DECIMATED WAVELET REPRESENTATION OF IMAGES - APPLICATION TO COMPRESSION

Riccardo Leonardi, Andrea Mazzarri and Alberto Signoroni

Signals & Communications Lab., Dept. of Electronics for Automation
University of Brescia, I-25123, Italy
E-mail: leon@bsing.ing.unibs.it

ABSTRACT

In this work a new way to improve the representation of images using a discrete wavelet transform for coding purposes is presented. The idea lies in combining all wavelet coefficients related to detail information at a same resolution level but along different orientations (horizontal, vertical, and diagonal), into a single image. Given that detail information is located for all subband images in the neighborhood of high frequency textures or edge locations, the pattern of significant coefficients remains unchanged after the combination process. This process allows further to reduce the number of transformed coefficients by 2/3, while preserving the multiresolution structure. This information can thus be efficiently coded using a multiresolution embedded coding scheme, such as Shapiro's zerotree coder. Overall, a higher coding efficiency can be reached while preserving the cross-scale prediction of significance among coefficients. Ultimately, approximate detail information must be recovered from the combined and coded data for each subband of the original wavelet, so as to reconstruct a decoded image.

1. INTRODUCTION

The introduction of the wavelet representation has arisen much interest to the image coding community. This fact is due to the capability of the wavelet transform to achieve both a frequency compaction and a spatial localization of the transformed coefficients. Following the initial image subband coding scheme [3], the first examples of image coders based on wavelet transform used only the energy compaction capability, achieving compression through an adequate scanning and a proper bit allocation of the different subband coefficients [4].

Second generation wavelet coders take into account the second feature as well, i.e. the spatial localization of the transformed coefficients, and the consistent occurrence of high energy coefficients across the multiresolution representation of detail information that is provided by the wavelet transform. Prediction of significance or non-significance across scales has therefore been exploited to further increase the coding efficiency [2, 1, 5]. This approach makes still use of the capability of wavelets to compact the coefficients energy by distributing them over a set of different frequency subbands, but also considers the fact that there is always a spatial localization of the most meaningful coefficients in each subband, and this localization is often repeated across different levels of resolution. These relationships between transformed coefficients, if efficiently exploited, can give rise to coding algorithms with significantly higher performance with respect to the first subband coders generation. Perhaps the simplest example of such a coder is Shapiro's embedded algorithm [1]: the zerotree structure is an efficient way to exploit the self-similarity across subbands and the decreasing spectrum model for the wavelet transformation.

The characteristics of a wavelet transform which has never been considered so far, for coding purposes, is to efficiently exploit the spatial orientation or directionality of patterns that occur in natural images. For instance, for synthetic images which are all composed by objects with vertical, horizontal and diagonal edges (i.e. the three preferential filtering directions in a 2-D separable wavelet transform), the associated wavelet transform shows a coefficient energy distribution with the following characteristic: higher coefficients in subbands other than the lowest frequency one (LL subband) are localized in correspondence to edges in the original image. Besides, and this is perhaps the most interesting feature, when attention is placed at a precise resolution level of detail information (corresponding to high frequency subbands), it can be noticed that regions with most significant coefficients corresponding to horizontal and vertical orientations rarely overlap. The fact that the edge regions in the original image are well separated reflects into the separation of the meaningful regions of transformed coefficients in the three subbands corresponding to each orientation.

The fact that there is almost no overlap in the meaningful sets of coefficients at a given resolution level can be used to obtain better coding performance. The first idea is that it is possible to consider the eventually non linear superposition of the three subbands at a given resolution level, during the coding procedure, instead of the original subband information. A simple but effective way must then be found to reconstruct in an approximate but reliable way the original subband information during the decoding stage, by a local pattern analysis of the decoded information.

The corresponding separation process may be eased with the use of a small coding overhead. This however remains significantly balanced by the reduction in the number of coefficients to be scanned during the coding stage.

The paper is organized as follows: Section 2 provides ad-

This work was supported by the Italian Ministery of the University and of Scientific and Technological Research (MURST)

ditional motivation to use a combination of wavelet detail information prior to any coding phase. Section 3 describes in detail the proposed combination-separation technique, while section 4 indicates some preliminary compression results, before some concluding remarks are drawn in section 5.

2. MOTIVATION

If all information corresponding to the same level of detail but with different orientation is combined, only one third of the original discrete multiresolution representation must be quantized, with a substantial reduction in the number of symbols that need subsequently to be entropy encoded. Considering a coding environment like the zerotree coder [1], this fact is of great relevance, because the adoption of the described approach improves the effectiveness of the prediction of the non-significance condition: in fact, three zerotree symbols in the three different subbands with the same spatial location become only one zerotree symbol (with the same associated quantization information) after the combination of the various subbands (see Fig. 2). Initial experimental simulations have shown that for distortions in the range of 36 dB (PSNR), a reduction of roughly 30% of the bit-stream to code the lenna image can be expected (see section 4).

If any given edge in an image results in a concentration of significance in the corresponding neighborhood of each detail subband, the concentration of energy for each orientation subband will depend upon the correspondence between the edge orientation and the subband orientation, and upon the type of wavelet transform. If all data corresponding to a given level of detail are combined into a single image it may be difficult to recover the original data from the superposition (which we may initially consider a simple linear combination of wavelet coefficients, i.e. a projection operation) of the subbands, due to the spatial overlapping of the meaningful regions for edge reconstruction. Obviously if the original image contains only horizontal and vertical edges, only the corresponding orientations subbands will contain significant wavelet coefficients, so that even if combined the orientation dependent information may be recovered with minimum loss.

Hence, without considering coding, in the scheme proposed there is an inherent loss in the quality of the image reconstruction. However, the key idea is that it is possible to achieve better coding results (in a rate-distortion sense) than a standard wavelet coder (in our case, a zerotree coder), if we consider the coding operation combined with the wavelet data reduction operation described above, thanks to the lowering in the number of transformed coefficients to scan and quantize. In the decoding stage, the fundamental operation corresponds to the inversion of the subband combination of detail information.

To show that this can be achieved (without using any additional information) we first propose to use a non linear filtering of the linear combination subband image (simply considered as the sum of the three differently oriented subband images for eagc resolution level). More precisely, the sum image is filtered morphologically with a rectangular structuring element (one by six samples), first horizontally and then vertically oriented; the samples are classified as belonging to vertical or horizontal orientations if there is at least a certain percentage of greater samples than a given threshold in the subband domain covered by the current position of the element. If it is the case that a coefficient is assigned to both the horizontal and vertical subband, its energy is equally distributed among them. From the sum subband the diagonal subband is derived by simply subtracting the other two ones and weighting the result with an appropriate coefficient. As noticed above, for real images this procedure cannot allow a perfect reconstruction (even in the absence of quantization and coding), because the condition of separation of the meaningful edge regions in the transformed domain is not very effective. The complete wavelet decomposition with combination of the different orientation subbands and decomposition using the morphological filters as described before is shown in Fig. 1. While horizontal and vertical edge regions are well preserved, artefacts appear along diagonal edges. Accordingly, the next section provides for a much more robust scheme for combination and separation of the different subband images.

3. THE DECIMATED WAVELET REPRESENTATION

Let us consider the problem of how to efficiently reduce the data set represented by the wavelet decomposition of a generic gray level image. We consider each level of decomposition separately, and the procedure described in the following is applied in the same manner at each resolution in the transformed domain.

At a given resolution level N , we have the three subbands HH-N or $w_d^{(N)}$, HL-N or $w_v^{(N)}$, LH-N or $w_h^{(N)}$, (except at the lowest resolution); the HH band contains the high pass information, approximately associated to 45 degrees oriented edges (diagonal edges), while the LH and HL bands contain the needed information to reconstruct horizontal and vertical edges. To reduce the data set composed by these subbands, we can apply a linear transformation to the subbands that tries to maintain the supposed spatial separation of the most meaningful coefficients regions. The simplest combination is obtained by the sum of the wavelet coefficients that occur at the same spatial locations in the three subbands. Where there is no significance in the subband coefficients (i.e. they are lower than a certain threshold), no significance is present in the sum, and where there is "separate significance" (i.e. the coefficient is over the threshold only in one subband), the result of the sum contains virtually only the coefficient of one of the three subbands.

Unfortunately, even when located in correspondence to an edge the degree of significance separation in a subband of a given orientation depends on the edge direction. In particular for certain angular values, significant coefficients may appear for all three orientations, i.e. horizontal, vertical and diagonal subband images. Moreover, coefficients at the same spatial location but from different subbands may be opposite in sign, so that a simple sum of the three subbands may become insignificant at this location. What one can ensure however is the co-spatial location of significance along the pattern of the edge, across scales, and within subbands of different orientations. A combination that preserves the significance pattern can be obtained instead by an absolute summation of the three subbands:

$$w_s^{(N)} = |w_v^{(N)}| + |w_h^{(N)}| + |w_d^{(N)}| \tag{1}$$

where $w_s^{(N)}$ is the combined image, that we may also call because of the reduction in the number of coefficients, the decimated wavelet image.

The problem lies now in defining a proper decomposition procedure to recover the three subband images. Given a series of sample images containing directional edge elements, we propose to define for every level N of the decomposition three sets of coefficients, $c_d^{(N)}(\theta)$, $c_h^{(N)}(\theta)$, and $c_v^{(N)}(\theta)$, such that

$$c_i^{(N)}(\theta) = w_i^{(N)}(\theta)/w_s^{(N)}(\theta) \quad i = d, h, v \tag{2}$$

where θ corresponds to the edge orientation. These coefficients are non-zero only in the vicinity of the edge element. Additional properties which may be efficiently used are:

- they are highly structured (i.e., they show some periodicity along the edge orientation),
- they depend on the type of wavelet transform,
- they do not depend on the edge amplitude, nor significantly on the edge profile.

When dealing with an arbitrary image a local analysis of the w_s image provides for an estimation of local gradient direction, thus allowing to select the proper set of coefficient patterns to estimate each individual orientation subband, by using eq. (2). Clearly, if the pattern in the w_s image does not correspond to an edge element, the reconstruction procedure will fail (i.e. it leads to a wrong estimate of the different orientation subband coefficients at the corresponding location). We claim however that this can be perceptually tolerated because of the masking effect provided by high frequency unstructured texture areas where this may happen.

The decomposition procedure to reconstruct from a given decimated wavelet image the three orientation-dependent subband images for any given scale N, can be performed by any decoder provided that the structure of the decimated wavelet image be preserved. The decoder will simply use the a-priori coefficient patterns $c_i^{(N)}(\theta)$.

4. CODING PERFORMANCE

At present no complete coding procedure has been implemented, because of the complexity of the decomposition process. Also, because of the inherent loss that this process introduces improvement in performance can be expected only at low rates (high distortion), unless side information is transmitted to the decoder for correcting the initial estimates given by the decomposition process.

Therefore we can only provide at present for preliminary estimates of the compression performance. For this purpose, we are showing in Table 1 the compression results for the lenna image by adjusting the level of significance of coefficients using a zerotree codec (which affects the number of dominant passes of Shapiro's coder).

PSNR	# smb.	# kbits	# smb.	# kbits	Gain
36.7 dB	210k	218	178k	109	50%
33.6 dB	108k	103	63k	47	54%
30.6 dB	53k	50	49k	35	30%

Table 1: Comparison of coded streams for the lenna image generated by Shapiro's zerotree (col. 1-3) and the Decimated WT coder (col. 4-5).

Improvements in compression performance can be expected because of two reasons: a) the reduction in the number of coefficients to encode (by a factor of 2/3); b) the fact that after the absolute summation of the wavelet images, the resulting decimated wavelet image will only contain positive terms, eliminating one symbol from Shapiro's original alphabet (the one defining the sign of the significant coefficient used in each dominant pass). Clearly the results shown here give only an indication of the expected performance. Only once the effects of the lossy decomposition procedure are taken into account can one figure out the exact benefit of the proposed combination process. Nevertheless, significant gains can be expected as the decimated wavelet representation allows to compress the bit-stream from 30% to 50% depending on the distortion level.

5. CONCLUDING REMARKS

In this work, we have introduced the concept of combining wavelet coefficients corresponding to a same level of detail but along different spatial orientations while maintaining the significance patterns across scale and space. Though the combination process is lossy, a relatively good estimate of each orientation subband is possible by analyzing the orientation of patterns of the decimated wavelet. Rhis approach brings a substantial coding gain, both in terms of number of coefficients to be coded and in terms of the number of symbols needed to represent the pattern of coefficient values. Unfortunately, it is not obvious to estimate the real advantage in terms of performance, given the complexity of the decomposition stage that must be applied to reconstruct the different orientation subbands. We designed to strategies to perform this decomposition: A first one is based on the use of morphological horizontal and vertical structuring elements; a second one is based on a local analysis of orientation of patterns of the transmitted decimated wavelet information.

Research is being conducted to tackle an efficient representation and normalization strategy of the decimated wavelet using local gradient direction. Further studies should include also scalable techniques to allow to reconstruct the original image at any level of predetermined distortion, starting from an initial guess provided by the decimated wavelet representation.

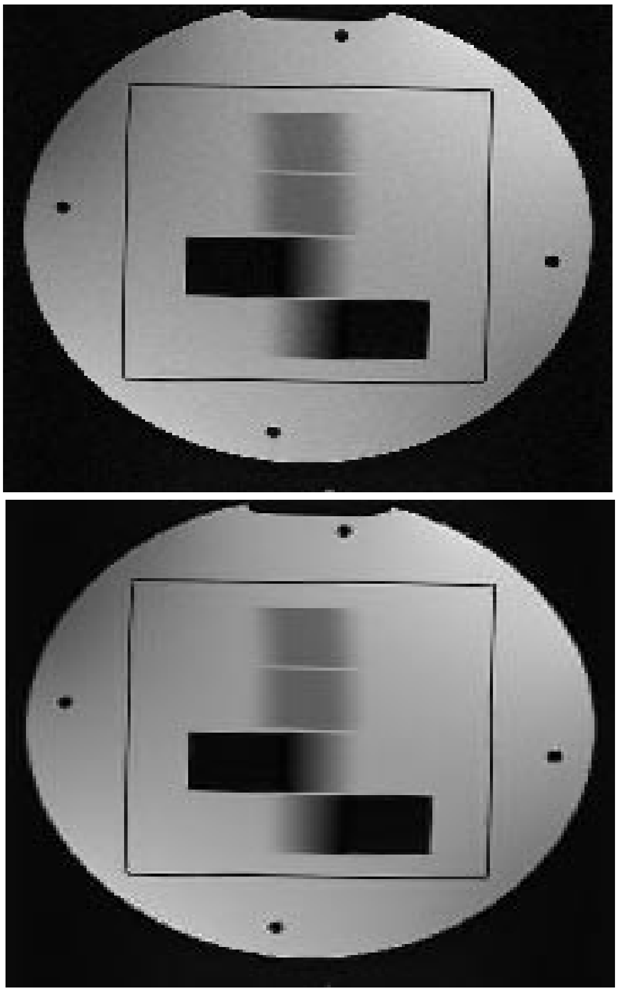

Figure 1: Sample image coded using the decimated wavelet representation and a morphological structuring element for decomposition (original-coded).

6. REFERENCES

[1] J.M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Signal Proc.*, SP-41(12):3445-3462, Dec. 1993.

[2] R. Rinaldo and G. Calvagno, "Image Coding by Block Prediction of Multiresolution Subimages", *IEEE Trans. on Image Proc.*, IP-4(7):909-920, Jul. 1995.

[3] J.W. Woods and S.D. O'Neil, "Sub-band Coding of Images", *IEEE Trans. on Acoust., Speech, and Signal Proc.*, ASSP-34(5):1278-1288, May 1986.

[4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform", *IEEE Trans. on Image Proc.*, IP-1(2):205-220, Apr. 1992.

[5] S.D. Servetto, K. Ramachandran, and M.T. Orchard, "Wavelet Based Image Coding Via Morphological Prediction of Significance", *Proc. of the 1995 ICIP Conference*, I:530-533, Oct. 1995.

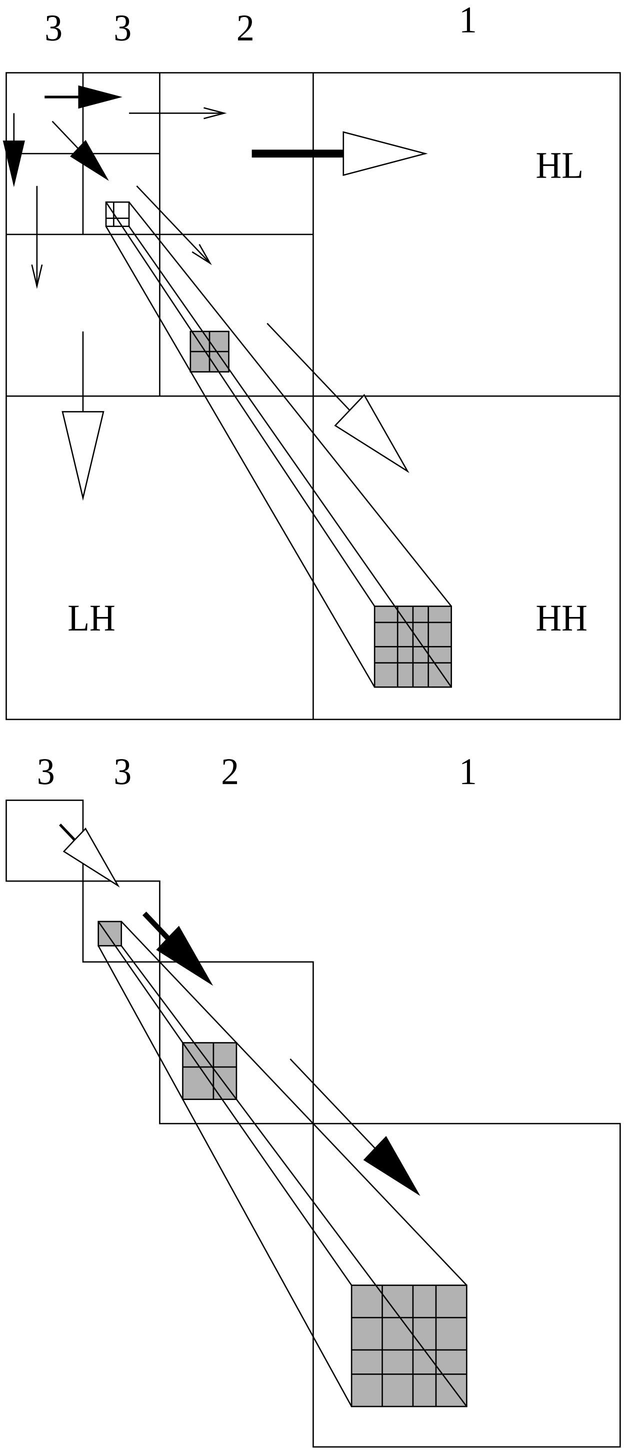

Figure 2: Prediction of non significance with zerotree: a) Traditional; b) Decimated Wavelet.