LIFTING SCHEMES FOR JOINT CODING OF STEREOSCOPIC PAIRS OF SATELLITE IMAGES)

M. Kaaniche[1], *A. Benazza-Benyahia*[1], *J.- C. Pesquet*[2] *and B. Pesquet-Popescu*[3]

[1] URISA, SUP'COM
Cité Technologique des
Communications, 2083, Tunisia
mounir.kaaniche@gnet.tn
benazza.amel@supcom.rnu.tn

[2] IGM and UMR-CNRS 8049,
Université de Paris-Est Marne-la-Vallée,
Champs-sur-Marne, 77454
Marne-la-Vallée, France
pesquet@univ-mlv.fr

[3] GET-ENST Paris,
Signal and Image Proc. Dept.
37-39, rue Dareau, 75014 Paris, France
pesquet@tsi.enst.fr

ABSTRACT

Stereo data compression is an important issue for the new generation of vision systems. In this paper, we are interested in lossless coding methods for stereo images allowing progressive reconstruction. Most of the existing approaches account for the mutual similarities between the left and the right images. More precisely, the disparity compensation process consists in predicting the right image from the left one based on the disparity map. Then, the disparity map, the reference image, and the residual image are encoded. In this work, we propose a novel approach based on the concept of vector lifting scheme. Its main feature is that it does not generate one residual image but two compact multiresolution representations of the left and the right views, driven by the underlying disparity map. Experimental results show a significant improvement using this technique compared with conventional methods.

1. INTRODUCTION

Stereo Images (SI) are associated with the same scene captured from two slightly different perspectives: a pair of a left and right view is generated. When each image of the stereo pair is viewed by the respective eye, the scene is perceived in three dimensions. This ability of obtaining depth perception and information has a wide range of applications such as medical surgical environments, telepresence in videoconferences, geographical information systems, computer games [1]. The increasing interest in SI has led to the constitution of image databases that require huge amounts of storage. Consequently, the compression of SI has become mandatory. Recently, an increased interest was paid to the investigation of this topic due to very promising openings especially for multimedia applications. Compressing SI is motivated by the existence of both intra-image redundancy due to the scene smoothness and inherent similarities between the two images resulting from the observation of a common scene from 2 viewpoints [2]. Indeed, if the two images are superimposed, some area in the first image can usually be found in the second image but at a different spatial position. Generally, the reported methods tend to exploit such cross-image redundancies by first estimating the displacement (called disparity or binocular parallax). Then, the disparity map and one image only of the SI pair (say the left image) is coded in addition to a residual information. This joint coding is similar to those encountered in interframe video coding. In this context, disparity compensation can be considered as a prediction technique [3]. Pioneering techniques worked in the spatial domain [2]. More recent methods are based on the Wavelet Transform (WT) [4, 5]. In [4], the disparity compensation takes place in the wavelet domain, the coding of the wavelet coefficients being performed through a subspace projection technique where a different basis is built for the representation.

The objective of this paper is to design a novel joint coding method for stereoscopic still images that allows both a gradual and exact decoding. Progressive reconstruction is a desired functionality for telebrowsing applications since the refinement of the decoded images depends on the user's needs. Furthermore, the decoder should be able to recover exactly the SI pair. Such constraint of lossless reconstruction may be required in some applications such as medical imaging or remote sensing imaging: an even minor distortion in the reconstructed images could affect the interpretation of the scene.

This paper is organized as follows. Section 2 is devoted to a short overview of disparity compensated coding methods. In Section 3, the proposed coding schemes are presented. Two variants are considered, the second one being shown to provide better analyses. In Section 4, we also conduct a theoretical analysis of the proposed schemes in terms of prediction error. Finally, in Section 5, experimental results are given and some conclusions are drawn in Section 6.

2. DISPARITY COMPENSATED CODER

2.1 Binocular imaging and disparity estimation

For the sake of both simplicity and efficiency, a binocular imaging system makes use of two sensors (the left and the right one) that have parallel optical axes. Furthermore, it is useful to consider central (or perspective) projections, whose centers are denoted by $O^{(l)}$ and $O^{(r)}$. As a result, a stereoscopic pair is formed by the image $I^{(l)}$ and $I^{(r)}$. Generally, a point P in the 3D space is visible in $I^{(l)}$ (resp. $I^{(r)}$) if its perspective projection on the left (resp. right) image plane is a point $P^{(l)}$ (resp. $P^{(r)}$) of $I^{(l)}$ (resp. $I^{(r)}$). If $I^{(l)}$ is superimposed on $I^{(r)}$, the position matching points $P^{(l)}$ and $P^{(r)}$ do not coincide. The distance between $(P^{(l)},P^{(r)})$ is called the disparity of point P (or binocular parallax). Therefore, a joint coding of $(I^{(l)},I^{(r)})$ attempts to exploit the similarities between $I^{(l)}$ and $I^{(r)}$ through the disparity estimation. The latter operation consists in pairing the respective points of $I^{(l)}$ and $I^{(r)}$ in such a way they correspond to the projection of the same 3D point. Many methods have been investigated [6] for this purpose. The most simple Disparity Compensation (DC) process relies on a block-matching technique similar to those employed in motion estimation for video coding. It consists in first partitioning $I^{(r)}$ into nonoverlapping blocks $b^{(r)}$ of size $h \times w$ and, for each block, finding the most "similar" block within a given search area S in $I^{(l)}$. More precisely, the disparity vector $\mathbf{v} = (v_x, v_y)$ for a current block in $I^{(r)}$ minimizes a dissimilarity criterion D:

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

where (m_x,m_y) are the spatial coordinates associated with the top leftmost pixel in the block. Often, the Sum of Square Differences (SSD) is the selected criterion:

$$D(v_x,v_y) \stackrel{\triangle}{=} \sum_{(m_x,m_y)\in b^{(r)}} (I^{(r)}(m_x,m_y) - I^{(l)}(m_x+v_x, m_y+v_y))^2. \quad (2)$$

Therefore, it is possible to predict $I^{(r)}(m_x,m_y)$ by $I^{(l)}(m_x+v_x,m_y+v_y)$. Further improvements can be achieved by employing overlapped blocks $b^{(r)}$ with an adaptive search window S [7]. Another basic disparity estimation is the correlation method that exploits the epipolar line constraint [8]. Indeed, given a current point $P^{(r)}$ in $I^{(r)}$, the intersection of the plane $(O^{(r)},P^{(r)},O^{(l)})$ with the image plane of the right camera gives rise to the so-called epipolar line. As the 3D point P associated to $P^{(r)}$ should lie somewhere on the line $(O^{(l)}P^{(l)})$, the projection $P^{(l)}$ of P onto the left image should also lie on the epipolar line. In practice, $P^{(l)}$ is not rigorously on the epipolar line because of the noise camera, 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 that strip can be considered as potential candidates $P^{(l)}$ to be paired with the current point $P^{(r)}$. The retained candidate is the point $P^{(l)}$ that maximizes the correlation between the two windows centered respectively at $P^{(r)}$ and $P^{(l)}$. Finally, a dense disparity field is generated since a disparity vector $\mathbf{v}$ is obtained for each current point $P^{(r)}$. More sophisticated density estimation methods were presented and evaluated as those using dynamic programming [9] or advanced convex optimization methods [10]. Once the disparity vectors are generated, several schemes can be envisaged for the coding of the SI pair.

2.2 State of the art

Coding of SI pairs is generally based on exploiting the inter-image similarities and hence, designing *joint* coding schemes. Inspired by the works in video, several disparity compensated techniques were developed. Most existing algorithms operate in the spatial domain. Basically, they proceed as follows. After estimation of the disparity $\mathbf{v}$, the prediction error $I^{(e)}$ is computed:

$$I^{(e)}(m_x,m_y)\stackrel{\triangle}{=}I^{(r)}(m_x,m_y)-I^{(l)}(m_x+v_x,m_y+v_y). \quad (3)$$

The disparity field, the left image and the residual one are then encoded. It is worth pointing out that a lossless coding is required for the $I^{(e)}$ signal in order to ensure an exact reconstruction. As many choices can be made in the settings of each of the 3 coders, many joint coding methods have been already proposed in the spatial domain [2, 11, 12]. The residual images can also be coded in the discrete cosine transform domain [13, 14]. However, a great attention was paid to the wavelet transform domain to meet the scalability requirement. For instance, the wavelet transform is applied to both the reference image $I^{(l)}$ and to the residual image in [15]. A more sophisticated approach was presented in [4]: both the estimation and the disparity compensation take place in the wavelet domain, subspace projection techniques being used for the coding of the wavelet coefficients. In [5], an efficient exploitation of the zerotree algorithm is carried out to reduce short embedded bitstreams of the wavelet coefficients of both the reference and the residual images.

3. PROPOSED JOINT CODING SCHEMES

3.1 Motivations

In this paper, we design new compression methods for stereo pairs. Our approach relies on a joint coding of $I^{(l)}$ and $I^{(r)}$ that exploits judiciously the available disparity map. Such a coding technique is achieved via a *Vector* Lifting Scheme (VLS) [16]. Unlike existing methods, the proposed approach does not explicitly generate a residual image but two multiresolution representations of $I^{(l)}$ and $I^{(r)}$. This can be viewed as an appealing property. Indeed, the proposed coding approach provides a more symmetric processing of the right and left images. Furthermore, the pyramidal structure of the generated coefficients allows the use of efficient encoding strategies e.g. *vector*-embedded zerotree algorithms [17].

3.2 Disparity based VLS

The wavelet coefficients of an image are usually obtained by a 2-band filter bank structure [18]. If an exact reconstruction is required, lifting schemes are often employed since they allow to construct an integer-valued version of the wavelet coefficients whatever the underlying operators are [19]. For the sake of simplicity, a separable decomposition is considered in this paper. Therefore, it is enough to address the decomposition of a given line m_x. 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 respectively $I^{(l)}$ and $I^{(r)}$ are the input coefficients of the lifting scheme. The objective of the vector lifting 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 $\widetilde{I}_{j+1}^{(l)}$, $\widetilde{I}_{j+1}^{(r)}$. Similar lifting structures operating along the image columns allow us to generate $I_{j+1}^{(l)}$ and $I_{j+1}^{(r)}$ as well as the associated detail coefficients in the horizontal, vertical and diagonal directions at resolution level $j+1$. A wide range of nonlinear operators can be applied to extract the correlations. 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 $(j+1)$ expressed as:

$$\widetilde{d}_{j+1}^{(l)}(m_x,m_y)=I_j^{(l)}(m_x,2m_y+1)-\lfloor\sum_{k\in\mathscr{P}_j^{(l)}}p_{j,k}^{(l)}I_j^{(l)}(m_x,2m_y-2k)\rceil, \quad (4)$$

where $\lfloor\cdot\rceil$ designates the rounding operation. The set $\mathscr{P}_j^{(l)}$ and the coefficients $p_{j,k}^{(l)}$ denote respectively the support and the weights of the predictor of $I_j^{(l)}(m_x,2m_y+1)$. Then, in the update step, the approximation coefficients are computed as follows:

$$\widetilde{I}_{j+1}^{(l)}(m_x,m_y)=I_j^{(l)}(m_x,2m_y)+\lfloor\sum_{k\in\mathscr{U}_j^{(l)}}u_{j,k}^{(l)}\widetilde{d}_{j+1}^{(l)}(m_x,m_y-k)\rceil, \quad (5)$$

where the set $\mathscr{U}_j^{(l)}$ is the support of the update operator whose coefficients are $u_{j,k}^{(l)}$. The reversibility of the basic lifting scheme is ensured since the prediction in (4) makes use of even indices only. 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. The location of the latters is given by the transformed disparity vector $\mathbf{v}_j=(v_{x,j},v_{y,j})$ associated with the pixel $(m_x,2m_y+1)$ to be predicted:

$$\widetilde{d}_{j+1}^{(r)}(m_x,m_y)=I_j^{(r)}(m_x,2m_y+1)-\lfloor\sum_{k\in\mathscr{P}_j^{(r,r)}}p_{j,k}^{(r,r)}I_j^{(r)}(m_x,2m_y-2k)$$
$$+\sum_{k\in\mathscr{P}_j^{(r,l)}}p_{j,k}^{(r,l)}I_j^{(l)}(m_x+v_{x,j}(m_x,2m_y+1),2m_y+1+v_{y,j}(m_x,2m_y+1)-k)\rceil, \quad (6)$$

where $\mathbf{v}_j$ is obtained by filtering and subsampling the initial (and full resolution) disparity field $\mathbf{v}$. For example, one can choose:

$$\mathbf{v}_j(m_x,2m_y+1)=\mathbf{v}(2^jm_x,2^j(2m_y+1)). \quad (7)$$

In other words, the vector lifting makes use of a hybrid predictor that exploits at the same time the intra and inter-image redundancies in the stereo pair. The update step for $\widetilde{I}_{j+1}^{(r)}$ can be performed

similarly to Eq. (5). The decomposition is iterated on the columns m_y of the resulting subbands resulting in 2×4 sub-images for the left and right images at each resolution level j and the decomposition is 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. An appropriate choice of the involved prediction and update operators remains however necessary in order to generate compact representations of $(I^{(l)},I^{(r)})$.

3.3 Optimization of the predictors

As the detail coefficients can be viewed as prediction errors, the prediction operators can be optimized so as to minimize the variance of these coefficients at each resolution level. If the rounding operators are omitted, it is straightforward to check that the minimum variance predictors must satisfy Yule-Walker's equations. Concerning the update, it is possible to generalize the optimization procedure described in [20] in order to adapt the underlying operators to the statistical properties of the input image. A simpler solution that we have retained in our experiments consists in choosing the same update operator at all resolution levels. Indeed, in our simulations it has been observed that the decrease of the entropy is mainly due to the optimization of the predict operators.

3.4 Example I

We provide a simple example (designated by VLS-I) of the considered lifting structure. The image $I^{(l)}$ is decomposed following the well-known $5/3$ scheme [19] described by: $\mathscr{P}_j^{(l)}=\{-1,0\}$, $\mathscr{U}_j^{(l)}=\{0,1\}$. The prediction and update weights are fixed: $p_{j,-1}^{(l)}=p_{j,0}^{(l)}=\frac{1}{2}$, $u_{j,0}^{(l)}=u_{j,1}^{(l)}=\frac{1}{4}$. The hybrid prediction step related to $I^{(r)}$ is expressed via the following supports: $\mathscr{P}_j^{(r,r)}=\{-1,0\}$, $\mathscr{P}_j^{(r,l)}=\{0\}$. In other words, the prediction mask contains the same spatial prediction indices as those used in the $5/3$ scheme and the collocated position in the left image. The prediction coefficients are obtained by the resolution of Yule-Walker's equations. The update operator is the same as the two-tap filter employed for $I^{(l)}$.

3.5 An improved lifting structure

One of the potential drawbacks of the previous structure is that it generates an update leakage effect in the sense that the information coming from the left view, which is used for prediction, is fed back to compute the approximation coefficients of the right view.

An alternative solution is given by the P-U-P lifting structure described as follows:

$$\widetilde{d}_{j+1}^{(r)}(m_x,m_y)=I_j^{(r)}(m_x,2m_y+1)-\lfloor\sum_{k\in\mathscr{P}_j^{(r,r)}}p_{j,k}^{(r,r)}I_j^{(r)}(m_x,2m_y-2k)\rceil, \quad (8)$$

$$\widetilde{I}_{j+1}^{(r)}(m_x,m_y)=I_j^{(r)}(m_x,2m_y)+\lfloor\sum_{k\in\mathscr{U}_j^{(r)}}u_{j,k}^{(r)}\widetilde{d}_{j+1}^{(r)}(m_x,m_y-k)\rceil, \quad (9)$$

$$\begin{aligned}d_{j+1}^{(r)}(m_x,m_y)&=\widetilde{d}_{j+1}^{(r)}(m_x,m_y)-\lfloor\sum_{k\in\mathscr{Q}_j}q_{j,k}\widetilde{I}_{j+1}^{(r)}(m_x,m_y-k)\\&+\sum_{k\in\mathscr{P}_j^{(r,l)}}p_{j,k}^{(r,l)}I_j^{(l)}(m_x+v_{x,j}(m_x,2m_y+1),2m_y+1+v_{y,j}(m_x,2m_y+1)-k)\rceil,\end{aligned} \quad (10)$$

where $\mathscr{Q}_j$ is the support of the second intra-image predictor for the right view and the corresponding prediction weights are denoted by $q_{j,k}$. It is worth noting that a prediction and an update as in (8) and (9) (with the same weights) are applied to $I^{(l)}$. In addition, at the last resolution level $j=J$, instead of coding directly the approximation $I_J^{(r)}$, we code the residual image:

$$\begin{aligned}e_J^{(r)}(m_x,m_y)&=I_J^{(r)}(m_x,m_y)\\&-\lfloor\sum_{k\in\mathscr{P}_J^{(r,l)}}p_{J,k}^{(r,l)}I_J^{(l)}(m_x+v_{x,J}(m_x,m_y),m_y+v_{y,J}(m_x,m_y)-k)\rceil.\end{aligned} \quad (11)$$

An interesting property of the proposed decomposition is the following: 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 $d_{j+1}^{(r)}$ (resp. $e_J^{(r)}$) at each resolution level j (resp. at the coarsest resolution level J). In the ideal situation when $I^{(l)}=I^{(r)}$, the obtained multiresolution representation of $I^{(r)}$ reduces to zero provided that $\{0\}\cup\{2k+1,k\in\mathscr{P}_j^{(r,r)}\}\subset\mathscr{P}_j^{(r,l)}$, when $j<J$, and $0\in\mathscr{P}_J^{(r,l)}$. In contrast, this property which may appear desirable in order to get a consistent joint representation of $I^{(l)}$ and $I^{(r)}$ is not satisfied by the decomposition presented in Section 3.2.

3.6 Example II

A new VLS (designated by VSL-II) can be built by adding a prediction stage to the conventional 5/3 lifting structure. This amounts to choosing $\mathscr{P}_j^{(r,r)}=\{-1,0\}$, $\mathscr{U}_j^{(r)}=\{0,1\}$, and $p_{j,-1}^{(r,r)}=p_{j,0}^{(r,r)}=\frac{1}{2}$, $u_{j,0}^{(r)}=u_{j,1}^{(r)}=\frac{1}{4}$, while the last prediction stage is performed by choosing $\mathscr{Q}_j=\{-1,0\}$ and $\mathscr{P}_j^{(r,l)}=\{-3,\ldots,3\}$ for $j\in\{0,\ldots,J-1\}$ and $\mathscr{P}_J^{(r,l)}=\{0\}$. The coefficients $q_{j,k}$ and $p_{j,k}^{(r,l)}$ are determined by solving Yule-Walker's equations (still omitting the rounding operations) and imposing the following symmetry properties: $q_{j,-1}=q_{j,0}$ and $p_{j,k}^{(r,l)}=p_{j,-k}^{(r,l)}$.

4. THEORETICAL ANALYSIS

In this section, we perform a theoretical analysis of the performances of VLS-II in terms of prediction efficiency. As the underlying schemes are separable, it is enough to develop our analysis in the case of 1D signals. More precisely, let (m_x,m_y) be a given pixel, we consider the pair of 1D signals defined by: for all $n\in\mathbb{Z}$,

$$i_j^{(r)}(n)=I_j^{(r)}(m_x,n), \quad (12)$$

$$i_j^{(l)}(n)=I_j^{(l)}(m_x+v_{x,j}(m_x,2m_y+1),n+v_{y,j}(m_x,2m_y+1)). \quad (13)$$

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

$$\left\{\begin{array}{ll}i_j^{(r)}(n) & =\alpha_j a_j(n)+\beta_j b_j(n)\\ i_j^{(l)}(n) & =\beta_j a_j(n)+\alpha_j b_j(n)\end{array}\right., \quad (14)$$

where $(\alpha_j,\beta_j)\in\mathbb{R}^2\setminus\{(0,0)\}$, and a_j and b_j are two autoregressive signals of order 1 which are mutually independent. For the sake of simplicity, we assume that they are zero-mean and they have the same (nonzero) variance and the same correlation factor $\rho_j\in[-1,1]$. Then, it is easy to show that:

$$\mathsf{E}[i_j^{(r)}(n)i_j^{(r)}(n-k)]=\mathsf{E}[i_j^{(l)}(n)i_j^{(l)}(n-k)]=\sigma_j^2\rho_j^{|k|}, \quad (15)$$

$$\mathsf{E}[i_j^{(r)}(n)i_j^{(l)}(n-k)]=\sigma_j^2 s_j\rho_j^{|k|}, \quad (16)$$

where $\sigma_j\in\mathbb{R}_+^*$, $s_j\stackrel{\triangle}{=}\sin(2\theta_j)$ and $\theta_j\stackrel{\triangle}{=}\arg(\alpha_j+\imath\beta_j)$. At this point, it is worth noticing that the spatial similarities between samples of $i_j^{(r)}$ (resp. $i_j^{(l)}$) are related to the correlation factor ρ_j. The factor θ_j controls the cross-redundancies between samples of $i_j^{(r)}$ and $i_j^{(l)}$. Thus, one can expect that the proposed lifing scheme will efficiently

reduce these correlations. If the rounding operators are omitted, it can be checked that

$$d_j^{(r)}(m_x,n) = r_j(n) - q_{j,0}\frac{2u_j(n)+v_j(n)}{4} - p_{j,0}^{(r,l)} i_j^{(l)}(2n+1) - \sum_{k=1}^{3} p_{j,k}^{(r,l)}(i_j^{(l)}(2n+1-k)+i_j^{(l)}(2n+1+k)) \quad (17)$$

where $r_j(n) \stackrel{\triangle}{=} i_j^{(r)}(2n+1) - \frac{1}{2}(i_j^{(r)}(2n) + i_j^{(r)}(2n+2))$, $u_j(n) \stackrel{\triangle}{=} 2(i_j^{(r)}(2n+1) + i_j^{(r)}(2n) + i_j^{(r)}(2n+2)) + i_j^{(r)}(2n-1) + i_j^{(r)}(2n+3)$ and $v_j(n) \stackrel{\triangle}{=} i_j^{(r)}(2n) + i_j^{(r)}(2n+2) - i_j^{(r)}(2n-2) - i_j^{(r)}(2n+4)$. From Eq. (17), $d_j^{(r)}(m_x,n)$ can be viewed as the error in the prediction of $r_j(n)$ by the reference signals grouped into the vector $\tilde{\mathbf{r}}_j(n)$ given by

$$\tilde{\mathbf{r}}_j(n) \stackrel{\triangle}{=} (\frac{2u_j(n)+v_j(n)}{4}, i_j^{(l)}(2n+1), i_j^{(l)}(2n) + i_j^{(l)}(2n+2), i_j^{(l)}(2n-1) + i_j^{(l)}(2n+3), i_j^{(l)}(2n-2) + i_j^{(l)}(2n+4))^\top. \quad (18)$$

The prediction weight vector $\mathbf{p}_j = (q_{j,0}, p_{1,0}^{(r,l)}, p_{1,1}^{(r,l)}, p_{1,2}^{(r,l)}, p_{1,3}^{(r,l)})^\top$ minimizes the variance of $d_j^{(r)}(m_x,n)$. Consequently, the normal equations must be solved:

$$\mathsf{E}[\tilde{\mathbf{r}}_j(n)(\tilde{\mathbf{r}}_j(n))^\top]\mathbf{p}_j = \mathsf{E}[r_j(n)\tilde{\mathbf{r}}_j(n)]. \quad (19)$$

By taking into account the adopted statistical model, the optimal weights can be calculated:

$$q_{j,0} = 2\gamma_j(\rho_j-1)(\rho_j^2-4\rho_j+1), \quad (20)$$

$$p_{j,0}^{(r,l)} = \gamma_j s(\rho_j^5 - 5\rho_j^4 - 3\rho_j^3 + 23\rho_j^2 + 8\rho_j + 40), \quad (21)$$

$$p_{j,1}^{(r,l)} = -\gamma_j s(\rho_j-3)(\rho_j^4 - 2\rho_j^3 - 2\rho_j^2 - 18\rho_j - 11)/2, \quad (22)$$

$$p_{j,2}^{(r,l)} = -\gamma_j s(\rho_j-1)(\rho_j^2-4\rho_j+1), \quad (23)$$

$$p_{j,3}^{(r,l)} = \gamma_j s(\rho_j-1)(\rho_j^2-4\rho_j+1)/2. \quad (24)$$

with $\gamma_j = (\rho_j^5 - 5\rho_j^4 - \rho_j^3 + 13\rho_j^2 + 18\rho_j + 38)^{-1}$. Furthermore, the minimal value of the variance ε_j of the prediction error is:

$$\begin{aligned}\varepsilon_j(\rho_j,\theta) &= \mathsf{E}[r_j^2(n)] - \mathbf{p}_j^\top \mathsf{E}[r_j(n)\tilde{\mathbf{r}}_j(n)] \\ &= \frac{1}{2}\sigma_j^2\gamma_j\cos^2(2\theta_j)(1-\rho_j) \\ &\quad \times (3\rho_j^4 - 16\rho_j^3 + 4\rho_j^2 + 24\rho_j + 113). \end{aligned} \quad (25)$$

It is interesting to note that ε_j has a separable form in ρ_j and θ_j. A small gain is achieved by taking into account the spatial redundancies (controlled by ρ_j) compared with the one obtained by the inter-image prediction terms (controlled by θ_j). Figure 1 shows the variations of ε_j with respect to θ_j. This study carried out with a relatively simple statistical model allows us to show the benefit which can be drawn from the use of the VLS-II structure.

5. EXPERIMENTAL RESULTS

We have applied the proposed decompositions to many natural stereo images and 3 pairs of size 512×512 which were extracted from a SPOT5 scene. These scenes are shown in Fig. 3. The retained measure of compression is the entropy of the multiresolution representations. The latter is the weighted average of the entropies of the approximation and the detail subbands. The advantages are two fold: this measure is easily computed and it is independent of the performance of any (embedded) coder. The disparity map is computed using a block-matching technique with a 4×4 block size and a search area of 64 pixels in the horizontal direction and ± 3 in the vertical direction, the SSD being the retained matching criterion. The resulting disparity vectors are losslessly encoded using DPCM with arithmetic encoding. In order to show the benefit of the joint coding by VLS, we compare VLS-I and VLS-II with two up-to-date SI wavelet-based coding methods. The first one consists in coding the left image $I^{(l)}$ and the DC-residual one $I^{(e)}$ through a 5/3 transform over $J = 3$ levels [5]. We have also tested a version of a stereo JPEG2000 coder (Annex I of Part II). It consists in first applying a reversible transform to exploit the redundancies existing between the SI pair. Thus, in addition to $I^{(e)}$, a new image $\tilde{I}$ is produced. Then, the spatial redundancies are exploited by separately applying the 5/3 transform to $I^{(e)}$ and $\tilde{I}$. In our experiments, we have considered the following transform:

$$\tilde{I}(m_x,m_y) = \lfloor (I^{(r)}(m_x,m_y) + I^{(l)}(m_x+v_x, m_y+v_y))/2 \rfloor. \quad (26)$$

It can be noticed that the average coding cost of the disparity vectors cost is equal to 0.55 bpp which is not negligible w.r.t. the wavelet coefficient coding cost. As the disparity map $\mathbf{v}$ is coded in the same way for all the tested methods, it is enough to compute the average entropy of the image data as illustrated in Table 1.

Our simulations indicate that VLS-I leads to lower entropy values than the 5/3 transform applied to $(I^{(l)}, I^{(e)})$ and $(\tilde{I}, I^{(e)})$. More precisely, the use of VLS-I results in an average gain of about 0.2-0.3 bits/pixel (bpp) over the conventional method that encode the $(I^{(l)}, I^{(e)})$ pair and about 0.1-0.15 bpp over the JPEG2000 scheme. If we now compare the performance of VLS-II to those provided by VLS-I, our simulations show that the optimized VLS-II leads to a further improvement of 0.1-0.15 bpp.

The good performance of the proposed approach is also confirmed by looking at the final bit rates obtained by respectively applying the Embedded Zerotree Wavelet (EZW) [21] and the Embedded ZeroBlocks of wavelet coding based on Context modeling (EZBC) encoder [22] as depicted in Table 2. In addition, the scalability in quality of the reconstruction procedure is illustrated in Fig. 2 showing the variations of the peak-signal-to-noise ratio (PSNR) versus the bit rate for the SI pair "spot5-2" using the EZBC encoder. It is worth pointing out that at the last iteration of the EZW algorithm (which corresponds to the final bit rate), the PSNR is infinite (since the Mean Square Error is zero).

6. CONCLUSIONS AND PERSPECTIVES

In this paper, we have presented a novel joint coding method for stereo pairs. Unlike conventional methods, a pair of multiresolution representations of the original images is generated. Two simple examples of the proposed decompositions have been considered. Our simulations indicate that they yield more compact representations than classical approaches. Due to the versatility of the proposed framework, there are still some aspects of these methods that can be improved, in particular by designing more sophisticated prediction/update operators. Also, further developments for choosing the most appropriate disparity estimation method could be investigated as performed in [23].

REFERENCES

[1] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. on PAMI- l.*, vol. 25, no. 8, pp. 993-1008, 2003.

[2] M. G. Perkins, "Data compression of stereo pairs," *IEEE Trans. Commun.*, vol. 40, no. 4, pp. 684-696, April 1992.

[3] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Trans. CSVT*, vol. 8, pp. 327-344, June 1998.

[4] Q. Jiang, J. J. Lee and M. H. Hayes, "A wavelet based stereo image coding algorithm," *Proc. of the ICASSP'99*, pp. 3157-3160, Phoenix, Arizona, USA, 1999.

[5] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Trans. CSVT*, vol. 12, no. 10, pp. 898-903, October 2002.

[6] V. R. Dhond and J. K. Aggarwal, "Structure from stereo - a review," *IEEE Trans. SMC*, vol. 19, no. 6, pp. 1489-1510, November 1989.

[7] O. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Trans. CSVT*, vol. 10, no. 2, pp. 194-200, March 2000.

[8] G. Xu and Z. Zhang, *Epipolar geometry in stereo, motion and object recognition*, Kluver Academic Publishers, 1996.

[9] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. PAMI*, vol. 7, no. 2, pp. 139-154, March 1985.

[10] W. Miled, J.-C. Pesquet and M. Parent, "Wavelet-constrained regularization for disparity map estimation," *European Signal and Image Processing Conference, EUSIPCO'06*, 5 p., Firenze, Italy, 4-8 Sept. 2006.

[11] T. Ozkan and E. Salari, "Coding of stereoscopic images," *Image and Video Processing, SPIE*, vol. 1930, pp. 228-235, 1993.

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

[13] J. Han and Z. Lu, "DCT-based embedded coding scheme for stereo image," *Proc. of the IEEE Aerospace Conf.*, vol. 4, February 1997.

[14] M. S. Moellenhoff and M. W. Maier, "Transform coding of stereo image residuals," *IEEE Trans. IP*, vol. 7, no. 6, pp. 81-92, June 1998.

[15] J. Xu, Z. Xiong and S. Li, "High performance wavelet-based stereo image coding," *Proc. of the IEEE* , vol. II, pp. 273-276, 2002.

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

[17] A. Benazza-Benyahia, M. Hamdi and J.-C. Pesquet, "Efficient lossless and progressive compression scheme for multispectral images," *Proc. of the SPIE Conf. on Mathematics of Data/Image Coding, Compression and Encryption*, vol. 4475, pp. 161-171, San Diego, California, USA, August 2001.

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

[19] W. Sweldens, "The lifting scheme: a new philosophy in biorthogonal wavelet constructions," *Proceedings of SPIE*, San-Diego, CA, USA, 2569, pp. 68-79, September 1995.

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

[21] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on SP*, vol. 41, pp. 3445-3462, December 1993.

[22] S.-T. Hsiang and J. W. Woods, "Embedded image coding using zeroblocks of subband/wavelet coeffcients and context modeling," *IEEE International Symposium on Circuits and Systems, ISCAS'00*, pp. 662-665, Geneva, Switzerland, May 2000.

[23] A. Aksay, M. O. Bici and G. B. Akar, "Evaluation of disparity map characteristics for stereo image coding," *Proc. of the IEEE ICIP*, Genoa, Italy, September 2005.

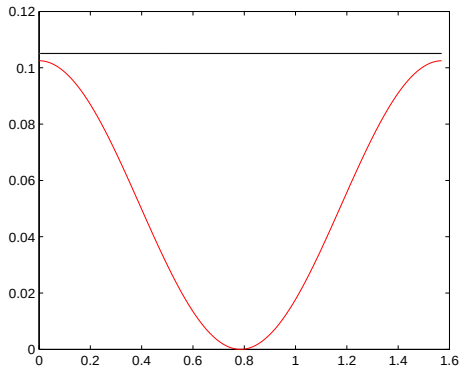

Figure 1: Influence of θ_j: $\varepsilon_j(\rho_j,\theta_j)$ (in red), $\mathsf{E}[r_j^2(n)]$ (in black) *versus* θ_j when $\sigma_j = 1$ and $\rho_j = 0.9$.

Table 1: Performance of SI wavelet-based coders in terms of entropies (in bpp).

Transform	Original	5/3 for $(I^{(l)},I^{(e)})$	5/3 for $(\tilde{I},I^{(e)})$	VLS-I	VLS-II
spot5-1	5.81	4.01	3.87	**3.71**	**3.57**
spot5-2	5.82	4.11	3.95	**3.80**	**3.66**
spot5-3	5.98	4.42	4.23	**4.12**	**3.97**
pentagon	6.68	5.06	4.87	**4.83**	**4.74**
fruit	6.36	3.85	3.68	**3.62**	**3.52**

Table 2: Performance of VLS-II in terms of bit rate (bpp).

	5/3 for $(I^{(l)},I^{(e)})$		5/3 for $(\tilde{I},I^{(e)})$		VLS-II	
	EZW	EZBC	EZW	EZBC	EZW	EZBC
spot5-1	4.20	3.78	4.04	3.71	3.93	**3.41**
spot5-2	4.30	3.93	4.16	3.88	4.04	**3.53**
spot5-3	4.56	4.29	4.37	4.14	4.28	**3.89**
fruit	4.02	3.77	3.90	3.66	3.78	**3.54**

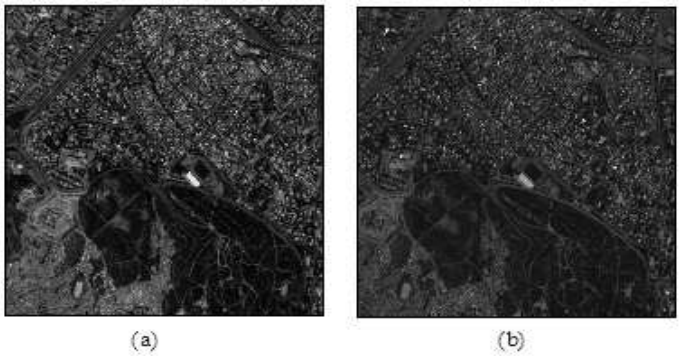

Figure 2: Original SI pair "spot5-2": (a) left image, (b) right image.

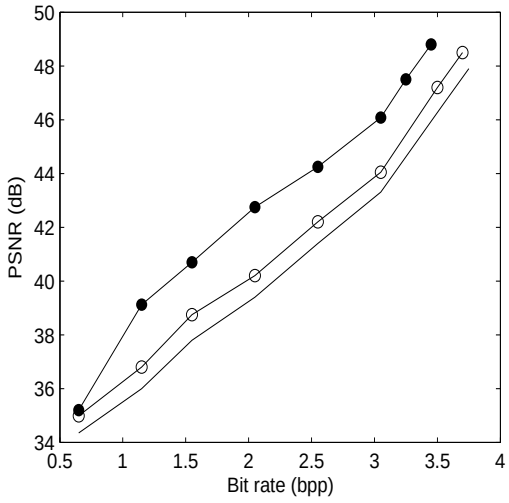

Figure 3: PSNR (in dB) versus the bit rate (bpp) after EZBC encoding for the SI pair "spot5-2": 5/3 for $(I^{(l)},I^{(e)})$ in solid line, 5/3 for $(\tilde{I},I^{(e)})$ in line with circle , VLS-II in dots.