ℓ_1-ADAPTED NON SEPARABLE VECTOR LIFTING SCHEMES FOR STEREO IMAGE CODING

M. Kaaniche[1], B. Pesquet-Popescu[1] and J.- C. Pesquet[2]

[1] Telecom ParisTech,
Signal and Image Proc. Dept.
37-39, rue Dareau, 75014 Paris, France
mounir.kaaniche@telecom-paristech.fr
beatrice.pesquet@telecom-paristech.fr

[2] Université Paris-Est,
LIGM and UMR-CNRS 8049,
Champs-sur-Marne, 77454
Marne-la-Vallée, France
jean-christophe.pesquet@univ-paris-est.fr

ABSTRACT

Due to the great interest of stereo images in several applications, it becomes mandatory to improve the efficiency of existing coding techniques. For this purpose, many research works have been developed. The basic idea behind most of the reported methods consists of applying an inter-view prediction via the estimated disparity map followed by a separable wavelet transform. In this paper, we propose to use a two-dimensional non separable decomposition based on the concept of vector lifting scheme. Furthermore, we focus on the optimization of all the lifting operators employed with the left and right images. Experimental results carried out on different stereo images show the benefits which can be drawn from the proposed coding method.

Index Terms— Stereo image coding, disparity estimation, vector lifting schemes, non separable transforms, adaptive transforms.

1. INTRODUCTION

Recent advances in acquisition and display technologies have contributed to a widespread usage of stereo images. These data correspond to 2 views, called left and right images, obtained by recording the same scene from two slightly different positions. One of the main advantages of these images consists of providing a three-dimensional perception to the users. Such a 3-D representation enables various functionalities like 3DTV, telepresence in videoconferences [1], computer vision and remote sensing. The increasing demand in stereo images have motivated many researchers to design efficient compression techniques for both storage and transmission purposes. A straightforward approach consists of separately coding each image by using existing still image coders. However, this method is not so efficient since the images are often highly correlated. Therefore, more efficient coding schemes have been designed to take into account the inter-image redundancies [2]. The state-of-the-art coding approach is a combination of inter-view prediction and transform coding. More precisely, the generic stereo image coding scheme involves three steps. In the first step, one image (say the left one) is selected as a reference image, and the other image (the right one) is selected as a target image. After that, the disparity map between the right and the left images is estimated [3]. In the second step, the target image is predicted from the reference one along the disparity field, and the difference between the original target image and the predicted one, called residual image, is generated. Finally, the reference image, the residual one and the disparity map are encoded. Generally, the disparity map is losslessly encoded using DPCM with an entropy coder whereas the residual and the reference images are encoded in different transform domains. Pioneering techniques have been developed for the Discrete Cosine Transform [4, 5]. However, a great attention was paid to the wavelet transform domain to achieve the quality scalability and guarantee a lossy-to-lossless reconstruction [6, 7]. To this end, lifting schemes have been already used to encode the reference and the residual images [7]. In a recent work [8], an adaptive lifting scheme is also presented. The direction of prediction is selected according to the local horizontal and vertical gradient information of the reference image. While this approach can achieve good results in terms of bitrate, it is not efficient in a lossy coding context (especially at low bit rate) since it is very sensitive to the quality of the reference image. In [9], the disparity map and the residual image are generated by applying a bandlet transform [10] to the left and the right images. In [11], a hybrid coding scheme is designed where DCT is employed for the best matching blocks and the Haar wavelet transform for the occluded ones. Recently, we have proposed a novel approach based on the Vector Lifting Schemes (VLS) [12]. It consists of coding the reference image in intra mode whereas the other image is coded according to a hybrid mode driven by the estimated disparity map. Its main feature is that it does not explicitly generate a residual image, but two compact multiresolution representations of the left and right images. We should note that the proposed joint multiscale decomposition is handled in a separable way by cascading one dimensional (1D) VLS along the horizontal direction, then along the vertical one. However, it is well known that such a separable processing may not be well-suited for images with edges which are neither horizontal nor vertical. To overcome this drawback, some works on still image compression have been devoted to the development of 2D non separable lifting schemes in order to offer more flexibility in the design of the transform [13, 14, 15].

Due to the advantages of using non separable structures as shown in [15], we propose to perform the joint coding of the stereo image by adopting an extension of the previous VLS structure to 2D Non Separable schemes. The resulting decomposition will be denoted in what follows by NS-VLS. Another objective of this work is to design adaptive decomposition, well adapted to the characteristics of the images, through an optimization of *all* the filters used with the reference and the target images. While the proposed design strategy is inspired from our recent nonsmooth optimization technique used for still image coding [16], it is worth pointing out that this paper aims at extending this technique to the context of stereo image coding.

The remainder of this paper is organized as follows. In Sec. 2, we present the principle of the considered NS-VLS decomposition. The proposed optimization strategy is described in Sec. 3. Finally, in Sec. 4, experimental results are given and some conclusions are

drawn in Sec. 5.

2. 2D NS-VLS STRUCTURE

Let $I^{(l)}$ and $I^{(r)}$ denote the left and right images to be coded. At each resolution level j and each pixel location (m,n), the approximation coefficient of the left image $I_j^{(l)}$ (resp. right image $I_j^{(r)}$) has four polyphase components $I_{0,j}^{(l)}(m,n) = I_j^{(l)}(2m,2n)$, $I_{1,j}^{(l)}(m,n) = I_j^{(l)}(2m,2n+1)$, $I_{2,j}^{(l)}(m,n) = I_j^{(l)}(2m+1,2n)$, and $I_{3,j}^{(l)}(m,n) = I_j^{(l)}(2m+1,2n+1)$ (resp. $I_{0,j}^{(r)}(m,n) = I_j^{(r)}(2m,2n)$, $I_{1,j}^{(r)}(m,n) = I_j^{(r)}(2m,2n+1)$, $I_{2,j}^{(r)}(m,n) = I_j^{(r)}(2m+1,2n)$, and $I_{3,j}^{(r)}(m,n) = I_j^{(r)}(2m+1,2n+1)$). The proposed analysis NS-VLS structure is shown in Fig. 1. As mentioned before, the reference image $I^{(l)}$ is generally encoded in intra mode. Thus, it can be seen that a non separable structure, comprising *three* prediction steps and an update step, is employed to generate the diagonal detail coefficients $I_{j+1}^{(HH,l)}$, the vertical detail coefficients $I_{j+1}^{(LH,l)}$, the horizontal detail coefficients $I_{j+1}^{(HL,l)}$, and the approximation coefficients $I_{j+1}^{(l)}$ of the left image:

$$I_{j+1}^{(HH,l)}(m,n) = I_{3,j}^{(l)}(m,n) - \lfloor (\mathbf{P}_{0,j}^{(HH,l)})^\top \mathbf{I}_{0,j}^{(HH,l)} + (\mathbf{P}_{1,j}^{(HH,l)})^\top \mathbf{I}_{1,j}^{(HH,l)} + (\mathbf{P}_{2,j}^{(HH,l)})^\top \mathbf{I}_{2,j}^{(HH,l)} \rfloor, \quad (1)$$

$$I_{j+1}^{(LH,l)}(m,n) = I_{2,j}^{(l)}(m,n) - \lfloor (\mathbf{P}_{0,j}^{(LH,l)})^\top \mathbf{I}_{0,j}^{(LH,l)} + (\mathbf{P}_{1,j}^{(LH,l)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,l)} \rfloor, \quad (2)$$

$$I_{j+1}^{(HL,l)}(m,n) = I_{1,j}^{(l)}(m,n) - \lfloor (\mathbf{P}_{0,j}^{(HL,l)})^\top \mathbf{I}_{0,j}^{(HL,l)} + (\mathbf{P}_{1,j}^{(HL,l)})^\top \overline{\mathbf{I}}_{j+1}^{(HH,l)} \rfloor, \quad (3)$$

$$I_{j+1}^{(l)}(m,n) = I_{0,j}^{(l)}(m,n) + \lfloor (\mathbf{U}_{0,j}^{(HL,l)})^\top \mathbf{I}_{j+1}^{(HL,l)} + (\mathbf{U}_{1,j}^{(LH,l)})^\top \mathbf{I}_{j+1}^{(LH,l)} + (\mathbf{U}_{2,j}^{(HH,l)})^\top \mathbf{I}_{j+1}^{(HH,l)} \rfloor, \quad (4)$$

where for every $i \in \{0,1,2\}$ and $o \in \{HL, LH, HH\}$,

- $\mathbf{P}_{i,j}^{(o,l)} = (p_{i,j}^{(o,l)}(s,t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,l)}}$ is the prediction weighting vector whose support is denoted by $\mathcal{P}_{i,j}^{(o,l)}$
- $\mathbf{I}_{i,j}^{(o,l)} = (I_{i,j}^{(l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,l)}}$ is a reference vector used to compute $I_{j+1}^{(o,l)}(m,n)$
- $\underline{\mathbf{I}}_{j+1}^{(HH,l)} = (I_{j+1}^{(HH,l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{1,j}^{(LH,l)}}$ and $\overline{\mathbf{I}}_{j+1}^{(HH,l)} = (I_{j+1}^{(HH,l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{1,j}^{(HL,l)}}$ are used in the second and the third prediction steps
- $\mathbf{U}_{i,j}^{(o,l)} = (u_{i,j}^{(o,l)}(s,t))_{(s,t)\in\mathcal{U}_{i,j}^{(o,l)}}$ is the update weighting vector whose support is designated by $\mathcal{U}_{i,j}^{(o,l)}$
- $\mathbf{I}_{j+1}^{(o,l)} = (I_{j+1}^{(o,l)}(m+s,n+t))_{(s,t)\in\mathcal{U}_{i,j}^{(o,l)}}$ is the reference vector containing the samples used in the update step.

It is important to note that the main difference between a vector lifting scheme and a basic one is that for the target image $I_j^{(r)}$, the prediction step involves samples from the same image and also some matching samples taken from the disparity-compensated reference image. To this end, we firstly apply Eqs (1)-(4) to generate three intermediate detail subbands and an approximation one denoted respectively by $\widetilde{I}_{j+1}^{(HH,r)}$, $\widetilde{I}_{j+1}^{(LH,r)}$, $\widetilde{I}_{j+1}^{(HL,r)}$ and $I_{j+1}^{(r)}$. After that, we add a *second prediction stage* composed of three steps, which involves a hybrid prediction exploiting at the same time the intra and inter-image redundancies in the stereo pair. This is achieved by using the estimated disparity field denoted by $v_j = (v_{x,j}, v_{y,j})$. In the following, the disparity compensated left image on a given matching sample (m,n), given by $I_j^{(l)}(m+v_{x,j}(m,n), n+v_{y,j}(m,n))$, will be simply replaced by $I_j^{(c)}(m,n)$ for notation concision. Similarly to the left image, let us denote by $I_{0,j}^{(c)}(m,n)$, $I_{1,j}^{(c)}(m,n)$, $I_{2,j}^{(c)}(m,n)$ and $I_{3,j}^{(c)}(m,n)$ the four polyphase components of $I_j^{(c)}(m,n)$. Therefore, the final detail subbands of the right multiresolution analysis can be expressed as:

$$\begin{aligned} I_{j+1}^{(HH,r)}(m,n) = \widetilde{I}_{j+1}^{(HH,r)}(m,n) - \lfloor & (\mathbf{Q}_{0,j}^{(HH,r)})^\top \widetilde{\mathbf{I}}_{0,j+1}^{(HH,r)} + (\mathbf{Q}_{1,j}^{(HH,r)})^\top \widetilde{\mathbf{I}}_{1,j+1}^{(HH,r)} + (\mathbf{Q}_{2,j}^{(HH,r)})^\top \widetilde{\mathbf{I}}_{2,j+1}^{(HH,r)} \\ & + (\mathbf{P}_{0,j}^{(HH,r,l)})^\top \mathbf{I}_{0,j}^{(HH,c)} + (\mathbf{P}_{1,j}^{(HH,r,l)})^\top \mathbf{I}_{1,j}^{(HH,c)} \\ & + (\mathbf{P}_{2,j}^{(HH,r,l)})^\top \mathbf{I}_{2,j}^{(HH,c)} + (\mathbf{P}_{3,j}^{(HH,r,l)})^\top \mathbf{I}_{3,j}^{(HH,c)} \rfloor, \end{aligned} \quad (5)$$

$$\begin{aligned} I_{j+1}^{(LH,r)}(m,n) = \widetilde{I}_{j+1}^{(LH,r)}(m,n) - \lfloor & (\mathbf{Q}_{0,j}^{(LH,r)})^\top \widetilde{\mathbf{I}}_{0,j+1}^{(LH,r)} + (\mathbf{Q}_{1,j}^{(LH,r)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,r)} + (\mathbf{P}_{0,j}^{(LH,r,l)})^\top \mathbf{I}_{0,j}^{(LH,c)} \\ & + (\mathbf{P}_{1,j}^{(LH,r,l)})^\top \mathbf{I}_{1,j}^{(LH,c)} + (\mathbf{P}_{2,j}^{(LH,r,l)})^\top \mathbf{I}_{2,j}^{(LH,c)} \\ & + (\mathbf{P}_{3,j}^{(LH,r,l)})^\top \mathbf{I}_{3,j}^{(LH,c)} \rfloor, \end{aligned} \quad (6)$$

$$\begin{aligned} I_{j+1}^{(HL,r)}(m,n) = \widetilde{I}_{j+1}^{(HL,r)}(m,n) - \lfloor & (\mathbf{Q}_{0,j}^{(HL,r)})^\top \widetilde{\mathbf{I}}_{0,j+1}^{(HL,r)} + (\mathbf{Q}_{1,j}^{(HL,r)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,r)} + (\mathbf{P}_{0,j}^{(HL,r,l)})^\top \mathbf{I}_{0,j}^{(HL,c)} \\ & + (\mathbf{P}_{1,j}^{(HL,r,l)})^\top \mathbf{I}_{1,j}^{(HL,c)} + (\mathbf{P}_{2,j}^{(HL,r,l)})^\top \mathbf{I}_{2,j}^{(HL,c)} \\ & + (\mathbf{P}_{3,j}^{(HL,r,l)})^\top \mathbf{I}_{3,j}^{(HL,c)} \rfloor, \end{aligned} \quad (7)$$

where for every $i \in \{0,1,2,3\}$ and $o \in \{HL, LH, HH\}$,

- $\mathbf{Q}_{i,j}^{(o,r)} = (q_{i,j}^{(o,r)}(s,t))_{(s,t)\in\mathcal{Q}_{i,j}^{(o,r)}}$ is an intra prediction weighting vector whose support is denoted by $\mathcal{Q}_{i,j}^{(o,r)}$
- $\mathbf{P}_{i,j}^{(o,r,l)} = (p_{i,j}^{(o,r,l)}(s,t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,r,l)}}$ is an inter prediction weighting vector whose support is denoted by $\mathcal{P}_{i,j}^{(o,r,l)}$
- $\widetilde{\mathbf{I}}_{0,j+1}^{(o,r)} = (I_{j+1}^{(r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{0,j}^{(o,r)}}$ is a reference vector used to compute $I_{j+1}^{(o,r)}(m,n)$
- $\widetilde{\mathbf{I}}_{1,j+1}^{(HH,r)} = (I_{j+1}^{(HL,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(HH,r)}}$ and $\widetilde{\mathbf{I}}_{2,j+1}^{(HH,r)} = (I_{j+1}^{(LH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{2,j}^{(HH,r)}}$ are two reference vectors used to compute $I_{j+1}^{(HH,r)}(m,n)$
- $\underline{\mathbf{I}}_{j+1}^{(HH,r)} = (I_{j+1}^{(HH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(LH,r)}}$ and $\overline{\mathbf{I}}_{j+1}^{(HH,r)} = (I_{j+1}^{(HH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(HL,r)}}$ are two intra prediction vectors used to compute $I_{j+1}^{(LH,r)}(m,n)$ and $I_{j+1}^{(HL,r)}(m,n)$
- $\mathbf{I}_{i,j}^{(o,c)} = (I_{i,j}^{(c)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,r,l)}}$ is a reference vector containing the matching samples used to compute $I_{j+1}^{(o,r)}(m,n)$.

Finally, at the last resolution level $j = J$, instead of directly coding the approximation subband $I_J^{(r)}$, we predict it from the approximation of the left image using disparity compensation. As a result, the following residual subband $e_J^{(r)}$ is generated:

$$e_J^{(r)}(m,n) = I_J^{(r)}(m,n) - I_J^{(c)}(m,n). \quad (8)$$

Once the considered NS-VLS has been defined, we address in the next section the issue of the optimal design of its lifting operators.

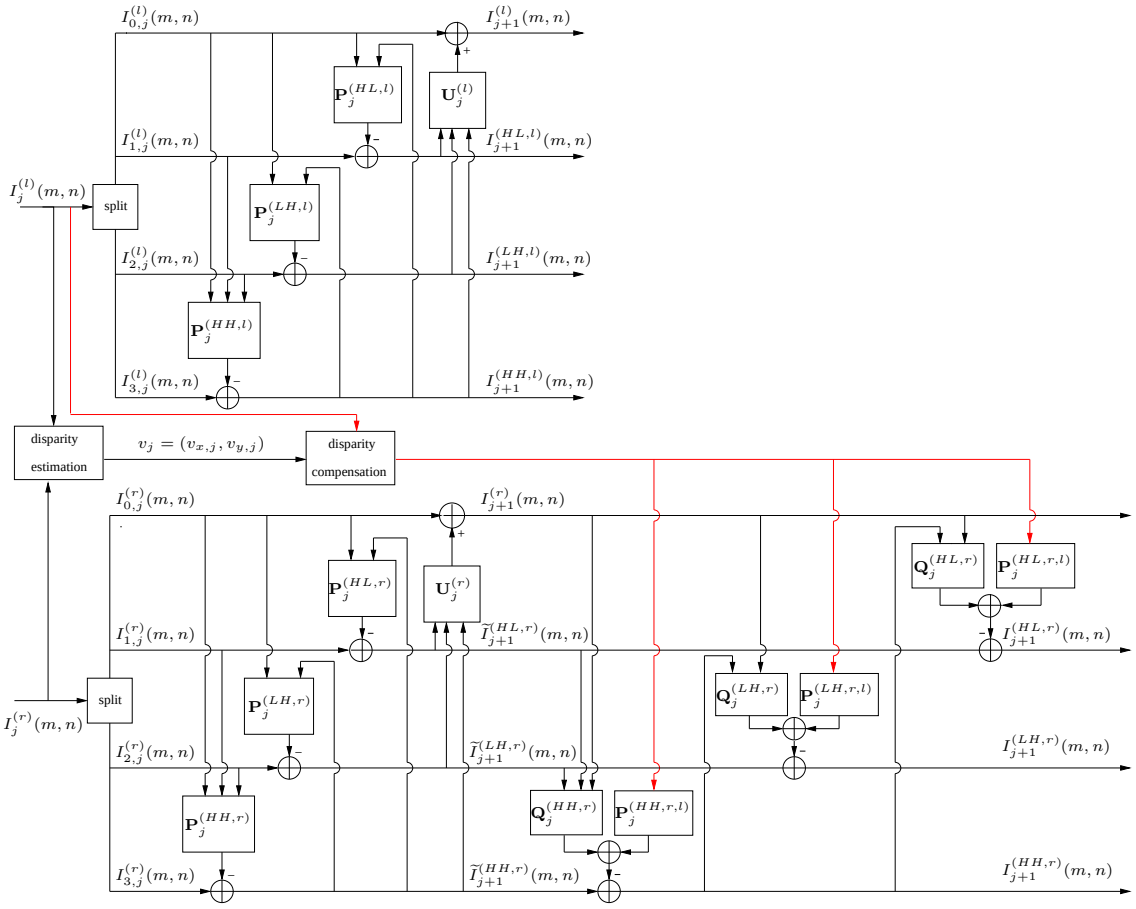

Fig. 1. NS-VLS decomposition structure.

3. DESIGN OF A FULLY-ADAPTIVE STRUCTURE

3.1. Design of the filters used for the reference image $I^{(l)}$

With the ultimate goal of producing sparse wavelet coefficients, we propose to optimize the prediction filters $\mathbf{P}_j^{(o,l)}$ of the left image by minimizing the ℓ_1-norm of the detail coefficients:

$$\forall\, o \in \{HL, LH, HH\}, \forall\, i \in \{1,2,3\},$$
$$\mathcal{J}_{\ell_1}(\mathbf{P}_j^{(o,l)}) = \sum_{m=1}^{M_j}\sum_{n=1}^{N_j} \left| I_{i,j}^{(l)}(m,n) - (\mathbf{P}_j^{(o,l)})^\top \mathbf{X}_j^{(o,l)}(m,n) \right| \quad (9)$$

where $I_{i,j}^{(l)}(m,n)$ is the sample to be predicted, $\mathbf{X}_j^{(o,l)}(m,n)$ is the reference vector containing the samples used in the prediction step, $\mathbf{P}_j^{(o,l)}$ is the prediction operator vector to be optimized, M_j and N_j corresponds to the dimensions of the input subband $I_{j+1}^{(l)}$. To minimize such a criterion, the Douglas-Rachford algorithm can be employed, which is an efficient optimization tool in this context [17]. However, it can be noticed from Fig. 1 that the diagonal detail signal $I_{j+1}^{(HH,l)}$ is used as a reference signal in the second and the third prediction steps to generate the detail signals $I_{j+1}^{(LH,l)}$ and $I_{j+1}^{(HL,l)}$ respectively. Therefore, it is interesting to optimize the prediction filter $\mathbf{P}_j^{(HH,l)}$ by minimizing the following *weighted* sum of the ℓ_1-norm of the three detail subbands $I_{j+1}^{(o,l)}$:

$$\mathcal{J}_{w\ell_1}(\mathbf{P}_j^{(HH,l)}) = \sum_{o\in\{HL,LH,HH\}} \sum_{m=1}^{M_j}\sum_{n=1}^{N_j} \frac{1}{\alpha_{j+1}^{(o,l)}} \left| I_{j+1}^{(o,l)}(m,n) \right| \quad (10)$$

where $\alpha_{j+1}^{(o,l)}$ can be estimated by using a classical maximum likelihood estimate. We should note that (10) is related to the approximation of the entropy of an i.i.d. Laplacian source. To solve this minimization problem, we can also use the Douglas-Rachford algorithm, reformulated in a *three-fold product space* [18]. For more details about the minimization algorithm, the reader is referred to [16]. By minimizing the weighted criterion (10), it can be noticed that the optimization of the filter $\mathbf{P}_j^{(HH,l)}$ depends on the optimization of the filters ($\mathbf{P}_j^{(LH,l)}$,$\mathbf{P}_j^{(HL,l)}$) and vice-versa. As a result, it appears interesting to use a *joint* optimization method which iteratively optimizes the prediction filters $\mathbf{P}_j^{(HH,l)}$, $\mathbf{P}_j^{(LH,l)}$ and $\mathbf{P}_j^{(HL,l)}$. For this purpose, we start by optimizing separately each prediction filter $\mathbf{P}_j^{(o,l)}$ based on the ℓ_1 criterion (9). Then, the update filter $\mathbf{U}_j^{(l)}$ is optimized by minimizing the error between the approximation signal $I_{j+1}^{(l)}$ and the decimated version of the output of an ideal low-pass filter, and the resulting weighting terms $\frac{1}{\alpha_{j+1}^{(o,l)}}$ are evaluated. After that, we iteratively repeat the following three steps: re-optimize the filters $\mathbf{P}_j^{(HH,l)}$, $\mathbf{P}_j^{(LH,l)}$ and $\mathbf{P}_j^{(HL,l)}$ by minimizing respec-

tively $\mathcal{J}_{w\ell_1}(\mathbf{P}_j^{(HH,l)})$, $\mathcal{J}_{\ell_1}^{(LH,l)}(\mathbf{P}_j^{(LH,l)})$ and $\mathcal{J}_{\ell_1}^{(HL,l)}(\mathbf{P}_j^{(HL,l)})$, re-optimize the update filter $\mathbf{U}_j^{(l)}$ and re-compute the weighting terms. Note that the convergence of the proposed joint optimization algorithm is achieved during the early iterations (after about 5 iterations) where each one takes about 4 seconds for an image of size 512×512 using a Matlab implementation [16].

3.2. Design of the filters used for the target image $I^{(r)}$

Let us denote by $\widetilde{\mathbf{P}}_j^{(o,r,l)}$ the sum of the two filters $\mathbf{Q}_j^{(o,r)}$ and $\mathbf{P}_j^{(o,r,l)}$ used at the second prediction stage of the NS-VLS structure. Intuitively, one can again optimize all the prediction filters $\mathbf{P}_j^{(HH,r)}$, $\mathbf{P}_j^{(LH,r)}$, $\mathbf{P}_j^{(HL,r)}$, $\widetilde{\mathbf{P}}_j^{(HH,r,l)}$, $\widetilde{\mathbf{P}}_j^{(LH,r,l)}$ and $\widetilde{\mathbf{P}}_j^{(HL,r,l)}$ by minimizing the weighted sum of the ℓ_1-norm of the three details subbands $I_{j+1}^{(o,r)}$. However, since the left and right images contain nearly similar contents, we propose to set the filters used at the *first* lifting stage, applied to the right image, to those corresponding to the reference image:

$$\begin{aligned} &\mathbf{P}_j^{(HH,r)} = \mathbf{P}_j^{(HH,l)}, \mathbf{P}_j^{(LH,r)} = \mathbf{P}_j^{(LH,l)}, \\ &\mathbf{P}_j^{(HL,r)} = \mathbf{P}_j^{(HL,l)}, \mathbf{U}_j^{(r)} = \mathbf{U}_j^{(l)}. \end{aligned} \tag{11}$$

The advantages of this strategy is two fold. First, it simplifies the optimization process. Furthermore, it reduces the transmission cost of the filter coefficients. Once the optimal operators of the first stage are determined, the other prediction filters $\widetilde{\mathbf{P}}_j^{(HH,r,l)}$, $\widetilde{\mathbf{P}}_j^{(LH,r,l)}$ and $\widetilde{\mathbf{P}}_j^{(HL,r,l)}$ will be designed by an alternating optimization approach similar to that addressed in the previous section.

4. EXPERIMENTAL RESULTS

Simulation results are performed on five real stereo pairs downloaded from[1] . In order to show the benefits of the proposed scheme, we provide the results for the following decompositions carried out over three resolution levels. The first one consists of coding independently the left and right images using the 9/7 transform which was selected for the lossy mode of the JPEG2000 standard. This scheme will be designated by "Independent". The second method, which will be denoted by "Scheme-B", is the state-of the-art method where the reference and the residual images are encoded using also the 9/7 transform [7]. The third one corresponds to our previous joint stereo coding scheme based on a *separable* optimized VLS decomposition. Finally, we consider the proposed extension of this method to a *non separable* structure where a *joint* optimization approach is performed. The two latter methods will be designated respectively by "SEP-VLS-OPT" and "NS-VLS-OPT". Fig. 2 displays the scalability in quality of the reconstruction procedure by providing the variations of the average PSNR versus the average bitrate of the "houseof" stereo images. These plots show that the proposed method achieves an average gain of about 0.1-0.15 dB compared to our recent work "SEP-VLS-OPT". The gain becomes more important (up to 0.65 dB) compared with the state-of-the art methods. Fig. 3 displays the reconstructed target image of the "aerial" stereo pairs for "Scheme-B" and "NS-VLS-OPT". We notice that the coding of the residual image leads to blocking artefacts whereas our approach reduces significantly this problem. It is important to emphasize here that the blocking artefacts appearing with the state-of-the-art method are not related to the wavelet codec and result mainly from the limitations of the generic scheme where a residual image is generated using a block-based approach. Finally, in order to measure the relative gain of the proposed method, we used the Bjontegaard metric [19]. The results are illustrated in Table 1 for low, middle and high bitrates corresponding respectively to the four bitrate points $\{0.15, 0.2, 0.25, 0.3\}$, $\{0.5, 0.55, 0.6, 0.65\}$ and $\{1.25, 1.3, 1.35, 1.4\}$ bpp. Table 1 gives the gain of the method "NS-VLS-OPT" compared with "Scheme-B". Note that a bitrate saving with respect to the reference method corresponds to negative values. It can be observed that the proposed approach outperforms the classical one by about -20% and 0.2-1.4 dB in terms of bitrate saving and quality of reconstruction.

[1] http://vasc.ri.cmu.edu/idb/html/stereo/index.html, http://vasc.ri.cmu.edu/idb/html/jisct/index.html

5. CONCLUSION

In this paper, we have exploited the flexibility offered by non separable vector lifting schemes to perform a fully-optimized structure for joint coding of stereo images. Experiments have shown the benefits of the proposed method. In a future work, a new criterion defined simultaneously on the reference and the target images could be envisaged.

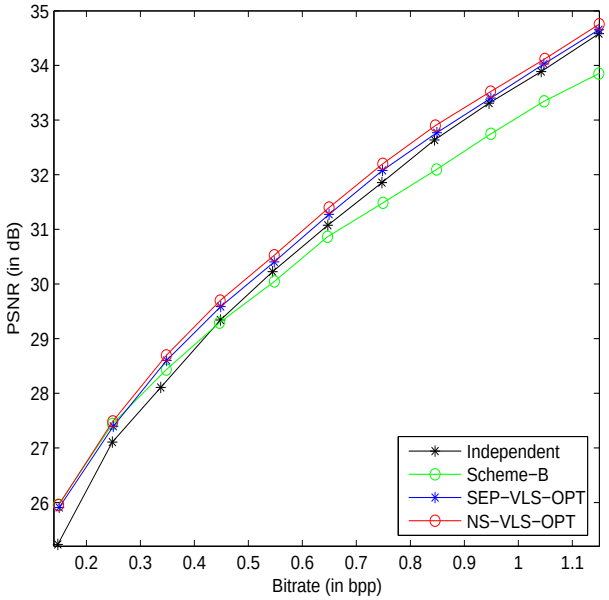

Fig. 2. PSNR (in dB) versus the bitrate (in bpp) after JPEG2000 progressive encoding of the stereo pair 'houseof'.

Table 1. The average PSNR differences and the bitrate saving at low, medium and high bitrates. The gain of "NS-VLS-OPT" w.r.t Scheme-B.

Images	bitrate saving (in %)			PSNR gain (in dB)		
	low	**middle**	**high**	**low**	**middle**	**high**
houseof	-1.07	-10.25	-11.31	0.04	0.46	0.87
pentagon	-6.91	-21.68	-27.51	0.22	0.92	1.89
ball	1.91	-10.99	-18.29	-0.03	0.30	0.78
birch	-15.17	-39.79	-23.70	0.81	1.13	2.12
aerial	-0.45	-17.12	-19.44	0.02	0.75	1.43
average	-4.34	-19.96	-20.05	0.21	0.71	1.41

(a) Original target image (b) PSNR=27.10 dB, SSIM=0.728 (c) PSNR=27.81 dB, SSIM=0.744

Fig. 3. Reconstructed target image of the "aerial" stereo pair at 0.3 bpp using (b) Scheme-B (c) NS-VLS-OPT.

6. REFERENCES

[1] I. Feldmann, W. Waizenegger, N. Atzpadin, and O. Schreer, "Real-time depth estimation for immersive 3D videoconferencing," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, June 2010, pp. 1–4.

[2] M. G. Perkins, "Data compression of stereo pairs," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, April 1992.

[3] D. Tzovaras, N. Grammalidis, and M.G. Strintzis, "Disparity field and depth map coding for multiview 3D image generation," *Signal Processing: Image Communication*, vol. 11, no. 3, pp. 205–230, January 1998.

[4] O. Woo and A. Ortega, "Stereo image compression based on disparity field segmentation," in *SPIE Conference on Visual Communications and Image Processing*, San Jose, California, February 1997, vol. 3024, pp. 391–402.

[5] M. S. Moellenhoff and M. W. Maier, "Transform coding of stereo image residuals," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 804–812, June 1998.

[6] J. Xu, Z. Xiong, and S. Li, "High performance wavelet-based stereo image coding," in *IEEE International Symposium on Circuits and Systems*, May 2002, vol. 2, pp. 273–276.

[7] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 898–903, October 2002.

[8] R. Darazi, A. Gouze, and B. Macq, "Adaptive lifting scheme-based method for joint coding 3D-stereo images with luminance correction and optimized prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, April 2009, pp. 917–920.

[9] A. Maalouf and M.-C. Larabi, "Bandelet-based stereo image coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, United States, March 2010, pp. 698–701.

[10] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, April 2005.

[11] T. Frajka and K. Zeger, "Residual image coding for stereo image compression," *Optical Engineering*, vol. 42, no. 1, pp. 182–189, January 2003.

[12] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, "Vector lifting schemes for stereo image coding," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2463–2475, November 2009.

[13] O. N. Gerek and A. E. Çetin, "Adaptive polyphase subband decomposition structures for image compression," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1649–1660, October 2000.

[14] V. Chappelier and C. Guillemot, "Oriented wavelet transform for image compression and denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2892–2903, October 2006.

[15] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, "Non separable lifting scheme with adaptive update step for still and stereo image coding," *Elsevier Signal Processing: Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations*, vol. 91, no. 12, pp. 2767–2782, January 2011.

[16] M. Kaaniche, B. Pesquet-Popescu, A. Benazza-Benyahia, and J.-C. Pesquet, "Adaptive lifting scheme with sparse criteria for image coding," *EURASIP Journal on Advances in Signal Processing: Special Issue on New Image and Video Representations Based on Sparsity*, vol. 2012, 22 pages, January 2012.

[17] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting methods and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[18] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer-Verlag, New York, 2010.

[19] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," Tech. Rep., ITU SG16 VCEG-M33, Austin, TX, USA, April 2001.