ACTA STEREOL 1995; 14/2: 137-146
ORIGINAL SCIENTIFIC PAPER

AN EFFICIENT IMAGE CODER BASED ON THE COMBINATION OF VISUAL AND PROBABILISTIC MODELS

Hocine Cherifi, Thierry Eude[1]

IIV URA-CNRS 842 site GIAT 3 Rue Javelin Pagnon 42000 St-Etienne, France
[1]LIESIB Université de Bourgogne 6 Boulevard Gabriel BP138 21004 Dijon CEDEX, France

ABSTRACT

To maintain high image quality with low bitrate, an effective coding algorithm should not only remove statistical correlation but also perceptual redundancy from image signals. In this paper, a coding scheme based on statistical and perceptual models is presented. Roughly speaking, the compression process is performed in two stages. In the first stage we derive a perceptual measure called the just noticeable distortion (JND). This measure provides each signal being coded with an error visibility threshold, below which reconstruction error are rendered imperceptible. Using this threshold the perceptually insignificant signals can be screened out. In the second stage, we perform a quantization thresholding step using statistical information on the DCT coefficient distribution. Such that statistical redundancy is minimized. Using this system, high quality is obtainable at rates lower than 1 bit/pixel for a wide variety of images without user intervention.

Keywords: compression, human visual system, JPEG, quantization, statistical redundancy.

INTRODUCTION

Data compression has become an essential component as the demand of both data rate and volume increases dramatically. Also the necessity to develop efficient coding techniques to reduce the number of data required for image storage and transmission is clear. The need to inter-operate different realizations of signal encoding devices and signal decoding devices has led to the formulation of international image and video compression standards (Pennebecker et al, 1992). The compression standard developed by Joint Photographic Experts Group (JPEG) has been widely used not only because of its algorithmic simplicity but also of its performance in terms of visual losslessness. Within the definition of this standard there is opportunity to develop alternative coding algorithms and naturally there is a great deal of interest in defining coding procedures which reduce information redundancy without loss of image quality. To tackle this problem, one can adopt a statistical or perceptual point of view.

In the statistical approach, many authors have proposed quantization and thresholding methods based on various strategies and measures. In the perceptual approach "subjective" measures based on human visual system (HVS) knowledge are used to remove perceptual redundancy (Jayant et al, 1993). The compression ratios of the current generation of low bitrate JPEG coders based on perceptual or statistical models appear to have reach a saturation level of around 1 bit per pixel. Many applications would benefit from higher compression rate. To achieve this level of performance we believe that the coder design must combine elements from these complementary approaches. In other words, one has to take into account the two parts of the information pertinence contained in an image. In this paper, we propose a compression algorithm that achieves good image quality with minimal redundancy in a statistical and perceptual meaning. Our goal in this work was to develop a perceptually and statistically tuned image coder which would provide nearly transparent quality to a coded image. First an overview of the baseline JPEG encoder is presented. Then the statistical and perceptual models is described. We discuss how the proposed encoder using both models is defined. Finally results of the comparisons with the baseline JPEG are given.

THE BASELINE JPEG ENCODER

This overview will concentrate on the aspects of the baseline encoder that are needed to understand the compression scheme developed. The JPEG standard defines a family of encoding/decoding algorithms for continuous tone images and a data stream architecture for encapsulating and describing the compressed data. The algorithm used in this study is the so-called baseline sequential algorithm. It can be divided into the three major components shown in Fig. 1.

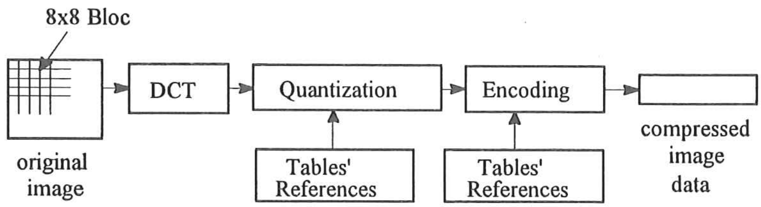

Fig. 1. General principle of DCT based compression method.

The image is divided into 8x8-pixel blocks which are each transformed into a set of coefficients. The transform is an 8x8 Discrete Cosine Transform (DCT). Its purpose is to perform an energy compaction on the signal. Since most images have a low pass spectrum, transforming the data into the frequency domain results in a fewer significant samples clustered at the low frequencies.

Quantization of the DCT coefficients achieves image compression, but also causes distortion in the decompressed image. This step causes information loss, but provides for the majority of the data rate reduction in the system. The 64 scalar-quantizer steps are collectively termed the quantization matrix. In the JPEG norm, these coefficient thresholding and quantizing stages have been combined into a single quantization matrix.

The DCT coefficients $DCT(u,v)$ are quantized to $DCT_q(u,v)$ using:

$$DCT_q(u,v) = IntegerRound\left[\frac{DCT(u,v)}{f \times q_{u,v}}\right] \qquad u,v = 0,1,2,\ldots,7 \tag{1}$$

where f is a parameter that determines the compression ratio and $q_{u,v}$ is an element of the quantization table.

Two example quantization matrices have been included in the JPEG standard. One of these matrixes is commonly used for gray-level images, and for the luminance component image of color images. It is commonly called the visibility matrix. The other matrix is used for color images. These matrices were designed for a particular compression/viewing scenario, and it is not clear how they should be changed when used under different viewing conditions. Indeed, acceptable performance rate/distortion performance depends upon proper design of the quantization matrix. Then, quantized coefficients are converted from matrix to vector form using a so-called zigzag scan that roughly orders the coefficients according to increasing frequency. The resulting vectors of coefficients are runlength and Huffman coded.

STATISTICAL OPTIMIZATION

Modeling probabilistic distributions of DCT coefficients can play an important role in designing an optimal quantizer for the transform-based compression system. There has been several studies on the distributions of the DCT coefficients.

Pratt (Pratt, 1978) used the central limit theorem to justify the gaussian model. Reininger (Reininger et al, 1983) claimed that the DC coefficient has a Gaussian distribution and AC coefficients can be modeled by Laplace laws. Murakami (Murakami et al, 1982) and Bellifemine (Bellifemine et al, 1992) end in the same conclusion. Eggerton (Eggerton et al, 1986) has concluded that no one density function can be used to approximate each coefficient. He stated that Laplacian distributions provide the best fit for most of the coefficients, but if all of the coefficients must be lumped into one density function, a Cauchy distribution is best. At last, Zang (Zang et al, 1990) introduces the general gaussian distribution which provides a better fit. His works are confirmed by Lee (Lee et al, 1993). Performing an extensive statistical study (Eude et al, 1994), we showed that the distribution of the DCT coefficients is characterized by a high kurtosis variability as shown in figure 2.

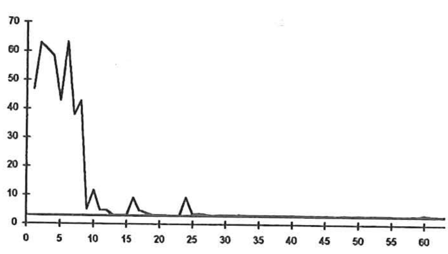

Fig. 2. Kurtosis values of DCT coefficients of a radiological image.

This lead to the conclusion that all the distributions which have a fixed kurtosis value are inadequate to fit to this behavior. That's why the generalized gaussian has been introduced. Indeed this distribution has two parameters that allows to control the kurtosis value. Nevertheless it has no physical justification and identification of the parameters is uneasy. In a previous work we showed (Eude et al, 1994), that finite mixture of 1, 2, or 3 Gaussian laws are more appropriate to fit the DCT coefficients distribution. The mixture of Gaussian is defined as follows:

$$\Phi(\sigma_i, p_i, x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{2\pi}\ \sigma_i} \exp-\left(\frac{x^2}{2\sigma_i^2}\right) \quad (2)$$

where p_i, σ_i are respectively the prior probability and the standard deviation of the ith component.

For that, we tested the sample data with reference functions of Cauchy, finite mixture of Laplacians and Gaussians up to 3 components. To conduct the experiment, we used a medical image bank from the French Lecturer Radiology College (CERF) (Lucas et al, 1992). For all of the images, we obtain similar results. These results exhibited the general trend that high frequency coefficients are well approximated by a single Gaussian while low frequency coefficients are much better approximated by a two or three components mixture.

To remove the statistical redundancy from the image, we use these results to determine a quantization matrix. The idea is to keep "less probable" values which have most of the signal information. The thresholding and quantization process are performed in a single operation using the same operation than JPEG but the tables are deduced from a statistical study of a set of images using the strategy:

$$q_{u,v}^{-1} \text{ is such that } \int_{-\infty}^{q_{u,v}^{-1}} \Phi(x, \hat{k}, \hat{p}_i, \hat{\sigma}_i)dx = 0.95 \quad (3)$$

where $\Phi(x, \hat{k}, \hat{p}_i, \hat{\sigma}_i)$ is the estimated probability density function of the coefficient DCT(u,v) assuming a Gaussian mixture distribution model.

The coding scheme is illustrated by the Fig.3.

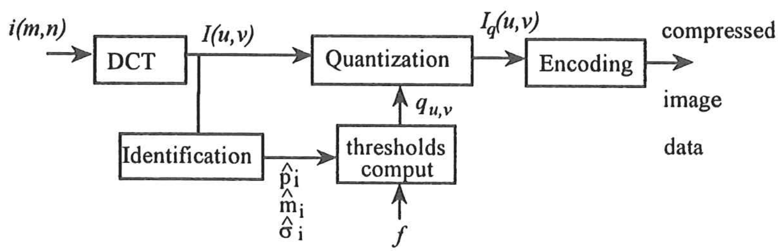

Fig. 3. Statistical quantization process methodology block diagram.

PERCEPTUAL OPTIMIZATION

The removal of subjective redundancy involves discarding information which the designer feels can be removed without any change being noticed by the human observer. The sensitivity of the human visual system to stimuli of varying levels of contrast, luminance and different spatial and temporal frequencies varies greatly, and these inconsistencies can be exploited to determine how information can be discarded without subjectively degrading the final image. A number of methods have already been proposed for including certain psychovisual properties of the human visual system (HVS) (Watson, 1994). These studies have attempted to derive a computational model of visual masking. From these models, it is possible to determine a masking threshold for each DCT coefficient.

In this work, we use the just noticeable distortion (JND) concept recently introduced by Jayant (Jayant, 1992). This measure incorporates contrast sensibility and spatial masking effect of the human visual system. It is defined by:

$$JND(x,y) = \max\left\{f_1\left[m_{bg}(x,y), \sup(m_{ld}(x,y,n_{x,y})\right], f_2\left[m_{bg}(x,y)\right]\right\} \tag{4}$$

where $m_{bg}(x,y)$ and $m_{ld}(x,y)$ denote respectively the average background luminance and the maximum weight average of luminance differences around the pixel at (x,y). The spatial masking effect is modeled by $f_1(x,y)$ in which the relationship between visibility threshold and luminance edge height is linear. The visibility threshold due to the background luminance is empirically modeled by $f_2(x,y)$ as a root equation when $m_{bg}(x,y) < 127$, and as linear equation when $m_{bg}(x,y) \geq 127$. In case of low bitrate coding, a minimally noticeable distortion (MND) can be used. The MND profile is obtained from lifting JND by a scale factor. This allows to reach a minimally perceptible distortion which is expected to be uniformly distributed over the reconstructed image.

According to the sensitivity of human perception to spatial frequency, the thresholds are derived using :

$$\gamma(u,v) = |DCT(JND(x,y))| H(f) \tag{5}$$

where $H(f)$ denotes the energy weight of the DCT coefficient which is obtained from normalizing the relative sensitivity given by the modulation transfer function (MTF) of the HVS.

ENCODING SCHEME VISUALLY AND STATISTICALLY OPTIMIZED

The problem now becomes how to use the visual masking model in the JPEG environment. Most authors (Watson, 1994) incorporate a perceptual model through a perceptually optimal quantization table. The main drawback of this approach is that whatever strategy is used to deduce a quantization table from these values, it will result on a globally optimal matrix that is not necessarily locally optimal. Remember that the perceptual model allows to determine a perceptually optimal quantization table for each block of DCT coefficients but unfortunately, JPEG allows only one quantization table for each image.

Addressing the problem in such a way neglect the fact that many masking phenomena take place at a local level rather than on the level of the entire image. In fact the JND threshold of a given coefficient within a block represents a measure of the subjective pertinence of the observed value. If it's lower than the threshold it has no perceptual meaning and therefore can be discarded. It can be interpreted as part of the redundancy of the signal because it is non informative on a perceptual sense. This observation is the key to reduce perceptual redundancy in the JPEG process. Rather than looking for a perceptually optimal matrix our strategy is to use the JND thresholds as decision thresholds of pertinence of the DCT coefficients. All the coefficients. which have value lower than the JND threshold are not perceptually significative and can be set to zero. All the remaining coefficients correspond to an essential visual information and should be properly encoded. The overall encoding scheme is shown in figure 4. It is the combination of the perceptual processing followed by a compression process using a statistical quantization.

The perceptual model drives a local thresholding stage of the image in the transformed domain using the following formula:

$$I_p(u,v) = \begin{cases} I(u,v) & \text{if } I(u,v) > \gamma(u,v) \\ 0 & \text{if } I(u,v) \leq \gamma(u,v) \end{cases} \tag{6}$$

This stage identifies the meaningful information while reducing the bitrate required to encode the image. Then in the second stage, the parameters identification of the DCT coefficient distribution allows to construct the quantization matrix. Finally these quantized coefficients are entropy coded as in standard JPEG.

Not that for very low bitrate application the same coding scheme can be used by computing MND instead of JND.

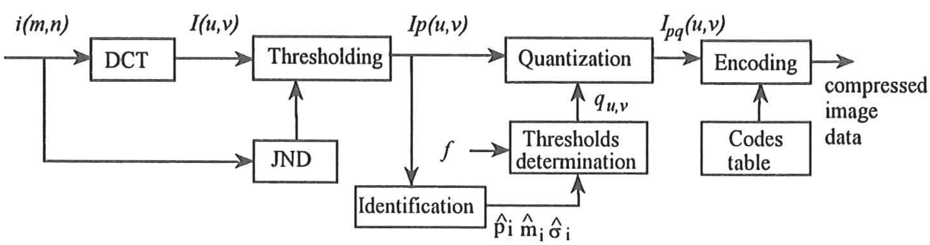

Fig. 4. Encoding scheme

RESULTS

We investigate the performance of the compression process proposed in the previous section on a set of 50 test images ranging from simple (low-resolution scenery) to complex (strongly contrasting textures). Baseline JPEG using the visualization matrix and our coder were run on the database. For each image an optimized matrix is properly defined from the probabilistic model. The parameters identification has been carried out using the expectation maximization algorithm (Redner 1984). Two performance measures are used in the evaluation of the algorithms. The first one is the peak signal-to-noise ratio (PSNR), the second criterion is subjective quality obtained after the reconstruction process. For a grey-scale NxN image the PSNR is defined by :

$$PSNR = 10\log_{10}\frac{255^2}{\left(\frac{1}{N}\right)^2\sum_{k=1}^{N}\sum_{l=1}^{N}\left(i(k,l)-\hat{i}(k,l)\right)^2} \tag{7}$$

Comparisons at fixed bitrate have been carried out. The results clearly show that a significant gain in visual quality is obtained. The image used to illustrate this behavior is shown in Fig.5 a). Both the reconstructed and difference (between original and reconstructed) images are displayed in Fig. 5b and 5c, and 5d and 5e, respectively for the two coding process.

Table 1. Results of the subjective image-quality experiments.

Subject	Times preferred	Times Baseline JPEG preferred
1	25	25
2	30	20
3	28	22
Total	83	67

Typical results of the PSNR versus the measure of the compression rate are shown in Fig.6 for the same image. Data points are given for each of the compression scheme. Note that optimized JPEG performs always better than the baseline JPEG. The PSNR gain is at least of 2 DB. In fact, at the highest bitrate tested the proposed coding scheme has significantly higher average PSNR than does baseline JPEG.

Table 1 summarizes the results of the subjective image quality experiments that compared the two encoding processes at an average bitrate of 0.45 bit/pixel for all the database. Each inexperimented subject made 50 forced choices selecting the decompressed image that they felt exhibited better image quality. It can be said that no individual exhibited a strong preference for one compression algorithm over the other. Nevertheless, in total, the group of viewers selected Baseline JPEG 67 times and our codec 83 times out of 150. All the subjects commented on the annoying blocking effect that are sometimes present in baseline JPEG.

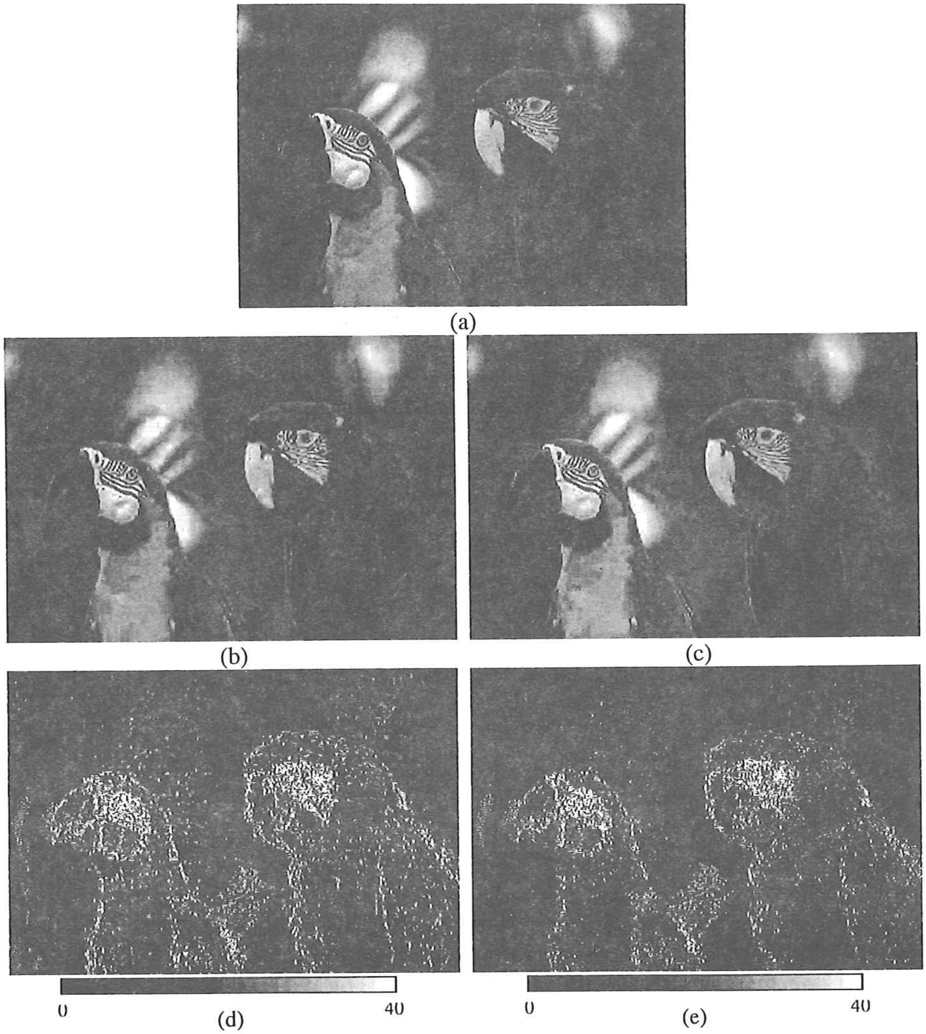

Fig. 5. Original (a), reconstructed (b,c) and difference images (d,e) respectively for baseline JPEG (left) and the optimized coding process (right).

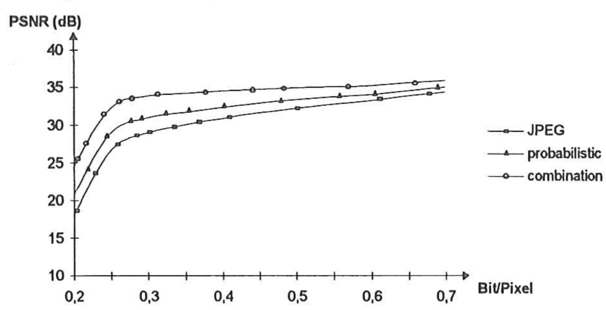

Fig.6. Peak signal-to-noise ratio as a function of the bitrate.

CONCLUSION

The removal of statistical or perceptual redundancy for JPEG coders is an important area of study. However no coding scheme as yet adequately combined this effect to produce an efficient and optimum method of removing subjectively redundant information and statistical redundancy. The combination of standard transform coding techniques statistically optimized through the quantization step and psychovisually optimum thresholding has resulted in a high compression algorithm. Its compression gains are a result of a combination of all the compression methods which work cooperatively to automatically provide good compression results and quality over a variety of images without user intervention. Of course, these gains are obtained at the cost of an increased encoder complexity. Because of the general nature of the models exploited in the compression scheme, the same techniques can be incorporated into almost any communication system involving still or moving images. It should be remembered that the bitrate could be further reduced if a limited amount of suprathreshold distortion was allowed by using the MND instead of the JND.

REFERENCES

Bellifemine F, Capellino A, Chimienti A, Picco R, Ponti R. Statistical analysis of the 2D-DCT coefficients of the differential signal for images. Signal Processing Image Communication 1992; 4(6): 477-88.

Eggerton J D, Srinath M D. Statistical distribution of image DCT coefficients. Computer and Electrical Engineering 1986; 12: 137-45.

Eude T, Grisel R, Cherifi H, Debrie R. On the distribution of the DCT coefficients. Proceed of IEEE ICASSP 1994; 5: 365-68.

Jayant N. Signal compression: technology targets and research directions. IEEE Selected Areas Communication 1992: 314-23.

Lee H, Kim Y, Rowberg A H, Riskin E A. Statistical Distributions of DCT Coefficients and their Application to an Interframe Compression Algorithm for 3-D Medical Images. IEEE Transaction on Medical Imaging 1993; 12(3): 835-39.

Lucas C et al, Evaluation subjective de la compression d'images médicales issues d'un vidéodisque. Revue d'Imagerie Medicale, Springer-verlag, 1992, 4: 53-59.

Murakami H, Hatori Y, Yamamoto H. Comparison between DPCM and Hadamard transform coding in the composite coding of the NTSC color TV signal. IEEE Transaction Communication 1982, COM-30: 469-79.

Pennebecker W B, Mitchell J L. JPEG Image data compression standard. Van Norstrand Rienhold 1993, New York.

Pratt W K. Digital Imaging Processing 1978; New York:Wiley-Inter-science.

Reininger R C, Gibson J C. Distribution of two-dimensional DCT Coefficients for Images. IEEE Transaction Communication 1983; COM-31(6): 835-39.

Redner R A, Walker H F. Mixture densities, Maximum likelihood and the EM algorithm. SIAM Review 26 1984: 195-240.

Watson A B. Perceptual optimization of DCT color quantization matrices. Proceed of IEEE ICASSP 1994; 5: 100-4.

Zang Y, Loew M H, Pickholtz. A methodology for modeling the distributions of medical images and their stochastic properties. IEEE Trans Medical Imaging 1990; 9: 376-83.

Presented at the 9th International Congress for Stereology, Copenhagen, August 20th to 25th, 1995.