

Optimal Design of Update and Predictions of Adaptive Directional Lifting based on CDF9/7 or SPL 5/3 for lossy to lossless Image Compression

Sanjay Hindurao Dabhole, Richa Verma

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. Further, this strategy modified to new adaptive lifting scheme that not only locally adapts the filtering directions to the orientations of image features, but also adapts the lifting filters to the statistic properties of image signal. The proposed approach refines previous adaptive directional lifting-based wavelet transform (ADL) by combining directional lifting and adaptive lifting filters to form a unified framework. The prediction step is designed to minimize the prediction error of the image signal, and the update step is designed to minimize the reconstruction error. Experimental results show that the proposed ADL-based on CDF9/7 or SPL 5/3 wavelet transform for image coding outperforms the conventional lifting-based wavelet transform up to 4.12 dB in PSNR and significant improvement in subjective quality is also observed. Compared with the previous ADL approach, up to 1.08 dB improvement in PSNR is reported.

Index Terms—Adaptive Directional Lifting, Cohen–Daubechies–Feauveau 9/7 Generalized Lifting, spine5/3.

I. INTRODUCTION

A. Introduction To ADL Description

The ADL is a modification of the classical lifting proposed in [Pie01a]. In the description below and for simplicity, it is assumed without loss of generality that the adaptive lifting step is an ADL. In the adaptive scheme, at each sample a lifting filter is chosen according to a decision function $D(x[n], y)$, which may be a scalar-valued, a vector-valued, or a set-valued function of R^n . The decision $D(x[n], y)$ depends on y , as in the space-varying lifting case, but it also depends on the sample $x[n]$ being modified by the ADL. The decoder knows the coefficient $xt[n]$, which is an updated version of $x[n]$ through an unknown lifting filter. Coder and decoder have different information to take the same decision. A goal in adaptive lifting design is to find a decision function and a set of filters that allow to recover the coder decision $D(x[n], y)$ at the decoder side, i.e.,

$$D(x[n], y) = Dt(xt[n], y). \tag{1.1}$$

This is the decision conservation condition. If the decision is recovered, then the decomposition scheme may be reversible. The decision function domain is the sample domain of the approximation signal X and a set of k times the detail signal domain Y , since a set of k detail samples in a window around $x[n]$ is employed for the decision,

$$D: X \times Y^k \rightarrow D(x[n], y[n]) \rightarrow d. \tag{1.2}$$

Usually, X and Y are the real numbers R , e.g. (1.3), the integer numbers Z , or a finite set of integers Z_n . The function $D(\cdot)$ maps the input samples to the decision range, which is the real positive numbers or the binary numbers. The result d is used to choose the update filter and the addition operator that merges the update filter output with the sample $x[n]$ (usually through a linear combination). The range of D may indicate whether there exists an edge at $x[n]$ if D is the l_1 -norm of the gradient or whether $x[n]$ resides in a textured region or any other geometrical constraint. Depending on the detected signal local characteristics, a suited lifting filter for these characteristics is chosen.

$$D: R \times X \times R^k \rightarrow R_+$$

$$(x[n]), (y[n]) \rightarrow \sum |y_{i-x}| \tag{1.3}$$

A relevant feature of the adaptive scheme is that it does not require any book-keeping to enable PR at the decoder side despite the filter may vary at each location using non-causal information. In this context, non-causal information is referred to information available at the coder to perform the filtering but not available at the decoder side at the time of performing the inverse filtering.

In [Pie01a], the proposed adaptive ADL employs two detail samples, i.e., $k = 2$ in expression (3.2). The restriction to $k = 2$ is also satisfied in the classical 1-D lifting with the ADL with SPL5/3 filter. The approximate signal sample $xt[n]$ is found through the update coefficients (α_d , β_d , and γ_d) for the given decision,

$$xt[n] = \alpha_d x[n] + \beta_d y[n-1] + \gamma_d y[n]. \tag{1.4}$$

The sum of the filter coefficients is defined as

$$\kappa_d = \alpha_d + \beta_d + \gamma_d.$$

Decision maps are restricted to be based on the gradient vector, noted in the following form

$$vT[n] = (v[n] \ w[n]) = (x[n] - y[n-1] \ y[n] - x[n]),$$

$$D(x[n], y[n-1], y[n])[n] = d(v[n], w[n]), \text{ where } d: R \times R$$

Manuscript published on 30 October 2016.

*Correspondence Author(s)

Mr. Sanjay Hindurao Dabhole, PhD. Scholar, Department of Electronics and Computer Engineering, University Institute of Engineering and Technology, CSJMU, Kanpur (U.P.) India.

Dr. Richa Verma, Department of Electronics and Computer Engineering, University Institute of Engineering and Technology, CSJMU, Kanpur (U.P.) India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

→ D. Observe that $v[n] + w[n] = y[n] - y[n - 1]$ does not depend on $x[n]$. Therefore, if $d(v[n], w[n])$ depends only on $v[n] + w[n]$, the scheme is reduced to the non-adaptive case.

Scenario 1.1: Consider a gradient-based decision map. In order to have perfect reconstruction it is necessary that κd is constant on every subset $D(c) \subseteq D$ given by $D(c) = \{d(v, w) \mid v + w = c\}$, where $c \in \mathbb{R}$ is a constant.

Proof: Assume that for some $c \in \mathbb{R}$ there exist $d_1, d_2 \in D(c)$ such that $\kappa d_1 \neq \kappa d_2$. Also, assume that (v_j, w_j) is such that $d(v_j, w_j) = d_j$ for $j = 1, 2$. Let the signals x_j, y_j be such that $y_j[n - 1] = q, x_j[n] = q + v_j$, and $y_j[n] = q + v_j + w_j = q + c$.

From (1.4), it is obtained $x_j[n] = \alpha d_j (q + v_j) + \beta d_j q + \gamma d_j (q + c) = \kappa d_j (q + v_j) - (\beta d_j + \gamma d_j)(q + v_j) + \beta d_j q + \gamma d_j (q + c) = \kappa d_j q + \kappa d_j v_j - \beta d_j v_j + \gamma d_j w_j$.

If q is chosen in such a way that

$\kappa d_1 q + \kappa d_1 v_1 - \beta d_1 v_1 + \gamma d_1 w_1 = \kappa d_2 q + \kappa d_2 v_2 - \beta d_2 v_2 + \gamma d_2 w_2$, which is possible since it has been assumed that $\kappa d_1 \neq \kappa d_2$, then $x_1[n] = x_2[n]$. Since $y_1[n - 1] = y_2[n - 1]$ and $y_1[n] = y_2[n]$, this implies that PR is not possible.

Definition 1.1 (Injection): Let f be a function defined on a set A and taking values in a set B . Then, f is said to be an injection (or injective map, or embedding) if, whenever $f(x) = f(y)$, it must be the case that $x = y$. equivalently, $x \neq y$ implies $f(x) \neq f(y)$.

Note that the proof of scenario 1.1 is based on the injection of the gradient based decision map. The PR condition on κd is established by assuring that whenever $x_1[n] = x_2[n]$, it is not possible that $x_1[n] = x_2[n]$ (being $y_1[n - 1] = y_2[n - 1]$ and $y_1[n] = y_2[n]$). In other words, the value $x[n]$ should be derived without ambiguity from the values of $x[n - 1]$, $y[n - 1]$, and $y[n]$. Assume now that the decision is given by the l_1 -norm of the gradient, i.e.,

$$d(v, w) = |v| + |w|. \quad (1.5)$$

In this case, the following scenario holds.

Scenario 1.2: If the decision is given by (1.5), then it is necessary that κd is constant for all $d \in D$ in order to have PR.

This result is derived from scenario 3.1. It states that κd has to be constant for every subset $D(c)$. If the decision is (3.5), then the subset $D(c = 0)$ is the whole \mathbb{R}^+ . In consequence, κd must be constant for every decision $d \in \mathbb{R}^+$. Assume in the following $\kappa d = 1$. Sufficient conditions on the filter coefficients $\alpha d, \beta d$, and γd that guarantee PR are found:

Proposition 1.1: PR is possible with previous assumptions in each of the following two cases:

1. For $\alpha d > 0$ and $\beta d, \gamma d$ non-increasing w.r.t. d .
2. For $\alpha d < 0$ and $\beta d, \gamma d$ non-decreasing w.r.t. d .

Adaptive (update) lifting has some drawbacks. Firstly, $x[n]$ is weighted with a real number, requiring quantization and thus, the decision recovery becomes in practice more difficult than stated. Also, in lossy compression this may be the cause of more difficulties to achieve PR. Secondly, the described approach imposes severe constraints on the FIR filter coefficients.

B. Generalized Discrete Prediction Design

This section discusses three approaches for a discrete ADL design. In the case of prediction, a column (1.10) is defined as

$$C_{i \in Z_{255}^k} = \{y[n], x[n - n_1] = i_1, \dots, x[n - n_k] = i_k\}. \quad (1.6)$$

The filter design problem amounts to find a mapping from every column of the $Z_{255} \times Z_k$ space to the transformed column (noted $C_{i \in Z_{255}^k}$),

$$C'_{i \in Z_{255}^k} = \{y'[n], x[n - n_1] = i_1, \dots, x[n - n_k] = i_k\}. \quad (1.7)$$

Every column mapping should be bijective for the transform to be reversible according to the considerations established in §1.4. The restriction to $k = 2$ holds. Therefore, two neighbors are considered for the prediction of the sample in between, as in the classical lifting with ADL with SPL5/3 wavelet filter.

C. Geometrical Design of the Prediction

The design proposed in [Sol04a] is outlined in this section. The approach is quite intuitive and shows the GL flexibility because the design reduces to manipulate the mapping from a three dimension space to itself according to three simple rules depending only on geometrical distances. Every point on the left space of figure 3.10 is mapped (or transformed) to a point on the right space following the three rules, which do not explicitly try to minimize any criterion but that are based on intuitive arguments. In the $Z_{255} \times Z_{255}^2$ space, the line $l: x[n] = x[n + 1] = y[n]$ plays a special role because every point p on l should be mapped to the point $(0, x[n], x[n + 1])$ to have a zero detail output if the input signal is a constant. Then, the mapping of a point $p: (y[n], x[n], x[n + 1])$ is based on its relative position and the distance w.r.t. the line l . The distance between a point and a line is defined as the minimum of the distances between the point and any point of the line. The square distance of the point p to the line l is given by distance

$$(p, l)^2 = \alpha y[n]^2 - (x[n] + x[n + 1])y[n] + (x[n]^2 + x[n + 1]^2) - x[n]x[n + 1]. \quad (1.8)$$

Note that (1.10) is the equation of a parabola respect to $y[n]$. The prediction is constructed by reordering the points of a column $C_{i,j}$ according to their distance to l and the following rules, which impose conditions on the filter.

D. Vanish the detail signal first moment.

This condition and the restriction to $k = 2$ completely specify a linear filter. In the nonlinear case, this rule only means that the nearest point of every column to l should be mapped to 0. The nearest point y_{min} is the average of the two neighbors,

$$y_{min} = y[n] = x[n] + x[n + 1]/2 \quad (1.9)$$

The rest of the points are specified by the two other rules that try to employ in the most effective way the additional degrees of freedom obtained from relaxing the constraint of linearity.

1.4.1 Continuity. A desired property is that similar inputs give similar outputs. Therefore, the prediction should be a function of $y[n]$ as continuous as possible. This is attained inside the so called linear zone (figure 5.1), where the values below y_{min} are mapped to negative integers maintaining their order, and in the same fashion, values over y_{min} is mapped to positive integers.

1.4.2 Logical nonlinear mapping. Beyond the linear zone, values are alternatively mapped to the positive and negative remaining integers. This nonlinear zone generally exists in mappings from a finite discrete space to itself. The proposal is logical because in natural images it maps the more probable remaining points to the minimum output values, thus minimizing the output energy. Obviously, other “logical” mappings exist. For instance, it is also interesting to preserve continuity by using the mapping with minimum discontinuities. It is possible to construct a mapping with only one discontinuity, with the tradeoff that it does not minimize the output energy for a wide range of images. In practice, the coding results that we have obtained are alike. The prediction based on the geometrical design (geometrical prediction for short) is equivalent to the classical ADL with SPL5/3 wavelet ADL inside the linear zone. This is verified by observing that both mappings are linear, that the output of both mappings is zero when $y[n] = y_{min}$ (1.10), and that the output varies in the same way as a function of $y[n]$ (cf. input-output relation in figure 1). Outside the linear zone, the geometrical prediction does not correspond to a simple linear filter. However, this mapping offers several advantages:

The mapping is easily computed through the distance function, avoiding the look up table usage.

1. The resulting detail samples have the typical “high-pass meaning” within the linear zone.
2. Wavelet-type output coefficients, which amount to:
 - The possibility to attain a multi-resolution decomposition.
 - The employment of usual entropy coders.
3. For most of images, the geometrical prediction detail signal energy is smaller than using the ADL with SPL5/3 wavelet, as it has been experimentally verified.
4. A resulting detail sample with typical high-pass meaning is important if an update filter follows the prediction, because it can operate as in the classical lifting. An update is useful for a multi-resolution decomposition, since posterior processing of the approximation signal in the next resolution level performs better when signal is a low-pass version of original data than when it is a simple down-sampled version.

Figure 1 illustrates the point. Then, assuming that the pdf is known, a column map is created by constructing a vector with input values sorted by their probability in descending order. The first element of this vector, which is the more probable input sample for the given context, is assigned (mapped) to a 0 output value (the minimum energy output). Following, the output value of -1 is assigned to the vector second element (corresponding the input values of second highest probability), 1 is assigned to the third element, 2 to the fourth, and so on. In practice, an ADL is performed by column mappings which are look-up-tables that reorder input values according to their probabilities. These look-up-tables are more practical representations of permutation matrices.

E. Optimized Prediction Design

This section addresses the work presented in [Sol04c]. The ADL design is formulated as an optimization problem that depends on the signal probability density function. The

resulting lifting step is applied to biomedical images (like lena, barbara) and remote sensing images (sea surface temperature) with good results.

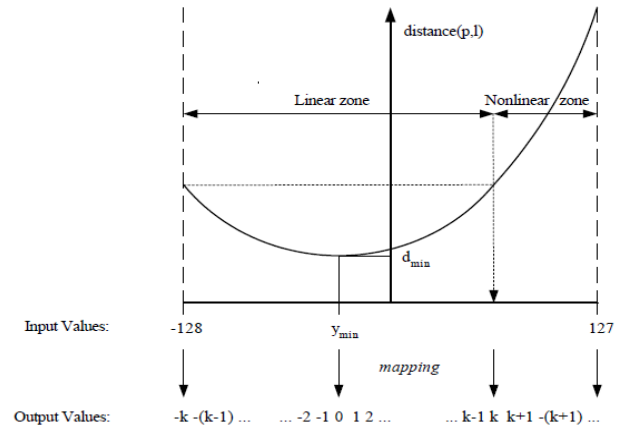


Figure 1: Distance between the points of a column to the line l and the proposed geometrical mapping for the generalized prediction.

The transform is reversible if every column mapping is bijective. Columns form a partition of the space $Z_{255} \times Z_{255}^k$, so the prediction mappings are independent. Accordingly, every column mapping $P_i(\cdot)$ is independently designed from each other,

$$y'|_n = P(y|n, x|n) = \bigcup_{\forall i \in Z_{255}^k} P(y|n, x|n)|_{x|n|=i} = \bigcup_{\forall i \in Z_{255}^k} P_i(y|n, x|n)|_{x|n|=i} \tag{1.10}$$

Given $i \in Z_{255}^k$, the transform relates every input value $y[n] \in Z_{255}$ one-to-one to every output value $y_t[n] \in Z_{255}$. Therefore, output values for each i are related to input values simply through a permutation matrix. A prediction step P is seen as the union of $|Z^k|$ permutation matrices, noted P_i . Consequently, the complexity associated to this formulation grows exponentially with k . In practice, one has to use a low value of k (i.e., a reduced number of context values $x[n]$) or to take advantage of the similarities between permutation matrices that may arise. State-of-the-art entropy coders benefit from several characteristics of wavelet coefficients. Specifically, they tend to increase their performance when coefficients energy is minimized. Therefore, a reasonable goal is to design a mapping that minimizes the expected energy of the detail signal. Such an optimal prediction is

$$P_{opt} = \underset{P_i}{\operatorname{argmin}} E|y'^2| = \bigcup_{\forall i \in Z_{255}^k} E|y|^2 | X = i \tag{1.11}$$

The second equality in (1.11) is due to the independency between columns. As a result, the design of the prediction function reduces to the definition of the optimal column mapping $P_i(\cdot)$ (or permutation matrix P_i) for all columns:

$$E|y'^2|_{x=i} = \sum_{n=-128}^{127} n^2 P_r(y' = n | x = i)$$



$$= \sum_{n=-128}^{127} n^2 P_r(P_i(y) = n) \\ = \sum_{n=-128}^{127} n^2 P_r(y = P_i^{-1}(n)) \quad (1.12)$$

Note that $Pr(\cdot)$ stands for the probability function. Expression 1.12 can be formulated as because $P_i(\cdot)$ is bijective.

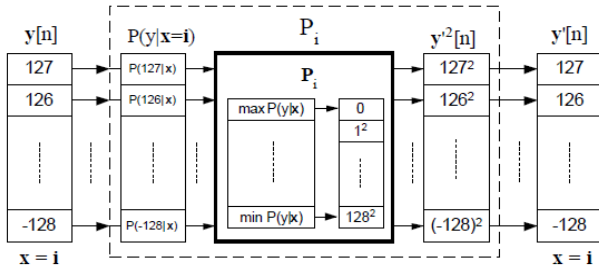


Figure 2: Optimized prediction design.

$$E[y'^2|_{x=i}] = \sum_{n=-128}^{127} P_i^{-1}(n)^2 P_r(y = n|_{x=i}) \\ = (P_i^{-1}(-128)^2 \dots P_i^{-1}(127)^2) (P_r(y = -128|_{x=i}) \dots P_r(y = 127|_{x=i}))^T \quad (1.13)$$

Expression 1.13 is obtained by introducing the permutation matrix in (1.14):

$$E[y'^2|_{x=i}] = ((-128)^2 \dots (127)^2 P_i(P_r(y|_{x=i}) \dots (P_r(y = 127|_{x=i}))^T) \quad (1.14)$$

The energy expectation in (1.14) is minimized when the permutation matrix relates input values of high probability with small energy output values. The permutation matrix optimizing that relates input conditional probabilities with output energies is used in the discrete sample space to relate each input with the corresponding output.

II. EXPERIMENTS AND RESULTS

This design strategy is applied to three classes of images: the natural images and two classes of specific images, like lena, barbara and sea surface temperature images. The last two classes are chosen because of their pdf, which differs significantly from that of natural images. Like in the previous section, the restriction to $k = 2$ holds. The decomposition is followed by an ASPIHT.

Experiment 1: Gray scale Images The probability distribution function of a sample $y[n]$ conditioned to the value of its two vertical neighbors, $x[n]$ and $x[n + 1]$, is extracted from a set of seven natural images (those of table 1.1). Figure 3 partly represents the average pdf: it is a histogram depicting the frequency of apparition of a sample value in function of the mean of its two neighbors, $m = (x[n] + x[n + 1])/2$. Note that a complete representation would have 4 dimensions because the histogram depends on both neighbors and not only their mean, but this simplified representation will allow us to analyze the system behavior. A common pattern is observed for all contexts in this pdf. Concretely, it has a maximum at the mean value m and decreases monotonically and symmetrically on both sides.

This structured pdf allows the avoidance of the ADL implementation by means of the look-up-table design previously described. Once m is computed, the conditional probability order only depends on the difference $dy[n] = y[n] - m$. The value of $dy[n]$ is related to the number of input values with higher probability than $y[n]$, which have to be mapped to lower energies than $y[n]$. Therefore, $dy[n]$ indicates the output value corresponding to $y[n]$.

For testing purposes, the four natural images are compressed with the 1-D 2-taps optimized prediction and the ASPIHT coder. No update step is used. Images are first filtered vertically and then the approximation signal is filtered horizontally, resulting in three-band decomposition. The same ADL is employed vertically and horizontally for all resolution levels. Optimized prediction performs better than the ADL with SPL5/3 wavelet for 2 resolution levels and marginally better for 8 resolution levels.

$$y'[n] = y[n] \frac{x[n] + x[n+1]}{2} \quad (1.15)$$

However, very similar results are obtained for both decompositions using the ASPIHT coder. This fact suggests that the design strategy in the case of natural images does not provide a prediction significantly different from Spine's 5/3 linear case. In order to clarify this point, let us analyze the prediction resulting from the optimization strategy. The kind of prediction mapping that arises from the natural images pdf has two differentiated parts, which are a linear and a nonlinear part. Figure 3 shows the prediction mapping when the context is $x[n] = x[n + 1] = -28$. The context value is indicated by a vertical line at -28. Input values between -128 and 72 are linearly mapped to output values between -100 and 100. This mapping is almost equivalent to the linear combination of the linear part of the mapping is due to the pdf that has a maximum in m and decreases monotonically and symmetrically on both sides. In fact, when the conditional pdf has this shape (figure 3) the optimized design coincides with the geometrical design of 1.5. For input values above 72, the mapping is highly nonlinear but it arises from the choice to work with a discrete finite output space, with values between -128 and 127. As a result of this analysis, it can also be deduced that the mapping is the same as the ADL with SPL5/3 prediction filter for most probable input values.

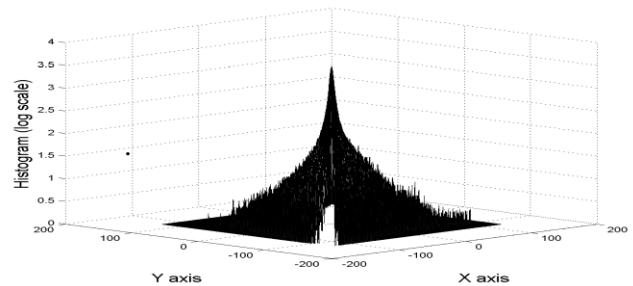


Figure 3: A pdf approximation (logarithmic scale histogram) of $y[n]$ (Y axis) conditioned to the mean value of its two vertical neighbors (X axis) for the set of natural images.

Therefore, a powerful coder like ASPIHT returns practically the same results for both decompositions. On the other side, there is a potential compression gain for those images belonging to a class with a pdf that significantly differs from that of the natural images.

III. EXPERIMENTS AND RESULTS

This experiment is presented as in [Sol04a]. Geometrical ADL performance is assessed in a multi-resolution

Table 1.1: Natural images rate for 4 and 8 resolution levels using ADL with SPL5/3 wavelet (column headed by “5/3”), and the geometrical prediction (column “G. Pr.”) followed by ASPIHT coder. Results are given in bits per pixel (bpp).

Rate (bpp)	4 resolution levels			8 resolution levels		
	CDF9/7	SPL5/3	G. Pr.	CDF9/7	SPL5/3	G. Pr.
Lena	7.034	7.207	7.827	6.243	5.882	5.831
Baboon	8.418	8.202	8.784	7.740	7.419	7.446
Barbara	7.828	7.498	8.602	6.787	6.449	6.481
Peppers	7.648	7.475	8.058	6.513	6.265	6.276
Plane	7.235	7.973	7.565	5.896	6.549	5.451
Mean	7.633	7.671	8.167	6.635	6.512	6.297

Framework. To this end, the scheme is completed with a space-varying update which varies according to the modulus of the detail signal samples.

$$x[n]' = \begin{cases} x[n] & \text{if } \max(y'[n-1]) > T, \\ x[n] + [(y'[n-1] + y'[n])/4], & \text{Otherwise,} \end{cases} \quad (1.17)$$

A sample $x[n]$ is updated with two detail samples $y[n-1]$ and $y[n]$. If the modulus of these detail samples is small, then, as a first approximation, they are high-pass coefficients and can be directly used by the ADL as classical updates do.

Detail samples with large values mean, also as a first approximation, that $y[n]$ comes from an edge. If a smooth x_t is desired, edges should not flow to lower resolution levels and consequently, no update is performed. Values have small or large value according to a threshold T , fixed to 12 as the best value after several experiments with natural images. Since values in the ensuing resolution levels may not have the same dynamic range, the discrete generalized prediction is modified to handle an arbitrary range of values. The algorithm is the same, but the range of values has to be sent to the decoder to recover the original data. One resolution level is obtained by first filtering every row and then only the columns of the approximation image. This leads to three-band decomposition. The method is applied to 7 natural images and compared to two non-adaptive wavelet filters: the CDF and the ADL with SPL5/3 wavelets. Decompositions are followed by the SPIHT coder. Resulting bit-rates are shown in table 1.1. For two resolution levels and the tested images, the proposed scheme performs around 4.5% better than ADL with SPL5/3 wavelet. For three levels, results are only slightly better than Spine’s. This decrease of gain is possibly due to the worse multi-resolution performance of the prediction and update filter,

which are not the best choice for obtaining a good approximate signal for further processing.

The geometrical ADL based decomposition is also applied to the MRI group of images through the three dimensions. The decomposition has 4 bands per resolution level, instead of the usual 8 bands. The transformed coefficients are coded with ASPIHT, which is detailed in

Rate (bpp)	ADL with SPL5/3	Geom. Pred.
2 res. lev.	5.980	4.943
4 res. lev.	4.798	3.731
8 res. lev.	4.667	3.597

Table 1.2 contains the final bit-rates. In contrast with the natural images experiment, the MRI is better compressed for all resolution levels with the geometrical approach. For this set, the geometrical ADL reduces the detail signal energy w.r.t. its linear counterpart. Meanwhile, the space-varying ADL is effective enough to obtain approximation signals which are good for its further decomposition.

Table 1.2: MRI set compressed with ASPIHT using ADL with SPL5/3 and the geometrical prediction.

The generalized ADL is optimized w.r.t. the image probability density function (pdf). The assumption of an underlying image pdf has revealed useful in our practice. There is a typical pdf which leads to an optimized prediction mapping equivalent to the geometrical prediction, thus explaining the given compression results.

A. EXTENSIONS

The transform support of the proposed decomposition is 3×1 and 5×3 pixels per transformed coefficient for the H1 and HL1 bands (the first detail bands). This support is smaller than the 2-D ADL with SPL5/3 wavelet support, which is 3×3 , 3×5 , and 5×3 for the HH1, HL1, and LH1 bands, respectively. Therefore, the proposed geometrical prediction scheme obtains better compression results using less information (i.e., less input samples contribute to obtain an output sample). However, difficulties arise in the generalization of the proposal to larger supports. The mapping between 3-D spaces becomes a less intuitive higher dimensional mapping. In each case, a geometrical place that plays a role similar to the line l in the proposed scheme has to be found. For instance, let us analyze the case $k = 4$, in which four neighbors are used for the generalized prediction. In a similar manner to $k = 2$, the line l : $x[n-1] = x[n] = x[n+1] = x[n+2] = y[n]$ formed by the points that have all the components equal may be considered. In this case, the nearest point of every column to l is $ymin = y[n] = x[n-1] + x[n] + x[n+1] + x[n+2]/4$, which is also the average of the neighbors, as for $k = 2$. The mapping with l is interesting in two dimensional image grids: it is logical to predict a pixel with the mean of its left, right, up, and down neighbors. However, the resulting mapping does not vanish four moments of the detail signal, which is a property attainable with $k = 4$.

To this goal the line $l2: -x[n-1] = 9x[n] = 9x[n+1] = -x[n+2] = 41/4 y[n]$ is the appropriate. The line $l2$ is related to the Lagrange interpolating polynomial of degree three with equidistant points. The mapping arising from $l2$ vanishes four moments.

The following experiments 2 and 3 illustrate results for two classes of biomedical and remote sensing images with such type of pdf.

Experiment 2.4: Colorful Images: the set of the first 11 like lena, Barbara, etc images from the database is selected to realize the experiment for biomedical images. The size of these images is about 1 MB without compression. Six like lena, barbara are used to estimate the pdf for this class of images. The resulting pdf does not exhibit a regular pattern as in the case of natural images. As figure 3 shows, like lena, barbara images pdf is not as structured as natural images pdf. The histogram neither is nor does symmetrical neither decreases monotonically. Usually, several maxima appear and also, darker values are rather probable for most of the contexts. Figure 4 depicts the mapping when $x[n] = x[n+1] = -88$. As it may be observed, the mapping of the most probable input values (around -88) is quite nonlinear. The decomposition is performed for the 5 like lena, barbara not used for the pdf estimation. They are compressed with ASPIHT.

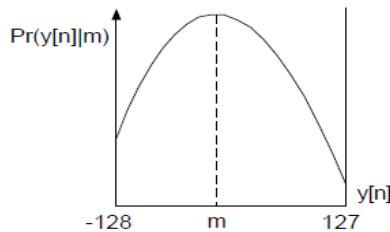


Figure 4: Example of an optimized prediction mapping (solid line) for natural images and ADL with SPL5/3 prediction (dash-dot line) for the same context (vertical dot line indicates both neighbors value).

For comparison, images are also decomposed with the ADL with SPL5/3 transform and followed by ASPIHT Results (table 1.3) are 5% to 6% better for the generalized prediction for all resolution levels.

B. ADAPTIVE OPTIMIZED PREDICTION DESIGN

This section explains a modification of the optimized prediction presented in [Sol05] that avoids the necessity of the previous knowledge of the image pdf and thus, also avoids the storage of a LUT for every image class at the coder and decoder side. Furthermore, it may avoid any LUT storage at all if the application at hand requires it, but at the cost of further processing. Indeed, in this approach the LUT may be different from level to level and for each filtering direction, which may result in compression gains w.r.t. one fixed LUT per image class. The drawback is the computation cost of adaptive pdf estimation.

The pdf estimation should be updated at each sample n in a way that permits the coder and the decoder reaching the same results, i.e., a synchronized iterative estimation. Therefore, the prediction is adapted to image statistics and even, the pdf may be independently estimated for each resolution level and each direction reaching finer

optimization than using a fixed LUT.

Non-parametric density estimation methods are suited for this application because they model data without making any assumption about the form of the distribution. A kernel-based method is a subclass of these methods which construct the estimation by locating weighted kernel functions at the index position of the samples. Experiments using different kernel shapes and bandwidths have been carried out leading to similar results for a wide range of values. The delta function has been chosen as the kernel. It is the simplest kernel and amounts to the computation of the histogram. The delta kernel is the choice because its results are not worse w.r.t. other kernels and it has two interesting properties for our purpose. First, it is demonstrated that histogram pdf estimation converges to the optimal pdf that minimizes the detail signal energy for the image at the given resolution level and filtering direction. Second, in practice, the choice of delta 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.

Initial pdf estimation is required when no data is available. Different initial estimations may be considered. 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.

Table 1.3: Bit-rates comparison. Mean Values for gray scale classes and for 2 synthetic/compound images using 4resolution levels.

Rate (bpp)	JPEG2000	Fixed Pred.	Adap. Pred.	JPEG-LS
Lena	3.874	3.326	3.356	3.322
Barbara	3.444	3.302	3.333	3.355
plane	3.082	3.431	2.352	2.242
peppers	4.088	3.579	4.038	3.836

For the following experiments, 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.

IV. EXPERIMENTS AND RESULTS

For testing purposes, several images are compressed with the proposed 1-D adaptive optimized prediction with 2-taps and followed by the ASPIHT coder. Note that no ADL is used. The image is first filtered vertically and then, only the approximation signal is filtered horizontally (resulting in a three-band decomposition) because it is observed that applying the horizontal filter on the detail signal damages results. The pdf is estimated twice at each resolution level: vertically and horizontally.

For comparison, images are coded with lossless JPEG2000 using ADL with SPL5/3 filter and with the fixed prediction (assuming the pdf is available for this image class) and followed by ASPIHT. Table 1.3 shows results for 4 resolution level decompositions.

Optimized prediction when applied to natural images tends to perform slightly worse than ADL with SPL5/3 filter for all resolution levels. Adaptive optimized prediction performs 4.5% better than JPEG2000 for like lena, barbara and 18% better for GRAYSCALE images, that is, only slightly worse than the 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 images) the adaptive prediction gives compression rates up to 80% better than Spine’s 5/3. As an example, results for two images from the official JPEG2000 test set (cmpnd1 and chart) are given in table 1.4. The adaptive optimized ADL is also applied to the MRI group of images through the three dimensions. The transformed coefficients are coded with ASPIHT. Table 1.4 shows the bit-rates, compared to those obtained with Spine’s 5/3, the geometrical ADL based decomposition and the optimized fixed prediction for the barbara image, which is the point of departure for the adaptive prediction rate (bpp). These images are composed of text, figures, and natural images, so they tend to significantly worsen results of adaptive prediction with respect to “pure” synthetic images. JPEG-LS bit-rates are also given. These results show how the conditional pdf does not need to be known in advance: for a wide range of images it may be adaptively estimated.

Table 1.4: MRI set compressed with ASPIHT 3-D using ADL with ADL with SPL5/3, the geometrical prediction optimized for the images (Fixed Pr.), and the adaptive prediction.

Barbara 2D Image	ADL with SPL5/3	Geom. Pr.	Fixed Pr.	Adaptive Pr.
2 res. lev.	5.980	4.943	4.740	4.632
4 res. lev.	4.798	3.731	3.745	3.618
8 res. lev.	4.667	3.597	3.635	3.508

The fixed prediction behaves better than ADL with SPL5/3 despite the fact that the chosen pdf is that of the natural images, which does not correspond to the MRI pdf. On the contrary, the geometrical ADL based decomposition attains better results than the fixed optimized ADL for 2 and 4 resolution levels. This means that for the MRI set the space-varying ADL is beneficial for the multi-resolution decomposition. The adaptive prediction starts with the natural pdf, but it successfully captures the underlying MRI pdf, since the final bit-rate is the best. The huge size of this set and the similarity of all images help to reach finer adaptation.

A. CONVERGENCE ISSUES

The following experiments are performed for the assessment of the adaptive pdf estimation convergence. Prediction mappings are constructed using the pdf estimated at different image points. The initial prediction coincides with the natural images prediction. Then, the pdf is progressively estimated and thus, the prediction at the end of the process

employs the optimal image pdf. Since the prediction goal is to minimize the energy of the detail coefficients, mappings constructed at different points are used to decompose the whole image and the resulting detail coefficients energy is computed.

Figure 5 depicts the evolution of the normalized detail energy depending on the percentage of the image used to construct the optimized prediction. Energy decreases in almost all cases. The convergence curve with two slopes of the Barbara image (appendix A) is due to the evident non-stationarity of this image. In the right-half of the image highly textured regions appear, like the striped trousers, which belong to contexts never seen before. When the probability conditioned to these contexts is learned, prediction that poorly performed in such difficult regions becomes quickly adapted to the new pattern. Therefore, detail energy decreases strongly after the 50% of the image is analyzed. Because of the variety of contexts and the small image size, the adaptive algorithm is able to capture Barbara image behavior. However, notice that this knowledge is obtained a posteriori. In practice, the adaptation to patterns that are initially difficult to predict is effective when they appear repeatedly. For the like lena, barbara, the curve is smoother because all contexts are quite similar and the pdf does not differ considerably from that of natural images. This difference is more remarkable for the gray scale image, fact that causes the strong decrease in detail signal energy, which is also due to the higher number of different contexts (for the land, sea, and cloud regions).

Previous considerations lead to the conclusion that the adaptive algorithm is able to learn the image statistics. It remains to establish if this learning implies energy minimization. The experiment to prove this is summarized in figures 2 and 3. It shows the relative detail signal energy obtained by the decomposition of the image using the adaptive optimized prediction w.r.t. the detail energy obtained using the initial pdf optimized prediction. The energy is the mean of each column. A 10-tap low-pass filter has been applied to the plot in order to avoid an annoying jitter and thus making easier the observation of the more global trend. The visual inspection of figure 1.15 reveals that the learning is much more effective for the gray scale image. Meanwhile, the adaptive estimation efficiency is less obvious for the like lena, barbara. In mean, the detail signal energy is lower. However, for some regions the adaptive algorithm performance is worse than the non-adaptive case. This is due to the fact that the statistics vary in a way that produces a better performance of the initial pdf than the adaptive estimated pdf, i.e., the region statistics match better to the initial pdf assumption. It may be concluded that the adaptive algorithm attains its best performance for those images with slow-varying statistics which differ from the natural image pdf.

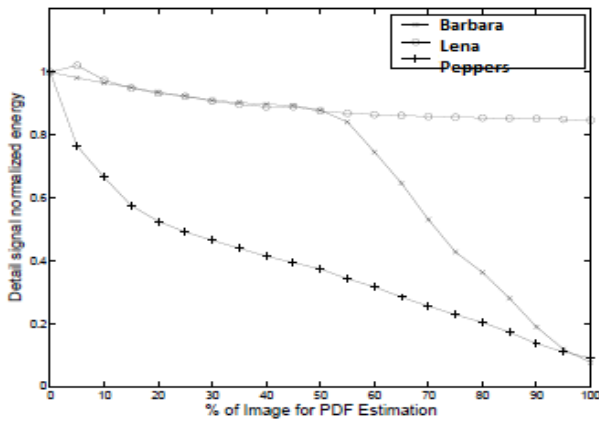


Figure 5: Adaptive prediction convergence for 3 images. Vertical axis is normalized by the energy obtained using natural images pdf mapping on the image.

Figure 5 seems to confirm the conclusion. The adaptive scheme worsens the energy obtained for the natural image Barbara w.r.t. the initial pdf. Meanwhile, the image lena attains a considerable energy reduction due to the statistics of the letter region. Note the relative energy is one for the white regions on both sides of the image: energies are equal because both filters are the same, since the optimized prediction is that given by the initial pdf. The results for these two images highlight the relation between energy reduction and compression performance. The adaptive scheme reaches very good compression rates for image lena, whereas it does not improve results for Barbara.

B. UPDATE STEP OBJECTIVES

The most usual lifting structure is the prediction-then-update (or update-last) structure, in which the polyphase decomposition is directly followed by an ADL. First, even samples channel is used to extract redundancy from the odd channel. The differences, which are the detail or wavelet coefficients, are left in this odd channel. Details are small, except at significant features (those difficult to predict from neighboring data). Then, wavelet coefficients are used to update the even channel in order to obtain a coarse scale version of the input signal that approximates this input as accurately as possible. The ADL can be seen as an anti-aliasing filter after the data splitting, i.e., it has the same objective as the low-pass filter in the classical filter bank implementation. In these prediction-then-update schemes the function of the ADL is twofold: to obtain an accurate approximation signal for embedded coding and to ensure that this signal is useful for the next resolution level processing.

A problem arises when the ADL is nonlinear, because it is not clear how to construct an ADL that preserves signal for further processing. Frequency localization is a main property for filter design that is lost. In consequence, powerful linear signal processing tools (as Fourier or z transforms) are not available any more. The non-existence of their nonlinear equivalent seriously limits the ability to face the challenge. Possibly for this reason, there are no works in compression applications (to the author's knowledge) with an update after a nonlinear prediction. In consequence, a down-sampled version of the original signal (that is, the approximation subsignal without update) seems

to have better multi-resolution properties than any output of two nonlinear filter stages. In lossy subband Image coding, some authors, e.g. [Luo01] have reported that the ADL degrades rate-distortion performance and should be omitted altogether, leading to a truncated wavelet transform. This is a controversial assert since other works [Gir05, Til05] suggest that an accurate design of the ADL produces better rate-distortion curves. To sum up, in the video coding field the appropriate motion compensation for an update step is not obvious at all. Two ways may still be open for the prediction-then-update architecture with nonlinear filters. The first one is to assume signal frequency interpretation despite it is false. As a first approximation it may be useful. However, the assumption forces to remain near the linear restrictions. A second way is to definitively free the scheme from linear ties at the cost of resigning oneself to the reduced remaining tools at hand. This path, fully interpreting signal from the statistical point of view. The ADL assumes that the image pdf is known and minimizes the signal entropy, which is a common optimization criterion in compression applications. In some proposals the update is the first of the steps after the splitting, an update-then- prediction (or update-first) lifting scheme. In this architecture, while ADL goal remains the same, the ADL purpose may change. The following analyzes the ADL purpose in an update-first structure.

Linear filter banks may be reversed, interchanging analysis and synthesis stages and so, the corresponding underlying biorthogonal basis. From this point of view, the linear space is equally partitioned independently of the order of the bases. Therefore, if the linear prediction- then-update structure is reversed the result is still a wavelet filter bank. For instance, the 5/3 wavelet becomes a 3/5 wavelet that has an update as first lifting step in the analysis side. Also, there exist families of wavelets which have larger high-pass filter, for example the (1, N) Cohen-Daubechies-Feauveau family [Coh92]. All these examples fit in the update-then-prediction lifting structure. In this situation, the update is a low-pass filter that pretends to conserve signal running average. However, a scaling factor is required in the even channel in order to maintain the expected signal mean. The coefficients after the scaling factor are real-valued. The inclusion of a rounding operator at this point is not possible without losing information, so these schemes are unsuitable for lossless coding applications. They are used in lossy compression, but such a filter bank produces more ringing artifacts in the decoded image at low bit-rate. A second ADL purpose is to preserve singularities in the approximation signal. Image salient structures that may carry most significant information are preserved at coarser scales. The drawback is precisely that through these structures the prediction is difficult and a singularity preserving ADL makes this difficulty to be found in the approximation signal throughout all the resolution levels, thus possibly damaging the global performance. However, this approach is interesting for embedded coding.

Finally, the update-then-prediction structure has an advantage if the ADL is linear and the transform is only iterated on the low-pass channel because then all low-pass coefficients throughout the entire decomposition linearly depend on the original data and so, they are not affected by the prediction step, which may be nonlinear and freely designed. Moreover, the ADL is only based on low-pass coefficients. Once surveyed these objectives, the possibility to include them in the design of an ADL within the discrete generalized lifting framework remains to be analyzed. Some of the reviewed hints are retaken in the following sections for an ADL design. Perhaps the most damaging consequence of the nonlinear processing choice is that the multi-resolution analysis of nested subspaces is $x[n]$ abandoned far behind, so output signals may not possess, in the classical sense, any multi-resolution property.

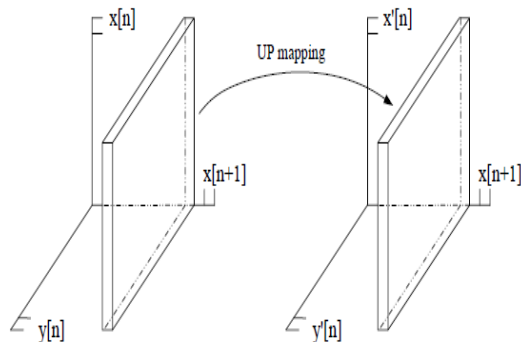


Figure6: Joint update-prediction (UP) mapping between 3-D spaces.

The importance of this point is illustrated in the joint update-prediction design, in which despite of the very low energy of the detail coefficients, the algorithm may be hardly iterated on the approximation signal.

C. JOINT UPDATE-PREDICTION DESIGN

The goal of this scheme is to permit a joint design of both lifting steps with the same objective: to minimize the detail signal energy. The idea is to extend the GL to let two samples be modified at the same time. In this situation, the column mappings become a plane mapping in which an approximation and a detail sample are transformed at the same time (figure 2). This permits to avoid a separate design of the lifting steps when in fact their common objective is to obtain good transforms for compression. Therefore, the set of a prediction and the successive update mapping is embedded in one joint update-prediction (UP) mapping. The construction of the UP step is analogous to that of the optimized prediction and the same knowledge about the image pdf is assumed. The conditional probabilities of the points within a plane $Pr(x[n], y[n]|x[n + 1])$ are ordered and, according to this order, the points are mapped to the transformed plane. Like in the optimized prediction case, most probable inputs are mapped to smallest energy outputs. In this 2-D map, a second criterion is required to distinguish the Z255 outputs with same detail $y[n]$, that is, the value of the approximation sample $x[n]$. The choice is to select $x[n]$ in order to improve the horizontal (vertical) filtering that follows the vertical (horizontal) filtering. According to this, the most probable input is mapped to the output value that maximizes the probability of $x[n]$ conditioned to its

horizontal neighbor $x[n]$, the second most probable input is mapped to the second horizontally conditioned output, and so on. Both criteria and mappings are mixed in an UP step with the lexicographical order.

The scheme is applied to the like lena, barbara set. Images are decomposed in two and three resolution levels, iterating the algorithm on the approximation bands. The energy of each band is computed. Energy for the first high band (H1) is from 2 to 6 times smaller than using the optimized prediction. Most of the detail coefficients are zero. On the other hand, LH1 bands have higher energy than in the optimized prediction case. Compression results with ASPIHT are slightly worse than JPEG2000. When the algorithm is iterated resolution level, new detail bands energy dramatically increases. This is due to the bad multi-resolution properties of the UP step, which only aims to minimize detail bands without creating a good signal for further processing it. The approximation band entropy is elevated, and its aspect is similar to noise. Results with ASPIHT for 3 resolution levels get worse with respect to those of JPEG2000. However, notice that this entropy coder is not suited for the kind of approximation signal supplied by the UP transform.

D. UPDATE-LAST DESIGN

The entropy minimization criterion is proposed in this section for the ADL design with the prediction-then-update structure.

Experiments and Results: This last option has been chosen to perform an experiment with the like lena, barbara image classes. Three pdf are estimated for constructing three LUT using an image training set: one LUT for the optimized vertical prediction, one for the vertical optimized update minimizing the entropy, and the last one for the optimized horizontal prediction. As done with previous experiments $k = 2$, i.e., two approximation neighbor samples are employed for each ADL and two detail neighbor samples for the ADL. With such LUT, a two level decomposition is computed for the test images set. Results are given in table 1.5. The entropy descends for training images, but only sometimes for the test set. The compression rate does not improve w.r.t. the no-update case. This is a shocking result, but it may be explained. Output samples have little relation among them, since their entropy is minimized without any other consideration. The value of one sample gives little information about their neighbor's value, so the following prediction performs poorly. Also, the gains in entropy are quite small because detail samples do not partition probability space efficiently: approximation samples are almost independent of the values of the detail samples. Furthermore, because of the low degree of dependence, the update pdf is very image specific, considerably varying from one image to other even within the same image class, so the estimated pdf usefulness is very restricted. These drawbacks are the cost to pay for the chosen nonlinear prediction. However, ASPIHT entropy coder is not suited for the kind of signals supplied by the entropy-minimizing ADL transform.



ASPIHT expects approximation coefficients that are quite different from the ones supplied by the transform. Surely, an entropy coder specifically created for such a transform and its output signal statistics would improve results.

E. UPDATE-FIRST DESIGN

The use of a geometrical approach similar to the one like lifting structure for the construction of an ADL leads to an update-first lifting step equal to the identity because of the restrictions of the problem: the operator is integer-to-integer by definition and it should preserve the ranks of the discrete input output space. If a linear part is included in order to attain multi-resolution properties, then the only reasonable ADL respecting such restrictions seems to be the identity operator.

Table 1.5.: Like lena, Barbara and GRAYSCALE images first vertical approximation image entropy. Comparison between down-sampled images (No. Up.) and the entropy optimized update output (Opt. Up.). Images are divided into the training set at the top and the test set at the bottom.

Entropy	Vertical Approximation Image					
	Image	No Up.	Opt. Up.	Image	No Up.	Opt. Up.
	lena	7.213	7.012	GRAYSCALE	7.097	6.940
	barbara	6.880	6.721	GRAYSCALE	6.767	6.619
	plane	7.242	7.029	GRAYSCALE	6.302	6.182
	Lena	6.834	6.678	COLOR	5.341	5.127
	barbara	6.647	6.722	COLOR	6.380	6.096
	plane	5.902	5.925	GRAYSCALE	5.512	5.532
	peppers	7.131	7.004	GRAYSCALE	5.160	5.250

Alternatively, the update-first design problem may be seen from the pdf point of view with the goal of minimizing the detail signal energy arising from the subsequent prediction. In this case, for natural images it happens that the most probable value is already at the optimal position in the space for the ensuing prediction, then the second most probable is at the second most probable point, and so on. In conclusion, the resulting optimal ADL is also the identity. Another criterion is required. This section proposes to employ the knowledge of the pdf to minimize the approximation signal gradient, which is a nonlinear version of the ADL designs proposed.

F. GRADIENT MINIMIZATION (GM)

In the entropy-optimized prediction design 1.6, somehow the multi-resolution image properties have been destroyed in return for a small gain in entropy terms, which is not enough for obtaining compression improvements using ASPIHT. While entropy may decrease, the difference between neighbor samples tends to be more random. This leads to design a lifting that may preserve multi-resolution properties; in this case, by minimizing the gradient between the samples which are neighbors in the coarser resolution level. In the proposed scheme, the update is the first of the lifting steps and it creates a 2-D approximation image. Update acts on one of every two samples of the coarser resolution level: approximation image is partitioned in two quincunx grids and one of these two grids is modified by the

update. The other rests unmodified in order to retain some of the original image statistics.

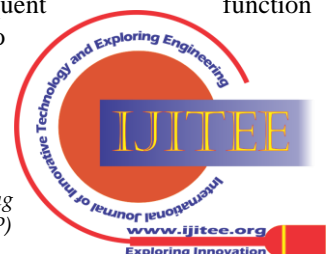
The update mapping is the equivalent to that of section 2.4 minimizing the entropy. In this case, the labels are chosen to minimize the gradient of the updated sample with respect to its four-connected neighbors at the coarser level (which are not updated). Then, the ADL is performed vertically and horizontally leading to multi band decomposition.

V. CONCLUSIONS

The ADL algorithm can be slightly modified. If the transform is performed backwards, i.e., starting the prediction process at the coarsest approximation band and estimating the pdf, and then computing the coefficients from coarse to fine scales, the coder and decoder can be kept perfectly synchronized. By these means, finer bands, which are also the larger, are coded with a pdf estimated from coarser resolution levels that may lead to better results than using a “blind” initial pdf. Experiments show that this gain exists, but it is marginal. This ADL-based on CDF9/7 or SPL 5/3 wavelet transform approach is tested with the like lena, Barbara, plane, peppers set. Images are decomposed in three resolution levels in order to establish whether the algorithm can be iterated on the approximation signal keeping reasonable results. The approximation image is observed to be smoother than without the use of the update. In the opposite, detail energy is higher than employing only the optimized prediction. In mean, compression results are outperforms better than JPEG2000 with 8 resolution levels, up to 4.12 dB in PSNR and significant improvement in subjective quality is also observed. Compared with the previous ADL approach, up to 1.08 dB improvement in PSNR is reported.

A. FUTURE SCOPE

- Adaptive directional lifting may be improved in several ways. Usually, the underlying SPL 5/3 transform support is smaller than the 2-D other wavelet family support. This fact leads to the extension of the proposal by enlarging the generalized lifting support, possibly improving the current results. A proportional increase of computation and memory requirements should be avoided if possible. Eventually, 2-D transforms may be developed, both non-separable and 2-D direction-selective. Finally, an important effort should be focused on the update step development. One possibility is to identify appropriate update steps for the predictions proposed in chapter three. A second way is to study variations on the gradient minimization generalized structure, which seems promising.
- The CDF 9/7 framework deserves more attention. For instance, this transforms may be evaluated in lossy compression. Perhaps, the objective structure may be changed in order to design specific transforms for lossy compression. The formulation accepts many other design modifications: a possible approach is to design different adaptive lifting for the even and odd samples, which makes sense because the subsequent function of these samples is also different.



Another working topic within the linear framework is the search and use of other image models than the auto-regressive one.

- The discrete adaptive directional lifting version is a fruitful approach that has demonstrated its richness. However, it may restrict the generalized lifting scope since it is a concrete case of the continuous version. The continuous generalized lifting requires a further study which may lead to appropriate mappings for lossy compression applications. There are some practical implementation issues that could help the verification of the potential of this work and that would extend its range of application:

- The entropy coders used in this work are devoted to specific linear wavelet transforms. Better results are expected for the spine/CDF scheme by using an entropy coder that takes into account the characteristics of the nonlinear coefficients. The development of a simple JPEG-LS-like entropy coder for the nonlinear transforms would be interesting.

REFERENCES

1. G.Wallace, "The JPEG still picture compression standard", IEEE TCE, 38, 2012.
2. I. Daubechies, W. Sweldens, 2012, "Factoring wavelet into lifting steps", J. Fourier Anal. in Proceedings of Advanced Concepts for Intelligent Vision Systems Wavelets By K.P. Soman, K.I. Ramachandran, N.G. Resmi, Third Edition, PHI Publication, ISBN: 978-81-203-4053-4
3. O.Strome et al, 2013, "Study of wavelet decompositions for image compression by software codecs", IEEE, pp.125–132.
4. N.V. Boulgouris, DimitriosTzavaras and Michael Gerassimos Strintzis, 2011, "Lossless image compression based On Optimal prediction, adaptive lifting and Conditional arithmetic coding", IEEE Transaction on Image Proc., Vol.10 (1), pp.1–14.
5. Omer N.Gerek and Enis Cetin. A,(2013) ," A 2-DOrientation Adaptive Prediction Filter in Lifting Structures for Image Coding",IEEE Transactions on Image Proc., vol.15, pp.106- 111
6. ÖmerNeziğGerek, A.Enişçetin, 2010, "Adaptive polyphase decomposition structure for image compression", IEEE Trans. OnImage Processing, Vol.9 (10).
7. Piella G.andHeijmans, H. J. A. M. (July. 2012) "Adaptive lifting Schemes with perfect reconstruction." IEEE Transactions on Sign Processing, vol. 50, no. 7, pp 1620– 1630,.
8. R.L. Caypoole, G.M. Davis, W. Sweldens, and R.Gboranuk, 2013, "Nonlinear wavelet transforms for image coding via lifting", IEEE Trans. on Image Processing, Vol.12,pp.1149–1459.
9. Said A. and William A. Pearlman.,(2009), "A New, Fast Efficient Image Codec Based Set Partitioning Hierarchical Trees", IEEE Transaction on Circuit and Systems for Video Technology,vol.6 No.3,pp 243-250
10. Sanjay H. Dabhole, , Johan Potgieter, An efficient modified structure of cdf 9/7 wavelet based on adaptive lifting with spiht for lossy to lossless image compression. IEEE explore digital library of Signal Processing, Image Processing and Pattern Recognition - 2012.
11. Sanjay H. Dabhole, Johan Potgieter, Performance evaluation of traditional and adaptive lifting based wavelets with spiht for lossy Image compression, IEEE explore digital library of Signal Processing, Image Processing and Pattern Recognition-2012.
12. Second Generation Wavelets and applications by marten H. Jansen, Patrick J. Ooninx, Springer Publication, 2010. ISBN: 1849969582, 9781849969581.
13. ShaoyuZheng, Fang Xu, Deqing Wang (2009) "An Improved Adaptive Lifting Scheme Combining Gradient Operator for Image Coding 1st International conference on information science and engineering vol.1, pp-1133-1136.
14. Sweldens, W. (1996)"The lifting scheme: A custom-design construction of biorthogonal wavelets," Applied and computational Harmonic Analysis" vol. 3, no. 2, pp. 186–200
15. The Essential Guide to Image Processing By Alan Conrad Bovik, Elsevier Inc. Publication, ISBN: 978-0-12-374457-9
16. W. Sweldens, 1997, "The lifting scheme: A construction of second generation wavelets", SIAM J. Math. Anal.,Vol.29 (2), pp.511–546.
17. Wavelets and Filter Banks by Gilbert Strang, Truong Nguyen, Wellesley-Cambridge Press, ISDN: 0-9614088-7-1
18. Xiaoyuan Yang, Zhipin Zhu, Bo Yang,(2008) "Adaptive Lifting Scheme forImage Compression, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 1, pp.547-551



Prof. Sanjay H. Dabhole, last eight years working as Principal in Sant Gajanan Maharaj Rural Polytechnic, Mahagaon, Gadhinglaj Tahsil, dist-Kolhapur. (M.S), INDIA, obtained his Bachelor's Degree of Electronics Engineering in 2002 from Walchand College of Engineering Sangli at Shivaji University, Kolhapur M.S., and took his Masters Degree of Electronics and Telecommunication Engineering in 2006 from KIT C.O.E. of Shivaji University Kolhapur with first rank, M.S. Currently, he is pursuing PhD from Electronics and Communication Dept. of University Institute of Engineering and Technology, CSJMU, Kanpur-INDIA.I have total 14 years of teaching experience at various technical institutes and setup new Sant Gajanan Maharaj Technical campus in all aspects within 4 years. National level Best Teacher award on occasion of World teachers' day on 5thOct. 2014 from Aviskar Social was Received and Educational Foundation Kolhapur (M.S.), INDIA. I am life time member of The Institution of Engineers, India (IEI) apex body. And his area of interest is Signal & Image Processing. To his credit 12 papers have been published in International & National Conferences and 08 papers have been published in International journals including IEEE explore.

Dr. RichaVerma, last twelve years working as Sr. Lecturer in Electronics and Communication Engg. Dept. of UIET, CSJMU, Kanpur (U.P), INDIA, and took his Masters from HBTI, Kanpur with first rank and also completed her Ph.D from MNIT, Allahabad. Her area of interest is VLSI& Image Processing. To her credit 15 papers have been published in International journals.