MODELING OF CONTOURS IN WAVELET DOMAIN FOR GENERALIZED LIFTING IMAGE COMPRESSION

Julio C. Rolón[†‡], *Antonio Ortega*[§] *and Philippe Salembier*[†]

[†]Technical University of Catalonia (UPC), Dept. of Signal Theory and Communications, Spain
[‡]National Polytechnic Institute (IPN), CITEDI Research Center, Mexico
[§]University of Southern California (USC), Signal and Image Processing Institute, USA
jcrolon@gps.tsc.upc.edu, antonio.ortega@sipi.usc.edu, philippe@gps.tsc.upc.edu

ABSTRACT

This paper introduces the design of context-based models of contours in the wavelet domain, which are used to construct generalized lifting (GL) mappings for image compression. The GL context-based mapping may significantly reduce the signal energy and the resulting bitrate. Here, we propose a strategy to define a reduced set of structured models to design the GL. The models capture the contour structures and are contrast-invariant. Initial experimental results applying the strategy on a wavelet subband exhibit potential gains. Iterations of the GL scheme as well as an adaptive entropy coding strategy may increase the coding gain.

Index Terms— Generalized lifting, wavelets, image coding, contour model, pdf model.

1. INTRODUCTION

One of the known drawbacks of the discrete wavelet transform (DWT) when applied to images is its limited ability to decorrelate coefficients that correspond to contours. As correlation persists, large magnitude coefficients that usually represent contours remain spatially co-located. Decorrelation of these coefficients to minimize their energy is of particular interest in wavelet-based image coding [1-5,12].

The Generalized Lifting (GL) approach [8] is a signal decomposition method derived from classical lifting [9]. GL enables the implementation of linear and non-linear operators in the analysis and synthesis stages, with perfect signal reconstruction. It is conceptually different from classical lifting because the GL operator that produces the detail signal observes the context provided by a set of approximation signal samples; thus, the detail signal results from a mapping operation that may be highly non-linear but invertible. In [6], a GL mapping which minimizes the energy of wavelet detail signals for lossy image coding was proposed. This mapping can drastically reduce the signal energy, but it was obtained under the ideal but unrealistic assumption that accurate context-dependent pdfs are available for wavelet-domain data (e.g., for every possible causal context we have a histogram of wavelet coefficients values after that context). 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 belongs to an image class. The pdf of each context is estimated through training on the image class and is known by the decoder, which is assumed optimized for this image class. This allows us to maintain

a large number of context models, but the resulting gains highly depend on the class itself and the similarity of the images within the class. If we want to be able to deal with arbitrary images, the goal should be to reduce the number of model parameters to estimate (e.g., by structuring all possible context models into classes that share some common parameters). This way, it will become possible to adaptively learn the model parameters or alternatively to send the model parameters to the decoder as overhead (this can only be done if the number of parameters to learn is small).

In this paper, we present a first approximation to the use of models for GL design. We propose a limited set of structured models that reduce, through the GL mapping, the energy of wavelet coefficients in particular around contours. The key point of the strategy is to cluster contexts based on their structure in a contrast invariant framework. As a result, four models are proposed and their associated parameters can be sent to the receiver without severe penalty on the resulting bitrate. These models will allow us to approximate the ideal pdf of each context and still provide very significant gains in terms of energy. The philosophy behind the use of these context-based models is in some sense similar to the one used in JPEG 2000 (in wavelet domain) [10] or CALIC (in image domain) [11], where contexts are also clustered in order to improve adaptation performance.

Next section reviews the Generalized Lifting method; Section 3 introduces the model design and Section 4 presents the study of one of the models and the resulting GL mapping. The same strategy is used to define the remaining models. 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 [8] 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.
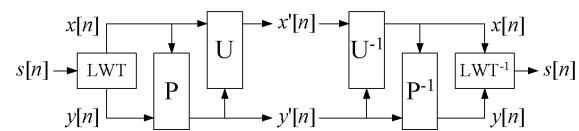

Fig. 1. Generalized Lifting Scheme

The generalized predict operator P may be viewed as a mapping between $y[n]$ and $y'[n]$ that takes into account a context

represented by samples from $x[n–i]$ for $i \in C$, C being the set of sample positions that constitutes the context.

Formally, the generalized predict operation can be written as

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

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

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

Apart from the injectivity that is required to achieve perfect reconstruction, the generalized predict (P) and update (U) operators may be arbitrary.

3. CONTOUR MODELING IN WAVELET DOMAIN

Contours in wavelet domain are characterized by large amplitude coefficients along a certain direction, surrounded by low amplitude, zero, or opposite sign large coefficients. The steepness of contours depends both on the image contents and on the wavelet filter used to transform the image.

From a generalized lifting perspective, we will select a specific context to predict a certain detail sample whenever the context elements fit a model. In order to improve image coding efficiency, we seek to use prediction based on mappings that minimize the energy in the detail signal produced by the GL decomposition.

The design of the models involves: *1)* definition of the context shape and dimension; *2)* decomposition of the wavelet subband through the LWT (Fig. 1) in context and detail samples; *3)* observation of the characteristics of the context and the pdf of the detail samples; *4)* identification of those contexts types associated to contours that would produce large coding gains.

We use a quincunx sampling grid, with contexts defined as shown in Figure 2. Wavelet coefficients belonging to the approximation phase after the LWT $\{x_1, x_2, x_3, x_4\}$ constitute the context that will predict y, the detail sample. The mapping is defined as $y\Big|_{\text{with context } \{x_1, x_2, x_3, x_4\}} \xrightarrow{P} y'$.

After the DWT, there are three subbands per wavelet scale. In this paper, we use encoding of the HL subband of the first wavelet scale as an example to illustrate our proposed method, but results can be easily extended to other subbands and scales.

In the HL subband, edges in horizontal direction prevail, structures such as $\{x_1, x_2, x_3, x_4\} = \{A, 0, A, 0\}$, with $A \gg 0$ would indicate the possible presence of an edge along $\{x_1, y, x_3\}$, cases such as the more general $\{A, -A, A, -A\}$, or the pair $\{A, 0, A, A\}$, $\{A, A, A, 0\}$ may also indicate so. Observation of these types of context structures may lead to proper modeling and mapping.

4. MODEL-BASED GENERALIZED LIFTING

4.1. Model definition

If we study the histogram of $\{x_1, x_2, x_3, x_4\} \rightarrow y$ for the HL subband of a natural image, we would find that, for the context
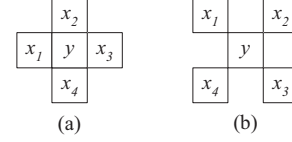
Fig. 2. Sample positions for (a) odd and (b) even GL decomposition scales in a quincunx sampling grid.

type $\{A, 0, A, 0\}$, the histogram is unimodal and centered at $y = A$, indicating the presence of a horizontal edge.

To simplify notation, in what follows we will refer to the central position of the histogram of y and its relation to the type of context as $\{A, 0, A, 0\} \rightarrow A$ when $y = A$, or $\{A, 0, A, 0\} \rightarrow 0$ when $y = 0$.

As the rule described by $\{A, 0, A, 0\} \rightarrow A$ tends to be true for all A values, the model is contrast invariant. In order to know if a particular arbitrary context $\{x_1, x_2, x_3, x_4\}$ can be approximated as belonging to this model, we use a mean squared error (MSE) criterion to determine how well it approximates a $\{A, 0, A, 0\}$ structure, i.e., we first scale it to $\{kx_1, kx_2, kx_3, kx_4\}$ and then compare it with a prototype of the model:

$$mse = (kx_1 - A)^2 + (kx_3 - A)^2 + (kx_2)^2 + (kx_4)^2$$

where A has to be given a specific value, for example A=100. The value of the scaling parameter k is chosen as the one which minimizes the *mse*. Of course, a threshold T_{mse} on the *mse* has to be defined to finally decide whether the initial context belongs to the $\{A, 0, A, 0\}$ model or not. Note that if the context $\{0, 0, 0, 0\}$ is observed, it is treated as a special case and directly assigned to the *zero* model.

Figure 3 shows the pdf of model $\{A, 0, A, 0\} \rightarrow A$, that is the pdf of the scaled values of the detail samples ky for the contexts that have been assigned to this model. This pdf has been estimated on the HL subband of DWT scale 1 of the *barbara* image (Figure 5(a)) with $A = 100$ and $T_{mse} = 6000$. As can be seen, the ky pdf is centered at $A = 100$. Moreover, the shape resembles an asymmetric Laplacian. For the time being, we will neglect the pdf asymmetry and assume that the pdf is simply unimodal, symmetric and centred at A. Under these assumptions, the best GL mapping, which is the mapping that minimizes the detail energy reduces to a linear mapping: $y' = y - A/k$.

In practice, the parameter A in this last formula may be replaced by a different parameter, θ, in order to partially take into account the asymmetry of the pdf while still retaining the simplicity of the linear mapping (in the future we plan to study nonlinear mappings that would take full benefit of pdf asymmetry). The final mapping is therefore $y' = y - \theta/k$.

As in the case of context type $1. \{A, 0, A, 0\}$; other interesting structures in HL subband which are most likely to model horizontal edges are: $2. \{A, -A, A, -A\}$, which may be seen as an extension of context type 1; $3. \{A, 0, A, A\}$ and $4. \{A, A, A, 0\}$. For all the four models just defined, we have observed pdfs similar to the one shown in Figure 3. The corresponding mappings for these models are therefore similar to the one defined for the first model possibly with different values of T_{mse} and of θ.
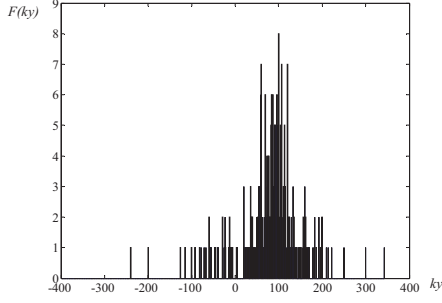
Fig. 3. Histogram of ky for model $\{A,0,A,0\} \to A$, $A=100$.

In practice, of course, a given context may produce *mse* values that are below the threshold T_{mse} of several models. In this case, the context is assigned to the model which gives the lowest *mse*.

4.2. Generalized lifting implementation

The GL decomposition has been implemented using only the predict operator (P) of Figure 1. We focus on the detail signal y produced by the context $\{x_1, x_2, x_3, x_4\}$.

The mappings derived from the four models described above have been implemented on a quincunx sampling grid, we have set $A=100$ initially, and have evaluated each model individually to optimize the threshold T_{mse} to obtain the smallest bitrate. With the set of thresholds $\mathbf{T}_{mse}$ we optimize the values of A and θ for each model and obtain $\mathbf{A}, \mathbf{\theta}$. The subband is encoded with these parameter vectors. The vector is encoded as overhead, see Fig. 4.

As an initial experiment, we have used a Haar wavelet filter to implement the DWT, other filters like Le Gall 5/3 or CDF 9/7 will be considered in the future. Two images, *barbara* and *boat* have been used to test the model-based strategy. They are shown in Figures 5 (a) and (c); Their corresponding HL wavelet subbands are shown in Figures 5 (b) and (d).

5. CODING SCHEME

The complete coding scheme is shown in Figure 4. 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 models, thus, we have used $Q=1$ as the step size of the quantizer. We apply the model-based generalized lifting (GL) operation described in Section 4.2 to the quantized coefficients with the mapping described in Section 4.1.

Finally, an arithmetic encoder serves as entropy coder. Parameter vectors $\mathbf{A}, \mathbf{\theta}$, and $\mathbf{T}_{mse}$ of the four models are sent as overhead. The overhead is around 0.2% of the total bitrate. The spatial location of the coefficients that are mapped is preserved.
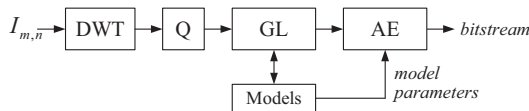

Fig. 4. Model-based GL encoder

Fig. 5. Test images, (a) *barbara*, (c) *boat*, and their wavelet-domain HL subbands (b) *barbara*, (d) *boat*

The GL decomposition is iterated as shown in Fig. 6. Context and predicted sample positions are alternated according to Figure 2. Note that mappings discussed previously are used for iteration GL-1 and GL-3. Indeed, the study of the same models for GL-2, that relies on the sampling pattern of Figure 2(b), has produced pdfs which were centered on $\theta=0$. So the resulting mapping was $y'=y$. This result is consistent with intuition as the context defined by Fig. 2(b) is not really appropriate to detect the presence of horizontal contours. Finally for each iteration GL-1 or GL-3, the set of parameters of the models must be optimized.
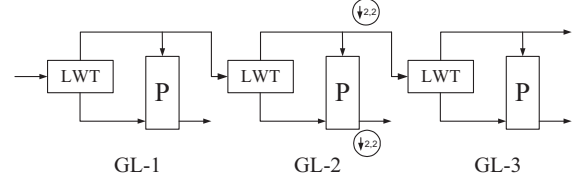

Fig. 6. Three dyadic iterations of the GL decomposition

Figure 7 shows the effect of the mapping applied to scales GL-1 and GL-3. Reduction of the magnitude of at least one out of each two coefficients along the edge can be seen in Figure 7(b); which corresponds to GL-1. GL-2 works in this example as a bypass which only separates samples through the LWT. Figure 7(c) shows the mapping effects of GL-3. Some additional coefficients along the edge have been reduced, producing longer runs of low-energy coefficients and eliminating in part the checkerboard pattern produced by GL-1 after mapping the subband. More iterations of the GL decomposition could eventually reduce to the minimum the number of coefficients with large magnitude along the edge. Implementation of a wavelet packet decomposition scheme may also improve results by successive decomposition of detail coefficients that are not mapped but are suitable of mapping after additional decompositions that involve downsampling. This will also be investigated in the future.

6. EXPERIMENTAL RESULTS

Experiments with the coding scheme of section 5 were performed with GL-1 decomposition scale, energy was measured in the detail subband before and after mapping. The results for the two test images of Figure 5 are shown in Table I. The four models (1 to 4) described in Section 4.1 were evaluated and are listed. The particular context type $\{0,0,0,0\}$ is listed as Z, and all those contexts that did not fit in any of the models are listed as NA.

Table I shows the percentage of contexts assigned to each model measured as

$$Cont_i = 100\, N_i \Big/ \sum_{i=1}^{6} N_i ,$$

with N_i the number of contexts found for model *i*. Table I also reports the energy gain, measured as
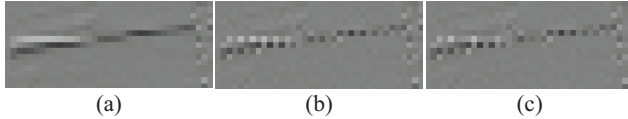
Fig. 7. Iterations of the GL decomposition for *barbara*,
(a) Quantized wavelet coefficients, (b) GL-1, (c) GL-3

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

Negative value indicates a reduction in energy.

Table I. Contribution of the models to the GL detail subband

Model	*barbara*		*boat*	
	Cont (%)	E_{gain} (%)	*Cont* (%)	E_{gain} (%)
1	7.89	-69.96	7.07	-83.69
2	16.43	-51.23	21.42	-35.71
3	13.48	-23.37	10.95	7.70
4	10.43	-14.99	9.29	-5.71
Z	1.7	–	0.93	–
NA	50.09	–	50.34	–

Models 1 to 4 account for about 50% of the total of contexts observed in the detail subband in both cases, while condition $\{0,0,0,0\}$ is non-relevant for these images. Models 1). $\{A,0,A,0\}$, and 2). $\{A,-A,A,-A\}$, exhibit the largest energy reductions. The total reduction in energy is quite significant, around 36% (Table II). In the ideal case approach of [6], complete knowledge of the image pdf was required on the received side. Energy gains were larger (about 50%) but the cost of sending the pdf to the receiver made this approach unrealistic. We now have a complete and realistic scheme which achieves comparable performance in energy reduction.

Table II. Coding and energy results for GL-1

Image	Bitrate (bpp)		Coding gain (%)	E_{gain}(%)
	DWT	GL		
barbara	4.7755	4.6859	1.91	-36.34
boat	5.2432	5.2224	0.40	-37.50

Table II presents coding results for HL subband of both test images. The detail signal produced by the mapping represents one half of the total coefficients of the subband; when the mapped coefficients are re-located at their corresponding spatial places in the subband, the effect shown in Figures 7(b) and 7(c) occurs. The energy of the detail signal is significantly reduced; however, this reduction is not reflected in bitrate. Entropy coding plays a crucial role here. DWT coefficients (Figure 7(a)) are encoded with an arithmetic encoder; the raster scan process finds series of highly correlated symbols, so each new encoded coefficient is highly predictable, reducing the overall bitrate. When we encode the mapped subband (Figure 7(b)) the series of mapped coefficients produce symbols that are uncorrelated; which in turn prevent the arithmetic encoder to achieve a large bitrate reduction. If we iterate the GL decomposition (Figure 7(c)), we may produce an increased correlation of low-energy mapped coefficients, at the expense of a larger overhead. The arithmetic coder is a straightforward approach that have enabled us obtain first coding results; however, the use of context-based or hierarchical entropy coders may greatly improve coding gain as they may take advantage of the increased number of zeros corresponding to the reduction in energy. Future research work should focus on designing an appropriate entropy coding strategy for the mapped signals.

7. CONCLUSIONS

We have introduced the use of context-based models of contours to implement a Generalized Lifting mapping that reduces the energy of the wavelet coefficients. The use of models in GL is an important improvement in the strategy of implementation of non-linear mappings for generic image coding. This first approach to the definition of models and the implementation of the coding scheme forms part of an ongoing research. Coding results are encouraging as they show it is possible to reduce significantly the wavelet coefficients energy and, to a moderate extent, the overall bitrate of the subband. Next steps in the research would involve improving the modeling of the contour contexts and the resulting mapping, the investigation of wavelet packet decomposition strategies and the design of specific entropy encoding that would better exploit the energy reduction.

8. ACKNOWLEDGMENTS

9. 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,* 6-9 Nov 2007, Lisbon, Portugal.

[7] Rolón, J.C., Salembier, P., Alameda, X., Image Compression with Generalized Lifting and partial knowledge of the signal pdf, *IEEE ICIP'08, oct 12-15, 2008*, San Diego, USA.

[8] Solé, J., Salembier, P., Adaptive discrete generalized lifting for lossless compression, in proc. *IEEE ICASSP 2004*, 17-21 May 2004 pp.iii - 57-60 vol.3, Montreal, Canada.

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

[10] Taubman, D., High Performance Scalable Image Compression with EBCOT, *IEEE Transactions on Image Processing,* vol. 9, no. 7, pp. 1158-1170, 2000.

[11] Wu, X., Nemon, N., Context-based, Adaptive, Lossless Image Coding, *IEEE Trans on Comms,* vol.45, no.4, pp. 437-444, 1997.

[12] 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.