SPARSE OPTIMIZATION WITH DIRECTIONAL DCT BASES FOR IMAGE COMPRESSION

A. Drémeau, C. Herzet, C. Guillemot and J.-J. Fuchs

INRIA Centre Rennes - Bretagne Atlantique, Campus universitaire de Beaulieu, 35000 Rennes, France

ABSTRACT

This paper proposes a new compression algorithm based on the directional DCT (DDCT) bases introduced in [1]. We first explain how to extend the DDCT concept to rectangular bases and exploit them to build a set of bases using a bintree segmentation. We then use dynamic programming to select a basis from this set according to a rate-distortion criterion. Comparisons in terms of rate-distortion performance are finally made with the current compression standards JPEG and JPEG2000.

Index Terms— Sparsity, Optimization, Anisotropic Directional DCT basis.

1. INTRODUCTION

Recent results in image compression tend to show that adaptation of the transform to the local characteristics of the image allows a gain in performance. In practice, the optimization of the transform to the image characteristics can mainly be made at two levels: *i)* in the spatial domain by adapting the support of the transform; *ii)* in the transformed domain by adapting the atoms of the projection basis to the signal characteristics we want to describe.

Several contributions considering these general approaches can be found in the literature. In [2], Chen suggests to use the DCT on blocks of variable size. In [3], Meyer compares different lapped DCT transforms. At the same time, a number of new transform bases have emerged. In [4], the authors replace the DCT in the JPEG scheme by other transforms, better adapted to the local statistics of each block. In a similar way, Sezer *et al.* optimize a library of bases on a training set to maximize the sparsity of the transformed vectors (see [5]).We can likewise refer to the well-known wavelets [6], curvelets [7], contourlets [8] and bandelets [9], which are very effective at describing the edges of an image. Finally, we find in [10] a hybrid approach, where the authors optimize both the size of the blocks and the direction of a bandelet basis.

In this paper, we study a different hybrid approach: on the one hand, we consider the library of directional DCT (DDCT) bases developed by Zeng *et al.* in [1]; on the other hand, we use a bintree block segmentation, which allows a better flexibility in the adaptation of the basis to the local properties of the image.

2. ANISOTROPIC DIRECTIONAL DCT

2.1. Low-bit rate compression and sparsity

We consider a compression scheme using an orthonormal basis $\mathcal{B} = \{b_k\}_{k=1}^N$ and a quantizer $\mathcal{Q}$. An image block f of N pixels is then approximated by $\hat{f}$ as

$$\hat{f} = \sum_{k=1}^N \mathcal{Q}(\langle f, b_k \rangle) b_k \qquad (1)$$

where $\langle f, b_k \rangle$ denotes the scalar product between vectors f and b_k.

In that context, Mallat and Falzon show in [11], that the number of nonzero quantized transform coefficients (*i.e.,* scalar products), say M, plays a crucial role in the characterization of the rate-distortion performance at low bit rates. On the one hand, they emphasize that the distortion D can be linked to the number of nonzero coefficients M via a function $\varphi(M)$ which modelizes the ability of the basis to approximate a signal with M projection coefficients, *i.e.,*

$$D = \varphi(M). \qquad (2)$$

On the other hand, they show that the rate required to represent the quantized coefficients is proportional to M, *i.e.,*

$$R = \gamma M. \qquad (3)$$

As a consequence, Mallat and Falzon's work leads to the following observation: at low bit rates, the rate-distortion performance depends on the ability of the basis to provide a good approximation of the signal with few coefficients.

In practice, different bases are often well-suited to different kinds of images. This suggests that enhanced rate-distortion performance could be achieved by adapting the projection basis to the image to compress.

In this paper, we propose to test the performance achieved with a set of bases constructed from the concatenation of local DDCT bases. We explain the construction of this set of bases in the remainder of this section.

2.2. Directional DCT bases

In [1], the authors introduce *square* directional DCT (DDCT) bases able to provide better coding performance for image blocks that contain directional edges. In this section, we extend this principle to rectangular bases.

The key idea of DDCT relies on the *separability* property of the standard 2D-DCT transform *i.e.,* a 2D-DCT can be realized by performing successively a 1D-DCT on the columns and a 1D-DCT on the rows of the considered image block. The construction of the directional DCT is based on the modification of the scan order of the block pixels to create bases with some "selected" directions.

For the sake of clarity, we expose the construction of the DDCT transform for a block of 8×4 pixels and the "diagonal down-left" directional mode represented in Fig. 1. The extension to other directional modes is straightforward.

To any directional mode we associate a scan order defined by a set of vectors $\{v_k\}_k$. Each vector v_k contains a subset of pixels taken according to a given direction. For the "diagonal down-left" mode illustrated in Fig. 1, the v_k's contain the pixels located on a same arrow. They have therefore different lengths.
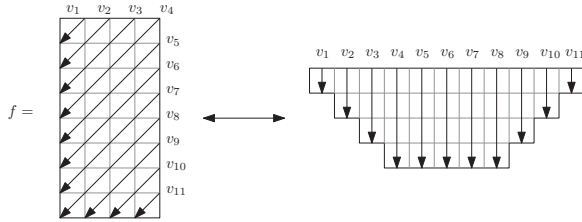

Fig. 1. Pixels ordering for the "diagonal down-left" mode on a rectangular block

Pixels can be reordered as illustrated in Fig. 1, resulting in a "pyramidal" pixel array, whose columns are the vectors v_k's. Considering this array, the process is then similar to the conventional 2D-DCT. A first DCT is primarily performed on the columns v_k's; a second DCT is then performed on the rows of the resulting "pyramidal" matrix, as illustrated in Fig. 2.
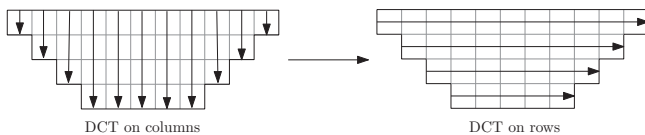

Fig. 2. "Diagonal down-left" DCT transform on a rectangular block

The transform block $\tilde{f}$ is finally obtained by rearranging the coefficients into their initial positions in f.

2.3. "Bintree" concatenation of local DDCT bases

A common way to form an image basis from local bases is to use a quadtree. This method leads to a set of bases which are the concatenation of square local bases whose support sizes can possibly be different (see [10] for example). In this paper, we want to exploit anisotropic (*rectangular*) local basis supports. This is achieved by the use of a bintree.

With this additional degree of freedom, a bintree segmentation increases the number of possible image bases. On the one hand, we can intuitively assume that the more bases we have at our disposal, the more likely we are to select a basis which properly catches the local properties of the image with few coefficients. On the other hand, this leads to an increase of the bit rate since the choice of the selected basis has to be transmitted.

Therefore, in addition to the bit rate required to code the transform coefficients (expressed in (3) and noted R_c here), we have to take into account the encoding cost of the image basis, *i.e.,* the cost associated to the encoding of the tree specifying the local basis supports (noted R_s), and the cost associated to the encoding of the local directional modes (noted R_m). The total bit rate can then be expressed as:

$$R = R_c + R_s + R_m. \tag{4}$$

In the next section we will discuss how to select the basis which leads to the best compromise in terms of rate-distortion. The use of a bintree instead of a quadtree increases R_s but we could observe experimentally that this is compensated by a better local adaptivity, resulting in a gain in performance.

3. DDCT BASED IMAGE COMPRESSION

3.1. Basis selection

In the transform coding paradigm, two parameters can impact the rate-distortion performance: the transform basis and the quantization of the transform coefficients. Therefore, the rate-distortion optimization problem can be formalized as follows:

$$(\mathcal{B}^\star, \mathcal{Q}^\star) = \arg \min_{\mathcal{B}, \mathcal{Q}} \ D(\mathcal{B}, \mathcal{Q})$$
$$subject \ to \quad R(\mathcal{B}, \mathcal{Q}) \leq R_t, \tag{5}$$

where $\mathcal{B}$ (resp. $\mathcal{Q}$) is a trial image basis (resp. quantizer) and R_t is a target rate we specify as a constraint.

Solving (5) is usually intractable. Instead, Shoham and Gersho showed in [12] that, under some conditions, this problem is equivalent to a simplified unconstrained problem depending on a Lagrangian multiplier λ:

$$(\mathcal{B}^\star_\lambda, \mathcal{Q}^\star_\lambda) = \arg \min_{\mathcal{B}, \mathcal{Q}} \ D(\mathcal{B}, \mathcal{Q}) + \lambda R(\mathcal{B}, \mathcal{Q}), \tag{6}$$

In our compression scheme, we consider a scalar uniform quantizer with a quantization step Δ and a deadzone equal to
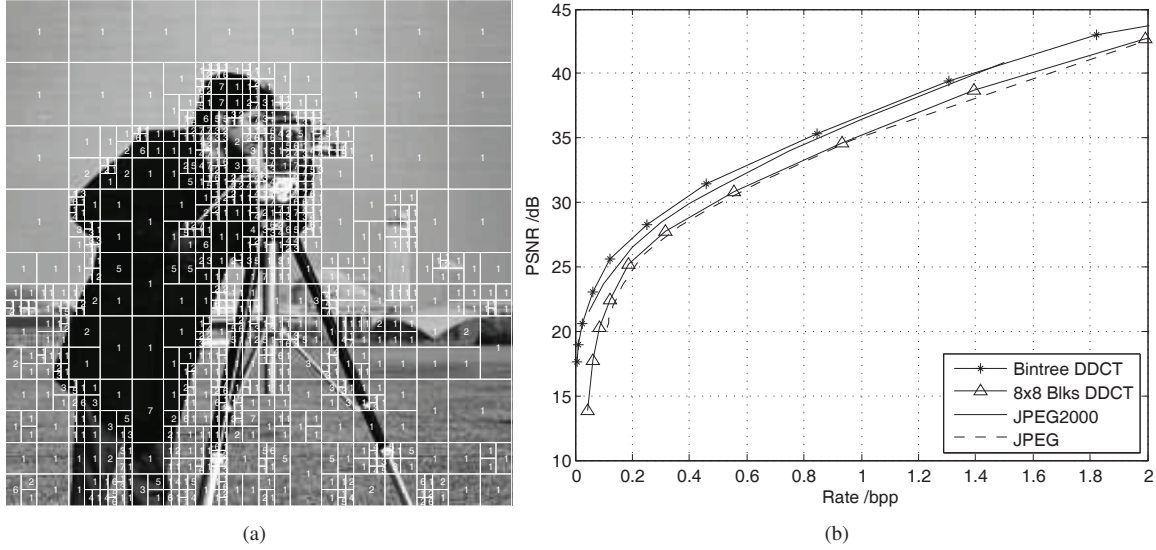
Fig. 3. (a) Bintree DDCT segmentation obtained for "Cameraman" at $R = 0.46$ bpp and PSNR= 31.4 dB. (b) Rate-distortion curves for the compression of "Cameraman" with the proposed coder, the DDCT applied on 8×8 blocks, the JPEG2000 standard and the JPEG standard

2Δ. Optimizing $\mathcal{Q}$ is then equivalent to optimizing Δ. In this context, Le Pennec and Mallat proved in [10] that, as long as assumption (3) is valid, the optimal quantization step, say $\Delta_\lambda^\star$, is related to λ as follows:

$$\Delta_\lambda^\star = \sqrt{\frac{4\,\gamma\,\lambda}{3}}. \tag{7}$$

With this result, problem (6) reduces to optimizing the rate-distortion function with regard to basis $\mathcal{B}$, *i.e.,*

$$\mathcal{B}_\lambda^\star = \arg\min_{\mathcal{B}} \; D(\mathcal{B}, \Delta_\lambda^\star) + \lambda\, R(\mathcal{B}, \Delta_\lambda^\star). \tag{8}$$

The solution of this problem can be found efficiently using dynamic programming [13] provided that: *i)* the set of image bases is the "tree" concatenation of local bases; *ii)* the distortion and the rate can be decomposed into local terms associated to each local basis. The first condition is obviously satisfied by our set of image bases, see section 2.3. The second one has to be verified for the "global" terms R_c, R_s and R_m.

Using assumption (3), the first term R_c obviously satisfies the condition. For the DDCT basis, we estimate[1] $R_c = 6.5\,M$ as a good approximation of the rate required to encode the transform coefficients. For the two following terms R_s and R_m, some additional coding assumptions are required: *i)* we adopt a simple implementation of the bintree by assigning "1"

---

[1]The value of γ actually depends on the practical coding implementation, see section 3.2.

to internal nodes and "0" to leaf nodes, this satisfies the condition for R_s; *ii)* we assume that the choice of the local directional mode is encoded by a fixed-length code (FLC) allowing to express R_m as a sum of local terms. In the next section, we will describe a more efficient way to encode the directional modes. Here, the resulting R_m can therefore be seen as an upper bound on the rate which can actually be achieved by our practical scheme.

3.2. Implementation and results

In this section, we detail the implementation of our image compression codec and illustrate its performance.

We consider a dictionary of image bases constructed as explained in section 2. The supports of the local bases range from 32×32 to 4×4 pixels. For each size of support, we consider 7 directional modes which correspond to the prediction modes of H.264, see [14]: mode "1" corresponds the conventional DCT; modes "2" to "7" correspond to "oriented" DCTs. The transform basis is selected with the optimization procedure described in section 3.1.

The quantized transform coefficients are encoded with Huffman codes. The Huffman tables are optimized according to the size of the support of the local transforms. The indices of the M nonzero projection coefficients are encoded with a run-length encoder.

The encoding of the directional modes is performed by means of a quadtree proceeding as follows. The image is segmented into 4 square blocks of equal dimension. If all the local bases in a block have the same directional mode, this block corre-

Rate (bpp)	Lena (512×512 pixels)			Barbara (512×512 pixels)			Roof (512×512 pixels)		
	Coder (dB)	JPEG2000 (dB)	JPEG (dB)	Coder (dB)	JPEG2000 (dB)	JPEG (dB)	Coder (dB)	JPEG2000 (dB)	JPEG (dB)
0.1	29.48	29.90	-	25.31	24.80	-	24.63	23.50	18.75
0.2	32.65	33.00	29.88	28.14	27.30	24.03	27.53	26.50	22.79
0.5	36.76	37.30	35.44	32.69	32.20	29.75	33.25	31.70	29.59
0.7	38.36	38.66	36.82	34.91	34.28	32.30	35.98	34.18	32.57
1.0	39.70	40.40	38.54	37.29	37.10	35.18	39.03	37.60	35.94
1.5	41.83	42.80	-	40.38	40.40	38.24	42.95	42.30	39.93

Table 1. Summary table of the rate-distortion performance for the compression of 3 different images with the proposed coder and the standards JPEG and JPEG2000.

sponds to a leaf of the quadtree and is labeled by the common directional mode of the local bases in the block; otherwise the block is subdivided into 4 square blocks and so one.

In Fig. 3, we illustrate the performance achieved by our codec for the compression of "Cameraman". Fig. 3(a) represents the supports of the local bases making up the image basis as well as their directional modes for $\lambda = 100$. We can notice that conventional DCT (mode "1") is selected in smooth areas whereas DDCT transforms are used in areas where directionality is more important (*e.g.*, oriented edges).

Fig. 3(b) compares the rate-distortion performance obtained by our compression scheme to the one obtained by the DDCT applied on 8×8 pixels blocks and the compression standards JPEG and JPEG2000. We can notice that the proposed codec outperforms JPEG by more than 1 dB at low bit rates. As shown in Table 1, this observation holds for all the images we have tested. As far as "Cameraman" is concerned, we can also observe that the proposed codec slightly outperforms JPEG2000 at low bit rates. Additional results in Table 1 show that the DDCT-based codec usually exhibits a good behavior with regard to JPEG2000.

4. CONCLUSION

In this paper, we studied the performance of a "basis-adaptive" image compression codec. The set of bases used in the codec is built by the concatenation of local anisotropic DDCT bases. The selection of the optimal basis is made by exploiting the bintree structure of the basis dictionary and using dynamic programming.

As far as the images tested are concerned, the proposed coder outperforms JPEG in terms of rate-distortion and is slightly superior to JPEG2000 in most cases.

5. REFERENCES

[1] B. Zeng and J. Fu, "Directional discrete cosine transforms - a new framework for image coding," *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 305–313, March 2008.

[2] C.-T. Chen, "Adaptive transform coding via quadtree-based variable blocksize dct," in *Proc. IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 23-26 May 1989.

[3] F. G. Meyer, "Image compression with adaptive local cosines: A comparative study," *IEEE Trans. On Image Processing*, vol. 11, no. 6, pp. 616–629, June 2002.

[4] M. Helsingius, P. Kuosmanen, and J. Astola, "Image compression using multiple transforms," *Signal Processing*, vol. 15, pp. 513–529, 2000.

[5] O. G. Sezer, O. Harmanci, and O. G. Guleryuz, "Sparse orthonormal transforms for image compression," in *Proc. IEEE Int'l Conference on Image Processing (ICIP), San Diego, CA.*, October 2008.

[6] S. Mallat, "A theory for multiresolution signal decomposition : the wavelet representation," *IEEE Trans. On Pattern Analysis Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[7] E. J. Cands and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," Tech. Rep., Stanford University CA - Dept of Statistics, 2000.

[8] M. N. Do and M. Vetterli, "Contourlets: a directional multiresolution image representation," in *Proc. IEEE Int'l Conference on Image Processing (ICIP).*, 2002, vol. 1, pp. 357–360.

[9] E. LePennec and S. Mallat, "Bandelet image approximation and compression," *SIAM MMS*, vol. 4, no. 3, pp. 992–1039, April 2005.

[10] E. LePennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. On Image Processing*, vol. 14, no. 4, pp. 423–438, April 2005.

[11] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. On Signal Processing*, vol. 46, no. 4, pp. 1027–1042, April 1998.

[12] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. On Acoustics, Speech and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, September 1988.

[13] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distorsion sense," *IEEE Trans. On Image Processing*, vol. 2, no. 2, pp. 160–175, April 1993.

[14] ITU-T Rec. H.264 ISO/IEC 14496-10 (AVC), *Advanced Video Coding for Generic Audiovisual Services*, March 2005.