IMPROVED LOCAL PDF ESTIMATION IN THE WAVELET DOMAIN FOR GENERALIZED LIFTING

Julio C. Rolón[†] and Philippe Salembier[‡]

[‡]Technical University of Catalonia (UPC), Dept. of Signal Theory and Communications, Spain
[†]National Polytechnic Institute (IPN), CITEDI Research Center, Mexico
jcrolon@citedi.mx, philippe.salembier@upc.edu

ABSTRACT

Generalized Lifting (GL) has been studied for lossy image compression in [2,3]. It has been demonstrated that the method leads to a significant reduction of the wavelet coefficients energy and entropy. The definition of the GL relies on an estimation of the pdf of the pixel to encode conditioned to a surrounding context. The objective of this paper is to present an improved method for the estimation of the pdf at the local level. Rather than assuming that the local pdf is monomodal, symmetric, and centered at the central value of the best context match within a neighborhood, as in [3], we follow the idea of self similarity proposed in [1] for denoising, and propose to estimate the pdf using all the causal contexts within a window. Therefore, all the available knowledge about the neighborhood is incorporated. No assumptions about the characteristics of the pdf are made. A generalized lifting operator that minimizes the energy is applied to each context during the encoding process. Experimental results show an important increment in the energy and entropy gains when compared to previous strategies [2,3].

Index Terms— Generalized lifting, wavelets, pdf estimation, lossy image coding.

1. INTRODUCTION

The Generalized Lifting (GL) approach [5] is a signal decomposition method derived from classical lifting [6]. GL enables the implementation of linear as well as non-linear operators in the analysis and synthesis stages while preserving the perfect reconstruction property. The fundamental difference between the classical lifting and the GL can be highlighted considering how the detail signal is generated. In the classical lifting, details samples are viewed as a prediction error: odd samples are predicted from even samples and the prediction error, that is the detail signal, is computed by subtracting the predicted values from the actual values of odd samples. The GL takes a different viewpoint as it produces the detail signal through an invertible and possibly nonlinear mapping that takes as input the odd samples themselves. The mapping is adaptive as it depends on a context defined by even samples spatially close to the detail sample that is computed. If the pixels pdf conditioned to their contexts is known, the GL can be designed to minimize the details sample energy.

In [2], a modeling approach was proposed to apply GL for generic image coding. In generic images, the set of possible context values is so large that the strategy is to cluster the contexts into a reduced set of six context classes based on their structure and to model the pdfs conditioned to the classes. The clustering process is contrast invariant and focuses on the features resulting from contours. Note that these features are dependent upon the wavelet filters used for the initial DWT. Once the context classes are defined, their pdf is estimated through a training process run over the entire subband. Only the parameters of the models necessary to define the inverse GL mapping are sent to the receiver. In [2], it is shown that this strategy provides significant gains in terms of energy reduction of the wavelet coefficients, in particular around contours.

In [3], a local pdf estimation approach has been proposed. The focus is on the local characteristics of the sample to encode and its context. The conditional pdf estimation is done assuming that the pdf is monomodal and symmetric. In order to design the mapping under this assumption, we only need to know the value of the pdf maximum. An adaptive estimation strategy that consists in searching in a causal neighborhood of the current sample to encode for contexts that are similar to the one being analyzed was proposed in [3]. The idea is that it is very likely that the most similar context has the most similar central value that minimizes the energy of the detail signal. The resulting mapping can be implemented by a subtraction. It is shown in [3] that the adaptive local pdf estimation approach produces results comparable to those of [2].

In this paper, we move a step forward in the estimation of the pdf for generic images. We have identified in [3] that the similarities that exist at the local level in the wavelet signal may be used to construct GL mappings that reduce the energy of the detail signal. Our goal here is to remove the assumption following which the local pdf is monomodal and symmetric. To that end, we will not use only the context in the causal neighborhood that provides the best match. In fact, all contexts within the causal neighborhood contribute to define the pdf.

Buades *et al* [1] introduced the notion of self-similarity of the pixels in an image, and its importance at the local level. They proposed a weighting function that depends on the similarity between two patches of the image to implement denoising algorithms. A similar idea is used in this paper to estimate the pdf at the local level. The contribution of each of the contexts within a window is weighted depending on the similarity that exists between the context of the pixel being encoded, and the context against which it is being compared. The higher the similarity between the contexts; the larger the contribution to the pdf.

The strategy proposed here makes a comprehensive exploration of the neighborhood. Moreover, the context positions used to estimate the pdf do not lie only over the approximation grid, but also over positions that belong to the causal part of the

detail grid, since these coefficients are known by the coder and the decoder. We compare the results of this strategy against those obtained with the methods described in [2] and [3]. We show that the comprehensive pdf estimation strategy presented here enables a significant improvement of the energy and entropy gains.

Next section reviews the Generalized Lifting method. Section 3 describes the new local pdf estimation approach. Section 4 describes the implementation of the comprehensive local pdf estimation method (CLPE). The coding scheme and the experimental results are presented in Sections 5 and 6. Finally, the conclusions are exposed in Section 7.

2. GENERALIZED LIFTING

The generalized lifting (GL) decomposition shown in Fig. 1 enables the implementation of linear and non-linear operations. The GL involves first a polyphase decomposition or Lazy Wavelet Transform (LWT), followed by a generalized predict (P) step. In classical lifting, detail samples $y'[n]$ are computed as a prediction error: a prediction $P(y[n])$ is computed by filtering approximation samples, $x[n]$, and this prediction is subtracted from $y[n]$ to get the detail sample: $y'[n] = y[n] - P(y[n])$. The generalized predict operator P takes a different perspective and may be viewed as a mapping between $y[n]$ and $y'[n]$. This mapping takes into account a context represented by samples from $x[n-i]$ for $i \in C_n$, C_n being the set of sample positions that constitutes the context. Formally, the generalized predict operation can be written as

$$y[n]|_{C_n} \xrightarrow{P} y'[n]$$

Assuming discrete signals, the mapping itself is discrete. To get perfect reconstruction, the mapping P should be invertible, that is injective. If the number of possible values for $y[n]$ and $y'[n]$ is the same, then the mapping should be bijective.

Apart from the injectivity that is required to achieve perfect reconstruction, the generalized predict (P) operator may be arbitrary. In order to improve image coding efficiency, we seek to use a prediction based on mappings that minimize the energy of the detail signal produced by the GL decomposition. This minimization can be achieved if the pdf of the pixel to encode, conditioned by its context composed of neighboring pixels, is known [5].
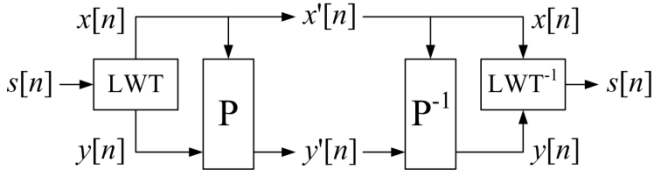

Fig. 1. Generalized Lifting Scheme

In the following section, we describe a specific strategy to estimate the pixel conditional pdf. The pdf is locally estimated for each pixel to encode taking into account the contributions of all locations within a window.

3. COMPREHENSIVE ESTIMATION OF THE LOCAL PDF FOR GL

For the local estimation of the pdf, the proposed approach assumes that, within a certain neighborhood, all the causal coefficients contribute to the statistics of the signal. Assume that the sample to

be mapped is denoted by $y[n]$ and its context C_n is defined by a set of neighboring samples. The local pdf may be defined as

$$\text{pdf}_{local}(y_n|C_n) = \bigcup_{k \in \Omega}^{y \in \varphi} Pr(y_n = y|C_k), \qquad n \notin \Omega.$$

The non-normalized pdf (or histogram) of the random variable y at position n (y_n) conditioned to the context C_n is the set of probabilities of the realizations of $y \in \varphi$ (with φ the set of possible values of y), subject to their respective contexts C_k; which occur in the neighborhood Ω, with the exception of position n, which is the location of the pixel to encode and does not belong to the neighborhood.

To construct the mapping P that minimizes the energy of $y[n]$, the set of probabilites of $\text{pdf}_{local}(y_n|C_n)$ is ordered in descending order $\{Pr(y_n = y_1), Pr(y_n = y_2), \dots, Pr(y_n = y_\ell)\}$, $\{y_1, y_2, \dots y_l\} \in \varphi$. The lowest energy value of the output mapped signal $y' = 0$, is assigned to the input value $y_n = y_1$, the coefficient with the largest probability. The first non-zero value $y' = 1$ is assigned to $y_n = y_2$, the coefficient with the second largest probability. $y' = -1$ is assigned to the third largest probability, and so on until $y[n] = k_\ell$ completes the set of input values. As can be seen the GL is entirely defined by the ordered sequence of pdf values, thus, the normalization of the histogram is not relevant (as it does not change the ordered sequence).

The pdf is estimated and the mapping P is constructed for each new sample that is encoded. In this sense, the approach is local, as it depends only on the neighboring positions that belong to a window that is defined around the sample. The method is also adaptive, since it grasps the local dynamics, which change from sample to sample. However, a major difference between this proposal and our previous local and adaptive proposal [3] is that the new local pdf estimation method does not assume that the pdf is monomodal and symmetric. As a result, it may lead to a non-linear mapping P; instead of the linear mapping of [3]. As we will show, the non-linear mapping of the improved estimation technique presented here leads to a very significant increase of the energy and entropy gains with respect to [2] and [3].

In this work, the LWT relies on a quincunx subsampling. The sampling pattern for odd scales is represented in Fig. 2(a). y represents the detail sample to be mapped. Its context may be composed of approximation as well as detail samples. Approximation samples can be chosen freely as they will be available in the decoder during the synthesis process. The four closest neighbors of y denoted by $\{x_1, x_2, x_3, x_4\}$ in Fig. 2(a) form the context. Experimental results have shown [3] that the introduction of causal detail samples $\{y_1, y_2\}$ in the context improves the scheme efficiency. The sampling pattern for even scales is shown in Fig. 2(b). We did not find any evidence that the inclusion of causal detail samples in the context introduced any advantage [3]. Therefore, only the four closest approximation samples will form the context for even scales.
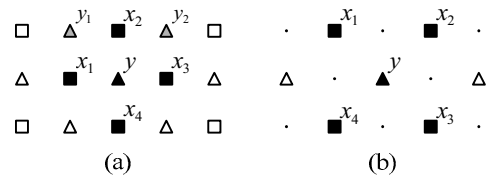

Fig. 2. Sample positions for, (a) odd and, (b) even GL decomposition scales in a quincunx sampling grid.

4. ADAPTIVE ESTIMATION OF THE LOCAL PDF

To estimate the pdf adaptively, we propose to define for each position n a causal window of size Ω. Of course, the window does not include the position n of the pixel to be mapped. All the coefficients that lie inside the window contribute to the estimation of the pdf. Fig. 3 shows an example of a window for the odd scales of the GL. The positions contributing to the estimation of the pdf are marked by circles. Contexts have the shape shown in Fig. 2(a). Fig. 4 shows an example of window for the even GL scales.

The contribution c of each context adds to the construction of the pdf according to its central value y,

$$Pr(y_n = y|C_k) = \sum_\Omega c(y, C_k).$$

Each contribution is a weighted function $(0 < c \leq 1)$ [1] of the distance, D, between the context of the sample being encoded, and the context whose contribution is being calculated

$$c = e^{-D/h}.$$

The higher the similarity between the contexts, the larger the contribution to the pdf. h helps to optimize the pdf.

In essence, the idea is similar to that of [3]. However, in this approach, the context with the smallest distance to the context of the pixel to encode is not the only one to contribute to the pdf
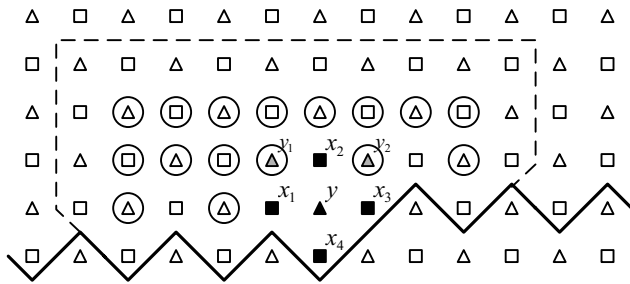

Fig. 3. Search window for odd scales of GL; context (black squares) and detail sample (black triangle).
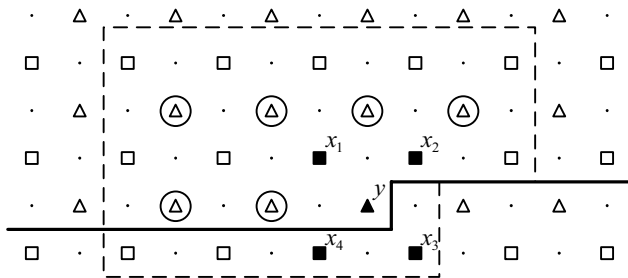

Fig. 4. Example of a search window, for even scales of GL; context (black squares) and detail sample (black triangle).

estimation. Instead, all contexts are taken into account; even if they are not so similar. The distance measure is the squared error of the context values:

$$D = \sum_{i=1}^{4}(x_i - x_i^*)^2 + \alpha\sum_{i=1}^{2}(y_i - y_i^*)^2 \, ,$$

where $\{x_i, y_i\}$ are the values of the context around the pixel being encoded; $\{x_i^*, y_i^*\}$ are the values of the context being compared against to estimate its contribution; and α weights the contribution of context samples y_i compared to x_i. For the second GL scale, the distance is $D = \sum_{i=1}^{4}(x_i - x_i^*)^2$.

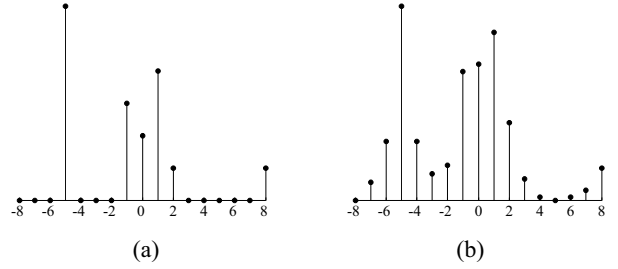(a)   (b)

Fig. 5. Filtering of a pdf through a convolution with a 5-sample Laplacian kernel. (a) original, (b) filtered.

The shape of the pdf depends of the size of the window Ω, and the local dynamics of the wavelet coefficients. The shape of the pdf may be arbitrary. Fig. 5(a) shows a typical pdf obtained for a certain window. It may be seen that $Pr(y_n = y|C_k) = 0$ for some values of y. This simply means that these values have not been observed within the window. To improve the estimation of the pdf, we may consider filtering the pdf through the convolution with a kernel, as proposed in [4]. The kernel function

$$K = e^{-\gamma\omega}; \quad -2 \leq \omega \leq 2$$

provided the best improvement in the estimation of the pdf. The result of the filtering of the pdf of Fig. 5(a) with the kernel K is shown in Fig. 5(b). Note how the main modes of the multimodal histogram are preserved, but the values that have not been observed are now represented adequately. The mapping is constructed with the filtered histogram.

Even filtered, it may happen that $Pr(y_n = y|C_k) = 0$ for some y within the dynamic range of the subband. The problem with this condition arises only if $Pr(y[n]|C_n) = 0.$, To deal with this situation, the mapping continues assigning output values to those values $Pr(y_n = y|C_k)$ whose probabilities are zero to cover the entire dynamic range. The allocation of the mapped values is done as follows: Starting from $y_n = y_1$, y_1 being the maxima of the pdf; we seek for $Pr(y_n = y|C_k) = 0$ alternating the search left and right sides away from $y_1$until the complete dynamic range of the subband is covered. A list of y values whose $Pr(\cdot) = 0$ is generated. This list is assigned output values y' that start at $y_n = y_{\ell+1}$ and continue increasingly; making certain that the assigned output values are all of higher magnitude than those assigned to the filtered pdf. In the case that the input value $y[n]$ has not been observed ($Pr(y[n]|C_n) = 0$); there will always be an unambiguous output assignment. The second condition that may be found is that $Pr(y_n = y_1|C_k) = Pr(y_n = y_2|C_k)$ for two different values y_1 and y_2. In this case, two different and consecutive output values are assigned, the smaller of the two is assigned to the larger of $\{|y_1|, |y_2|\}$, minimizing the energy, and avoiding the ambiguity.

5. CODING SCHEME

The coding scheme is shown in Fig. 6. The first step is to apply a discrete wavelet transform (DWT) to the image, and the GL scheme is applied in the wavelet domain to the details subbands only. After the DWT, the subband is quantized with a uniform scalar quantizer Q. We have selected a quantization step $\Delta = 12$, such that the quantized images have a reconstruction quality of about 36 dB of PSNR. Note that the quantization is applied before the GL in order to guarantee the invertibility of the mapping. The GL is applied using the CLPE. The overhead (OVH) consists of the window size, h, α, and γ.
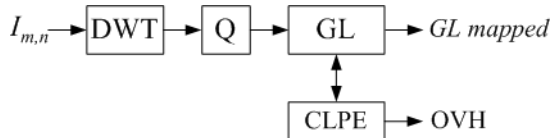
Fig. 6. GL encoder

Fig. 7. Test images. (a) *barbara*, (b) *bike*,
(c) *cameraman*, (d) *peppers*, (e) *boat*.

We have chosen to measure the entropy of the mapped subband to set the upperbound of the coding efficiency of the method. The set of generic test images is shown in Fig. 7. The GL decomposition is iterated twice in a dyadic way over a quincunx sampling grid. The filter used for the DWT transform is a CDF 9/7 biorthogonal filter.

6. EXPERIMENTAL RESULTS

For each image, the best window size in terms of energy minimization has been used. Parameters h, α, and γ have been optimized using two criteria: minimization of the energy; and minimization of the entropy of the whole subband. The experimental results presented here are for the HL wavelet subband, as a proof of the concept. The algorithm may be applied to the other details wavelet subbands as well.

Table I shows a comparison of the energy gain E_{gain} that results after comparing the energy of the wavelet subband against the energy of the GL-mapped subband using

$$E_{gain} = 100(1 - E_{DWT+GL}/E_{DWT}).$$

Table II shows a similar comparison for the entropy gain H_{gain}, computed with

$$H_{gain} = 100(1 - H_{DWT+GL}/H_{DWT}).$$

The entropy was measured over for the subband as whole using

$$H(S) = \sum_i p(s_i) \log_2 p(s_i).$$

We compare in Tables I and II, the use of a set of structured models of contours of [2] in the column *Models*; the use of the adaptive local pdf estimation method of [3] in the column *ALPE*; the comprehensive local pdf estimation (*CLPE*) method presented in this paper, optimized to minimize the energy of the detail signal; and the refinement of the method conducted to optimize the entropy *CLPE-H*. Positive values indicate a reduction in energy or entropy with respect to the wavelet subband. Positive entropy gain reflects a potential increment in the coding efficiency after GL.

Table I. Comparison of E_{gain} (%).

Image	Models	ALPE	CLPE	CLPE-H
barbara	19.58	11.43	**24.44**	**24.03**
bike	11.47	23.60	**32.43**	**31.76**
cameraman	21.25	15.38	**28.33**	**26.96**
peppers	9.48	7.68	**17.66**	**16.64**
boat	**13.47**	0.25	12.71	12.19
AVERAGE	15.05	11.67	**23.11**	**22.32**

Table II. Comparison of H_{gain} (%).

Image	Models	ALPE	CLPE	CLPE-H
barbara	3.55	-0.48	**3.67**	**4.01**
bike	4.36	7.03	**8.95**	**9.21**
cameraman	4.68	3.02	**6.82**	**7.34**
peppers	1.78	0.03	**2.15**	**2.76**
boat	**3.14**	-0.95	2.55	2.74
AVERAGE	3.50	1.73	**4.83**	**5.21**

From Tables I and II, it is observed the interesting improvement in energy and entropy gains over the methods of [2] and [3]. It is also evident how the energy gain of CLPE reduces slightly when the optimization of entropy is conducted (CLPE-H), however, this small reduction is compensated by the larger increase in entropy gain (Table II). In general, the average gain of the CLPE-H method is 22.32% for the energy and 5.21% for the entropy with respect to the DWT. Interestingly, the peak performance of over 31% in energy, and over 9% in entropy gains for *bike* image (Tables I and II resp.) may be due to the high local self-similarity that exists along the contours in the wavelet domain.

7. CONCLUSIONS

We have presented an improved adaptive local pdf estimation strategy that takes advantage from the similarities that exist at the local level in the wavelet domain. The method estimates the local pdf and applies a non-linear but invertible mapping that minimizes the energy (CLPE) and is refined to minimize the entropy (CLPE-H). The experimental results exhibit an important improvement in the performance of the GL method when compared to previous work for the case of generic images [2,3]. In order to translate this increased entropy gains to practical coding efficiency, further steps must be taken in the entropy coding part.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Buades, A. Coll, B., Morel, J.M., A review of image denoising algorithms, with a new one, *SIAM Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490-530, 2005.

[2] Rolón, J.C., Ortega, A., Salembier, P., Modeling of Contours in Wavelet Domain for Generalized Lifting Image Compression, *IEEE ICASSP 2009*, pp. 1025-1028, apr 19-24, Taipei, Taiwan.

[3] Rolón, J.C., Mendonça, E., Salembier, P., Generalized Lifting with Adaptive Local pdf Estimation for Image Coding, *Picture Coding Symposium PCS 2009*, pp.1-4, may 6-8, Chicago, USA.

[4] Scott, D.W., Sain, S.R., Multidimensional density estimation, *Handbook of Statistics: Data Mining and Comput. Stats.*, vol. 23, pp. 297-303, Elsevier, 2004.

[5] Solé, J., Salembier, P., Adaptive discrete generalized lifting for lossless compression, in *Proc. of the IEEE ICASSP 2004*, pp.iii - 57-60, may 17-21, Montreal, Canada.

[6] Sweldens, W., The Lifting Scheme: A Construction of Second Generation Wavelets, *SIAM J. Math. Anal.*, vol. 29, no. 2, pp.511-546, 1997.