ADAPTIVE DCT CODING BY ENTROPY GUIDED SEGMENTATION

Giovanni JACOVITTI and Alessandro NERI

INFOCOM DPT. -University of Rome "La Sapienza", Via Eudossiana 18, 00184 Rome, ITALY.

RÉSUMÉ

Le méthodes traditionnelles de compression d'image parmi transformée cosinus utilisent des règles de quantisation uniforme sur le plan d'image.

Dans cet article on expose un critère pour régler la quantisation suivant des éxigences générales d'intéret visuel.

ABSTRACT

Usual DCT based compression schemes are based on spatially uniform quantization rules. In this work a criterion for varying the quantization according to some general requirements of visual interest is illustrated

1. INTRODUCTION

Image coding is a very important area of bidimensional signal processing intended to reduce the number of bits necessary for image representation.

The worldwide growing multimedia activities require increasing resources for image storage and transmission over telecommunication channels. Image representation in terms of original samples is statistically very inefficient because of the inherent redundancy of information carried in this form.

For this reason, bits reduction rates of 25 for still images, and of 100 for live images (sequences of images) are reasonable objectives. However, these compression rate are attainable only if the adopted fidelity criterion properly takes into account the perceptive features of the human visual system.

This implies a loss of information which is judged not relevant for a human observer. As a matter of fact, it has been experienced that some important reduction systems, such as quantization of the Discrete Cosine Transform (DCT) spectral coefficients, are well tolerated and produce slight degradation of the subjective quality.

Still, compression rate achieved by efficient coding schemes are inadequate to deal with current tecnologies in data bases applications.

In fact, applications often require fast and frequent access to images. Browsing and navigation in image databases could imply large volumes of data flowing through server gates and transmission channels. Actual available rates are often insufficient especially in multiuser environment.

For this reason, progressive access schemes have been devised. The image access is broken and delivered throug several stages, corresponding to different visual information levels. The user begins with the lowest level corresponding to poorest quality and acquires more information passing to highers levels. He stops when the quality is sufficient for his purposes. In browsing activities this could arises immediatly at the lowest level , which implies the transfer of small size records.

For instance, in the JPEG proposal for ISO CCITT standard [1], two schemes of progressive access are considered.

This scheme, and other progressive methods such as subband coding, are based on a zoom or, equivalently, on a resolution image concept: the image is properly "focused" so that finer details are visible as low as higher stages are accessed.

This is a satisfiying procedure for many photografic (gray scale) images. On the other hand, in some significant applications the strategy based on resolution is not efficient. This occours for instance, when important details, such as letters or numbers, occupy a small portion of the image, and must be read at full resolution. To reduce the cost in such operative conditions, interesting details should be delivered immediately at low cost stages. However the interesting details and structures are generally user dependent, so that interactive methods such as selection of windows are employed.

If automatic operation is required, a measure of "relevance" statistically well related to the average user needs is necessary.

The problem of image segmentation based on visual relevance concept is not developed in current literature,

because it is believed that relevant features can not be extracted at signal level, and severe difficulties are present at semantic levels.

However, in an evolutional conceptual view, it can be argued that in general, biological sensorial systems are particularly oriented to detect and locate non-cahotic structures.

In fact, it appears that the attention of living beings is generally attracted towards well structured events whereas poor interest is devoted to noise-like signals which are globally perceived as "background" or "textures".

Starting from this observation, we accept the rater generic conjecture that " the relevance of image elements is related to their non-cahotic structure".

Many quantitative definitions of the cahotic nature of a given signal are possible, either in a deterministic or statistical sense.

We refer to the concept of entropy early employed in seismic deconvolution by Wiggins [2].

In this scheme, a signal is thought as a segment of a realization of a stocastic process. The measure of cahos is defined as the entropy of the source. For a given variance, the maximum cahos corresponds to a Gaussian source. The minimum of entropy corresponds to sparse spikes. In terms of images, Gaussian fields (wich are of minimal interest for any observer) are maximum entropy, whereas small bright spots in dark background (which are of outstanding interest) are minimum entropy.

These concepts have been tested in an operative framework of hierarchical image transmission, as reported in [3].

In this contribution, we show how the above conjecture and the subsequent entropy criterion can be applied in some specific contexts, where the problem is to optimize "one-shot" accesses to images characterised by small zones of paramount interest, rather than to progressive operation.

The results presented here apply to any compression method, when space-adaptivity is allowed. We refer in particular to the widely adopted ADCT technique.

2. THE ADCT TECHNIQUE

Any image compression/decomposition scheme can be regarded as a procedure involving three stages:
- decomposition
- coding
- imaging

In the decomposition stage, the image is represented through a set of new images where the coding takes place. Then, the imaging stage allows to reconstruct an image by inverting the decomposition process. The compression is performed in the second stage, and its effectiveness is strictly related to the kind of decomposition/imaging adopted. In fact, high compression rates can be achieved if some loss of information is accepted in the coding stage. The amount of acceptable loss depends on how the decomposition scheme is matched to some psychovisual mechanism.

In the ADCT compression method, the decomposition of an (NxM) image is made by generating a set of 64 ((N/8)x(M/8)) images containing each the (n,m)-th coefficient of the DCT transform of individual 8x8 image blocks. The reconstruction of the image (imaging) is performed by adding up in each 8x8 block the 64 kernels of the DCT weighted by the correspondent coefficients.

It has been demonstrated that the kernels do not have all the same importance from the human user viewpoint. This implies that rough quantization of some coefficients does not impair the final quality of the reconstructed image. More specifically, a 8x8 "visibility matrix" containing the quantization levels allowed for each coefficient in order to give good visual results has been experimentally found.

The outstanding performance of the ADCT scheme (25 compression factor) is based on this visibility matrix which implies heavy quantization.

Varying the visibility matrix, it is possible to regulate the quality and the corresponding compression factor to the wanted trade off.

The main limit of this regulation in conventional ADCT implementations is that the visibility matrix is kept constant over the whole image. This means that tolerable distortion into a zone becomes untolerable elsewhere.

As mentioned before, in many applications parts of an image have relevant visual importance for the user, while high distortion is tolerated for the remainder. In those cases, the visibility matrix assigned to visual relevant parts should be different from the one employed elsewhere, in order to optimize the access cost.

The entropy criterion provides an effective mean to select the proper matrix in many interesting cases, as demonstrated below.

3. THE ENTROPY GUIDED QUANTIZATION

The fact that the visual relevance is related to the non cahotic character of the image has been confirmed by several experiments.

Obviously, different behaviour must be expected from different users but the validity of our assumption is only of statistical concern.

It has been argued that, considering each zone of the image as a realization of a stochastic source, the related entropy could be used to classify the visual relevance.

However, the estimation of the entropy is rather

Fig. 1 - Original image quantized to 256 gray levels.

Fig. 2 - Image reconstructed after uniform quantization.

unpractical because of its complexity and unstability. Instead, the indication of cahotic behaviour can be more simply obtained by moments.

In particular, it has been found that the kurtosis is well suitable for image segmentation. As in seismic inversion, the kurtosis, defined as the fourth order central moment, normalized to the square of variance, is a suitable measure of non gaussianess and, indirectly, of the visual interest.

In fact, values of kurtosis close to 3 indicate gaussian behaviour, and deviations from 3 are well related to visual appearance [3].

However, in images the statistical dependence between samples is usually high so that the estimate of the "marginal" kurtosis is not the most adequate.

It is important to transform the image in such a way that the most part of the statistical dependence is removed.

The DCT transform yet employed for coding 8x8 blocks is well suited for this purposes. The spectral coefficients are loosely correlated, and this explains in part the good performances of such a coding technique.

Fig. 3 - Image reconstructed after entropy guided quantization.

The set of 64 spectral coefficients yelds a distribution depending on the input image. If the latter is gaussian as well, the coefficients are gaussian distributed, and the kurtosis value is 3.

Some experimental work has been carried out to properly define the visibility matrices to be selected by the kurtosis values.

An example taken from an image used for HDTV evaluations is reported in figs. 1-3.

In fig. 2 the image reconstructed after uniform application of the JPEG visibility matrix (see [1]) scaled by a factor of 2 is reported. The overall mean number of bits for representing such an image was 0.71 bit/pixel.

In fig. 3, the image reconstructed after entropy guided quantization is shown. The associated cost in terms of bits is 0.38 bit/pixel.

Of coarse, this saving is payed in terms of pleasance, because the blocking effect and the loss of details is evident.

However, if the scope of the access was to read some essential information, such as the calendar numbers, or the nature of the profiles into the foulard, the advantage becomes evident.

Conversely, it is possible to enhance the overall quality at the same cost.

REFERENCES

[1] JPEG Technical Specification, revision 5, Jan. 16, 1990.
[2] R. A. Wiggins "Minimum Entropy Deconvolution", *Proc. of Int Symp. on Computer Aided Seismic Analysis and Discrimination*, June 9-10, Falmouth, Mass., IEEE Computer Society, pp.7-14; also *Geoexploration* 16, pp. 21-35.
[3] M.A. Imperato, G. Jacovitti, A. Neri : "Space Variant Resolution ADCT Still Image Coding", *European Trans. on Telecommunication*, Vol. 2 - N.3, May- June 1991, pp. 299-304.