COSTS AND ADVANTAGES OF SHAPE-ADAPTIVE WAVELET TRANSFORM FOR REGION-BASED IMAGE CODING

M. Cagnazzo, G. Poggi, L. Verdoliva

Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni
Università Federico II di Napoli, via Claudio, 21 – 80125 Napoli, Italy
{cagnazzo,poggi,verdoliv}@unina.it

ABSTRACT

Region-based encoding techniques have been long investigated for the compression of still images and video sequences and have recently gained much popularity, as testified by the object-based nature of the MPEG-4 video coding standard. This work aims to analyze costs and advantages of implementing such an approach by shape-adaptive wavelet transform and shape-adaptive SPIHT. The analysis of several performance measures in a number of experiments confirm the potential of wavelet-based region-based approach, and provide insight about what performance gains/losses can be expected in various operative conditions.

Key-words: image coding, region-based coding, shape-adaptive wavelet transform.

1. INTRODUCTION

Object-based image coding has been an active area of research for almost two decades, starting from the seminal paper by Kunt et al. [1] on second generation coding techniques. The major conceptual reason to turn to object-based coding is that images are *naturally* composed by "objects". The usual pixel-level description is only due to the lack of a suitable language to efficiently represent objects, allowing for a compact description of images, their high-level manipulation and their editing. This philosophy inspires the recent MPEG-4 standard [2], which in fact has been a significant force behind research on object-based coding.

At a lower level of abstraction one can consider region-based coding, where the concept of object as elementary semantic unit is somewhat lost, but similar technical challenges arise. Image regions can be identified based on statistical rather than semantic properties by means of pixel-level segmentation, with the aim of exploiting their statistical homogeneity to improve coding efficiency, as is done, for example, in [3]. Likewise, the user himself can identify a region of interest (ROI) to be encoded at higher priority or with different techniques than the background, as envisaged in several applications and standards [4].

Both these approaches require some basic tools to efficiently represent component objects (or regions) within an image in order to provide a good coding performance. This breaks down to the description of the object shape, (segmentation map, region contours, etc.) and of its interior waveform (polygonal approximation, transform coding, etc.). Just as for rectangular images, also for generic regions wavelet-based techniques seem to offer the best waveform coding performance. In particular, the shape-adaptive wavelet transform (SA-WT) algorithm recently proposed by Li and Li [5] to extend the wavelet transform to arbitrarily shaped regions, is quite simple and general, and in fact has been adopted in the MPEG-4 standard. After SA-WT, one can resort to a suitable shape-adaptive version of well-known coding algorithms such as the EZW or EZT, as done in [5], or the well-known SPIHT, as done in [6].

Of course, before resorting to object-based coding, and to a particular suite of algorithms, one should be well aware that there are both advantages and costs to take into account. We have already remarked some of the high-level advantages of an object-based image description, which could justify its use even in the face of reduced efficiency, but here we only focus on measurable gains and losses in terms of rate-distortion performance implied by this approach. The goal of this work is to shed light on such trade-offs by means of numerical measures obtained with a specific, but quite popular, wavelet-based coding scheme. To this end we carry out a number of experiments in order to explore a wide range of operating conditions of interest and provide some guidelines about the convenience of the object-based approach.

In Section 2 we specify our encoding scheme, in Section 3 describe the elementary phenomena of interest and their related measures, in Section 4 present and discuss the experimental results, and finally in Section 5 draw conclusions.

2. CODING SCHEME

Our coding scheme comprises the following steps
- image segmentation;

- lossless coding of the segmentation map (object shapes);

- shape-adaptive wavelet transform of each object;

- shape-adaptive SPIHT coding of each object;

- optimal post-coding rate allocation among objects.

The problem of faithful image segmentation is not an issue here, and is not investigated. Moreover, to carry out our analysis we need to change rather freely the geometrical/statistical parameters of objects, therefore we resort to an "artificial" regular segmentation of the images in tiles having all the same shape and dimension but for tiles near image borders. The segmentation maps are encoded without loss by means of a slightly modified version of the RAPP algorithm [7], originally proposed for palette images, which is very efficient for this task. For the SA-WT we resort, as said before, to Li and Li's algorithm, with five levels of decomposition, Daubechies 9/7 filters, and the global subsampling option because of its superior performance. Each transformed object is then encoded by means of the SA-SPIHT algorithm [6]. This is very similar to ordinary SPIHT, with the differences that only active nodes, that is nodes belonging to the support of the SA-WT coefficients, are considered, and that in the lowest frequency band the tree structure is similar to EZW. Finally, the rate-distortion curves obtained for all objects are analyzed so as to optimally allocate bits among objects for any desired encoding rate.

3. MEASURING COSTS AND GAINS

We identify the following costs and gains of object-based coding (in the rate-distortion sense):

- the cost associated with the segmentation map;

- the cost due to the reduced efficiency of WT/SPIHT when applied to regions of arbitrary shape and size;

- the gain due to the increased efficiency of WT when region boundaries are removed.

The first cost depends only on the map complexity, that is the number, size and shape of the component regions. It is measured in a straightforward way, as the coding cost in bit per pixel (bpp) obtained using the modified RAPP algorithm. It could be observed that such a measure depends on the specific map coder, but state-of-the-art map coders have very close performances, and an objective measure, like map entropy, is practically unfeasible. In this cost we also include all other side information, like that for rate allocation, which is negligible anyway.

The second cost needs a deeper analysis since it depends on two distinct phenomena. First of all, WT's ability to compact energy depends on the length of the signal segments on which it operates (besides their statistics, of course): the shorter the signal, the smaller the transform
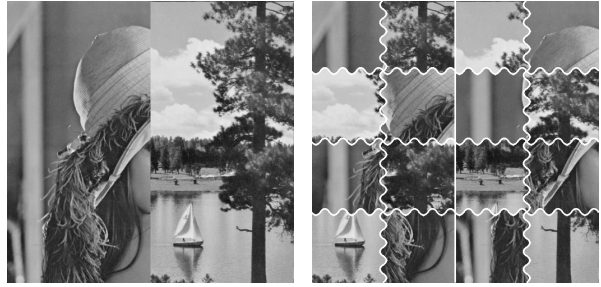
Fig. 1. Lena-Lake composite image (a) and an example mosaic, with 128×128 tiles and wavy boundaries (A=8, C=4).

gain. This effect is quite limited when segments are still relatively long, say, 64 samples, but can become significant when much shorter segments are considered. Note that short segments arise not only in the presence of small objects but also for large objects with rough contours. A second reason for reduced efficiency lies in the lower effectiveness of the zerotree coding approach. In fact, with full quaternary trees, each bit spent to describe an intermediate node conveys information about 4 children, 16 grand children and so on while this is no longer the case with arbitrarily-shaped regions. For example, in the case of a single-line horizontal object, the quaternary tree becomes a binary tree and the encoding efficiency reduces significantly.

We consider these phenomena together because we can only measure the cumulative rate-distortion loss produced by both. To this end, we will take natural images and object-encode them with arbitrary segmentation maps: no advantage can be expected in this situation, and the observed RD loss with respect to the "flat" (no segmentation) case can be surely imputed to the WT/SPIHT inefficiencies. However, in order to gain some insight also on the individual contribution to the loss we will also report two indirect cost figures. The compactation ability of WT is measured by the well-known coding gain (CG), defined as the ratio between the distortion D_{PCM} provided by direct quantization of the input signal and the minimum distortion D_{TC} achievable with transform coding. For orthogonal subband coding, at high resolution, this turns out to be the ratio between the variance of the input signal and the weighted geometric mean of subband variances. Some insight into the efficiency of SPIHT will be provided instead by computing the average number of pixels per tree (P/T), which is the number of symbols that is spared when a zerotree is encountered.

Let us finally turn to the gains. If segmentation accurately singles out the elementary regions in the image the above costs could be compensated by the advantage of removing edges, since most of the encoding resources, especially at low rates, are spent to describe the spatial discontinuities among neighboring regions. The segmentation map provides an alternative, and possibly more efficient,
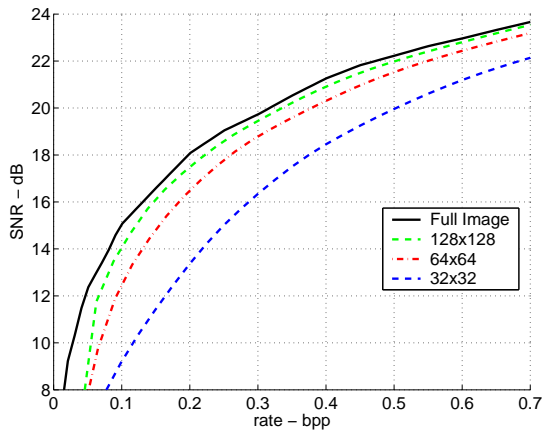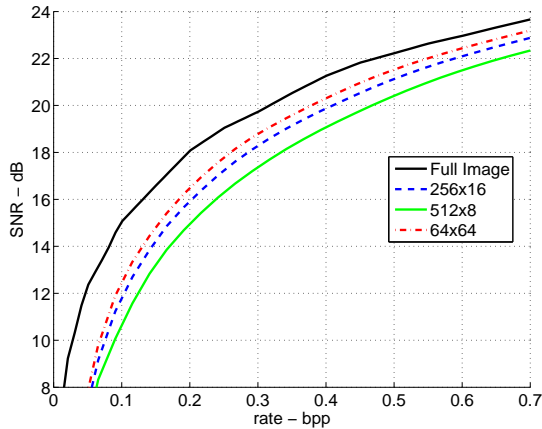
Fig. 2. Rate-distortion curves for square tiles

Fig. 3. Rate-distortion curves for rectangular tiles

Tiles	MC	CG	P/T	L@0.3
512×512	0	46.88	1024	0
128×128	.003	45.56	1024	0.24
64×64	.005	43.26	1024	0.83
32×32	.013	40.19	1024	2.77
128×32	.005	41.81	1024	0.95
256×16	.005	38.04	512	1.30
512×8	.005	32.43	256	2.24

Table 1. Synthetic parameters for rectangular tiles

sion to colour images does not present any significant problem, nor any special interest here.

In the first experiment the image Lena is segmented in square tiles of sizes 512×512 (whole image) down to 32×32 and encoded as described in Section 2. The RD curves in Fig.2 report the overall results, including the map coding cost which is always negligible (from 0.003 to 0.013 bpp). The 64×64 case seems to be a divide between the negligible loss of larger tiles and the significant loss of 32×32 tiles and smaller. This clearly indicates that region-based coding of small regions should be used with great care.

To begin investigating the effect of region shapes, we now consider rectangular tiles of fixed size (4096 pixels) but different aspect ratios, from 64×64 to 512×8. The RD curves, reported in Fig.3 together with the reference full-image curve, show that shape does matter when very short segments are considered. Even so, 256×16 tiles are still acceptable, much better than the 32×32 tiles seen before.

For these two experiments, some synthetic parameters are reported in Tab.1. For such regular tiles, the cost of map coding (MC) is always negligible. The coding gain (CG) decreases with decreasing size and increasing aspect ratio. The P/T ratio is always very large (the 1024 vs 256 difference is relevant only at very low bit rates) implying that large zero-trees are often discarded at once and SPIHT is rather efficient. Finally, the SNR loss w.r.t. the full image case at 0.3 bpp (L@0.3), gives a synthetic measure of the overall loss, and behaves coherently with the other meausres, but for a larger-than-expected sensitivity to the tile size.

Let us abandon now rectangular tiles, to consider instead tiles obtained from a 64×64 square by modeling the boundaries as sine-waves with adjustable amplitudes and number of cycles/tile. In Fig.4 we report some the RD curves obtained with 2 cycles/tile and various amplitudes from 4 to 16 pixels. Even when a low amplitude is considered we observe a significant loss (almost 2 dB at 0.3 bpp) with respect to the same-size square tile case, which further increases with increasing amplitude. This is only marginally due to the map coding cost (from 0.038 to 0.060 bpp) and to the coding gain, which remains in the 30-40 range, while it is probably due the the reduced SPIHT efficiency (P/T always
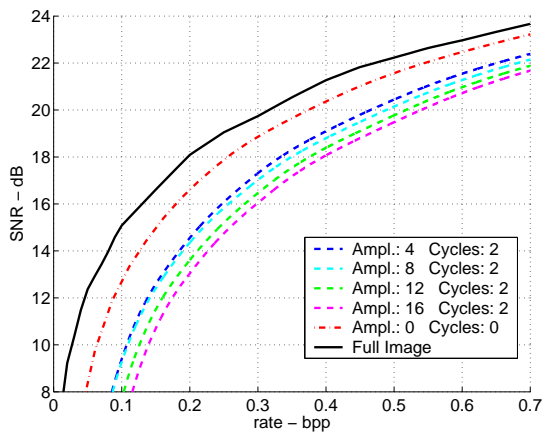
description of such edges, while the encoding of each region's texture becomes less expensive. Of course, the actual success of such an approach depends on many factors, such as the edge characteristics, the accuracy of segmentation, and the region textures, to mention the most obvious. To measure this gain, we build some mosaic images in which neighboring tiles are extracted from different original images, thereby simulating an image composed by regions with markedly varying statistics (see Fig.1(b) for an example). The flat encoding of such mosaics will be very challenging, with a sure performance loss with respect to the coding of the original images. Part of such loss could be recovered by region-based coding, and this will therefore be taken as the measure of the region-homogeneity gain.

4. EXPERIMENTAL RESULTS

All experiments are conducted on the well-known 512×512 8-bit grayscale test images of the USC database. The exten-

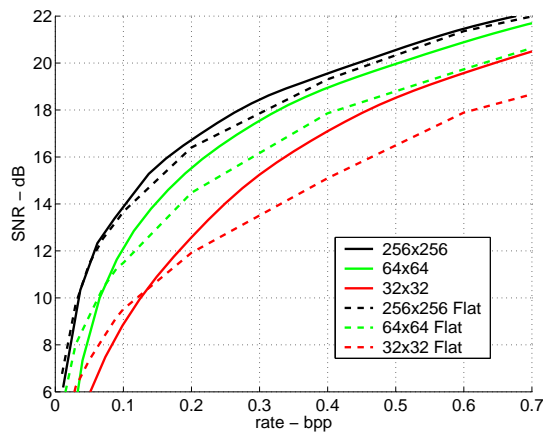**Fig. 4**. Rate-distortion curves for wavy tiles

Fig. 5. Rate-distortion curves for mosaic image

less than 100) due to the many "orphan" coefficients in high-frequency subbands that have to be scanned individually at each pass.

We complete this introductory analysis by considering the image of Fig.1(a), composed by half Lena and half Lake. This image is first scrambled in order to obtain a mosaic such as that shown in Fig.1(b) and then encoded, both by the flat (single tile) and region-based coders. In this case, flat coding is at a definite disadvantage because the ordinary WT produces large coefficients near the tile borders which will subtract significant resources to the encoding of region textures. Region-based coding does not suffer from this inconvenience. Note also that, in this experiment, the tile textures are not homogeneous anyway, and hence region-based coding does not take full advantage from the segmentation.

For lack of space, we report in Fig.5 only some results for square tiles, which however provide already clear indications, confirmed with the obvious differences in the other cases. When large tiles are used, flat and region-based coding are almost equivalent. With 64×64 tiles the situation is already different, the mosaic is quite busy and the flat coder shows a 2dB loss on the whole range of rates w.r.t. to the original image, while this loss is much smaller, from 0.5 to 1dB, when using the region-based coder. Going to 32×32 tiles the situation is similar, only worse, speaking even more in favor of region-based coding.

The synthetic measures show the main difference here to be in the coding gain. For 32×32 tiles, for example, CG equals 15.28 with ordinary WT while it raises to 27.17 with SA-WT, since in this case a large number of large coefficients do not appear anymore in the transform.

5. CONCLUSIONS

Many more experiments are necessary to assess wavelet-based region-based coding, and to gain insight about its

rate-distortion potential. However, these preliminary results indicate that region-based and flat coding have a very close performance when applied to large regions. When small regions are considered, a significant "geometrical" loss must be expected for region-based coding but if such regions do present significantly different statistics this can be the only way to obtain a good performance when ordinary waveform encoding is doomed to fail.

6. REFERENCES

[1] M. Kunt, M. Bernardi, and R. Leonardi, "Recent results in high-compression image coding," *IEEE Trans. Circuits Syst.*, pp. 1306–1336, 1987.

[2] ISO/IEC JTC1, *ISO/IEC 14496-2: Coding of audio-visual objects*, Apr. 2001.

[3] G. Gelli and G. Poggi, "Compression of multispectral images by spectral classification and transform coding," *IEEE Trans. Image Processing*, vol. 8, pp. 476–489, Apr. 1999.

[4] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Processing*, vol. 9, pp. 1158–1170, 2000.

[5] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 725–743, Aug. 2000.

[6] M. Cagnazzo, G. Poggi, L. Verdoliva, and A. Zinicola, "Region-oriented compression of multispectral images by shape-adaptive wavelet transform and SPIHT," in *Proceed. of IEEE Intern. Conf. on Image Proc.*, Singapore, Oct. 2004, pp. 2459–2462.

[7] V. Ratnaker, "RAPP: lossless image compression with runs of adaptive pixel patterns," in *Proc. of 32nd Asilomar Conf. Signals, System and Computers*, 1998.