NON-GEOMETRIC ENERGY FORMULATION FOR ADAPTIVE IMAGE COMPRESSION

Benjamin Le Guen[1,2], *Stéphane Pateux*[1], *Jacques Weiss*[2]

[1] France Télécom R&D, 4, rue du Clos Courtel, 35512 Cesson-Sévigné Cedex, France
[2] Supélec-SCEE/IETR-AC, avenue de la Boulaie, CS 47601, 35576 Cesson-Sévigné Cedex, France

ABSTRACT

This paper proposes a new adaptive scheme to cope with the limits of the conventional 2D wavelet. The principle is to search for the warped version of the original image which is best adapted to this isotropic kernel. Unlike most prior works, the scheme does not rely on geometry detection or exhaustive search for the best local directions. It formulates a *Description Length* of the warped Image which does not depend on any *a priori* about geometry. We demonstrate that its minimization can be compared to a motion estimation between two frames. The warped image can then be efficiently coded by any conventional wavelet coder. An application is presented where the warping is modelized by a regular 2D-mesh. At low rates, results show a significant reduction of ringing artefacts compared to JPEG2K.

***Index Terms*—** Adaptive Compression, Wavelet, Mesh

1. INTRODUCTION

Shortcomings of classical 2D-separable Wavelets have been put forward for more than a decade. This isotropic representation kernel cannot efficiently represent the geometrical structure of an image when this structure is neither horizontal nor vertical. This has a consequence on the visual quality of images approximated with a limited number of coefficients: features of high perceptual impact, mainly curved edges of objects, are corrupted by the appearance of a ringing effect in each direction. The motivation is then to add a touch of anisotropy to the representation.

Different approaches have been proposed to follow this goal. One approach is to project the image on a dictionary of *fixed* anisotropic atoms, such as Contourlets [1]. But the redundancy and the non-adaptivity of these techniques motivates the research on new *adaptive* basis. Adaptivity can be looked for in a variety of ways. We essentially distinguish between two kinds of approaches. On the one hand, adaptivity can be obtained by extracting from the signal a relevant geometrical content *beforehand*. Whether this extraction resorts on detection of contours [2] or regular curves [3], these techniques are based on geometrical *a priori*. This does not insure a fair modelization of energetical properties in the transform domain. On the other hand, recent approaches proposed to express an energy in the transform domain as a function of adaptivity parameters. Minimizing this energy leads to optimal parameters in a certain sense. For this purpose, most techniques [4, 5, 6] implement an exhaustive search. For complexity and compacity concerns, this imposes to segment the image into blocks and *independently* compute for each of these blocks a *limited* number of parameters. Hence, it leads to a discontinuous and low-level adaptivity. Note that some techniques, like [3, 7], propose regularization procedures to optimize the parameters in a Rate/Distorsion sense.

Fig. 1. Goal of the scheme: Pre-Warp the Image so as to adapt the resulting signal to any conventional wavelet coder. Warping parameters are transmitted independently.

This paper proposes a new energy formulation for the adaptivity issue. The principle is to search for the warped version $\tilde{I}$ of an image I which is best adapted to a conventional wavelet decomposition (see Fig. 1). A similar purpose is searched for in [2, 3]. The solution we propose is different because it is not based on any *a priori* about geometry. The second section describes our energy formulation considering any parametric warping transformation w. It demonstrates how the minimization process can be compared to a classical motion estimation between two frames. The third section outlines the specificities of an implementation where w is modelized by a 2D-regular mesh. Finally, the warping estimation layer is assembled to JPEG2K and results are shown.

2. ENERGY FORMULATION

2.1. Notations

Let us define a reversible transformation w that maps a position $\tilde{\mathbf{p}}$ in a Domain $\tilde{\mathcal{D}}(\subset \mathbb{Z}^2)$ - that we call the *Warped* Domain - to a position $\mathbf{p}$ in the *Image* Domain $\mathcal{D}(\subset \mathbb{Z}^2)$. Given w, the original Image I and the Warped Image $\tilde{I}$ are related as follows:

$$\tilde{I}(\tilde{\mathbf{p}}) = I(w(\tilde{\mathbf{p}})), I(\mathbf{p}) = \tilde{I}(w^{-1}(\mathbf{p})) \tag{1}$$

In this section, we will only assume w to be a parametric model composed of a set of N_p undefined parameters $\{\mathbf{p}_i\}_{i=1..N_p}$. In accordance with Fig. 1, the key issue of the approach is to optimize w such that the entropy of the resulting Warped Image $\tilde{I}$ in a conventional wavelet basis is minimized. Because geometry strongly contributes to the entropy of I, one can expect the computed w to have some relation with the image geometry.

We will now explain the construction of an energy function which depends on w.

2.2. Building the Energy as a Function of w

Let $\psi_{j\mathbf{n}}$ refer to a 2D discrete Wavelet scaled with a factor 2^{-j} ($j \in \mathbb{Z}_+$) and translated to a position $\mathbf{n} \in \mathbb{Z}^2$. The dot product of $\tilde{I}$ with $\psi_{j\mathbf{n}}$ gives a wavelet coefficient $c_{j\mathbf{n}}$.

Considering the set of wavelet coefficients $\{c_{j\mathbf{n}}\}_{j=0..J-1}$ for a user-fixed J, we define the *Description Length* DL_J as:

$$DL_J = \sum_{j=0}^{J-1} \gamma_j^2 \cdot \sum_{\mathbf{n}} c_{j\mathbf{n}}^2, \qquad (2)$$

where γ_j is a weight that can be adjusted in relation to the statistics of the coefficients at the scale 2^{-j}. This formulation is particularly well fitted to Gaussian distributions in the subbands. Indeed, if we assume that a coefficient $\{c_{j\mathbf{n}}\}$ - for a given j - follows a Gaussian probability law $P(c_{j\mathbf{n}}) \sim \mathcal{N}(0, 1/\gamma_j^2)$, then the *Description Length* of the transformed signal is $DL_J = -\sum_{j=0}^{J-1} log_2(P(c_{j\mathbf{n}}))$ which can be reduced as in Eq. (2).

To make the minimization of DL_J tractable, we need to express it as a function of w. Now, an inverse wavelet transform gives:

$$\tilde{I}(\tilde{\mathbf{p}}) = \sum_{j=0}^{j_{max}} \sum_{\mathbf{n}} c_{j\mathbf{n}} \cdot \psi^*_{j\mathbf{n}}(\tilde{\mathbf{p}}) \qquad (3)$$

where $\psi^*_{j\mathbf{n}}$ refers to the synthesis kernel corresponding to $\psi_{j\mathbf{n}}$, and j_{max} to the greatest possible decomposition level of $\tilde{I}$.

The right part of Eq. (3) can be decomposed to yield:

$$\sum_{j=0}^{j_0-1} \sum_{\mathbf{n}} c_{j\mathbf{n}} \cdot \psi^*_{j\mathbf{n}}(\tilde{\mathbf{p}}) = \tilde{I}(\tilde{\mathbf{p}}) - \sum_{j=j_0}^{j_{max}} \sum_{\mathbf{n}} c_{j\mathbf{n}} \cdot \psi^*_{j\mathbf{n}}(\tilde{\mathbf{p}}) \qquad (4)$$

for any $j_0 \in [0, j_{max}]$.

As the signal $\sum_{j=j_0}^{j_{max}} \sum_{\mathbf{n}} c_{j\mathbf{n}} \cdot \psi^*_{j\mathbf{n}}$ is the approximation of $\tilde{I}$ obtained by setting to 0 each coefficient $c_{j\mathbf{n}}$ for $j \in \{0..j_0-1\}$, we will refer to it as $\tilde{I}_{j_0}$. From Eq. (4), Parseval theorem gives:

$$\sum_{j=0}^{j_0-1} \sum_{\mathbf{n}} c_{j\mathbf{n}}^2 = \sum_{\tilde{\mathbf{p}}} (\tilde{I}(\tilde{\mathbf{p}}) - \tilde{I}_{j_0}(\tilde{\mathbf{p}}))^2 \qquad (5)$$

Therefore, it is possible to express the energy in (2) as:

$$DL_J(w) = \sum_{j_0=1}^{J} \eta_{j_0}^2 \cdot \sum_{\tilde{\mathbf{p}}} (I(w(\tilde{\mathbf{p}})) - \tilde{I}_{j_0}(\tilde{\mathbf{p}}))^2, \qquad (6)$$

with the new weights $\eta_{j_0}^2$ verifying the constraints:

$$\gamma_j^2 = \sum_{j_0=j+1}^{J} \eta_{j_0}^2 \Rightarrow \begin{cases} \eta_J^2 = \gamma_{J-1}^2 \\ \eta_{j_0}^2 = \gamma_{j_0-1}^2 - \gamma_{j_0}^2 \end{cases} \qquad (7)$$

This constraints show that the expression in (6) is only valid provided that the weights γ_j^2 are decreasing with j. Now, for most natural images, we observe in practice that the energy in a subband j is increasing with j. Under the Gaussian assumption, γ_j^2 is inversely proportional to this energy, and therefore the condition is verified.

Because DL_J is expressed as a function of w, we can now search for the set of parameters $\{\mathbf{p}_i\}_i$ which minimizes it.

2.3. The Minimization Algorithm

Each signal $\tilde{I}_{j_0}$ can be computed only if $\tilde{I}$ (hence, w) has been computed *beforehand*. For that reason, we decide to consider $\tilde{I}$ as a new variable in the minimization process. Then, we formulate the minimization issue as a joint optimization where the best couple $(w, \tilde{I})$ is looked for, subject to the constraint: $\tilde{I} = I(w(\tilde{\mathbf{p}}))$. We propose to solve this optimization problem through an iterative Expectation Maximization-like procedure. At each iteration, an estimate of w is updated given the current observation of $\tilde{I}$. In return, $\tilde{I}$ is refined given the updated observation of w:

COMPUTE-WARPING(I)
1 $n \leftarrow 0, w \leftarrow Id, \tilde{I}^{(0)} \leftarrow I$ // Initialization
2 **while** $(n < n_{max}$ or $\Delta w > \Delta w_{min})$
3   **do** $n \leftarrow n+1;$
4         $w^{(n)} \leftarrow$ UPDATE-WARPING$(I, w^{(n-1)}, \tilde{I}^{(n-1)})$
5         $\tilde{I}^{(n)} = w^{(n)} I$
6 **return** w

From the previous algorithm, it is clear that the optimization complexity is related to step 4. Knowing $\tilde{I}^{(n-1)}$, it is possible to compute each current approximation $\tilde{I}_{j_0}^{(n-1)}$ and minimize Eq. (6) taking $w^{(n-1)}$ as the initial guess for $w^{(n)}$. Further, it can be shown that:

$$\arg\min_{w^{(n)}} DL_J(w^{(n)}) = \arg\min_{w^{(n)}} \sum_{\tilde{\mathbf{p}}} (I(w^{(n)}(\tilde{\mathbf{p}})) - \tilde{I}_{ref}^{(n)}(\tilde{\mathbf{p}}))^2, \qquad (8)$$

with

$$\tilde{I}_{ref}^{(n)}(\tilde{\mathbf{p}}) = \frac{\sum_{j_0} \eta_{j_0}^2 \cdot \tilde{I}_{j_0}^{(n-1)}(\tilde{\mathbf{p}})}{\sum_{j_0} \eta_{j_0}^2}.$$

The problem expressed in the right hand side of Eq. (8) is the minimization of a *Displaced Frame Difference* between $\tilde{I}$ and a current *reference* frame $\tilde{I}^{(n)}_{ref}$. It is a well-known problem in the Video Coding Community, where w does not refer to a geometry but a motion. A variety of solutions has been proposed. They depend on their choice for the motion model. In the next section, we will present the specificities of an implementation using an active mesh as the model for w.

3. IMPLEMENTATION WITH A MESH

3.1. Geometry modelization with a Mesh

A geometry that is modelized with any mesh consists of two types of parameters: geometric parameters, i.e. the positions of its nodes, and connectivity parameters, i.e. how these nodes are linked to one another. In this work, we consider a *regular* mesh with N_p nodes. Hence, no connectivity parameters must be transmitted. Let $\{\mathbf{p}_i\}_{i=1..N_p}$ be the positions of these nodes in $\mathcal{D}$. Each position $\mathbf{p}_i$ is arbitrarily mapped to a position $\tilde{\mathbf{p}}_\mathbf{i}$ in $\tilde{\mathcal{D}}$. We decide to place the positions $\tilde{\mathbf{p}}_\mathbf{i}$ uniformly in $\tilde{\mathcal{D}}$. Therefore, the only geometric parameters to transmit are the size of the faces in $\tilde{\mathcal{D}}$ and the positions $\{\mathbf{p}_i\}_i$.

Letting $\phi(\tilde{\mathbf{p}})$ refer to a 2D shape function defined in $\tilde{\mathcal{D}}$ (e.g. the bilinear function), a *continuous* transformation w is defined as follows:

$$w(\tilde{\mathbf{p}}) = \sum_{i=1}^{N_p} \phi(\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_\mathbf{i}) \cdot \mathbf{p}_\mathbf{i} \tag{9}$$

3.2. Estimation of Nodes Displacements

The computation of the positions $\{\mathbf{p}_i\}_i$ follows the algorithm described in paragraph 2.3. As we said, the technical issue of this algorithm is the update of w at each iteration (step 4). In this work, we choose to implement this step using a bi-conjugate gradient descent algorithm. The linearization of Eq. (8) gives a sparse linear system. The solution of this system is a set of displacements for each position $\mathbf{p}_\mathbf{i}$. Details of the method are given in [8]. Initially, the nodes are placed uniformly in $\mathcal{D}$ so that $\mathbf{p}_\mathbf{i} = \tilde{\mathbf{p}}_\mathbf{i}\ \forall i \in \{1, .., N_p\}$ (or $w = Id$). At each iteration, the positions are updated *globally* by solving the linear system. $\tilde{I}$ is then updated by warping the original image. The process goes on until a maximum number of iterations has been reached, or the maximum displacement of the nodes is smaller than a threshold. Experiments have shown that no significant displacement occurs after 10 to 15 iterations. In term of complexity, this procedure can be compared to a motion estimation (with an active mesh) between a current frame and a reference frame, with the reference frame being updated at each iteration. The basic algorithm is greedy but can be sped up in numerous ways, which is not the topic of this article. Note that, in a coding scheme, the complexity of the decoding step remains comparable with a classical wavelet coder. Only the post warping in Fig. (1) must be added.

3.3. Mesh Warping and PSNR

As we work in a discrete setting, a non-isotropic transformation w, as defined in Eq. (9), cannot comply with the reversibility assumption because of the resampling implied. Indeed, such a transformation authorizes irreversible warpings, such as stretchings or contractions. In a coding scheme, only texture regions are visually affected by this loss at high rates. On the edges, this loss has no visual impact. As experiments have shown that the MSE was not a good indicator of this loss, the goal of such a coding scheme is not high SNR but high perceptual quality. A number of SNR-driven optimizations can be implemented but this is beyond the scope of this paper.

4. EXPERIMENTAL RESULTS

4.1. Codec Implementation

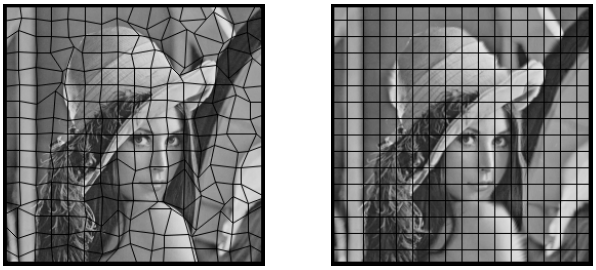
Our basic codec implementation is composed of three blocks:
Analysis: estimates w following our energy minimization. The results presented in this section were obtained by setting $J = 4$ and assuming a Gaussian model for the wavelet subbands. w was modelized by a *quadrangular* mesh with an initial size of 16x16 for its faces. The outputs are w and $\tilde{I}$.
Codec: $\tilde{I}$ was encoded and decoded using JPEG2K VM8.0 with its default parameters. w was quantized with a pixel precision and encoded using arithmetic coding.
Synthesis: takes as inputs the decoded warping transformation $\hat{w}$ and warped image $\hat{\tilde{I}}$ and reconstructs the original image by inverting the warping $\hat{I} = \hat{\tilde{I}}(\hat{w}^{-1})$.

4.2. Analysis Results

Fig. 2 shows the outputs of the analysis block taking *Lena* as the input image. On the left, the estimated mesh is overlayed upon the original image. Even though we did not inject any geometric *a priori* in the energy formulation, we see that the nodes have moved towards the geometric features, such as the contours of the shoulder or the hat. On the right, $\tilde{I}$ is represented. We notice that the warping has a tendency to align and smooth the contours on the horizontal or the vertical direction, like the right part of the shoulder. This is a result we could expect because conventional 2D wavelet can represent those features with fewer coefficients.

4.3. Compression Results

Numerical results of our coding scheme compared to JPEG2K are given in Fig. 3 for informative purpose. As we expected, the new scheme does not give high PSNR because of the loss



$\mathcal{D}$ $\tilde{\mathcal{D}}$

Fig. 2. Outputs of the Analysis Block.

introduced by successive interpolations. However, we believe that the subjective quality of our scheme is better than indicated from the PSNR values. Fig. 4 provides a qualitative assessment of perceptual quality by comparison with JPEG2K for low bit rates. The overall perceptual quality of the images obtained with the new scheme is significantly better than the one obtained with JPEG2K. Ringing artefacts characterizing the conventional wavelet are greatly reduced in the new scheme and the edges of objects are reconstructed with a high precision with very few coefficients. The irrelevance of using PSNR values to compare the two coders is particularly demonstrated by the encoding of *Cameraman* at 0.4 bpp. The PSNR obtained with the new scheme is 0.6 dB smaller than the one obtained with JPEG2K but the subjective quality is appreciably better. At high rates, some smoothing is perceptible on texture parts if no specific optimization is performed.

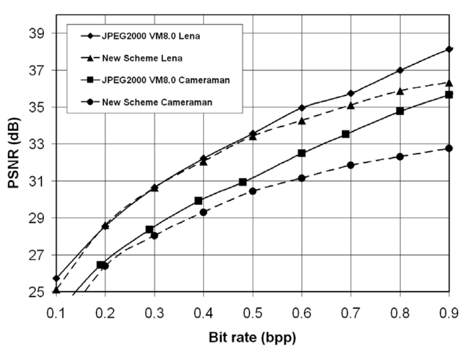


Fig. 3. Numerical Compression Results for *Lena* and *Cameraman* 256x256. The cost of the mesh is 0.02 bpp.

5. CONCLUSION

We have presented a new approach to adapt a signal to a conventional wavelet decomposition. The adaptation issue is solved by an energy formulation which is not based on any geometric *a priori*. Used as a pre-processing in a wavelet codec, its efficiency to reduce ringing artefacts has been underlined. Next studies include the relaxation of the regularity constraint put on the mesh to enable a closer adaptivity, as well as a generalization of the energy formulation to higher dimensional signals or different transforms (e.g. DCT).

JPEG2K: 0.19 bpp, 28.5 dB New Scheme: 0.2 bpp, 28.6 dB

JPEG2K: 0.40 bpp, 29.9 dB New Scheme: 0.40 bpp, 29.3 dB

Fig. 4. Visual Compression Results.

6. REFERENCES

[1] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[2] D. Taubman and A. Zakhor, "Orientation adaptive subband coding of images," *IEEE Trans. Image Process.*, vol. 3, pp. 421–436, July 1994.

[3] E. Le Pennec and S. Mallat, "Sparse Geometric Image Representations with Bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.

[4] W. Ding, F. Wu, and S. Li, "Lifting-based Wavelet Transform with Directionally Spatial Prediction," in *Proc. Pict. Coding Symp.*, San Francisco, USA, Dec. 2004.

[5] G. Peyré and S. Mallat, "Discrete bandelets with geometric orthogonal filters," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, USA, Sept. 2006, pp. 65–68.

[6] D. Wang, L. Zhang, A. Vincent, and F. Speranza, "Curved Wavelet Transform for Image Coding," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2413–2421, Aug. 2006.

[7] V. Chappelier and C. Guillemot, "Oriented wavelet transform on a quincunx pyramid for image compression," in *Proc. IEEE Int. Conf. Image Process.*, Sept. 2005.

[8] Y. Wang and O. Lee, "Active mesh - a feature seeking and tracking image sequence representation scheme.," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 610–624, Sept. 1994.