqrt}(2)/\sigma_{i,j}$;

$\sigma_{i,j}$ is the *(i,j)*-DCT standard deviation, directly measured over a large database;

Experiments confirm this hypothesis. It is interesting to note as in some recent papers [8], [12], these results has been applied in order to reduce the amount of computation in the DCT calculation and/or to improve MPEG encoder performances.

2.2. Human visual system

In order to customize and optimize the performance of the quantization step of JPEG compression algorithm, it is fundamental to be able to consider the perceptual redundancy presents in any digital image. Several attempts in this sense have been done in the past [5], [7], [15], [16], [23] with different results. Obviously the main target in this case is devoted to achieve better compression results preserving the perceptive quality of the image. We have chosen to weight each DCT frequency, in an 8x8 block, using the perceptual model based on both Modulation Transfer Function (MTF) and Contrast Sensitivity Function (CSF). MTF function was used for the first time by Mannos and Sakrison [9]. It describes the human eye's sensitivity to sine/cosine wave gratings at various frequencies. The classical functional expression of the MTF function is:

$$H(\omega)= a(b + c\,\omega)\ \exp(-c\omega)^{d}$$

where $a = 2.6$, $b = 0.192$, $c = 0.114$, $d = 1.1$ and ω is the radial frequency in cycle per degree of visual angle (e.g. based on both resolution display and viewing distance). It is possible to tune the different parameters of the function in order to obtain the most suitable target dependent behavior. The sensitivity of human eyes to sine/cosine wave gratings (e.g. DCT frequencies) can be also measured by the CSF contrast sensitivity function [15].

The two corresponding functions we used are:

$$S_{\lambda}(\omega) = -4.47\ \omega^{2} + 8.86\omega - 4.39;$$

$$S_{A}(\alpha) = -0.0222\ \alpha - 0.001\ \alpha^{2};$$

where ω is the natural logarithm of the number of cycles per degree appearing at the retina, α is simply obtained as $\theta = arctg(j\ /\ i)$ and $\alpha = min\{\theta, 90-\theta\}$ where *(i,j)* are the corresponding DCT basis. The final value used to suitable consider the HVS (Human Visual System) response functions is:

$$P_{i,j} = H(\omega) + ln(S_{\lambda}(\omega) + S_{A}(\alpha)) \qquad i, j = 0, 1, \ldots, 7$$

Obviously the parameters used above can be customized to the particular viewing conditions to achieve specific improvement.

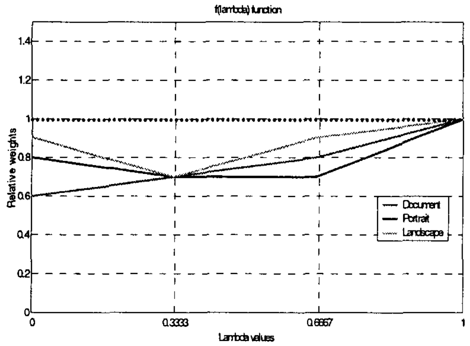

Figure 1 - Plot of the $f(\lambda)$ function for the three different image classes considered.

2.3. The model

Combining together both the statistical model and the perceptive weight associated with each one of the DCT basis we have derived suitable quantization tables, specific for a given class of digital images. The luminance Q-table we adopted is obtained in the following way:

$$q_{i,j} = 1 + round\,[(100/G * (G - P_{i,j}))*f(\lambda_{i,j})\,] \qquad i, j = 0, 1, \ldots, 7$$

where:

- $P_{i,j}$ is the measured weight associated with the i, j coefficient, for a given target display and viewing condition, obtained as a simple summation of the visual functions measured in the previous section with the exception of the DC coefficients however set to 1;
- G is a normalization constant;
- $f : R \rightarrow [0,1]$ is a function that captures the information theoretical data of the various classes. In particular it measures the relative weights for low, medium and high frequencies. Its analytical slope has been obtained, measuring the improvement, in terms of PSNR, with regard to the standard Q-tables, simulating the compression over a data set of images, after having fixed the relative target bit-rate.

In this way it has been possible to obtain quantization values modeling together the effective distribution of the data together with the relative perception of each single frequency. The particular slope of the f function is plotted in Fig. 1. The function has three areas of application: the λ-axis, is properly splitted in three part according to the max λ-value measured. The associated $f(\lambda_{i,j})$ is referred to the particular statistical variation of the (i,j) basis.

3. Experimental results

Our experiments have been performed independently for each one of the three classes of images. The intensive learning phase has been executed using about 100 RGB images for class, with different size and resolution, rigorously uncompressed. The main HVS parameter used for all results reported are: 3 times picture height as distance viewing and 72 *dpi* as resolution display but we have already noted how these values should be specifically evaluated with the real values. We have mainly compared our tables with the tables suggested in the Annex K of the standard JPEG [14]. Using *a priori* knowledge collected in the learning phase about the statistical DCT behavior, it has been possible derive the corresponding quantization tables for each one of the three specific image classes. At the moment our results are addressed versus the optimization of the luminance tables, but we claim that a similar analysis could be realized to for the chrominance tables, improving furtherly the final result. The evaluation of the real effectiveness of the tables obtained has been computed compressing a set of images for specific class, different from the training set, using the standard baseline JPEG compression. The final comparison which results are reported in Tables 1, 2, 3, respectively for *Document*, *Landscape* and *Portrait* images, shows a sensible improvement on the average gain obtained, in terms of PSNR (expressed in Db), between the original images and the compressed ones at various bit rates.

In Fig. 2 is plotted the improvement for a single image, where the typical PSNR-behavior for a single *Document* image compressed respectively with standard and new tables, is clearly showed. The improvement becomes evident for the main target of the particular experiment (2 bpp). Analogous plots have been obtained for *Portrait* and *Landscape* images. It is interesting to note how the best results have been obtained for *Document* images that tend to have, with respect to the other two specific classes considered, statistical regularities evident and particularly captured by the proposed method. The *Document* images present a heavier relative energy in the DCT domain just along the first raw and the first column that are known to be directly referred to the main edges in the block. The reported results suggest us to extend the strategy adopted in a more general sense, optimizing properly each DCT compression engine with the relative scalar dead-zone quantizer, present in almost consumer electronic devices (e.g. digital camera/video).

Bpp	Avg Gain (PSNR)	Avg Gain(%)
0.7	0.36	2.54%
2	2.12	3.77%
3	1.75	2.98%
4	1.60	2.8%

Table 1 - Results for *Document* images.

Bpp	Avg Gain (PSNR)	Avg Gain(%)
0.7	0.23	0.79%
2	0.7692	2.22%
3	0.6958	1.89%
4	0.5028	1.33%

Table 2 - Results for *Landscape* images.

Bpp	Avg Gain(PSNR)	Avg Gain(%)
0.7	0.0964	0.27%
2	0.5889	1.69%
3	0.6720	1.80%
4	0.5145	1.34%

Table 3 - Results for *Portrait* images

4. Q-Tables and camera sensors

The successive step of our optimization strategies has regarded the particular real environment where the compression engine is applied. The quantization process has typically designed to work in a sort of ideal environment without taking into account the real degradation process. A digital images is acquired by an optical engine and/or digital sensor (e.g. CCD/CMOS sensors in digital still cameras) followed by classical image pipeline. This process is intrinsically noisy. Our proposal here is related to the possibility to profile the error, introduced by the digital sensor that captures the image itself, trying to invert it into the successive quantization step, realized during the compression phase.

We have statistically measured, for large database of images, the relative error, in the DCT domain, block by block, between the input original images, and the final images obtained after the application of typical camera pipelines. In particular we have interested in the relative errors introduced by the specific color reconstruction/interpolation algorithm used inside the camera. Our primary goal is devoted to the statistical modeling of such kind of error. We have adopted a suitable strategy inspired by [21] that solves similar problems but in a different context using the relative error due to the color interpolation algorithm, between the input ideal image acquired in a typical bayer data format [1] and the images obtained after the reconstruction process.

The quantization coefficients relative to the basis having measured statistical error greater than a certain threshold have been properly augmented. Our experiments, even if in a preliminary analysis phase, show significant improvement in term of bit-rate, maintaining almost constant (both perceptually and with classical PSNR measures) the quality of the compressed images. Over a test performed over 30 different images we have obtained an average gain of about 15/20 % in term of compression ratio at the same quality level. We believe this strategy could be further refined keeping into account the several algorithmic components that works together, often independently, inside digital consumer devices

5. Conclusions and future works

We have proposed a novel approach to generate quantization tables, suitable addressed for baseline JPEG compression based on DCT transform. It uses both statistical analysis based on DCT coefficients distribution together with some heuristics based on HVS. The results show the effectiveness of the method with respect to the standard Q-tables. The methodology presented could be easily generalized and applied to each DCT compression engine having a dead-zone scalar quantizer. Further improvements have regarded the profiling of the specific error in the DCT domain measured in a common camera pipeline. Future works includes the design of a similar strategy/algorithm for the forthcoming JPEG2000 (Visual frequency weighting and quantization).

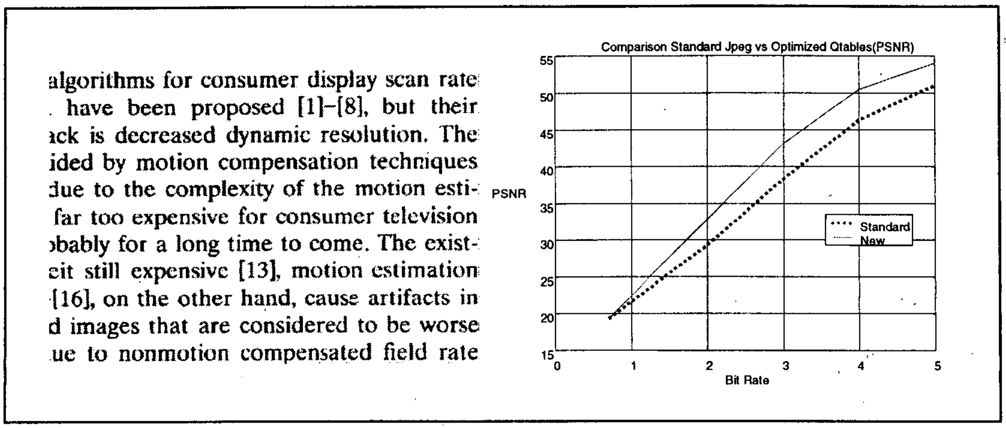
algorithms for consumer display scan rate
. have been proposed [1]-[8], but their
ıck is decreased dynamic resolution. The
ided by motion compensation techniques
due to the complexity of the motion esti-
far too expensive for consumer television
obably for a long time to come. The exist-
eit still expensive [13], motion estimation
[16], on the other hand, cause artifacts in
d images that are considered to be worse
ue to nonmotion compensated field rate

Figure 2 - Typical PSNR behavior for a *Document* JPEG-compressed with Standard and Optimized Q-tables.

References

[1] B.E. Bayer, *Color imaging array*, U.S. Patent 3,971,065 – 1995;

[2] G. Beretta, V. Bhaskaran, K. Kostantinides, B.K. Natarajan, *Method for selecting JPEG quantization tables for low bandwith applications*, Hewlett-Packard Company, U.S. Patent No. 5,883,979, 1999;

[3] N. Chadda, *Segmentation-Assisted Compression of Multimedia Documents*, In Proc. of ASILOMAR-29, IEEE, 1996;

[4] N. Chadda, R. Sharma, A. Agrawal, A. Gupta, *Text Segmentation in Mixed-Mode Images*, 1058-6393/95, IEEE 1995;

[5] L.W. Chang, C.Y. Wang, S.M. Lee, *Designing Jpeg Quantization Tables Based on Human Visual System*, In Proc. of ICIP 1999, pp. 376-380, 1999;

[6] M. Crouse, K. Ramchandran, *Joint Thresholding and Quantizer Selection for Transform Image Coding: Entropy-Constrained Analysis and Applications to Baseline Jpeg*, IEEE Trans. on Image Processing, Vol.6. No.2, 1997;

[7] B. Chitprasert, K.R. Rao, *Human Visual Weighted Progressive Image Transmission*, IEEE Transactions on Communications, Vol. 38, No. 7, July 1990;

[8] G. Lakhani, *Distribution-based restoration of DCT coefficients*, IEEE Trans. Circuits and Systems for Video Tech., Vol. 5 No. 6, 1995;

[9] H. Lohscheller, *A Subjectively Adapted Image Communication System*, IEEE Trans. Commun., Vol. 32, pp. 1316-1322, 1984;

[10] Mannos, D.J. Sakrisson, *The effects of visual fidelity criterion on the encoding of images*, IEEE Trans. on Information Theory, Vol. IT-20, pp. 525-536, 1974;

[11] D.M. Monro, B.G. Sherlock, *Optimal Quantisation Strategy for DCT image compression*, IEE Proc.-Vis. Image Signal Processing, Vol. 143, No 1. 1996;

[12] I.M. Pao, M.T. Sun, *Modeling DCT Coefficients for Fast Video Encoding*, IEEE Trans. Circuits and Systems for Video Technol., Vol. 9, No. 4, pp. 608-616, 1999;

[13] W.K. Pratt, *Digital Image Processing*, New York: Wiley-Interscience, Ch. 10, 1978;

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

[15] M.G. Perkins, T. Lookabaugh, *A Psychophysically Justified Bit Allocation Algorithm for Subband Image Coding Systems*, pp.1815-1818, ICASSP 1989;

[16] H.A. Peterson, A.J. Ahumada, A.B. Watson, *An Improved Detection Model for DCT Coefficient Quantization*, In Proceedings of SPIE Vol. 1913/91, 1995;

[17] C. Reininger, J.D. Gibson, *Distributions of the two-dimensional DCT coefficient for images*, IEEE Transactions on Communications, COM-31(6):835-839, June 1983;

[18] V. Ratnakar, M. Lyvny, *An Efficient Algorithm for Optimizing DCT Quantization*, IEEE Transactions on Image Processing, Vol. 9, No.2, Feb 2000;

[19] S.R. Smoot, L.A. Rowe, *Study of DCT Coefficient Distributions*, 1996;

[20] H.H.Y. Tong, A.N. Venetsanopoulos, *A Perceptual Model for Jpeg Applications Based on Block Classification, Texture Masking and Luminance Masking*, 0-8186-8821-1/98, 1998;

[21] R.A. Vander Kam, P.W. Wong, R.M. Gray, *JPEG-compliant Perceptual Coding for a Grayscale Image Printing Pipeline*, IEEE Trans. on Image Processing, Vol. 8, No. 1, Jan 1999;

[22] H. Yanagihara, *Image Compression Based on Pattern Fineness and Edge Presence*, Sony Corporation, U.S. Patent No. 5,374,958, 1994;

[23] K.A. Wen, C.Y. Lu, M.C. Tsai, *The Transform Image Codec Based on Fuzzy Control and Human Visual System*, IEEE Trans. on Fuzzy Systems, Vol. 3, No. 3, 1995;