Vector lifting schemes for stereo image coding

Mounir Kaaniche, Amel Benazza-Benyahia, Béatrice Pesquet-Popescu, Jean-Christophe Pesquet

▶ **To cite this version:**

Mounir Kaaniche, Amel Benazza-Benyahia, Béatrice Pesquet-Popescu, Jean-Christophe Pesquet. Vector lifting schemes for stereo image coding. IEEE Transactions on Image Processing, 2009, 8 (11), pp.2463-2475. 10.1109/TIP.2009.2026672 . hal-00621962

HAL Id: hal-00621962
https://hal.science/hal-00621962

Submitted on 22 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vector Lifting Schemes for Stereo Image Coding

M. Kaaniche, *Student Member IEEE,* A. Benazza-Benyahia, *Member IEEE,* B. Pesquet-Popescu, *Senior Member IEEE,* and J.- C. Pesquet, *Senior Member IEEE*

***Abstract*— Many research efforts have been devoted to the improvement of stereo image coding techniques for storage or transmission. In this paper, we are mainly interested in lossy-to-lossless coding schemes for stereo images allowing progressive reconstruction. The most commonly used approaches for stereo compression are based on disparity compensation techniques. The basic principle involved in this technique first consists of estimating the disparity map. Then, one image is considered as a reference and the other is predicted in order to generate a residual image. In this work, we propose a novel approach, based on Vector Lifting Schemes (VLS), which offers the advantage of generating two compact multiresolution representations of the left and the right views. We present two versions of this new scheme. A theoretical analysis of the performance of the considered VLS is also conducted. Experimental results indicate a significant improvement using the proposed structures compared with conventional methods.**

I. Introduction

The principle of stereoscopic imaging systems consists of generating two images by recording two slightly different view angles of the same scene. By presenting the appropriate image of a stereo pair to the left/right eye, the viewer perceives the scene in three dimensions (3D). The recent advances in acquisition and display technologies have allowed the widespread use of stereovision in various application fields such as entertainment, medical surgical environments, tele-presence in videoconferences [1], computer vision and remote sensing [2]. For instance, today's advances in satellite remote sensing technology provide the capability to collect Stereo Image (SI) pairs for several applications, such as cartography and urban planning. Satellite stereo images (such as those generated by IKONOS and SPOT5 sensors) are especially helpful to generate a digital elevation model which is a 3D representation of the topography of a given area [3]. The increasing interest in SIs has led to the constitution of image databases that require huge amounts of storage capacity. For example, the SPOT5 sensor covers areas of 60 Km $\times$ 60 Km at a resolution of 2.5 m and a single view requires more than 500 Megabytes. In addition to these stereo sensors, it is worth mentioning the Multi-angle Imaging SpectroRadiometer (MISR), which uses nine cameras to generate multiview data sets [4] at a data rate of 3.3 Mbps. Hence, the use of compression techniques is mandatory for image storage as well as for image transmission.

M. Kaaniche and B. Pesquet-Popescu are with the Ecole Nationale Supérieure des Télécommunications de Paris, 37-39 rue Dareau 75014 Paris, France. E-mail: kaaniche@telecom-paristech.fr, pesquet@telecom-paristech.fr. A. Benazza-Benyahia is with the Ecole Supérieure des Communications de Tunis (SUP'COM-Tunis), 2083, Tunisia. E-mail: benazza.amel@supcom.rnu.tn. J.-C. Pesquet is with the Université Paris-Est, Laboratoire d'Informatique Gaspard Monge, CNRS UMR 8049, 77454 Marne-la-Vallée, France. E-mail: jean-christophe.pesquet@univ-paris-est.fr.

Part of this work has been presented in [21].

To the best of our knowledge, the ongoing activity on stereoscopic *still* image coding is mainly carried out independently of any standardization activity [5]. Consequently, different approaches have been reported concerning still image coding. The most simple ones consists of separately coding each view by using existing still image coders. However, the resulting data rates may remain too high for some practical stereoscopic systems. As the two images have similar content, they are highly correlated. Therefore, more efficient coding schemes have been designed to exploit the cross-view redundancies [6], [7]. This is usually achieved by first estimating the disparity field between the SI pair [8]. Then, one image is considered as a reference (say the left one) and the other image (target) is predicted by disparity-compensating the reference one. A prediction error image, called residual image, is thus generated. Finally, the disparity field, the reference image and the residual one are encoded [7], [9]. This approach is known as disparity compensation due to its similarity with motion compensation techniques which are popular for video coding [10]. The goal of this paper is to design a novel joint coding approach enabling a gradual and finally exact decoding of the stereo pairs. Our main contribution is that the proposed coding scheme does not generate any residual image, but directly two compact multiresolution representations of the left and right images by exploiting the cross-view redundancies *via* the available disparity field. Furthermore, the proposed scheme is intrinsically flexible, as it allows the designer to optimize the number of prediction filter taps as well as the other parameters of the multiscale operators. In this way, we build a joint coding scheme which is adapted to the content of the stereo pair. Another advantage of the proposed method is that it guarantees a perfect reconstruction of the stereo images.

The remainder of this paper is organized as follows. Section II gives an overview of SI coding schemes based on disparity estimation and compensation techniques. In Section III, we propose a novel coding structure of which two examples are given. In Section IV, we conduct a theoretical analysis of the proposed schemes in terms of prediction efficiency. Section V describes how embedded binary streams can be produced to encode the resulting multiscale representations. Finally, in Section VI, experimental results are given and some conclusions are drawn in Section VII.

II. Stereo Image Coding

Generally, the reported stereo image coding methods rely on Disparity Estimation techniques (DE) followed by Disparity Compensation (DC) as discussed below. As mentioned earlier, DE is a key issue for exploiting the cross-view redundancies. This problem has been extensively studied in computer vision

and surveys of the different DE techniques proposed in the literature can be found in [2], [11], [12]. Two main approaches, pixel-based or block-based, can be used to estimate the disparity map. In what follows, we use a fixed size block matching DE, which consists of first partitioning the right image $I^{(r)}$ into nonoverlapping blocks of size $b_x \times b_y$. For each block, the objective is to find the most "similar" block within a given search area S in the left image $I^{(l)}$. The disparity vector $\mathbf{v} = (v_x, v_y)$ for a current block in $I^{(r)}$ minimizes a dissimilarity criterion $\mathcal{D}$:

$$(v_x, v_y)(m_x, m_y) \triangleq \arg \min_{(v_x, v_y) \in S} \mathcal{D}(I^{(r)}(m_x, m_y), I^{(l)}(m_x + v_x, m_y + v_y)) \quad (1)$$

where (m_x, m_y) are the spatial coordinates associated with the top leftmost pixel in the block. Very often, the Sum of Square Differences (SSD) or the Sum of Absolute Differences (SAD) is the selected criterion. It should be noted that in the ideal parallel-axis geometry, the displacement between the two views is restricted to the horizontal direction ($v_y = 0$) and it takes positive values ($v_x \geq 0$). However, in practice, the matching point of any current point of $I^{(r)}$ is not always rigorously on the epipolar line because of the sensor noise, the discretization errors and the deviation from the pinhole camera model. As a consequence, a strip along the epipolar line is considered and all the points falling within this strip are considered as potential matching candidates to be paired with the current point. It is worth mentioning that several works aimed at improving this block-based DE, e.g. by using overlapped block DE with adaptive windows [13], [14]. Once the disparity vectors are generated, a disparity compensation can be performed: the target image $I^{(r)}$ is predicted from $I^{(l)}$ along the disparity vectors. Then, the Disparity Compensated Difference (DCD) $I^{(e)}$ is computed as follows:

$$I^{(e)}(m_x, m_y) \triangleq I^{(r)}(m_x, m_y) - I^{(l)}(m_x + v_x, m_y + v_y) \quad (2)$$

where the dependence on (m_x, m_y) of v_x and v_y has been dropped for notation simplicity. Generally, the disparity vectors are losslessly encoded using DPCM followed by arithmetic encoding, whereas the reference and the residual images can be coded in different transform domains. Some works apply a discrete cosine transform [9], [15]. However, more recent works have preferred the wavelet transform, in order to meet the scalability requirement. In [16], an efficient exploitation of the zerotree algorithm [17] is performed to shorten the embedded bitstreams of the wavelet coefficients of both the reference image and the DCD. In [18], both the estimation and the disparity compensation take place in the wavelet domain, the coding of the wavelet coefficients being performed through a Subspace Projection Technique (SPT). Furthermore, we should note also that Annex I of Part II of the JPEG2000 standard is dedicated to multi-component image coding [19]. A decorrelation of the spectral components may be performed prior to the wavelet transform. In our case, each view of the stereoscopic image can be seen as a single component. Unlike the conventional methods, we propose a joint coding scheme that directly generates a pair of multiresolution representations of the left and the right images derived from a judicious lifting decomposition which will be described in the next section.

III. Proposed vector lifting schemes

A. Motivations

A novel approach that is based on a joint multiscale decomposition of $I^{(l)}$ and $I^{(r)}$ is developed in this section. It consists of coding the reference image $I^{(l)}$ in intra mode (purely spatial), whereas the other image is coded by exploiting cross-image redundancies via the available disparity map. The decomposition strategy is inspired from Vector Lifting Schemes (VLS) [20] and it has been briefly presented as a preliminary work in [21]. The main advantage of the proposed approach is that it does not explicitly generate a residual image, but two multiresolution representations of $I^{(l)}$ and $I^{(r)}$. Two versions of the VLS will be described in the following.

B. VLS decompositions

The wavelet coefficients of an image are usually obtained by a dyadic filter bank structure [22]. If an exact reconstruction is required, lifting schemes are often employed, since they allow to generate integer-valued versions of the wavelet coefficients whatever the underlying decomposition operators are [23], [24]. For the sake of simplicity, a separable decomposition is considered in this paper. Therefore, it is enough to address the decomposition in one dimension. The corresponding analysis structure is shown in Fig. 2. More precisely, at each resolution level j, the even and odd samples of the approximation (scaling) coefficients $I_j^{(l)}(m_x, 2m_y)$, $I_j^{(r)}(m_x, 2m_y)$, $I_j^{(l)}(m_x, 2m_y + 1)$ and $I_j^{(r)}(m_x, 2m_y + 1)$ of $I^{(l)}$ and $I^{(r)}$ respectively are the input coefficients of the lifting scheme. Furthermore, we denote by $\mathbf{v}_j = (v_{x,j}, v_{y,j})^\top$ the available disparity vectors which are obtained by sampling and dividing by 2^j the initial (full resolution) disparity vectors $\mathbf{v} = (v_x, v_y)^\top$, since the dimensions of the subbands at the j-th resolution level correspond to the dimensions of the initial images divided by 2^j:

$$\mathbf{v}_j(m_x, m_y) = \frac{1}{2^j} \mathbf{v}(2^j m_x, 2^j m_y). \quad (3)$$

It is important to note that (3) may yield non-integer values of the disparity vectors. Therefore, if the components of $\mathbf{v}_j$ are integer-valued, for any given pixel (m_x, m_y) in the right image corresponds a pixel in the disparity-compensated left image $I_j^{(l)}(m_x + v_{x,j}(m_x, m_y), m_y + v_{y,j}(m_x, m_y))$. Otherwise, the corresponding disparity-compensated intensity results from the usual bilinear interpolation. The objective of the vector lifting scheme is to simultaneously exploit the dependence existing between $I_j^{(l)}$ and $I_j^{(r)}$ by producing 2 kinds of outputs: the detail coefficients $\widetilde{d}_{j+1}^{(l)}$, $\widetilde{d}_{j+1}^{(r)}$ and the approximation ones $\widetilde{I}_{j+1}^{(l)}$, $\widetilde{I}_{j+1}^{(r)}$ for both images. Similar lifting structures operating along the image columns allow us to generate the approximation coefficients $I_{j+1}^{(l)}$ and $I_{j+1}^{(r)}$, as well as the associated detail coefficients in the horizontal, vertical and diagonal directions

at the resolution level $(j+1)$.[1] A wide range of nonlinear operators can be applied to reduce the intra and inter-image redundancies. However, for tractability purposes, we will only use combinations of shift operators, linear filters and rounding operations. For the reference image $I^{(l)}$, the detail coefficients can be interpreted as intra-image prediction errors at resolution $(j+1)$ and expressed as:

$$\widetilde{d}_{j+1}^{(l)}(m_x,m_y) = I_j^{(l)}(m_x,2m_y+1) - \lfloor (\mathbf{P}_j^{(l)})^\top \mathbf{I}_j^{(l)} \rfloor \quad (4)$$

where $\mathbf{P}_j^{(l)} = \left(p_{j,k}^{(l)}\right)_{k\in\mathcal{P}_j^{(l)}}$ is the prediction weighting vector, $\mathbf{I}_j^{(l)} = \left(I_j^{(l)}(m_x,2m_y-2k)\right)_{k\in\mathcal{P}_j^{(l)}}$ is the reference vector containing the even samples used in the prediction step, $\mathcal{P}_j^{(l)}$ is the support of the predictor of $I_j^{(l)}(m_x,2m_y+1)$ and $\lfloor . \rfloor$ is the integer-part operator. Then, at the update step, the approximation coefficients are computed as follows:

$$\widetilde{I}_{j+1}^{(l)}(m_x,m_y) = I_j^{(l)}(m_x,2m_y) + \lfloor (\mathbf{U}_j^{(l)})^\top \widetilde{\mathbf{d}}_{j+1}^{(l)} \rfloor \quad (5)$$

where $\mathbf{U}_j^{(l)} = \left(u_{j,k}^{(l)}\right)_{k\in\mathcal{U}_j^{(l)}}$ is the update weighting vector, $\widetilde{\mathbf{d}}_{j+1}^{(l)} = \left(\widetilde{d}_{j+1}^{(l)}(m_x,m_y-k)\right)_{k\in\mathcal{U}_j^{(l)}}$ is the reference vector containing the details coefficients used in the update step, and $\mathcal{U}_j^{(l)}$ is the support of the update operator. The reversibility of the basic lifting scheme is ensured since the prediction in (4) only makes use of even indexed samples. The main difference between a vector lifting scheme and a basic one is that for the image $I^{(r)}$, the prediction of the odd sample $I_j^{(r)}(m_x,2m_y+1)$ involves even samples from the same image *and* also neighbors of the matching sample taken from the reference image. For the sake of simplicity, the notation $I_j^{(l)}(m_x+v_{x,j}(m_x,m_y), m_y+v_{y,j}(m_x,m_y)-k)$ which corresponds to the compensated image on the neighbors of a given matching sample (m_x,m_y), will be replaced by $I_j^{(c)}(m_x,m_y,k)$. Thus, the detail signal $\widetilde{d}_{j+1}^{(r)}$ will be expressed as:

$$\begin{aligned}\widetilde{d}_{j+1}^{(r)}(m_x,m_y) = I_j^{(r)}(m_x,2m_y+1) - \lfloor (\mathbf{P}_j^{(r)})^\top \mathbf{I}_j^{(r)} \\ + (\mathbf{P}_j^{(r,l)})^\top \mathbf{I}_j^{(c)} \rfloor \quad (6)\end{aligned}$$

where $\mathbf{P}_j^{(r)} = \left(p_{j,k}^{(r)}\right)_{k\in\mathcal{P}_j^{(r)}}$ (resp. $\mathbf{P}_j^{(r,l)} = \left(p_{j,k}^{(r,l)}\right)_{k\in\mathcal{P}_j^{(r,l)}}$) is the prediction weighting vector of the intra-image (resp. inter-images), $\mathbf{I}_j^{(r)} = \left(I_j^{(r)}(m_x,2m_y-2k)\right)_{k\in\mathcal{P}_j^{(r)}}$ is the reference vector containing the even samples, $\mathbf{I}_j^{(c)} = \left(I_j^{(c)}(m_x,2m_y+1,k)\right)_{k\in\mathcal{P}_j^{(r,l)}}$ is the vector containing the neighbors of the matching sample associated with the pixel $I_j^{(r)}(m_x,2m_y+1)$ to be predicted, and $\mathcal{P}_j^{(r)}$ (resp. $\mathcal{P}_j^{(r,l)}$) is the spatial support of the intra-image (resp. inter-images) predictor. The update step for $\widetilde{I}_{j+1}^{(r)}$ can be performed similarly to (5). The decomposition is iterated on the columns m_y of the resulting subbands, leading to 2×4 sub-images for the left and right images at each resolution level j and the decomposition is again repeated on the approximation sub-images over J resolution levels. It is worth pointing out that the disparity based vector lifting scheme is perfectly reversible and that it maps integers to integers. However, an appropriate choice of the involved prediction and update operators remains necessary in order to generate compact representations of $(I^{(l)}, I^{(r)})$. To illustrate the ability of the considered vector lifting structure to produce a sparse representation, we provide a simple example (denoted by VLS-I) of the considered lifting structure. The image $I^{(l)}$ is first decomposed following the well-known integer-to-integer $5/3$ scheme employed for the lossless mode of JPEG 2000 [23]. According to our notations, the spatial supports for the prediction and update operators are: $\mathcal{P}_j^{(l)} = \{-1,0\}$, $\mathcal{U}_j^{(l)} = \{0,1\}$ and their related weights are: $p_{j,-1}^{(l)} = p_{j,0}^{(l)} = \frac{1}{2}$, $u_{j,0}^{(l)} = u_{j,1}^{(l)} = \frac{1}{4}$. The hybrid intra-inter prediction step related to $I^{(r)}$ is then expressed via the following spatial supports: $\mathcal{P}_j^{(r)} = \{-1,0\}$, $\mathcal{P}_j^{(r,l)} = \{0\}$. In other words, the prediction mask contains the same spatial prediction indices as those used in the $5/3$ scheme and the co-located position in the left image. As the detail coefficients can be viewed as prediction errors, the prediction coefficients $p_{j,k}^{(r)}$ and $p_{j,k}^{(r,l)}$ can be optimized at each resolution level by solving the well-known Yule-Walker equations. Concerning the update step, it is possible to generalize the optimization procedure described in [25], [26] in order to adapt the underlying operators to the statistical properties of the input image. A straightforward alternative solution that we preferred in our experiments consisted of choosing the same update operator at all resolution levels, the update employed for $I^{(r)}$ being the same as the two-tap filter employed for $I^{(l)}$ in Eq (5).

C. An improved VLS

One of the potential drawbacks of the previous VLS-I structure is that it generates an update leakage effect, in the sense that the information coming from the left view, which is used for the prediction of the right one, is also used, through the update operator, to compute the approximation coefficients of the right view. An alternative solution is given by the predict-update-predict (P-U-P) lifting structure shown in Fig. 3. The improved decomposition is described as follows:

$$\widetilde{d}_{j+1}^{(r)}(m_x,m_y) = I_j^{(r)}(m_x,2m_y+1) - \lfloor (\mathbf{P}_j^{(r)})^\top \mathbf{I}_j^{(r)} \rfloor, \quad (7)$$

$$\widetilde{I}_{j+1}^{(r)}(m_x,m_y) = I_j^{(r)}(m_x,2m_y) + \lfloor (\mathbf{U}_j^{(r)})^\top \widetilde{\mathbf{d}}_{j+1}^{(r)} \rfloor, \quad (8)$$

$$\begin{aligned}\check{d}_{j+1}^{(r)}(m_x,m_y) = \widetilde{d}_{j+1}^{(r)}(m_x,m_y) - \lfloor \mathbf{q}_j^\top \widetilde{\mathbf{I}}_{j+1}^{(r)} \\ + (\mathbf{P}_j^{(r,l)})^\top \mathbf{I}_j^{(c)} \rfloor, \quad (9)\end{aligned}$$

where notations similar to those used in Section III-B are used and $\mathbf{q}_j = (q_{j,k})_{k\in\mathcal{Q}_j}$ is the second intra-image predictor associated to the reference vector $\widetilde{\mathbf{I}}_{j+1}^{(r)} = \left(\widetilde{I}_{j+1}^{(r)}(m_x,m_y-k)\right)_{k\in\mathcal{Q}_j}$. It is worth noting that a prediction and an update as in (7) and (8) (with the same weights) are applied to $I^{(l)}$. In addition, at the last resolution level

[1] As we apply a separable decomposition, we denote by $\widetilde{I}_j$ the approximation coefficients after the first mono-dimensional processing at the j^{th} level, and by I_j the final approximation subband.

$j = J$, instead of directly coding the approximation $I_J^{(r)}$, we predict it from the approximation subband at the last level of the disparity-compensated image and only code the residual subband $e_J^{(r)}$ given by:

$$e_J^{(r)}(m_x, m_y) = I_J^{(r)}(m_x, m_y) - \lfloor (\mathbf{P}_J^{(r,l)})^\top \mathbf{I}_J^{(c)} \rfloor \qquad (10)$$

where

$$\mathbf{P}_J^{(r,l)} = \left(p_{J,k}^{(r,l)}\right)_{k \in \mathcal{P}_J^{(r,l)}} \text{ and } \mathbf{I}_J^{(c)} = \left(I_J^{(c)}(m_x, m_y, k)\right)_{k \in \mathcal{P}_J^{(r,l)}}.$$

Let the coefficients $q_{j,k}$ and $p_{j,k}^{(r,l)}$ (resp. $p_{J,k}^{(r,l)}$) be optimized so as to minimize the variance of $\check{d}_{j+1}^{(r)}$ (resp. $e_J^{(r)}$) at each resolution level $j < J$ (resp. at the coarsest resolution level J). An interesting property of the proposed decomposition is the following: in the ideal situation corresponding to $I^{(l)} = I^{(r)}$, the multiresolution representation of $I^{(r)}$ reduces to zero under some constraints that we are going to define in the following. Indeed, in the ideal case when $I^{(l)} = I^{(r)}$, the disparity vectors are zero: $(v_x, v_y) = (0, 0)$. Therefore, (9) becomes:

$$\begin{aligned}
\check{d}_{j+1}^{(r)}(m_x, m_y) &= \widetilde{d}_{j+1}^{(r)}(m_x, m_y) - \lfloor \mathbf{q}_j^\top \widetilde{\mathbf{I}}_{j+1}^{(r)} + (\mathbf{P}_j^{(r,l)})^\top \mathbf{I}_j^{(l)} \rfloor \\
&= I_j^{(r)}(m_x, 2m_y + 1) \\
&\quad - \lfloor \sum_{k \in \mathcal{P}_j^{(r)}} p_{j,k}^{(r)} I_j^{(r)}(m_x, 2m_y - 2k) \rfloor \\
&\quad - \lfloor \sum_{k \in \mathcal{Q}_j} q_{j,k} \widetilde{I}_{j+1}^{(r)}(m_x, m_y - k) \\
&\quad + \sum_{k \in \mathcal{P}_j^{(r,l)}} p_{j,k}^{(r,l)} I_j^{(l)}(m_x, 2m_y + 1 - k) \rfloor. \qquad (11)
\end{aligned}$$

It is worth pointing out that the coefficients $q_{j,k}$ and $p_{j,k}^{(r,l)}$ are optimized, at each decomposition level, by solving the Yule-Walker equations, the rounding operator being omitted. Thus, the detail coefficients $\check{d}_{j+1}^{(r)}(m_x, m_y)$ can be viewed as the errors involved in the prediction of $\widetilde{d}_{j+1}^{(r)}(m_x, m_y)$ by the signal $t(m_x, m_y) = \sum_{k \in \mathcal{Q}_j} q_{j,k} \widetilde{I}_{j+1}^{(r)}(m_x, m_y - k) + \sum_{k \in \mathcal{P}_j^{(r,l)}} p_{j,k}^{(r,l)} I_j^{(l)}(m_x, 2m_y + 1 - k)$. In this way, we can ensure that the detail coefficients of the right image $\check{d}_{j+1}^{(r)}(m_x, m_y)$ are zero if the prediction signal $t(m_x, m_y)$ is a linear combination of (at least) the same samples as those used by the reference signal $\widetilde{d}_{j+1}^{(r)}(m_x, m_y)$ to be predicted. This can be guaranteed provided that the support of the hybrid predictor $\mathcal{P}_j^{(r,l)}$ satisfies the two following conditions:

(i) The first term $I_j^{(r)}(m_x, 2m_y + 1)$ in the expression of $\widetilde{d}_{j+1}^{(r)}(m_x, m_y)$ in Eq. (7) can be found in the expression of the prediction signal $t(m_x, m_y)$ if $0 \in \mathcal{P}_j^{(r,l)}$.

(ii) The second term in the expression of $\widetilde{d}_{j+1}^{(r)}(m_x, m_y)$ involves the samples $(I_j^{(r)}(m_x, 2m_y - 2k))_{k \in \mathcal{P}_j^{(r)}}$. These samples can be found in the expression of the prediction signal $t(m_x, m_y)$ if $\{2k+1, k \in \mathcal{P}_j^{(r)}\} \subset \mathcal{P}_j^{(r,l)}$.

When the conditions (i) and (ii) are satisfied and $I_j^{(l)} = I_j^{(r)}$, the decomposition of $I_j^{(r)}$ first provides a detail subband $\check{d}_{j+1}^{(r)}(m_x, m_y)$ which is equal to zero and an approximation subband $\widetilde{I}_{j+1}^{(r)}$ which is equal to that of $I_j^{(l)}$: $\widetilde{I}_{j+1}^{(r)}(m_x, m_y) = \widetilde{I}_{j+1}^{(l)}(m_x, m_y)$. Then, while processing the image along the columns, the decomposition of $\widetilde{I}_{j+1}^{(r)}$ generates in the same way a detail subband which is equal to zero and an approximation subband $I_{j+1}^{(r)}$ which is equal to that of $\widetilde{I}_{j+1}^{(l)}$: $I_{j+1}^{(r)}(m_x, m_y) = I_{j+1}^{(l)}(m_x, m_y)$. Finally, the decomposition of $\check{d}_{j+1}^{(r)}$ provides two null detail subbands since $\check{d}_{j+1}^{(r)}(m_x, m_y) = 0$. Consequently, the resulting multiresolution representation of $I_j^{(r)}$ based on the new scheme allows us to generate an approximation subband which is identical to that of $I_j^{(l)}$ and three detail subbands equal to zero. Since at each resolution level the approximation subbands of $I_j^{(r)}$ and $I_j^{(l)}$ are equal, the residual sub-image $e_J^{(r)}$ in (10) becomes null if $0 \in \mathcal{P}_J^{(r,l)}$. Therefore, the P-U-P decomposition satisfies the property of cancelling the values of the wavelet coefficients of the multiresolution representation of $I^{(r)}$ provided that $\{0\} \cup \{2k+1, k \in \mathcal{P}_j^{(r)}\} \subset \mathcal{P}_j^{(r,l)}$, when $j < J$, and $0 \in \mathcal{P}_J^{(r,l)}$. This is a desirable property of the considered decomposition in order to get a consistent joint representation of $I^{(l)}$ and $I^{(r)}$. In contrast, this property does not hold for VLS-I.

Finally, as a supporting example, we design a scheme, which will be denoted in the following by VLS-II, by adding a prediction stage to the conventional 5/3 lifting structure. This amounts to choose $\mathcal{P}_j^{(r)} = \{-1, 0\}$, $\mathcal{U}_j^{(r)} = \{0, 1\}$, and $p_{j,-1}^{(r)} = p_{j,0}^{(r)} = \frac{1}{2}$, $u_{j,0}^{(r)} = u_{j,1}^{(r)} = \frac{1}{4}$, while the last prediction stage is performed by setting $\mathcal{Q}_j = \{-1, 0\}$, and $\mathcal{P}_j^{(r,l)} = \{-3, \ldots, 3\}$ for $j \in \{0, \ldots, J-1\}$ and $\mathcal{P}_J^{(r,l)} = \{0\}$. The coefficients $q_{j,k}$ and $p_{j,k}^{(r,l)}$ are determined by solving the Yule-Walker equations (still omitting the rounding operations) and imposing again the symmetry properties: $q_{j,-1} = q_{j,0}$ and $p_{j,k}^{(r,l)} = p_{j,-k}^{(r,l)}$ (which allows us to obtain linear phase filters often considered as desirable for image coding [27]).

IV. Theoretical analysis

In this section, we perform a theoretical analysis of the performances of VLS-I and VLS-II in terms of prediction efficiency, which is directly related to the coding efficiency [28], [29]. Firstly, we give the explicit expressions of the optimal prediction coefficients as well as the variance of detail signals for the two schemes. Then, we confirm that the prediction error variance of VLS-II is smaller than the one of VLS-I.

A. Notations

In the following, we will develop our analysis in the case of 1D signals, since we have considered a separable scheme. More precisely, let (m_x, m_y) be a given pixel, we consider

the pair of 1D signals defined for all $n \in \mathbb{Z}$ by:

$$\left\{ \begin{array}{l} i_j^{(r)}(n) = I_j^{(r)}(m_x, n) \\ i_j^{(l)}(n) = \\ I_j^{(l)}(m_x + v_{x,j}(m_x, 2m_y+1), n + v_{y,j}(m_x, 2m_y+1)). \end{array} \right. \tag{12}$$

We assume that, at a given resolution level j, these signals satisfy the following symmetric linear statistical model:

$$\left\{ \begin{array}{ll} i_j^{(r)}(n) & = \alpha_j a_j(n) + \beta_j b_j(n) \\ i_j^{(l)}(n) & = \beta_j a_j(n) + \alpha_j b_j(n) \end{array} \right. \tag{13}$$

where $(\alpha_j, \beta_j) \in \mathbb{R}^2$ such that $\alpha_j^2 + \beta_j^2 = 1$, and a_j and b_j are two stationary random processes which are mutually independent. For the sake of simplicity, we assume that they are zero-mean (which is always verified for wavelets coefficients) and they have the same autocorrelation function R_j with $R_j(0) > 0$. Then, it is easy to show that:

$$\mathsf{E}\left[i_j^{(r)}(n) i_j^{(r)}(n-k)\right] = \mathsf{E}\left[i_j^{(l)}(n) i_j^{(l)}(n-k)\right] = R_j(k), \tag{14}$$

$$\mathsf{E}\left[i_j^{(r)}(n) i_j^{(l)}(n-k)\right] = s_j R_j(k) \tag{15}$$

where $s_j \stackrel{\triangle}{=} \sin(2\theta_j)$ and $\theta_j \stackrel{\triangle}{=} \arg(\alpha_j + \imath\beta_j)$ (with $\imath^2 = -1$). At this point, it is worth noticing that the spatial similarities between the samples of $i_j^{(r)}$ (or $i_j^{(l)}$) are related to the autocorrelation function R_j. The factor θ_j controls the cross-redundancies between the samples of $i_j^{(r)}$ and $i_j^{(l)}$.

B. Minimum prediction error variance of VLS-I

By considering the support and the weights of the prediction operator involved in VLS-I (still omitting the rounding operators), the detail signal $\widetilde{d}_{j+1}^{(r)}(n)$ is expressed as follows:

$$\begin{aligned} \widetilde{d}_{j+1}^{(r)}(n) =& i_j^{(r)}(2n+1) - p_{j,0}^{(r)}\Big(i_j^{(r)}(2n) + i_j^{(r)}(2n+2)\Big) \\ & - p_{j,0}^{(r,l)} i_j^{(l)}(2n+1). \end{aligned} \tag{16}$$

Thus, $\widetilde{d}_{j+1}^{(r)}(n)$ can be viewed as the error in the prediction of $i_j^{(r)}(2n+1)$ by the multivariate reference signal $\mathbf{i}_j(n) \stackrel{\triangle}{=} \Big(i_j^{(r)}(2n) + i_j^{(r)}(2n+2), i_j^{(l)}(2n+1)\Big)^\top$. The prediction weight vector $\mathbf{p}_j = (p_{j,0}^{(r)}, p_{j,0}^{(r,l)})^\top$ satisfies the normal equations:

$$\mathsf{E}\left[\mathbf{i}_j(n)\mathbf{i}_j(n)^\top\right] \mathbf{p}_j = \mathsf{E}\left[i_j^{(r)}(2n+1)\mathbf{i}_j(n)\right]. \tag{17}$$

Hence, the optimal weights can be deduced as follows:

$$\left\{ \begin{array}{ll} p_{j,0}^{(r)} = & \gamma_{1,j}(\theta_j) R_j(0) R_j(1)(s_j^2 - 1) \\ p_{j,0}^{(r,l)} = & \gamma_{1,j}(\theta_j) s_j \Big(2R_j(1)^2 - Rj(0)^2 - R_j(0)R_j(2)\Big) \end{array} \right. \tag{18}$$

where $\gamma_{1,j}(\theta_j) \stackrel{\triangle}{=} \big(2s_j^2 R_j(1)^2 - R_j(0)^2 - R_j(0)R_j(2)\big)^{-1}$. Consequently, the minimum value $\varepsilon_{1,j}$ of the prediction error variance achieved by VLS-I is:

$$\begin{aligned} \varepsilon_{1,j}(R_j, \theta_j) =& \mathsf{E}\left[i_j^{(r)}(2n+1)^2\right] - \mathbf{p}_j^\top \mathsf{E}\left[i_j^{(r)}(2n+1)\mathbf{i}_j(n)\right] \\ =& \gamma_{1,j}(\theta_j)\cos^2(2\theta_j) R_j(0)\Big(2R_j(1)^2 - R_j(0)^2 \\ & - R_j(0)R_j(2)\Big). \end{aligned} \tag{19}$$

C. Minimum prediction error variance of VLS-II

Considering now VLS-II (still omitting the rounding operators), the detail signal $\check{d}_{j+1}^{(r)}(n)$ is given by:

$$\begin{aligned} \check{d}_{j+1}^{(r)}(n) =& i_j^{(r)}(2n+1) - \frac{1}{2}\Big(i_j^{(r)}(2n) + i_j^{(r)}(2n+2)\Big) \\ & - q_{j,0}\Big(\widetilde{i}_{j+1}^{(r)}(n) + \widetilde{i}_{j+1}^{(r)}(n+1)\Big) - p_{j,0}^{(r,l)} i_j^{(l)}(2n+1) \\ & - \sum_{k=1}^{3} p_{j,k}^{(r,l)}\Big(i_j^{(l)}(2n+1-k) + i_j^{(l)}(2n+1+k)\Big) \end{aligned} \tag{20}$$

where, as shown by (8), the signal $\widetilde{i}_{j+1}^{(r)}(n)$ can be expressed as:

$$\begin{aligned} \widetilde{i}_{j+1}^{(r)}(n) =& i_j^{(r)}(2n) + \frac{1}{4}\Big(\widetilde{d}_{j+1}^{(r)}(n-1) + \widetilde{d}_{j+1}^{(r)}(n)\Big) \\ =& \frac{1}{4}\Big(3i_j^{(r)}(2n) + i_j^{(r)}(2n+1) + i_j^{(r)}(2n-1)\Big) \\ & - \frac{1}{8}\Big(i_j^{(r)}(2n-2) + i_j^{(r)}(2n+2)\Big). \end{aligned} \tag{21}$$

Therefore, it can be checked that:

$$\begin{aligned} \check{d}_{j+1}(n) =& r_j(n) - q_{j,0} u_j(n) - p_{j,0}^{(r,l)} i_j^{(l)}(2n+1) \\ & - \sum_{k=1}^{3} p_{j,k}^{(r,l)}\Big(i_j^{(l)}(2n+1-k) + i_j^{(l)}(2n+1+k)\Big) \end{aligned} \tag{22}$$

where

$$r_j(n) \stackrel{\triangle}{=} i_j^{(r)}(2n+1) - \frac{1}{2}\Big(i_j^{(r)}(2n) + i_j^{(r)}(2n+2)\Big), \tag{23}$$

$$\begin{aligned} u_j(n) \stackrel{\triangle}{=}& \frac{1}{2} i_j^{(r)}(2n+1) + \frac{5}{8}\Big(i_j^{(r)}(2n) + i_j^{(r)}(2n+2)\Big) \\ & + \frac{1}{4}\Big(i_j^{(r)}(2n-1) + i_j^{(r)}(2n+3)\Big) \\ & - \frac{1}{8}\Big(i_j^{(r)}(2n-2) + i_j^{(r)}(2n+4)\Big). \end{aligned} \tag{24}$$

From (22), $\check{d}_{j+1}^{(r)}(n)$ can be viewed as the error in the prediction of $r_j(n)$ by the reference samples grouped into the vector $\tilde{\mathbf{r}}_j(n)$ given by:

$$\begin{aligned} \tilde{\mathbf{r}}_j(n) \stackrel{\triangle}{=}& \Big(u_j(n), i_j^{(l)}(2n+1), i_j^{(l)}(2n) + i_j^{(l)}(2n+2), \\ & i_j^{(l)}(2n-1) + i_j^{(l)}(2n+3), i_j^{(l)}(2n-2) + i_j^{(l)}(2n+4)\Big)^\top. \end{aligned} \tag{25}$$

The vector $\mathbf{p}_j = (q_{j,0}, p_{j,0}^{(r,l)}, p_{j,1}^{(r,l)}, p_{j,2}^{(r,l)}, p_{j,3}^{(r,l)})^\top$ is found by minimizing the variance of $\check{d}_{j+1}^{(r)}(n)$. Consequently, the following set of normal equations $\mathbf{\Gamma}_j \mathbf{p}_j = \mathbf{c}_j$ must be solved where $\mathbf{\Gamma}_j = \mathsf{E}\left[\tilde{\mathbf{r}}_j(n)\tilde{\mathbf{r}}_j(n)^\top\right]$ and $\mathbf{c}_j = \mathsf{E}\left[r_j(n)\tilde{\mathbf{r}}_j(n)\right]$. Once the auto-correlation matrix $\mathbf{\Gamma}_j$ and the cross-correlation vector

$\mathbf{c}_j$ are determined, the optimal weights are obtained as follows:

$$q_{j,0} = -4\gamma_{2,j}\Big(R_j(0) - 4R_j(1) + 4R_j(3) - R_j(4)\Big), \quad (26)$$

$$p_{j,0}^{(r,l)} = \gamma_{2,j}s_j\Big(40R_j(0) + 48R_j(1) + 31R_j(2) + 20R_j(3) - 8R_j(4) - 4R_j(5) + R_j(6)\Big), \quad (27)$$

$$p_{j,1}^{(r,l)} = -\gamma_{2,j}s_j\Big(33R_j(0) + 76R_j(1) + 31R_j(2) - 8R_j(3) - R_j(4) - 4R_j(5) + R_j(6)\Big)/2, \quad (28)$$

$$p_{j,2}^{(r,l)} = \gamma_{2,j}s_j\Big(R_j(0) - 4R_j(1) + 4R_j(3) - R_j(4)\Big), \quad (29)$$

$$p_{j,3}^{(r,l)} = -\gamma_{2,j}s_j\Big(R_j(0) - 4R_j(1) + 4R_j(3) - R_j(4)\Big)/2, \quad (30)$$

where $\gamma_{2,j} = \Big(38R_j(0) + 56R_j(1) + 31R_j(2) + 12R_j(3) - 6R_j(4) - 4R_j(5) + R_j(6)\Big)^{-1}$. Finally, the minimal value of the variance $\varepsilon_{2,j}$ of the prediction error generated by VLS-II is:

$$\begin{aligned}\varepsilon_{2,j}(R_j,\theta_j) &= \mathsf{E}\left[r_j^2(n)\right] - \mathbf{p}_j^\top \mathsf{E}\left[r_j(n)\tilde{\mathbf{r}}_j(n)\right] \\ &= \frac{1}{2}\gamma_{2,j}\cos^2(2\theta_j)\Big(113R_j(0)^2 - 240R_j(1)^2 + 31R_j(2)^2 \\ &- 16R_j(3)^2 - R_j(4)^2 - 4R_j(1)R_j(6) + 28R_j(0)R_j(3) \\ &- 16R_j(0)R_j(4) + 131R_j(0)R_j(2) - 4R_j(2)R_j(5) \\ &+ 3R_j(0)R_j(6) - 16R_j(1)R_j(3) + 16R_j(1)R_j(4) \\ &- 68R_j(1)R_j(2) + 16R_j(1)R_j(5) + 12R_j(2)R_j(3) \\ &+ R_j(2)R_j(6) + 8R_j(3)R_j(4) + 24R_j(0)R_j(1) \\ &- 6R_j(2)R_j(4) - 12R_j(0)R_j(5)\Big). \end{aligned} \quad (31)$$

D. Discussion

It should be noticed that the expressions of $\varepsilon_{1,j}$ and $\varepsilon_{2,j}$ are not restricted to a particular form of the autocorrelation function R_j, and so they are valid for any second-order stationary process. Furthermore, it is interesting to note that, unlike $\varepsilon_{1,j}$, $\varepsilon_{2,j}$ is separable in R_j and θ_j. In order to emphasize the advantages of VLS-I and VLS-II, we will consider a simple multivariate random process model driven by two autoregressive processes of order 1, $a_j(n)$ and $b_j(n)$ in (13). In this particular case, the autocorrelation function is given by:

$$\forall k \in \mathbb{Z}, \quad R_j(k) = \sigma_j^2 \rho_j^{|k|} \quad (32)$$

where $\rho_j \in [-1, 1]$ is the correlation factor. Therefore, the variances $\varepsilon_{1,j}$ and $\varepsilon_{2,j}$ of the prediction errors reduce to:

$$\varepsilon_{1,j}(R_j,\theta_j) = \sigma_j^2\tilde{\gamma}_{1,j}(\theta_j)\cos^2(2\theta_j)(\rho_j^2 - 1), \quad (33)$$

$$\varepsilon_{2,j}(R_j,\theta_j) = \frac{1}{2}\sigma_j^2\tilde{\gamma}_{2,j}\cos^2(2\theta_j)(1-\rho_j)(3\rho_j^4 - 16\rho_j^3 + 4\rho_j^2 + 24\rho_j + 113), \quad (34)$$

where $\tilde{\gamma}_{1,j}(\theta_j) = (2s_j^2\rho_j^2 - \rho_j^2 - 1)^{-1}$ and $\tilde{\gamma}_{2,j} = (\rho_j^5 - 5\rho_j^4 - \rho_j^3 + 13\rho_j^2 + 18\rho_j + 38)^{-1}$.
Furthermore, we can check that the variance of the intra prediction error generated by the 5/3 transform as indicated by (23) is given by:

$$\mathsf{E}\Big[r_j(n)^2\Big] = \frac{1}{2}(3-\rho_j)(1-\rho_j). \quad (35)$$

Fig. 4(a) shows the variations of $\mathsf{E}\Big[r_j(n)^2\Big]$, $\varepsilon_{1,j}$ and $\varepsilon_{2,j}$ with respect to ρ_j for a given value of θ_j. Thus, by taking into account the spatial redundancies (controlled by ρ_j), the variance $\varepsilon_{1,j}$ is smaller than $\mathsf{E}\Big[r_j(n)^2\Big]$. Lower values of the prediction error variance $\varepsilon_{2,j}$ are further achieved by the VLS-II transform for any value of ρ_j. We are also interested in comparing the variations of these three prediction errors with respect to θ_j for a given value of ρ_j, as depicted by Fig. 4(b). It can be noted that VLS-II gives also the best results by exploiting the inter-image redundancies (controlled by θ_j). This study has clearly shown the benefit that can be drawn from the use of VLS-II compared to VLS-I. This is explained by the proposed P-U-P structure in which the cross-view redundancies are exploited in the additional prediction step in order to avoid injecting the information coming from the reference image in the approximation of the target image.

V. Embedded coding of stereo images

A. Coding techniques

After applying a VLS to a stereo image pair, the generated coefficients must be encoded. However, the coding scheme should enable quality scalability for progressive reconstruction purposes. This is basically achieved by sending the coefficients in decreasing order of their importance. In other words, the most significant ones are first encoded at a reduced accuracy. So, a first approximation image is produced, which is further gradually refined by decoding the least significant coefficients. To this end, several scalable codecs have been developed [17], [30], [31], [32]. The main advantage of these embedded codecs is that the encoder can terminate the encoding at any point, thereby allowing a target bitrate to be exactly met. Similarly, the decoder can also stop decoding at any point resulting in the image that would have been produced at the rate corresponding to the truncated bitstream. In our experiments, we have employed the JPEG2000 codec, which yields excellent performance in terms of compression efficiency and quality scalability.

B. Transmission cost of the prediction coefficients

The prediction coefficients involved in the proposed VLS decompositions have to be transmitted to the decoder in order to proceed to the inverse transform with perfect reconstruction of the stereo pairs. The prediction weights correspond to an amount of $o_p = 3\,L\,J$ floating point coefficients, where L is the number of prediction weights in the VLS and J represents the number of resolution levels (the factor 3 stems from the fact that one horizontal prediction and two vertical predictions, one in the low-pass horizontal subband and the other in the high-pass horizontal subband, are performed). These weights are stored on 32 bits, inducing a negligible increase of the overall bitrate. More precisely, for a stereo pair of size $N_x \times N_y$, the transmission cost of the prediction coefficients will

increase the bitrate achieved by VLS-I and VLS-II, by $\frac{o_p}{2N_xN_y}$ bits per pixel. For example, when $N_x = N_y = 512$ and $J = 2$, the gain will be decreased by 0.0007 bpp (resp. 0.0018 bpp) in the case of VLS-I (resp. VLS-II) which is a very small fraction of the whole data bitrate.

VI. Experimental results

Simulations have been carried out on 6 image pairs of size 512×512 which have been extracted from a SPOT5 scene. The full scene, which corresponds to an urban zone, is shown in Fig. 1 and the six image pairs are represented in white squares. We have also used 4 pairs of natural stereo images ("fruit", "pentagon", "shrub" and "birch") downloaded from [2] and [3]. It should be noted that some stereo images have significant illumination variations between the views. For this reason, DC is performed by applying to the original SI the reversible remapping technique based on sorting permutations introduced in [33]. This preprocessing step is often used to improve the coding efficiency of pairs of images [34]. The disparity map is computed using a block-matching technique with a 8×8 block size and a search area that depends on the acquisition of the stereo pair (+50 pixels in the horizontal direction and ± 2 in the vertical direction for SPOT5 stereo images, and +30 pixels in the horizontal direction and ± 4 in the vertical direction for natural stereo images). The SSD is the chosen matching criterion. The resulting disparity vectors are losslessly encoded using a median prediction and DPCM with arithmetic encoding. In order to show the benefit of the joint coding by VLS, we compare VLS-I and VLS-II decompositions carried out over $J = 2$ resolution levels with some representative SI wavelet-based coding methods:

• The first one is the baseline coder which consists of coding the left image $I^{(l)}$ and the DC-residual $I^{(e)}$ with a 5/3 transform [16]. In the following, this method will be designated by scheme B.

• The second one is the subspace projection technique in the wavelet domain (SPT-WT) proposed by Jiang *et al.* [18]. This method consists of applying the DE and DC steps in the wavelet domain. More precisely, the method starts by applying the 5/3 transform to the original SI pair. We denote by $\{a_J^{(r)}, (d_j^{(r,o)})_{1\leq j\leq J},\ o \in \{1,2,3\}\}$ (resp. $\{a_J^{(l)}, (d_j^{(l,o)})_{1\leq j\leq J},\ o \in \{1,2,3\}\}$) the resulting approximation and detail subbands for the right (resp. left) image. A block-based DE is performed between the corresponding subbands $(a_J^{(r)}, a_J^{(l)})$ and $(d_J^{(r,o)}, d_J^{(r,o)})$. Then, a DC of each block of the image subbands is carried out, leading to the predicted subbands $\{\hat{a}_J^{(r)}, \hat{d}_j^{(r,o)},\ o \in \{1,2,3\}\}$. Finally, the computation of the DCD is obtained by projecting each block of the approximation subband of the target image $a_j^{(r)}$ onto the subspace $S = \text{span}\{\hat{a}_J^{(r)}, \hat{d}_J^{(r,o)},\ o \in \{1,2,3\}\}$, yielding the projection $\check{a}_J^{(r)} = \alpha_0 \hat{a}_J^{(r)} + \sum_{o=1}^{3} \alpha_o \hat{d}_J^{(r,o)}$ where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ are computed by a least squares approach. In our experiments, and in order to ensure a lossless reconstruction, we have encoded a rounded version of $\hat{a}_j^{(r)}$. Consequently, the approximation subband of the residual image is defined by $a_J^{(e)} = a_J^{(r)} - \lfloor \check{a}_J^{(r)} \rfloor$, whereas the other detail subbands are simply computed as: $d_j^{(e,o)} = d_j^{(r,o)} - \hat{d}_j^{(r,o)}$ $o \in \{1,2,3\}$.

• We have also tested a version of JPEG2000 (Annex I of Part II) dedicated to multicomponent images. It consists first of a decorrelation of the SI pair. Note that this decorrelation step must use a reversible transform in order to exactly recover the original SI pair. As a result, a pair $(\tilde{I}, I^{(e)})$ is produced by using a variation of the Haar transform [35]:

$$\begin{cases} I^{(e)}(m_x, m_y) = I^{(r)}(m_x, m_y) - I^{(l)}(m_x + v_x, m_y + v_y), \\ \tilde{I}(m_x + v_x, m_y + v_y) = \lfloor (I^{(l)}(m_x + v_x, m_y + v_y) \\ +I^{(r)}(m_x, m_y))/2 \rfloor \qquad \text{if } (m_x + v_x, m_y + v_y) \in \mathcal{S}, \\ \tilde{I}(m_x, m_y) = I^{(l)}(m_x, m_y) \quad \text{if } (m_x, m_y) \notin \mathcal{S}. \end{cases} \tag{36}$$

where $\mathcal{S}$ is the set of connected pixels in the left image. Then, the 5/3 transform is separately applied to $I^{(e)}$ and $\tilde{I}$. This method will be designated in the following by scheme C.

The compression measure is given by the final bitrates of the multiresolution representations. Let us denote by $R^{(v)}$, $R^{(l)}$, $R^{(r)}$ and $R^{(e)}$, respectively, the bitrate of the disparity vectors $\mathbf{v}$ and of the images $I^{(l)}$, $I^{(r)}$ and $I^{(e)}$. For the methods based on the coding of the residual image, we have computed the following average bitrate:

$$R_{\text{av}} = \frac{R^{(l)} + R^{(e)} + R^{(v)}}{2} \tag{37}$$

while the average bitrate for the proposed decompositions is given by:

$$R_{\text{av}} = \frac{R^{(l)} + R^{(r)} + R^{(v)}}{2}. \tag{38}$$

It can be noticed that the average coding cost $R^{(v)}$ of the losslessly encoded disparity vectors is around 0.07 bpp. Table I provides the final bitrates obtained in a lossless context by applying the JPEG2000 codec used only as an entropy codec on the produced subbands. Our simulations indicate that VLS-I results in an average gain of about 0.1 bpp over conventional methods. If we now compare the performance of VLS-II to those provided by VLS-I, our experiments show that VLS-II leads to a further improvement of about 0.1 bpp.

We have also tested the performance of our methods when applied as a lossy codec. In this case, the improved VLS are also compared in terms of Peak-Signal-to-Noise Ratio (PSNR) given by:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{(\text{MSE}^{(l)} + \text{MSE}^{(r)})/2} \right) \tag{39}$$

where $\text{MSE}^{(l)}$ and $\text{MSE}^{(r)}$ respectively correspond to the mean squared error of the left and right images reconstructed at the rates $R^{(l)}$ and $R^{(r)}$. We also used the SSIM quality metrics, which is based on models of visual perceptions, to evaluate the reconstruction quality of each compression method [36]. We are first interested in studying the evolution of the PSNR versus the bitrates achieved by VLS-I, VLS-II, the conventional schemes B and C, and the independent SI coder. In order to decode the SI pair, two alternatives can

[2] http://vasc.ri.cmu.edu/idb/html/stereo/index.html

[3] http://vasc.ri.cmu.edu/idb/html/jisct/

be envisaged. The most basic one consists of firstly decoding exactly the reference image. Then, the target image is decoded by using the original left image and the disparity vectors. However, in order to minimize the latency at the decoder side and to achieve the transmission of both images for a given bandwidth, we choose to simultaneously decode the SI pair. In other words, the decoding of the target image $I^{(r)}$ at a specified bitrate $R^{(r)}$ is achieved by using the decoded left image $\overline{I}^{(l)}$ at a bitrate $R^{(l)}$ without waiting for the final decoding of the reference image.

• More precisely, for the coding scheme B, the reconstructed target image $\overline{I}^{(r)}$ is obtained by using the reconstructed left image $\overline{I}^{(l)}$ and the residual image $\overline{I}^{(e)}$, decoded respectively at $R^{(l)}$ and $R^{(e)}$:

$$\overline{I}^{(r)}(m_x, m_y) = \overline{I}^{(e)}(m_x, m_y) + \overline{I}^{(l)}(m_x + v_x, m_y + v_y). \quad (40)$$

Then, by comparing the original images $I^{(l)}$ and $I^{(r)}$ with the reconstructed ones $\overline{I}^{(l)}$ and $\overline{I}^{(r)}$, we can evaluate the quality of reconstruction of the SI pair at the average bitrate defined by (37).

• Concerning the proposed methods, the reference image is decoded at different bitrates in the same way as in the previously mentioned methods. Then, the right image is decoded at some bitrate $R^{(r)}$ by using the reference image decoded at a bitrate $R^{(l)}$. Thus, we still evaluate the quality of reconstruction of the SI pair at the average bitrate given by (38). Figs. 5 and 6 show the scalability in quality with this reconstruction procedure by displaying the variations of the PSNR versus the bitrate for the SIs pair "shrub" and "spot5-6", using JPEG2000 as an entropy codec. These plots show that schemes B and C (based on the coding of the residual image) outperform the independent decomposition scheme, especially at low bitrates. VLS-I performs more poorly than these schemes at low bitrates but beyond some bitrate it is more performant. Finally, VLS-II outperforms all the schemes and improves the PSNR by at least 0.4 dB at high bitrate and the difference becomes much more important at low bitrates. Figs. 7 and 8 display a zoom applied on the reconstructed target image of the SI pairs "pentagon" and "spot5-5" for scheme B and VLS-II. We notice that the coding of the residual image leads to blocking artifacts at low bitrates. This problem is significantly reduced by resorting to VLS decompositions. Fig. 9 illustrates the reconstructed right image of the "shrub" pair at the decoder side corresponding to a progressive reconstruction. The quality of these images is compared both in terms of PSNR and SSIM. The difference in PSNR (resp. SSIM) between VLS-I and VLS-II ranges from 1.5 dB to 2 dB (resp. 0.05 to 0.1).

Finally, we propose to compare the different schemes in terms of execution time. Table II presents the encoding and decoding time of a Matlab implementation of the tested methods, at 0.2 bpp, for two stereo images of size 512×512. Simulations are carried out by using an Intel Core 2 (3 GHz) computer. We can note that the proposed methods VLS-I and VLS-II require respectively an additional average time of about 1.1 and 1.3 seconds compared to the residual image coding based method (scheme B). However, this difference in execution time is compensated by the good compression performance of the proposed VLS.

VII. Conclusions and perspectives

In this paper, we have presented a new technique for lossy-to-lossless compression of stereo image pairs. In order to take advantage of the correlations between the two images, we have proposed two schemes based on the vector lifting concept. Unlike conventional methods which generate a residual image to encode the stereo pair, the proposed schemes use a joint multiscale decomposition directly applied to the left and the right views. They exploit the intra and inter-image redundancies by using the estimated disparity map between the two views. Furthermore, the proposed decompositions guarantee the perfect reconstruction of the original stereo images. It is worth pointing out that these decompositions are also adapted to the content of the images. A theoretical analysis in terms of prediction error variance was conducted in order to show the benefits of the underlying VLS structure. Experimental results, carried out on a set of remote sensing and natural stereoscopic images, have indicated the good performance of the VLS over the conventional approaches in terms of bitrate and quality of reconstruction. In future work, we plan to improve the proposed decomposition by better taking into account the effect of occlusions. Also, an extension of the proposed scheme to multiview/video coding is currently envisaged.

References

[1] D. Tzovaras, I. Kompatsiaris and M. G. Strintzis, "3D object articulation and motion estimation in model-based stereoscopic videoconference image sequence analysis and coding," *Image Communication Journal*, vol. 14, no. 10, pp. 817-840, August 1999.

[2] M. Z. Brown, D. Burschka and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, August 2003.

[3] K. Tsutsui, S. Rokugawa, H. Nakagawa, S. Miyazaki, C.-T. Cheng, T. Shiraishi and S.-D. Yang, "Detection and volume estimation of large-scale landslides based on elevation-change analysis using DEMs extracted from high-resolution satellite stereo imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1681-1696, June 2007.

[4] G. W. Bothwell, E. G. Hansen, R. E. Vargo and K. C. Miller, "The multi-angle imaging spectroradiometer science data system, its products, tools, and performance," *IEEE Transactions on Geoscience and Remote Sensing*, no. 7, vol. 40, pp. 1467-1476, July 2002.

[5] R. Lau, S. Ince and J. Konrad, "Compression of still multiview images for 3-D automultiscopic spatially-multiplexed displays," *SPIE Conference on Stereoscopic Displays & Virtual reality Systems,* vol. 6490, pp. 0O1-0O9, San-Jose, CA, USA, 28 January- February 1, 2007.

[6] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 11, pp. 521-524, Tokyo, Japan, April 1986.

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

[8] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 3, pp. 328-344, June 1998.

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

[10] S. Wang and H. Chen, "An improved algorithm of motion compensation MPEG video compression," *International Vehicle Electronics Conference*, vol. 1, pp. 261-264, Changchun, China, September 1999.

[11] V. R. Dhond and J. K. Aggarwal, "Structure from stereo - a review," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489-1510, November 1989.

[12] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frames stereo correspondance algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7-42, June 2002.

[13] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 194-200, March 2000.

[14] S. Sodagari and E. Dubois, "Variable-block size disparity estimation in stereoscopic imagery," *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 3, pp. 2047-2049, Montreal, Canada, May 2003.

[15] H. Aydinoglu and M. Hayes, "Stereo image coding," *International Symposium on Circuits and Systems*, vol. I, pp. 247-250, Washington, USA, April 1995.

[16] 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.

[17] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445-3462, December 1993.

[18] Q. Jiang, J. J. Lee and M. H. Hayes, "A wavelet based stereo image coding algorithm," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3157-3160, Phoenix, Arizona, USA, March 1999.

[19] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publisher, 2004.

[20] A. Benazza-Benyahia, J.-C. Pesquet and M. Hamdi, "Vector lifting schemes for lossless coding and progressive archival of multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2011-2024, September 2002.

[21] M. Kaaniche, A. Benazza-Benyahia, J.-C. Pesquet and B. Pesquet-Popescu, "Lifting schemes for joint coding of stereoscopic pairs of satellite images," *European Signal and Image Processing Conference*, pp. 981-984, Poznán, Poland, September 2007.

[22] S. G. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, 1998.

[23] W. Sweldens, "The lifting scheme: a new philosophy in biorthogonal wavelet constructions," *In Wavelet Applications in Signal and Image Processing III, SPIE*, San-Diego, CA, USA, 2569, pp. 68-79, September 1995.

[24] F. J. Hampson and J.-C. Pesquet, "A nonlinear subband decomposition with perfect reconstruction," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1523-1526, Atlanta, GA, USA, 7-10 May 1996.

[25] A. Gouze, M. Antonini, M. Barlaud and B. Macq, "Design of signal-adapted multidimensional lifting schemes for lossy coding," *IEEE Transactions on Image Processing*, vol. 13, no. 12, pp. 1589-1603, December 2004

[26] B. Pesquet-Popescu, "Two-stage adaptive filter bank", first filling date 1999/07/27, official filling number 99401919.8, European patent number EP1119911.

[27] B. Belzer, J.-M. Lina and J. Villasenor, "Complex, linear-phase filters for efficient image coding," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 43, no. 10, pp. 2425-2427, October 1995.

[28] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1993.

[29] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, N.J: Prentice-Hall, 1984.

[30] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitionning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, June 1996.

[31] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158-1170, July 2000.

[32] S.-T. Hsiang and J. W. Woods, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling," *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 662-665, Geneva, Switzerland, May 2000.

[33] Z. Arnavut and S. Narumalani, "Application of permutations to lossless compression of multispectral thematic mapper images," *Optical Engineering*, vol. 35, no. 12, pp. 3442-3448, 1996.

[34] A. Benazza-Benyahia and J.-C. Pesquet, "Vector-lifting schemes based on sorting techniques for lossless compression of multispectral images," *SPIE Conference on Mathematics of Data/Image Coding, Compression and Encryption IV, with Applications*, vol. 4793, pp. 42-51, Seattle, WA, USA, July 7-11, 2002.

[35] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing* , vol. 3, pp. 1793-1796, Salt Lake City, UT, May 2001.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

LIST OF FIGURES

LIST OF TABLES

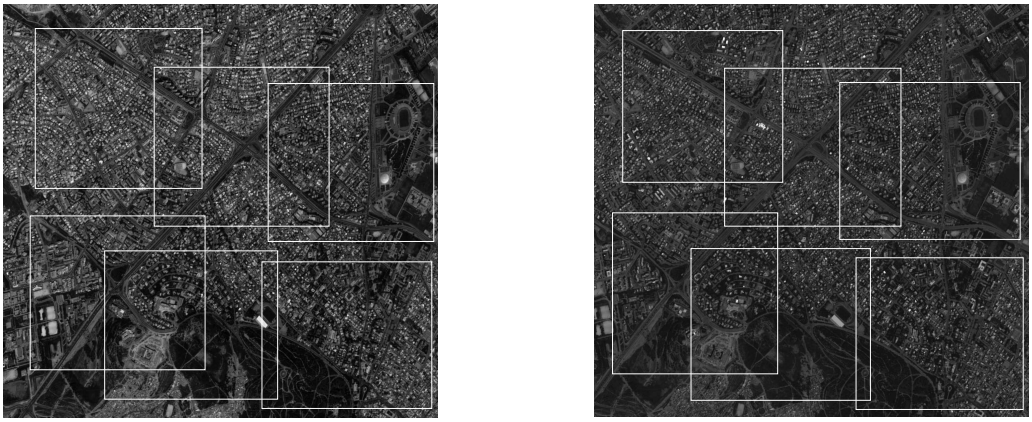


Fig. 1. Original SI pair "spot5": the left and right images.

TABLE I
PERFORMANCE OF SI WAVELET-BASED LOSSLESS CODECS IN TERMS OF AVERAGE BITRATE (IN BPP) USING JPEG2000.

Image	scheme B	SPT-WT	scheme C	VLS-I	VLS-II
spot5-1	3.63	3.59	3.58	**3.49**	**3.35**
spot5-2	3.85	3.80	3.78	**3.67**	**3.53**
spot5-3	4.27	4.21	4.24	**4.03**	**3.93**
spot5-4	4.22	4.18	4.21	**4.05**	**3.92**
spot5-5	3.91	3.87	3.89	**3.80**	**3.73**
spot5-6	3.89	3.84	3.81	**3.73**	**3.63**
fruit	4.05	3.99	3.97	**3.78**	**3.72**
shrub	3.73	3.69	3.69	**3.81**	**3.63**
birch	4.52	4.49	4.47	**4.44**	**4.37**
pentagon	5.37	5.32	5.20	**5.12**	**5.04**
Average	4.14	4.09	4.08	**3.99**	**3.88**

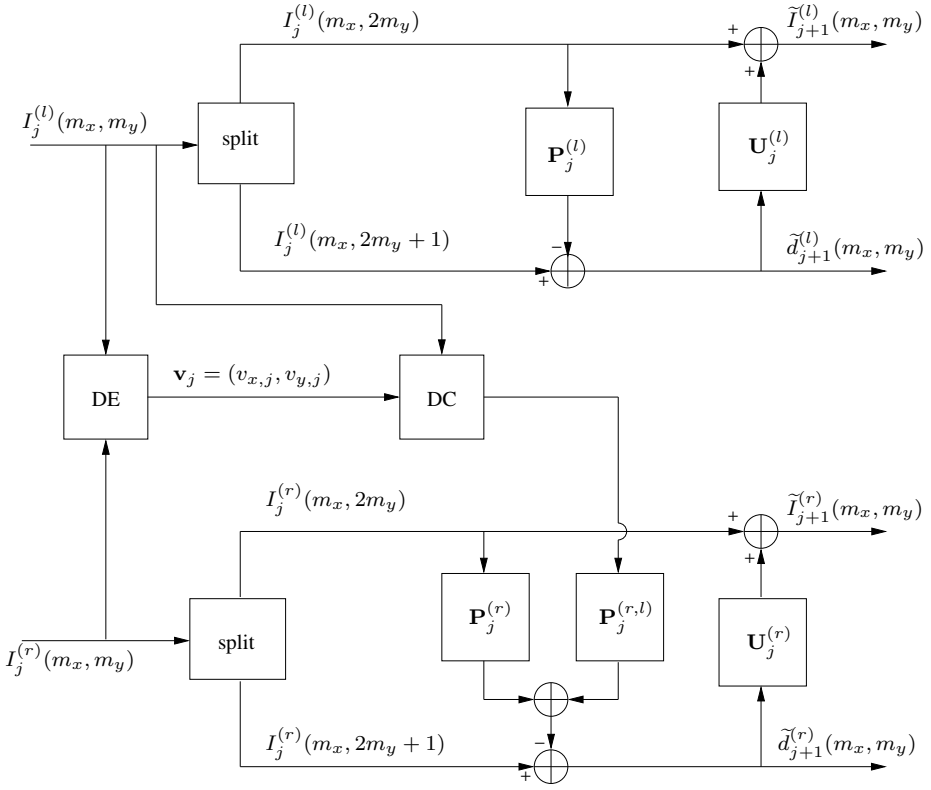


Fig. 2. Principle of the VLS-I decomposition.

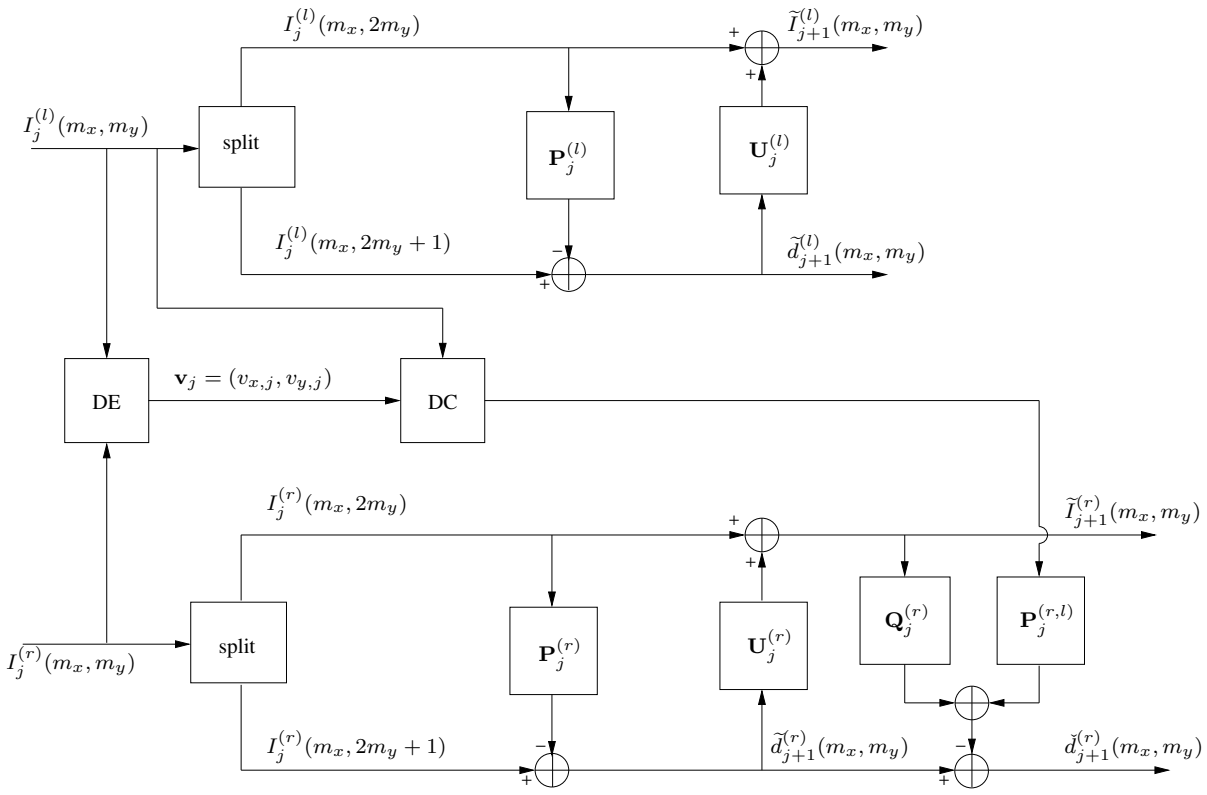
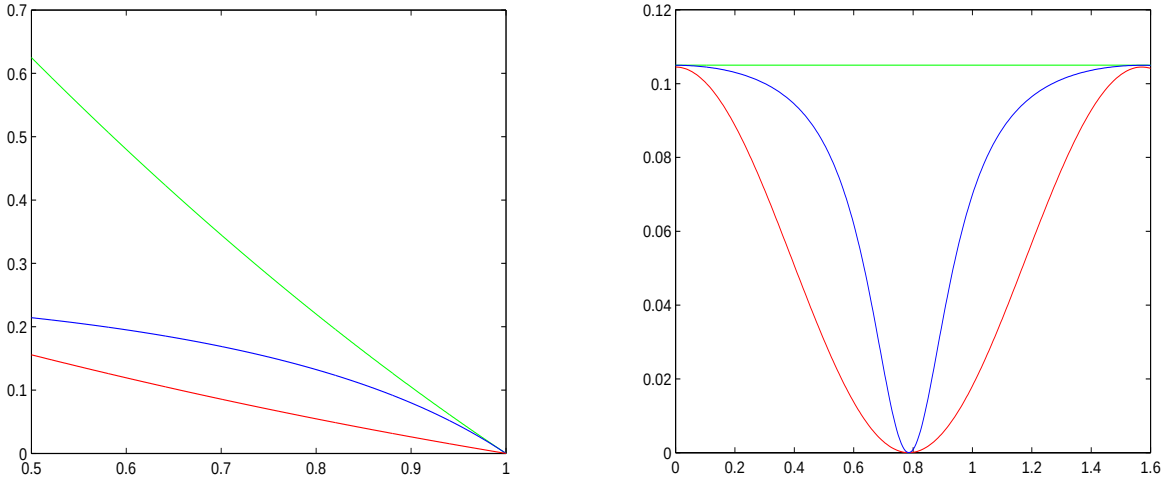


Fig. 3. Principle of the VLS-II decomposition.



(a): Variations w.r.t ρ_j when $\sigma_j = 1$ and $\theta_j = \frac{\pi}{6}$.

(b): Variations w.r.t θ_j when $\sigma_j = 1$ and $\rho_j = 0.9$.

Fig. 4. Prediction efficiency: $\mathsf{E}[r_j^2(n)]$ (in green), $\varepsilon_{1,j}(R_j, \theta_j)$ (in blue), $\varepsilon_{2,j}(R_j, \theta_j)$ (in red).

TABLE II
EXECUTION TIME OF THE PROPOSED METHODS (IN SECONDS).

Image	independent scheme		scheme B		VLS-I		VLS-II	
	encoding	decoding	encoding	decoding	encoding	decoding	encoding	decoding
spot5-6	0.57	0.15	0.83	0.49	2.29	1.20	2.44	1.46
fruit	0.55	0.15	0.84	0.50	2.31	1.25	2.58	1.48

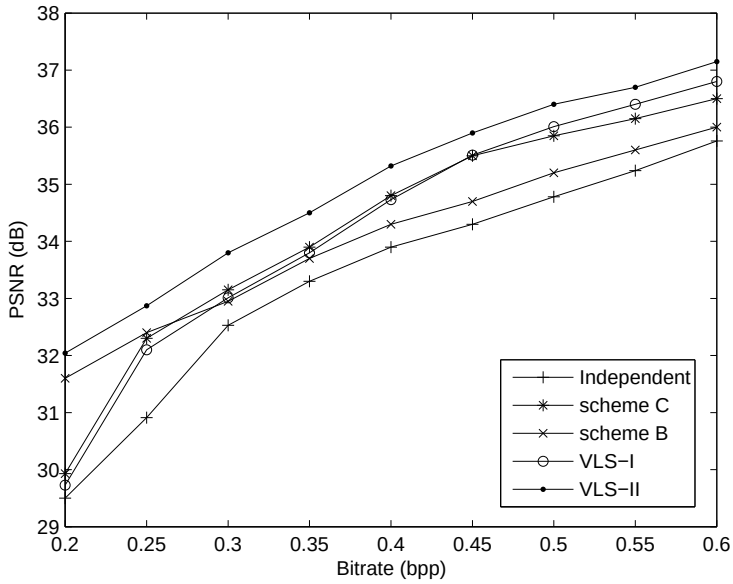


Fig. 5. PSNR (in dB) versus the bitrate (bpp) after JPEG 2000 encoding for the SI pair "shrub".

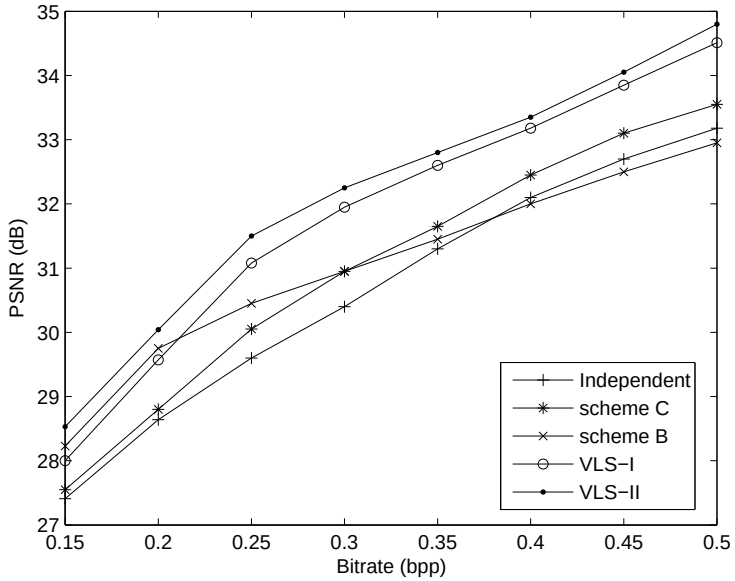


Fig. 6. PSNR (in dB) versus the bitrate (bpp) after JPEG 2000 encoding for the SI pair "spot5-6".

(a): PSNR=25.40 dB, SSIM=0.59

(b): PSNR=26.16 dB, SSIM=0.67

Fig. 7. Reconstructed target image $I^{(r)}$ of the "pentagon" pair at 0.2 bpp: (a) scheme B; (b) VLS-II.

(a): PSNR=30.03 dB, SSIM=0.80

(b): PSNR=31.48 dB, SSIM=0.83

Fig. 8. Reconstructed target image $I^{(r)}$ of the "spot5-5" pair at 0.13 bpp: (a) scheme B; (b) VLS-II.

Bitrate=0.15 bpp, PSNR=27.61 dB, SSIM=0.67

Bitrate=0.15 bpp, PSNR=29.04 dB, SSIM=0.76

Bitrate=0.2 bpp, PSNR=29.7 dB, SSIM=0.77

Bitrate=0.2 bpp, PSNR=31.45 dB, SSIM=0.83

Fig. 9. Reconstructed target image $I^{(r)}$ of the "shrub" pair at different bitrates: left column: VLS-I; right column VLS-II.