GENERALIZED LIFTING WITH ADAPTIVE LOCAL PDF ESTIMATION FOR IMAGE CODING

Julio C. Rolón[†‡], Eduardo Mendonça[†] and Philippe Salembier[†]

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

ABSTRACT

In this paper we introduce an adaptive local pdf estimation strategy for the construction of Generalized Lifting (GL) mappings in the wavelet domain. Our approach consists in trying to estimate the local pdf of the wavelet coefficients conditioned to a context formed by neighboring coefficients. To this end, we search in a small causal window for similar contexts. This strategy is independent of the wavelet filters used to transform the image. Experimental results exhibit interesting gains in terms of energy reduction comparable to those obtained in [8]. In order to take benefit from this energy reduction, specific entropy encoder should be designed in the future.

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

1. INTRODUCTION

One of the limitations of the discrete wavelet transform (DWT) when applied to images is its reduced ability to decorrelate coefficients along contours. As correlation persists, large magnitude coefficients representing contours remain spatially co-located. As a result, current research efforts are focusing on the design of schemes able to decorrelate these coefficients or to minimize their energy [1-5,11].

The Generalized Lifting (GL) approach [9] is a signal decomposition method derived from classical lifting [10]. 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.

In [6], a GL mapping minimizing the energy of wavelet coefficients obtained after an initial DWT was proposed for lossy image coding. This mapping is able to drastically reduce the coefficients energy, but it was obtained under the ideal but unrealistic assumption that accurate context-dependent pdfs are available for wavelet-domain data (e.g., the pdf of wavelet coefficients conditioned to a given neighborhood was assumed to be known). Clearly, in a coding application, this information would have to be sent to the receiver eliminating the gains in overall coding performance. In [7], one step towards the definition of a realistic scheme was taken by assuming that the image to encode belonged to an image class. The conditional pdf was estimated through training on the image class and was assumed to be known by the decoder, which was therefore dedicated to this image class. This strategy allowed us to define a realistic scheme, but the resulting gains highly depended on the class itself and the similarity of the images within the class.

To be able to deal with arbitrary images, a modeling approach was proposed in [8]. The strategy consists in clustering the initially huge set of possible context values into a reduced number of context classes and in modeling the pdfs conditioned to the classes. More precisely, contexts involving the four closest neighbors of the sample to be mapped through the GL are clustered into six classes based on their structure. 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. Finally, the parameters of the models necessary to define the inverse GL mapping are sent to the receiver. As the number of context classes and of parameters used to define the conditional subband pdfs is small, this overhead has only a minor impact on the final bitrate. In [8], we show that this strategy provides significant gains in terms of energy reduction of the wavelet coefficients, in particular around contours.

In this paper, we investigate an alternative approach and compare it with the one described in the previous paragraph. The fundamental difference between both approaches is that, here, we do not intend to describe the conditional pdf of the entire subband. Instead, we focus on the local characteristics of the sample to encode and its context. In order to maintain the analogy with the previous approach, we will still rely on the notion of conditional pdf but the description will be done from a pure deterministic viewpoint. When a given sample $y[n]$ to be encoded and its context are observed, we can say that, locally, the conditional pdf of the sample is one for the value $y[n]$ and zero for the remaining values. This trivial local pdf allows us to define a very simple GL mapping minimizing the energy of the output sample. Following this idea, the main issue is to be able to estimate the trivial local pdf. We propose here an adaptive estimation strategy that essentially consists in searching in a causal neighborhood of the current sample to encode for contexts that are similar to the one being

analyzed. The goal of this paper is to describe and analyze this approach and to compare its results with those reported in [8]. Note that the proposed approach eliminates the dependency of the GL mapping on the wavelet filters used in the initial DWT.

Next section reviews the Generalized Lifting method. Section 3 introduces the idea of local pdf and the resulting GL mapping. Section 4 presents the adaptive local pdf estimation technique. Sections 5 and 6 describe the coding scheme and the experimental results respectively. The conclusions are presented in Section 7.

2. GENERALIZED LIFTING

The generalized lifting (GL) decomposition shown in Figure 1 and introduced in [9] enables the implementation of linear and non-linear operations. The GL involves first a polyphase decomposition or Lazy Wavelet Transform (LWT), followed by generalized predict (P) and update (U) steps.
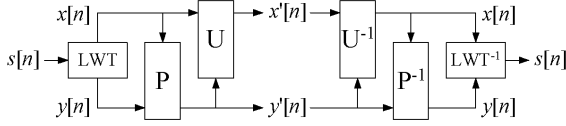
Fig. 1. Generalized Lifting Scheme

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

$$y[n]\Big|_{\text{with context } \{x[n-i]\}_{i \in C_n}} \xrightarrow{P} y'[n].$$

Assuming discrete signals, the mapping itself is discrete. To get perfect reconstruction, the mapping P should be invertible, i.e., injective. If the number of possible values for $y[n]$ and $y'[n]$ is the same, then the mapping should be bijective. The same reasoning can be applied to the generalized update operation. That is, it is an injective mapping

$$x[n]\Big|_{\text{with context } \{y'[n-i]\}_{i \in C_n}} \xrightarrow{U} x'[n].$$

Apart from the injectivity that is required to achieve perfect reconstruction, the generalized predict (P) and update (U) operators may be arbitrary. In order to improve image coding efficiency, we seek to use prediction based on mappings that minimize the energy of the detail signal produced by the GL decomposition. In [6], we have shown that, if the conditional pdf of each subband, $\text{pdf}_{subband}(y|C)$, is known, then the GL can provide a drastic reduction of the wavelet coefficients energy. This is an ideal assumption useful to assess the potential of the approach but unrealistic in practical situations. In the following section, we discuss the design of GL based on a local observation of the data to encode instead of a global characterization of the subband statistics.

3. GL BASED ON LOCAL PDF

As mentioned in the previous section, instead of taking a global perspective on the wavelet subband statistics, we consider now a local approach. Assume that the sample to be mapped is denoted by $y[n]$ and its context C_n is defined by the position of a set of neighboring samples. If we observe the value $y[n]$ conditioned to the value of the context C_n locally, we could say that the local pdf is described by

$$\text{pdf}_{local}(y = y[n] \| C = C_n) = 1, \ \text{pdf}_{local}(y \neq y[n] \| C = C_n) = 0,$$
$$\text{pdf}_{local}(y | C \neq C_n) = 0.$$

These equations simply state that the probability of the observation is one whereas the probability of all remaining configurations is zero. In other words, the local pdf consists of only one non-zero value. This is a trivial statement that will allow us to define a very simple GL mapping.

Let us assume that we have some means to estimate the local pdf and let us call $\text{pdf}_{local}^*(y|C)$ this estimate of $\text{pdf}_{local}(y|C)$. As it is a local pdf, it has also only one non-zero value. Let us call y^* the value for which this pdf is different from zero (therefore equal to one), $\text{pdf}_{local}^*(y = y^*|C) = 1$. As a result, the expected value of the sample to be mapped $y[n]$ is equal to y^*. In order to minimize the output detail samples energy, the best strategy is therefore to compute the difference: $y'[n] = y[n] - y^*$.

This equation is rather similar to the classical lifting approach, except that the computation of the y^* is not restricted to be obtained through a fixed convolution of the detail samples. Any strategy allowing us to estimate the local pdf, or equivalently the value of $y[n]$ based on its context, may be used. In the next section, we propose an adaptive strategy for this purpose.

In the sequel, the LWT relies on a quincunx subsampling. As a result, the sampling patterns for even and odd scales are different. 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. In the sequel, we will use the four closest neighbors of y denoted by $\{x_1, x_2, x_3, x_4\}$ in Fig. 2(a). If we want to increase this context to include detail samples, we have to select samples that will be available at the decoder when y will be processed; those are causal samples. Experimental results will show 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. Therefore, only the four closest approximation samples will form the context for even scales.
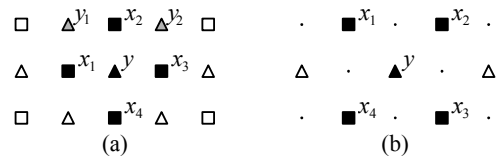
Fig. 2. Sample positions for (a) odd and (b) even GL decomposition scales in a quincunx sampling grid.

4. ADAPTIVE ESTIMATION OF THE LOCAL PDF

In this section, we define an adaptive strategy to estimate the local pdf of the detail sample y to be mapped. In the previous section, we have seen that this is equivalent to finding a strategy that defines the expected value of the sample y based on its context. Let us describe the approach for the odd scales of the GL decomposition. The context of the sample y is composed of 6 samples: $\{x_1, x_2, x_3, x_4, y_1, y_2\}$ (see Fig. 2(a)). Note that the values $\{y_1, y_2\}$ correspond to the values of the detail samples before the mapping.

The adaptive estimation consists in searching in a causal neighborhood around y for contexts $C^* = \{x_1^*, x_2^*, x_3^*, x_4^*, y_1^*, y_2^*\}$ that are similar to $C = \{x_1, x_2, x_3, x_4, y_1, y_2\}$. Fig. 3 shows an example of a search window and the corresponding contexts. The similarity between contexts is computed by the following expression:

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

Note that the samples that are farther away from the sample to be mapped are weighted by a factor α. Moreover, the weighted distance is normalized by the energy of the current context as we want to capture the context structure in a contrast invariant fashion.

Once the context minimizing D in the search window is found, its corresponding central detail sample y^* is used as the prediction of y and the resulting mapping is computed with $y' = y - y^*$.

Of course, we cannot expect to always find very similar contexts in the search window. As a result, the mapping is performed only if the similarity D is below a given threshold. In other circumstances, several different contexts produce exactly the same similarity. For the time being, we do not apply the mapping in these circumstances. For the even GL scales, Fig. 4 illustrates the search window, the context around the sample to be mapped and the context being searched. As we were not able to show the interest of using detail samples in the context, only four samples are included and the similarity D is computed with:

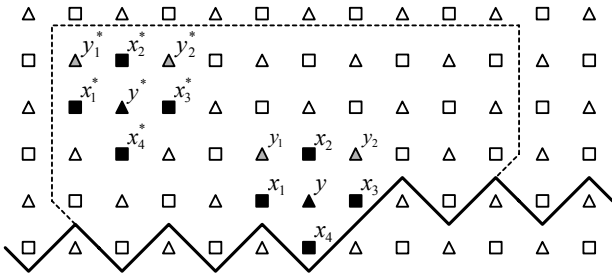$$D = \frac{\sum_{i=1}^{4}(x_i - x_i^*)^2}{\sum_{i=1}^{4} x_i^2} .$$

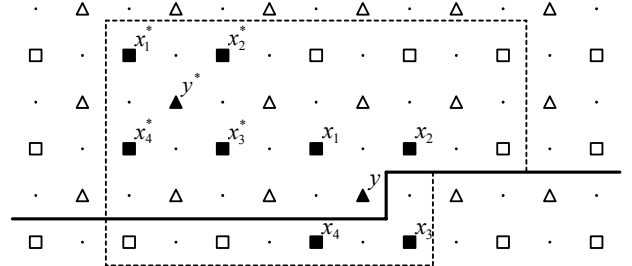**Fig. 3**. Example of a search window, contexts and detail samples for odd GL scales.

Fig. 4. Example of a search window, context and detail samples for even GL scales.

5. CODING SCHEME

The coding scheme is shown in Figure 5. First, we apply a discrete wavelet transform (DWT) to the image. Currently, we focus our analysis only on one of the three subbands produced by a decomposition of one scale of the DWT, namely HL or horizontal detail subband. Our strategy is easily applicable to other subbands and DWT scales. Next, subband is quantized with a uniform scalar quantizer. Our purpose at this moment is to evaluate the behavior of the approach, thus, we have used $Q = 1$ as the step size of the quantizer. We apply the adaptive local pdf estimation (ALPE). Parameters involved are search window size (ws), threshold T and α. These parameters are the only overhead the method introduces for each GL scale. Finally, the coding scheme is evaluated with two different entropy coders (EC): An arithmetic encoder (AE), and EBCOT. A comparison of the coding efficiency between them leads us to interesting conclusions.
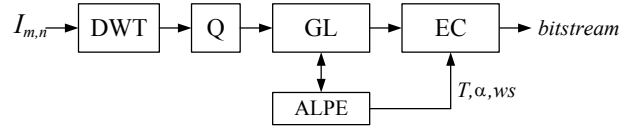

Fig. 5. GL encoder

Test images with their corresponding horizontal detail subbands are shown in Fig. 6. Wavelet domain subbands shown were obtained with a Haar filter. These images were selected because of their wavelet domain content; which varies from one to another in a way that is interesting for the evaluation of the local pdf; this different behavior would ease the analysis. In *barbara*, high energy coefficients are concentrated around contours; in *boat* we may appreciate non-structured textures at the bottom of the image; *bike* exhibits strong regular edges and large smooth regions. The GL decomposition is iterated twice in a dyadic way.
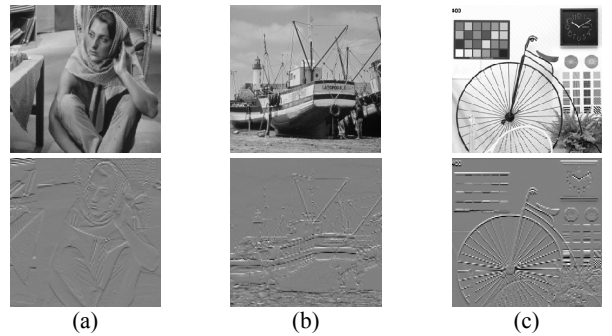

| (a) | (b) | (c) |

Fig. 6. Test images, (a) *barbara*, (b) *boat*, (c) *bike*, and their wavelet-domain HL subbands.

6. EXPERIMENTAL RESULTS

We conducted experiments with different window sizes. The best size in terms of energy minimization may depend on the image and filter used for the initial DWT. For the first GL scale and for *barbara* and *bike*, we found that a window size of 18×5 coefficients was the best for both Haar and CDF 9/7 filters of the DWT. For *boat,* we found that 18×3 and 18×7 were the best sizes for Haar and CDF 9/7 respectively. We attribute this difference to the presence of non-structured texture at the bottom of the image, whose behavior greatly differs from one filter to the other. For the second GL scale, we found the best window size of 14×5 for the three images in the case of Haar filter. For CDF 9/7 filter, we found window sizes of 10×5 for *barbara* and *boat*, and of 18×3 for *bike* as the best sizes. We attribute the difference of shape in *bike*, to the strong and uniform horizontal edges, whose similarity is best captured through the horizontally-elongated window. In the future, these optimum window sizes may be estimated but, for now, these parameters are sent to the receiver. We optimized the threshold T (also sent to the receiver), and computed the energy gain as

$$E_{gain} = 100\left(1 - E_{out}/E_{in}\right).$$

We have evaluated the influence of the weighting factor α. Best results were found for $0.3 \le \alpha \le 0.4$ depending on the image for Haar filter. In the case of CDF 9/7, $\alpha = 0.6$ for all images.

Table II summarizes experimental results. These results were obtained with the window sizes described above, optimized thresholds T and α for each of the two GL scales. The bitstream size in bits for the different configurations of the encoder is shown. Positive E_{gain} indicates a reduction in energy. Positive coding gain means an increase in coding efficiency after GL. The table compares the bitrates using Arithmetic Encoder or EBCOT after a wavelet decomposition (DWT) and possibly a GL (DWT+GL).

The average energy reduction is around 18%. However, table II shows that this energy reduction does not translate into a bitrate reduction. The best bitrates are obtained with EBCOT but, in this case, the average loss created by the GL is 2.59%. This loss is not totally surprising as the contexts used in EBCOT were designed to model the statistics of wavelet coefficients, not of the coefficients after applying the GL. If we use an arithmetic encoder that does not use contexts, then, although the global bitrates increase, the use of the GL is now positive and decreases the bitrates by an average of 1.56%. As a result, future research will focus on designing appropriate contexts to adapt EBCOT to the GL mapped signals.

7. CONCLUSIONS

Adaptive local pdf estimation approach to GL mapping construction has demonstrated to provide energy reductions comparable to those obtained through the use of models [8]; with the advantage of the independency of the DWT filter. As a counterpart, this approach is more sensitive to the wavelet domain subband structure.

Coding results are interesting. The mismatch between the statistics of GL signal and the statistics of the DWT prevents the EBCOT from exploiting its context classification scheme efficiently. The comparison of AE *vs* EBCOT shows the importance of having a good context classification scheme. Our future research will involve the design of a specific context classification scheme that would exploit better the energy reduction of GL.

Table II. Energy and Coding gain for the two entropy coders.

	barbara		*boat*		*bike*	
	Haar	CDF 9/7	Haar	CDF 9/7	Haar	CDF 9/7
E_{gain}(%)	18.19	18.13	11.42	5.12	29.21	28.96
Arithmetic Encoder						
DWT	78248	63144	89800	87584	86248	81648
DWT+GL	77448	63216	86688	83576	85992	81552
Coding gain (%)	1.02	-0.11	3.47	4.58	0.30	0.12
EBCOT						
DWT	73968	57176	81608	76536	81960	76704
DWT+GL	76488	60072	82792	78296	83512	77776
Coding gain (%)	-3.41	-5.07	-1.45	-2.30	-1.89	-1.40

8. REFERENCES

[1] Candès, E.J., *et al.*, Fast Discrete Curvelet Transforms, *SIAM Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861-899, 2006.

[2] Ding, W.P., *et al.*, Adaptive Directional Lifting-Based Wavelet Transform for Image Coding, *IEEE Transactions on Image Processing,* vol. 16, no. 2, pp. 416-427, 2007.

[3] Do, M.N., Vetterli, M., The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, *IEEE Trans on Im Proc,* vol. 14, no. 12, pp. 2091-2106, 2005.

[4] Mehrseresht, N., Taubman, D., Spatially Continuous Orientation Adaptive Discrete Packet Wavelet Decomposition for Image Compression, *IEEE Intl. Conf. on Image Processing, ICIP'06*, pp.1593-1596, 2006.

[5] Peyré, G., Mallat, S., Discrete Bandelets with Geometric Orthogonal filters, *Proc. of the ICIP,* vol.1, pp. 65-68, 2005.

[6] Rolón, J.C., Salembier, P., Generalized Lifting for Sparse Image Representation and Coding, *Picture Coding Symposium PCS 2007,* pp. 1-4, nov 6-9, 2007, Lisbon, Portugal.

[7] Rolón, J.C., Salembier, P., Alameda, X., Image Compression with Generalized Lifting and partial knowledge of the signal pdf, *Proc. of the IEEE ICIP,* pp. 129-132, oct 12-15, 2008, San Diego, USA.

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

[9] Solé, J., Salembier, P., Adaptive discrete generalized lifting for lossless compression, in *Proc. of the IEEE ICASSP 2004*, may 17-21, 2004, pp.iii - 57-60 vol.3, Montreal, Canada.

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

[11] Zhang, N., *et al.*, Directional Lifting-Based Wavelet Transform for Multiple Description Image Coding with Quincunx Segmentation, *Advances in Multimedia Information Proc.,* LNCS Vol. 3768, pp. 629-640, Springer-Verlag, 2005.