ADAPTIVE GENERALIZED PREDICTION FOR LIFTING SCHEMES

Joel Solé and Philippe Salembier

Technical University of Catalonia (UPC)
Dept. of Signal Theory and Communications, Campus Nord, Barcelona, Spain
{joel, philippe}@gps.tsc.upc.edu

ABSTRACT

The lifting scheme is a useful tool to create different types of wavelet decompositions, including adaptive and nonlinear. Generalized lifting is more flexible and can improve lifting results, but the design of generalized prediction and update steps remains difficult for a given application. A design strategy to optimize the prediction step according to the image statistics is established. The criterion aims to minimize the detail signal coefficients energy. The scheme is used for lossless compression of several classes of images without any book-keeping or side information requirements. Promising results are reported for certain classes of images.

1. INTRODUCTION

Adaptive and non-linear decompositions can describe images in a sparser way than classical wavelets do. The lifting scheme [1] is an excellent tool for developing such decompositions, mainly because the lifting structure itself assures reversibility. In [2], a generalization of the lifting scheme was proposed in order to allow more flexible nonlinear decompositions. However, nonlinear processing has a fundamental drawback in the context of compression: filter design for an embedded lossy-to-lossless code is very difficult. In this paper[1], we focus on lossless compression. There are many applications requiring compression in which the original image should be exactly recovered, as in medical imaging because of regulatory issues or in remote sensing.

In [3], we proposed an optimization criterion to design generalized prediction. The goal of the criterion is to minimize the energy of the detail coefficients energy. Promising results were demonstrated but the drawback of the approach is that the some statistical property of the image such as its probability density function (pdf) has to be known in advance. In some application, this is not a problem since the pdf can be fixed for a specific class of images and a specific lifting can be designed for this application. In this paper, we extend the previous approach [3] by avoiding the necessity of knowing the pdf beforehand. The solution consists of using adaptive generalized prediction steps in the lifting. More precisely, we propose a scheme that iteratively updates an optimized prediction design within the discrete generalized lifting framework. This prediction applied to natural images performs close to the LeGall wavelet via lifting used in the Jpeg2000 standard for lossy-to-lossless compression. For those images with a pdf diverging from that of natural images promising coding gains are obtained. Concretely, we apply the new prediction step to biomedical (mammography), remote sensing (sea surface temperature, SST) and synthetic images and evaluate the gains provided by the proposed prediction.

In section 2, the classical and the generalized lifting are briefly reviewed. Section 3 focuses on the design of the prediction step. Results are reported in section 4 and, finally, conclusions are established in section 5.

2. GENERALIZED LIFTING

The lifting scheme (Figure 1) introduced in [1] is a well-known method to create bi-orthogonal wavelet filters from other ones. Usually, a polyphase decomposition (or Lazy Wavelet Transform, LWT) of the input signal $\underline{x}_o$ is initially done, obtaining an approximation $\underline{x}$ and a detail signal $\underline{y}$. Then, lifting steps are performed by predicting the detail signal from the $\underline{x}$ samples (1) and updating the approximation signal with the $\underline{y}$ samples (2). The so-called prediction and update lifting steps improve the initial wavelet properties

$$y'[n] = y[n] - P[\underline{x}] \qquad (1)$$

$$x'[n] = x[n] + U[\underline{y}'] \qquad (2)$$

Inspired in [4], a generalization of the lifting scheme was proposed in [2]. As can be seen in Figure 2, the generalized prediction and update steps combine the filtering stage as well as the addition of classical lifting. This leads to a more general framework, allowing more complex and possibly nonlinear operations.

To establish a formal definition of the generalized steps, let A be the set of functions a from $\Re \times \Re^k$ to itself; $a \in A \Leftrightarrow a : \Re \times \Re^k \rightarrow \Re \times \Re^k$, such that:

[1] This material is based upon work partly supported by the IST program of the EU in the project IST-2000-32795 SCHEMA.

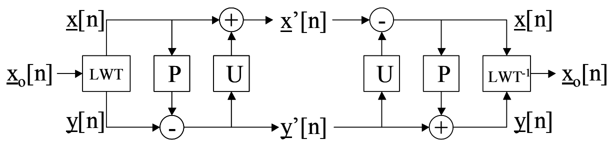

Fig. 1. Lifting Scheme

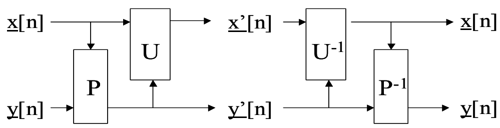

Fig. 2. Generalized Lifting Scheme

$\{z_1'[n], z_2'[n-n_1],.., z_2'[n-n_k]\}= a\{z_1[n], z_2[n-n_1],.., z_2[n-n_k]\}$

We denote the samples by z in order to maintain the same definition for both the prediction and the update steps. For the prediction (update) step, it is assumed that $z_1[n]=y[n]$ and $z_2[n]=x[n]$ ($z_1[n]=x[n]$ and $z_2[n]=y[n]$). Let A_0 be the subset of A containing all functions that do not modify $z_2[n]$, that is, for which the restriction to $\Re^k$ is the identity: $A_o = \{a \subseteq A \mid a_{|\Re^k \to \Re^k} = Id\}$. Then, a generalized lifting step is any function belonging to A_0.

In order to have a reversible scheme, the generalized prediction and update can not be chosen arbitrarily. To get reversibility the generalized steps must be bijective functions of A_0.

As presented, the scheme assumes that the values taken by $\underline{x}$ and $\underline{y}$ are real numbers. However, for lossless compression, it is useful to consider the discrete version of the generalized scheme in which it is assumed that the input and output values of the lifting steps are integers. Concretely, working in the framework of discrete gray-scale images where each pixel is represented by 8 bits, we assume that sample values range from -128 to 127. Let us call Z_{255} the set of integers that belong to [-128,127]. Then, the discrete generalized steps are functions from the $Z_{255} \times Z_{255}{}^k$ space to itself that can only modify the first component. Note that the output values are also restricted to the interval [-128,127]. The statements made for the real case are still valid. In particular, reversibility is obtained if the following a mappings are bijective:

$\{z_1'[n], z_2[n-n_1],\dots, z_2[n-n_k]\}=a\{z_1[n], z_2[n-n_1],\dots, z_2[n-n_k]\}$

For $z_2[n-n_1]$, …, $z_2[n-n_k]$ fixed, the set of all possible values of $z_1[n]$ describes a column in the $Z_{255} \times Z_{255}{}^2$ space. Let $C_{i \in Z_{255}{}^k}$ denote such a column:

$$C_{i \in Z_{255}{}^k} = \{z_1[n], z_2[n-n_1] = i_1, \dots, z_2[n-n_k] = i_k\}$$

As the generalized update and prediction can only modify the first component $z_1[n]$, they map any column $C_{i \in Z_{255}{}^k}$ to itself. In order to have a reversible scheme, this mapping should be bijective for all columns. Figure 3 illustrates the case where $k=2$.

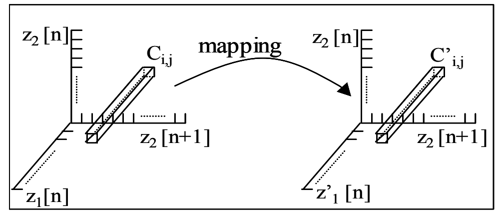

Fig. 3. Discrete mapping from $Z_{255} \times Z_{255}{}^2$ to itself. The lifting step is reversible if all mappings from every column $C_{j,k}$ ($C_{i \in Z_{255}{}^k}$) to itself are bijective.

3. ADAPTIVE OPTIMIZED PREDICTION

The optimized prediction is a transform applied to a sample $y[n]$ knowing its k neighbors, $\underline{x}[n]$. In this case, a column is defined as follows:

$$C_{i \in Z_{255}{}^k} = \{y[n], x[n-n_1] = i_1, \dots, x[n-n_k] = i_k\}$$

The filter design problem amounts to find a bijective mapping for every column $C_{i \in Z_{255}{}^k}$ of the $Z_{255} \times Z_{255}{}^k$ space to the transformed column, noted $C'_{i \in Z_{255}{}^k}$ .Columns form a partition of the $Z_{255} \times Z_{255}{}^k$ space, so the prediction (P) mappings are independent. Accordingly, every mapping (P_i) can be designed independently from each other:

$$P(y[n], \underline{x}[n]) = \bigcup_{\forall i \in Z_{255}^k} P_i(y[n])_{|\underline{x}[n]=i}$$

Given $i \in Z_{255}{}^k$, the transform relates every input value $y[n]$ one-to-one with every output value $y'[n]$. So, the output values for each i are related to the input values simply through a permutation matrix $\underline{P_i}$: $C'_i = \underline{P_i} C_i$.

The prediction step can be seen as the union of $card(Z_{255}{}^k)$ permutation matrices. Thus, the complexity associated to this formulation grows rapidly with k. In practice, one has to use a low value of k (i.e., a reduced number of context values $\underline{x}[n]$) or to take advantage of the similarities between permutation matrices that may arise.

3.1. Optimized Prediction Design

We aim at designing a mapping that minimizes the expected energy of the detail signal coefficients, $y'[n]$:

$$P_{opt} = \arg\min_{p} E\{y'^2\} = \bigcup_{i \in Z_{255}{}^k} \arg\min_{P_i} E\{y'^2 \mid \underline{x} = i\}$$

Second equality is due to the independency of columns. Then, the design of the prediction function reduces to the definition of the optimal column mapping P_i (or permutation matrix $\underline{P_i}$) for every column:

$$E\{y'^2 \mid \underline{x} = i\} = \sum_{n=-128}^{127} n^2 P(y' = n \mid \underline{x} = i) = \dots$$

$$\dots = \sum_{n=-128}^{127} n^2 P(P_i(y) = n) = \sum_{n=-128}^{127} n^2 P(y = P_i^{-1}(n))$$

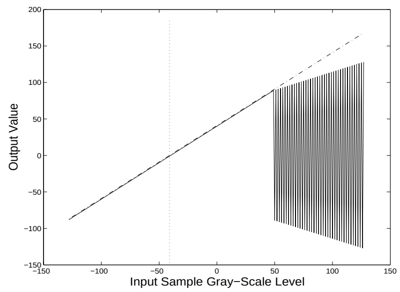

Fig. 4. Example of a mapping for natural images of the proposed prediction (solid line) and LeGall prediction (dash-dot line). Vertical dotted line indicates the mean value of both neighbors (the most probable input).

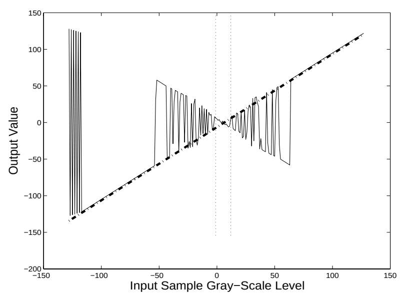

Fig. 5. An optimized mapping (solid line) for SST images class and the LeGall prediction (diagonal dash-dot straight line) for the same context (vertical dot lines indicate both neighbors values).

Because P_i is an isomorphism, this equation can be expressed as:

$$E\{y'^2 \mid \underline{x} = i\} = \sum_{n=-128}^{127} P(y = n \mid \underline{x} = i)\left(P_i^{-1}(n)\right)^2$$

$$= \underline{v}_i \left(P_i^{-1}(-128)^2 \ldots P_i^{-1}(127)^2\right)^T$$

Where $\underline{v}_i = \left(P(y = -128 \mid \underline{x} = i) \ldots P(y = 127 \mid \underline{x} = i)\right)$.
Introducing the permutation matrix, we obtain

$$E\{y'^2 \mid \underline{x} = i\} = \left((-128)^2 \ldots (127)^2\right)\underline{P}_i \, \underline{v}_i^{\,T}$$

which is minimized when the permutation relates input values of high probability with small output values.

As a result, assuming that the conditional pdf is known, a column mapping is created by constructing a vector with input values sorted by their probability in descending order. Then, the first element of this vector, which is the most probable input sample for the given context, is mapped to a *0* output value (the minimum energy output). Then, the output value *-1* is assigned to the vector second element, *1* to the third, *2* to the fourth and so on. In summary, a prediction step is performed by column mapping vectors which form look-up-tables (LUT) that re-order input values according to their probabilities. LUTs are practical representations of the permutation matrices.

The case of natural images illustrates the behavior of the proposed prediction. First, we restrict ourselves to $k=2$ (assumption which holds for the rest of the paper), and compute the pdf of a sample $y[n]$ given its two neighbors $x[n]$ and $x[n+1]$ for a set of natural images. Note that the neighbors are the same as the one used in a LeGall prediction step used in Jpeg2000 [5]. A clear pattern appears for all contexts: The conditional pdf has a maximum at the mean of the two neighbors and decreases monotonically and symmetrically on both sides. The resulting prediction has two parts, a linear one around the pdf maximum, and a nonlinear for the remaining values.

As Figure 4 shows, this prediction is the same as LeGall's, $y'[n] = y[n] - [(x[n]+x[n+1])/2]$ in the linear part (input values between -125 and 50), and differs only in the nonlinear part, which is indeed, the part of the less probable values. For this reason, no gains are obtained coding natural images, but compression improvements will be attained for images with less regular pdf. Figure 5 depicts an example where the optimized prediction mapping turns out to be very different from the LeGall's prediction mapping. In this case, the pdf is not symmetric with respect to the mean of the of the two neighbors and the optimized prediction significantly reduces detail coefficients energy compared to LeGall's prediction.

3.2. Adaptive Probability Estimation

For certain classes of images, like biomedical or remote sensing, a reasonable choice is to estimate the conditional pdf using several images of the class and then construct the LUTs, which should be available at coder and decoder in order to perform the transform and the inverse transform for other images. For $k=2$ and taking advantage of the symmetry observed in the pdf, such LUTs require around 4 Mbytes each one. Reference [3] reports compression gains up to 6% and 20% respectively for mammography and SST images. Considering that these images have a size up to several Mbytes, the coding of an even small database would justify the LUTs storage.

However, it is possible to avoid this *a priori* knowledge if a pdf estimation is performed from the actual image to code. The estimation should be updated at each sample n in a way that permits the coder and decoder to reach the same results, i.e., a synchronized iterative estimation. In this case, the prediction is matched to the image statistics. Furthermore, the pdf can be estimated for each resolution level and each direction reaching finer optimization than using a single LUT for all resolution levels and directions.

Non-parametric density estimation methods are suited for our application because they model the data without making any assumption about the form of the distribution. Kernel-based methods represent a subclass of these methods which construct the estimation by locating weighted kernel-functions at the samples index position. Experiments with different kernel shapes and bandwidths have been realized leading to similar results. We have chosen to use the simplest of them, the histogram, because it almost does not worsen results respect to other kernels and it has two other interesting properties for our purpose. First, it can be demonstrated that histogram pdf estimation converges to the optimal pdf which minimizes the detail signal energy for the image at the given resolution level and filtering direction. Secondly, in practice the choice of the histogram avoids an explicit pdf estimation that other choices would not allow: since at each sample only one histogram bin is modified, it is only necessary to re-order that bin in the vector that relates input probabilities with output values. In consequence, the time-consuming pdf re-estimation and the sorting pass of probabilities for constructing the input-output vectors are avoided.

An initial pdf estimation is required when no data is available. Different initial estimations could be envisaged, for example, an interesting approach is to use the LUT of the image class at hand and then refine the pdf on the fly for the specific image being coded. In this work, the chosen *a priori* is the pdf corresponding to natural images. At a given sample, the pdf estimation is done by adding the *a priori* (pdf of natural images) with the histogram of all samples seen until the current one. The estimated pdf is then used to optimize the prediction for the current sample.

4. RESULTS

For testing purposes, several images have been compressed with the proposed 1-D adaptive optimized prediction with 2-taps and followed by the EBCOT coder [5]. Note that no update stage is used. The image is first filtered vertically and then, only approximation signal is filtered horizontally, resulting in a three-band decomposition. It has been observed that there is no gain in applying the horizontal filter on the detail signal. The pdf is estimated twice at each resolution level, vertically and horizontally. For comparison, images are also coded with lossless Jpeg2000 (using LeGall filter with the classical lifting scheme) and with the fixed prediction (assuming the pdf is available for this image class) and followed by EBCOT. Table 1 shows results for 4 resolution level decompositions.

As explained, optimized prediction when applied to natural images tends to perform slightly worse than LeGall filter for all resolution levels. The fixed prediction method improves the results by 6% for mammography and 19% for SST images respect to Jpeg2000. Adaptive optimized prediction performs 4.5% better than Jpeg2000 for mammography and 18% better for SST images, that is, only slightly worse than fixed method but without the drawback of keeping a LUT in memory for every image class. For synthetic images (which cannot be treated as a class of image) the adaptive prediction gives compression rates up to 80% better than LeGall's. Both synthetic image examples in table 1 come from the official Jpeg2000 test set.

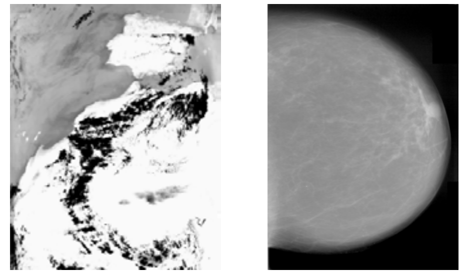

Fig. 6. Example of a Sea Surface Temperature image (left) and a mammography (right).

Bpp	Jpeg2000	Fixed Pred.	Adapt. Pred.
SST (3 Images)	2.874	2.325	2.356
Mamnography (5 Im.)	2.444	2.302	2.333
Cmpnd1	2.082	-------	1.352
Chart	3.088	-------	3.038

Table 1. Mean values for SST and Mammography classes and for 2 synthetic images using 4 resolution levels. Results in Bpp.

5. CONCLUSIONS

The generalized lifting scheme framework is used to derive a prediction step that aims to minimize the detail signal energy given an estimation of the conditional pdf of a sample. In this paper, we have shown how the conditional pdf does not need to be known in advance and can be estimated iteratively. The proposed prediction is nonlinear because of the mapping and adaptive as the pdf is progressively estimated while processing the image. Good results are obtained for images with a pdf considerably different from that of natural images, like biomedical or synthetic, and especially good results arise when the image is large enough to obtain a precise pdf estimation for most of the contexts. In our experiments, this happens for the SST images. Even if our filter support is smaller than LeGall's, compression gains are up to 20%. Note that larger supports seem to be difficult to handle in practice if the pdf does not show any structure. Moreover, to improve these results, it would be interesting to design an optimized update step following a strategy similar to the one described for the prediction. This will be the focus of our future research.

6. REFERENCES

[1] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets", *Appl. Comput. Harmon. Anal.*, vol. 3, num. 2, pp. 186-200, 1996.

[2] J. Solé and P. Salembier, "Adaptive discrete generalized lifting for lossless compression", in *Proc. of IEEE Int. Conf. on ASSP*, May 2004, vol. 3, pp. 57-60.

[3] J. Solé and P. Salembier, "Prediction design for discrete generalized lifting", in *Proceedings of Advanced Concepts for Intelligent Vision*, Sept. 2004, pp. 319-324.

[4] B. Pesquet-Popescu et al., "Adaptive update lifting with gradient criteria modeling high-order differences", in *Proc. IEEE Int. Conf. on ASSP*, May 2002, vol. 2, pp. 1417–1420.

[5] ISO/IEC 15444-1: JPEG 2000 image coding system, 2000.