Optimal Motion Estimation for Wavelet Motion Compensated Video Coding

Marco Cagnazzo, Filippo Castaldo, Thomas André, Marc Antonini, and Michel Barlaud, *Fellow, IEEE*

***Abstract*—Wavelet-based coding is emerging as a promising framework for efficient and scalable compression of video. Nevertheless, a number of basic tools currently employed in this field have been conceived for hybrid block-based transform coding. This is the case of motion estimation, which generally aims to minimize the energy or the absolute sum of prediction error. However, as wavelet video coders do not employ predictive coding, this is no longer an optimal approach. In this paper we study the problem of the theoretical optimal criterion for wavelet-based video coders, using coding gain as merit figure. A simple solution has been found for a peculiar but useful class of temporal filters. Experiments confirm that the optimally estimated vectors increase the coding gain as well as the performance of a complete video coder, but at the cost of an augmented complexity.**

***Index Terms*—Motion-compensated (MC) lifting schemes, MC wavelet transform, motion estimation (ME), wavelet video coding.**

I. Introduction

MOTION compensation (MC) is of crucial importance in order to obtain good performances in video compression [1], be it the classical hybrid coding [2] or one of the newer wavelet-based algorithms [3], and then a good motion estimation (ME) is equally very important. Usually, the criterion for ME is the minimization of some function of the prediction error (i.e., the difference between current and predicted frame), as the squared sum (energy) or the absolute sum of the prediction error. This approach is justified as far as hybrid coding is concerned, but in the case of a wavelet video coder, a deeper analysis is needed since we are no longer coding the prediction error but the transform coefficients. More precisely, in the so-called $t+2d$ coding paradigm [3]–[7], the input signal is filtered along the temporal axis following objects trajectories as to reduce the correlation and concentrate the energy into a few coefficients. Then, a further spatial filtering is performed, usually with the 9/7 Daubechies filters [8]. The resulting 3-D-transform coefficients are then encoded with a suitable algorithm. Of course, nothing assures that a ME algorithm which is based on prediction error, will give the best performances in this case as well.

The need of an optimal approach to ME for a video coder based on wavelet transform (WT) was early recognized by Choi and Woods [4]. They asserted that while for hybrid coders the objective of ME is to minimize the mean squared prediction error, for MC temporal WT this should be changed to maximization of the coding gain (CG). They claimed that, as the filter they used along temporal direction is orthogonal (Haar), the CG is expressed as the ratio between arithmetic and geometric mean of subband variances, and the maximum CG is nearly achieved by minimizing the energy (variance) of the temporal high frequency subband, since the variance of the temporal low frequency subband is relatively constant. Using the Haar filter for temporal analysis, the temporal high frequency subband is just a scaled version of the prediction error signal. Thus, the minimization of mean-square error (MSE) turns to be nearly optimal in their case.

However, some new temporal lifting schemes (LS) [9] have been recently proposed for WT video coding, and they seem to have quite better performances than Haar's. These filters are mainly the (2,2) LS [6], [7] (corresponding to the 5/3 filter) and the (2,0) LS [10], [11] (corresponding to the 1/3 filter). For these cases, and in the general case as well, there is no reason to assume *a priori* that prediction error-related criteria are the best solution for ME. This has been recognized in [12], where a generic criterion based on high frequency subband content is proposed. In that work the focus is specifically on the (2,2) LS, and on implementation complexity issues. Their experimental results confirm that an *ad hoc* designed estimator can improve the performances of MC WT video coders. In this work instead, we primarily look for a theoretical justification of the ME criterion, considering the optimality issue for a generic temporal filter, and taking into account the problem of multiple decomposition levels. Then we propose an estimator for the (2,0) filter which is optimal in this framework. The gain in performances and some complexity issues are considered as well.

The paper is organized as follows. In Section II we introduce a suitable definition of CG for nonorthogonal WT-based coding, which is used to define an optimal estimator. With such a definition of CG, Section III depicts a new criterion for ME, valid in the general case, and states the problems related to its computation. Further developments for the special (2,0) case are shown in Section IV, where we derive the optimal criterion for this filter. Experimental results are reported in Section V. Section VI concludes the paper, summarizing the main results and outlining future work.

II. Coding Gain for Biorthogonal WT

CG is a quantitative efficiency measure for a given transformation. It is defined [13] as the ratio between the distortion resulting from quantization of the input signal (or PCM distortion, D_{PCM}) and the minimum distortion achievable with transform coding (or transform coding distortion D_{TC}). For orthogonal subband coding, in the high resolution hypothesis, this turns out to be the ratio between arithmetic and geometric mean of subband variances. We want a suitable expression for this quantity in the case of generic spatiotemporal wavelet decomposition as well.

Manuscript received May 10, 2005; revised April 13, 2006. This work was supported in part by the Centro Regionale di Competenza ICT of Benevento, Italy. This paper was recommended by Associate Editor J.-R. Ohm.

M. Cagnazzo and F. Castaldo are with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazion DIET at the University of Naples, Naples 80125, Italy.

T. André, M. Antonini, and M. Barlaud are with the I3S Laboratory, University of Nice, Sophia Antipolis 06903, France.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.897110

After L levels of WT, the input signal is subdivided into M subbands. We will consider in the following only biorthogonal filters for WT. For this kind of wavelet basis, Usevitch [14] showed that reconstruction MSE is related to subband MSE: $D_{\mathrm{TC}} = \sum_{i=1}^{M} a_i w_i D_i$

Now, in [13] it is shown that, under the hypothesis of high resolution, $D_i = H_i \sigma_i^2 2^{-2b_i}$, where b_i is the number of bits per sample for encoding the ith subband, σ_i^2 is its variance, and H_i is the so-called *shape* factor, which depends only on the shape of ith band probability density function (pdf), f_i: $H_i = (1/12)\{\int_{-\infty}^{+\infty} [\sigma_i f_i(\sigma_i x)]^{1/3} dx\}^3$.

Asymptotically, the lowest transform coding distortion can be found by solving a constrained minimization problem. We have to minimize

$$D_{\mathrm{TC}} = \sum_{i=1}^{M} a_i w_i H_i \sigma_i^2 2^{-2b_i}$$

under the constraint: $\sum_{i=1}^{M} N_i b_i = B$ where B is the available bit budget (in bits). The coding gain can be computed as $\mathrm{CG} = D_{\mathrm{PCM}}/D_{\mathrm{TC}}^*$, where D_{TC}^* is the solution of this constrained minimization problem.

Defining $\bar{b} = B/N$ in bits per pixel (bpp), $W_{\mathrm{GM}} = \prod_i (w_i^{a_i})$, $H_{\mathrm{GM}} = \prod_i (H_i^{a_i})$, $\rho^2 = \prod_i [(\sigma_i^2)^{a_i}]$ respectively, it can be shown [15], [16] that

$$D_{\mathrm{TC}}^* = W_{\mathrm{GM}} H_{\mathrm{GM}} \rho^2 2^{-2\bar{b}} \tag{1}$$

and

$$\mathrm{CG} = \frac{H}{H_{\mathrm{GM}}} \frac{\sum_{i=1}^{M} a_i w_i \sigma_i^2}{\prod_{i=1}^{M} (w_i \sigma_i^2)^{a_i}} \tag{2}$$

where H is the shape factor of input (nontransformed) data.

In the hypothesis of Gaussian signal, $H_i = H$ and then $H_{\mathrm{GM}} = H^{\sum_i a_i} = H$. It is known that if each subband has the same number of coefficients ($a_i = 1/M \; \forall i \in \{1, 2, \ldots, M\}$), and if orthogonal filters are employed ($w_i = 1 \; \forall i \in \{1, 2, \ldots, M\}$), the coding gain can be expressed as the ratio of the arithmetic and geometric means of subband variances: this result, reported in [13] is valid only for orthogonal subband coding. Equation (2) extends it to the more general case of arbitrary wavelet decomposition with nonorthogonal filters. In this case we can read $\sum_{i=1}^{M} a_i (w_i \sigma_i^2)$ as a a_i-weighted arithmetic mean of normalized variances $w_i \sigma_i^2$, where the normalization accounts for nonisometry of the wavelet transform. Likewise, we can interpret $\prod_{i=1}^{M} (w_i \sigma_i^2)^{a_i}$ as a "weighted" geometric mean (by the same weights a_i) of the normalized variances.

However, we are interested in deriving a criterion which allows to find motion vectors (MVs) maximizing the CG. As MVs do not affect D_{PCM}, we can switch our attention to the quantity D_{TC}^*: the optimal ME criterion should then minimize this quantity as expressed in (1).

III. Generic Optimal Criterion

When we consider the minimum transform coding distortion D_{TC}^*, we should refer to three-dimensional (i.e., spatiotemporal) SBs. Actually we will consider only temporal SBs since some earlier studies on this problem showed that considering spatiotemporal instead of temporal SBs gives little or no gain

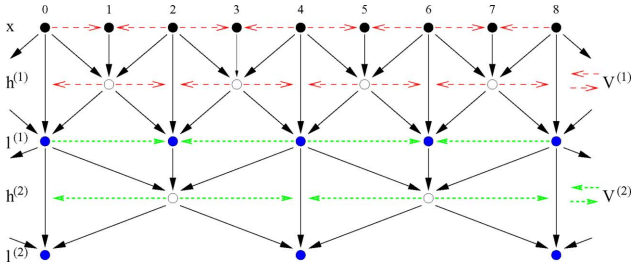

Fig. 1. Scheme of a two-levels temporal decomposition with the (2,2) LS. In this case $V^{(1)}$ and $V^{(2)}$ are needed to compute $h^{(2)}$.

[4]. Moreover we make the hypothesis that varying MVs does not affect too much shapes of subband pdf, so that the only term depending on MVs in (1) is ρ^2, that is the weighted geometric mean of temporal subband variances.

Unfortunately, in the general case, the minimization of ρ^2 is not an easy task because MVs computed for a generic decomposition affect all subsequent levels of WT transform. Let us define a few notations in order to gain insight about this dependence. We use $\mathbf{v}_{k \to k'}(\mathbf{p})$ to refer to the MV accounting for the displacement that the pixel $\mathbf{p}$ in frame k will have in frame k'. The kth frame of the input sequence is referred to as $x_k(\mathbf{p})$. We indicate with $l^{(1)}$ and $h^{(1)}$ the first level low (L) and high frequency (H) temporal subband sequences, with a possible subscript to indicate a specific frame, and an argument to indicate the pixel position. Likewise, $l^{(i)}$ and $h^{(i)}$ are the low and high frequency subband produced by the ith level decomposition. We consider L levels of dyadic temporal decomposition, resulting in $M = L + 1$ temporal subbands $h^{(1)}, h^{(2)}, \ldots, h^{(L)}, l^{(L)}$, and we indicate with $V^{(1)}$ the set of all MVs needed to compute the first temporal decomposition $\{l^{(1)}, h^{(1)}\}$ from the input sequence, and with $V^{(i)}$ the set of vectors needed to compute $\{l^{(i)}, h^{(i)}\}$ from $l^{(i-1)}$. These vectors are shown in Fig. 1 for the case of (2,2) LS with two levels of temporal decomposition. We see that in this case $V^{(1)}$ is made up of vectors $\mathbf{v}_{k \to k+1}$ and $\mathbf{v}_{k+1 \to k}$ for all k, while $V^{(2)}$ is constituted by vectors $\mathbf{v}_{2k \to 2k+2}$ and $\mathbf{v}_{2k+2 \to 2k}$ for all k. It is clear that in order to compute $h^{(2)}$ from input, we need not only $V^{(2)}$, but also $V^{(1)}$ as $l^{(1)}$ depends on this set of vectors.

In order to compute a single level of temporal decomposition in the case of a (N, M) LS, N vectors per frame are required for the high frequency subband and M vectors per frame for the low frequency subband. In order to compute $h^{(i)}$ from the input sequence, we need all the vectors from previous decomposition levels $\{V^{(1)}, V^{(2)}, \ldots, V^{(i)}\}$. This means that the optimization of $V^{(i)}$ must take into account the influence of this MVs set on all subsequent temporal subbands. In other words, we cannot simply choose $V^{(i)}$ such that σ_i^2 is minimized, but we should jointly optimize all level MVs in order to minimize ρ^2. This problem is difficult to approach analytically and extremely demanding in terms of computational complexity. However, it is possible to simplify it remarkably with a suitable choice of temporal filters. In particular, we consider the class of $(N, 0)$ LS. In this case, the low pass output of WT is just the temporally subsampled input sequence, which does not depend on MVs. Another crucial consequence is that the ith high frequency subband is computed directly from the input sequence, independently from other SBs. An example is

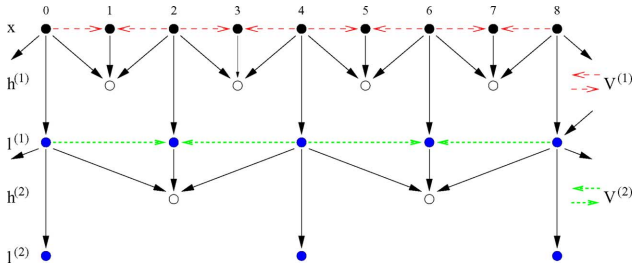

Fig. 2. Scheme of a two-levels temporal decomposition with the (2,0) LS. In this case, we need only $V^{(2)}$ to compute $h^{(2)}$ as it is obtained directly from x independently from $h^{(1)}$.

shown in Fig. 2 for the (2,0) LS and two level of temporal decomposition. We see that we can compute $h^{(2)}$ directly from the input sequence and from the vectors $V^{(2)}$. For a higher number of decomposition levels, as we compute $h^{(i)}$ from $l^{(i-1)}$ which in turn is just the input sequence under sampled by a factor 2^i, this subband depends only on $V^{(i)}$ instead of all the $V^{(1)}, V^{(2)}, \ldots, V^{(i)}$ for the (N, M) case. This means also that all the subband variances are actually independent from one another and that they can be minimized separately. In other words, each $V^{(i)}$ can be optimized separately providing that it minimizes the ith high frequency SB variance.

IV. Developing the Criterion for the $(N, 0)$ Case

When $(N, 0)$ LS are employed, MVs optimization can be carried out independently for each decomposition level. For this reason, we will refer from now on to the first one, and will drop the superscript from $h^{(i)}$ and from $V^{(i)}$ for the sake of simplicity. For higher decomposition levels, the same development is still valid, except that a suitably subsampled version of the input sequence should be considered.

Further analytical developments are possible if we refer to a specific $(N, 0)$ LS as the (2,0). Let us recall the equation for this special case

$$\begin{aligned} h_k(\mathbf{p}) &= x_{2k+1}(\mathbf{p}) \\ &\quad - \frac{1}{2}\left[x_{2k}\left(\mathbf{p} + \mathbf{v}_{2k+1\to 2k}(\mathbf{p})\right)\right. \\ &\quad \left. + x_{2k+2}\left(\mathbf{p} + \mathbf{v}_{2k+1\to 2k+2}(\mathbf{p})\right)\right] \\ l_k(\mathbf{p}) &= x_{2k}(\mathbf{p}). \end{aligned}$$

Therefore, for this LS the vector set to optimize is

$$V = \{\mathbf{v}_{2k+1\to 2k}, \mathbf{v}_{2k+1\to 2k+2}\}_{k\in\mathcal{N}}.$$

As h_k depends on both $\mathbf{v}_{2k+1\to 2k}$ and $\mathbf{v}_{2k+1\to 2k+2}$, and only on them, these vectors have to be jointly minimized. Without losing generality, we will refer from now on to the optimization of a vector couple, instead of the whole set V. By introducing the backward and forward MVs, $B_k = \mathbf{v}_{k\to k-1}$ and $F_k = \mathbf{v}_{k\to k+1}$, we can rewrite the equations of the high frequency subband as follows:

$$\begin{aligned} h_k(\mathbf{p}) = x_{2k+1}(\mathbf{p}) - \frac{1}{2}\left[x_{2k}\left(\mathbf{p} + F_{2k+1}(\mathbf{p})\right)\right. \\ \left. + x_{2k+2}\left(\mathbf{p} + B_{2k+1}(\mathbf{p})\right)\right]. \end{aligned}$$

TABLE I
Estimated Vector Entropy, kbps

Sequence	Precision	SAD	SSD	New
foreman	full-pixel	**66.3**	69.1	66.9
	half-pixel	84.8	89.5	**84.3**
	quarter-pixel	**102.6**	106.6	103.5
akiyo	full-pixel	5.37	6.10	**5.33**
	half-pixel	**10.1**	11.4	**10.1**
	quarter-pixel	**18.0**	20.6	20.4

TABLE II
Coding Gain for Different ME Techniques

Sequence	Precision	SAD	SSD	New
foreman	full-pixel	8.59	9.00	**10.17**
	half-pixel	9.66	10.67	**11.41**
	quarter-pixel	9.94	10.70	**11.84**
akiyo	full-pixel	57.8	59.4	**71.6**
	half-pixel	64.2	66.6	**77.6**
	quarter-pixel	66.0	68.5	**79.2**

The optimal couple B^*_{2k+1}, F^*_{2k+1} is the one minimizing the variance of h_k. Since it has zero mean, this is equivalent to minimizing the energy of h_k, indicated with $\mathcal{E}(h_k)$.

$$\left(B^*_{2k+1}, F^*_{2k+1}\right) = \arg \min_{B_{2k+1}, F_{2k+1}} \mathcal{E}\left\{h_k(\mathbf{p})\right\}.$$

We elaborate the expression of h_k

$$\begin{aligned} h_k(\mathbf{p}) &= x_{2k+1}(\mathbf{p}) - \frac{1}{2}\left[x_{2k}\left(\mathbf{p} + F_{2k+1}(\mathbf{p})\right)\right. \\ &\qquad \left. + x_{2k+2}\left(\mathbf{p} + B_{2k+1}(\mathbf{p})\right)\right] \\ &= \frac{1}{2}\left[x_{2k+1}(\mathbf{p}) - x_{2k}\left(\mathbf{p} + F_{2k+1}(\mathbf{p})\right)\right. \\ &\qquad \left. + x_{2k+1}(\mathbf{p}) - x_{2k+2}\left(\mathbf{p} + B_{2k+1}(\mathbf{p})\right)\right] \\ &= \frac{1}{2}(\epsilon_F + \epsilon_B) \end{aligned}$$

where $\epsilon_F[\epsilon_B]$ is the forward [backward] motion-compensated prediction error. This means that the optimal trajectory minimizes the energy of the sum of these errors. Moreover

$$\mathcal{E}(\epsilon_B + \epsilon_F) = \mathcal{E}(\epsilon_B) + \mathcal{E}(\epsilon_F) + 2\langle \epsilon_B, \epsilon_F \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product (correlation) between two images. In conclusion

$$B^*_{2k+1}, F^*_{2k+1} = \arg \min_{B_{2k+1}, F_{2k+1}} \left[\mathcal{E}(\epsilon_B) + \mathcal{E}(\epsilon_F) + 2\langle \epsilon_B, \epsilon_F \rangle\right]. \tag{3}$$

Equation (3) defines the ME criterion that we propose in this paper. We can compare it to the usual MSE-based criterion. When this latter is used, we independently minimize $\mathcal{E}(\epsilon_B)$ and $\mathcal{E}(\epsilon_F)$, so we probably attain a low value of $\mathcal{E}(\epsilon_B + \epsilon_F)$. This explains why MSE-based ME criteria often perform well with WT video coders. However, they do not necessarily achieve the minimum of criterion (3) as the mixed term is not taken into account. This term grows larger when the two error images are more similar, meaning that the optimal backward and forward vectors are

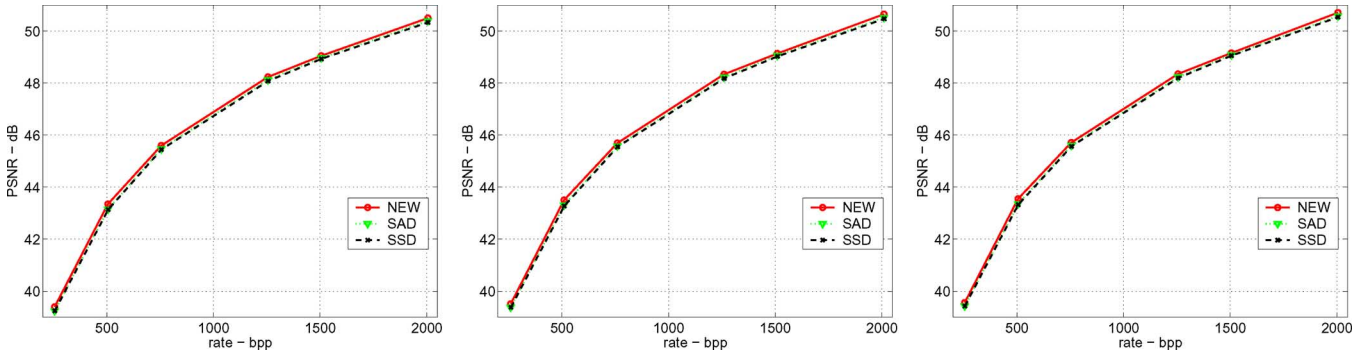

Fig. 3. RD comparison among ME techniques for the "akiyo" sequence. (Left) full-pixel, (center) half-pixel and (right) quarter-pixel precision.

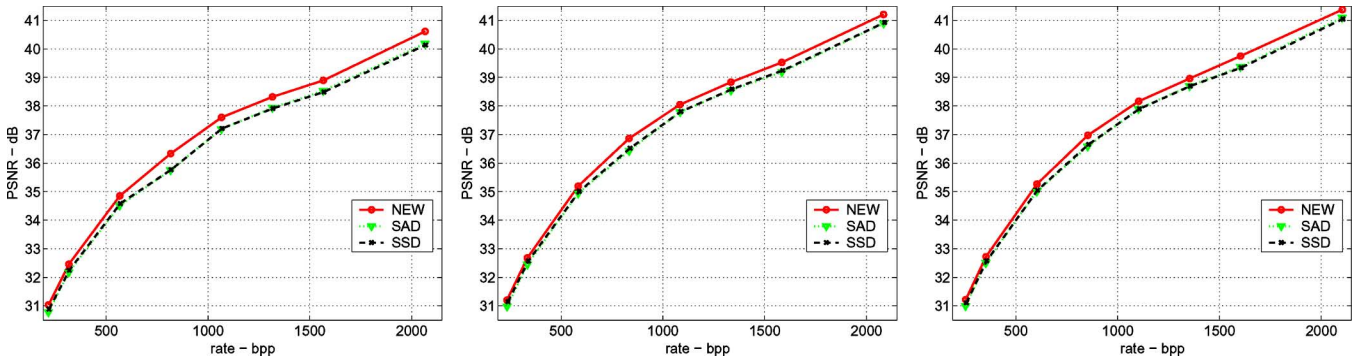

Fig. 4. RD comparison among ME techniques for the "foreman" sequence: (Left) full-pixel, (center) half-pixel and (right) quarter-pixel precision.

not independent: they should produce error images as different as possible, being not enough to minimize error images energies. In other words, regions affected by a positive backward error should have a negative forward error and viceversa.

V. Experimental Results

In a first experiment we use the ME criterion (3) in order to find MV in two test sequences, *foreman* and *akiyo*. We compare the resulting MVs to what we obtain by using common ME criterion such as the sum of absolute differences (SAD) or the sum of squared differences (SSD). We use a block-based motion estimator for these ME criteria with the following settings: the block size is 16×16 pixels; the maximum displacement is ± 12 pixels; the ME precision is full-pixel, half-pixel or quarter-pixel, with bilinear interpolation in the latter cases. The complexity not being an issue in this work, a 4-D full search of the best vector couple is performed for the proposed technique. In order to have a fair comparison, the same search strategy is employed in the case of SAD and SSD as well. The MVs obtained by using the different criteria are then compared. First of all, we take a measure of their encoding cost by computing the zero order entropy. Results are shown in Table I. We see that, for both the sequences the proposed technique produces MVs with a relatively small entropy. Anyway, the estimated coding costs are quite close to one another, with SAD criterion usually better than the SSD as it is more robust to outliers. From this first experiment, we conclude that the proposed method does not remarkably change the encoding cost of estimated vectors.

Once assessed the cost figure by the zero-order entropy, we turn to the merit figure, represented by the coding gain. The results of this experiments are shown in Table II. We see that our method achieves consistently the best coding gain for both test sequences and for all the precisions.

The results summarized in Tables I and II are very general since they do not depend on subsequent encoding steps. They suggest that the proposed technique can produce better MVs than classical methods, and at comparable costs. Nevertheless, in order to clearly assess the benefits of the proposed method, we must compare the global effect in a complete video codec. These results are less general than the previous ones, as they depend on the coding technique, but they are interesting as they shed light on the possible benefits of the proposed ME criterion. The video coder used for our experiments first performs a dyadic MC WT temporal filtering with three decomposition levels, then a dyadic 9/7 wavelet transform on four levels in the spatial domain, followed by a SPIHT [17] encoding of the resulting coefficients. This encoder is quite similar to the one proposed in [18], except that we use the motion-compensated (2,0) lifting scheme [11], [19].

We encode the test sequences with this encoder, using the different MVs. The experimental results shown in Figs. 3 and 4 prove that the proposed criterion provides the best global performances for all the test cases. More precisely, we see that the gain is not very large for the slow-motion *akiyo* sequence (up to 0.2 dB) while we achieve up to 0.5 dB of peak signal-to-noise ratio (PSNR) improvement for *foreman*. We conclude that optimal ME is not as critical for a sequence with little motion content as for more complex sequences. We also remark that the proposed criterion has a larger gain at higher bit-rates, i.e., where the high resolution hypothesis (which is at the basis of our estimator) is more reliable. However, as it can be seen from Fig. 5, even at

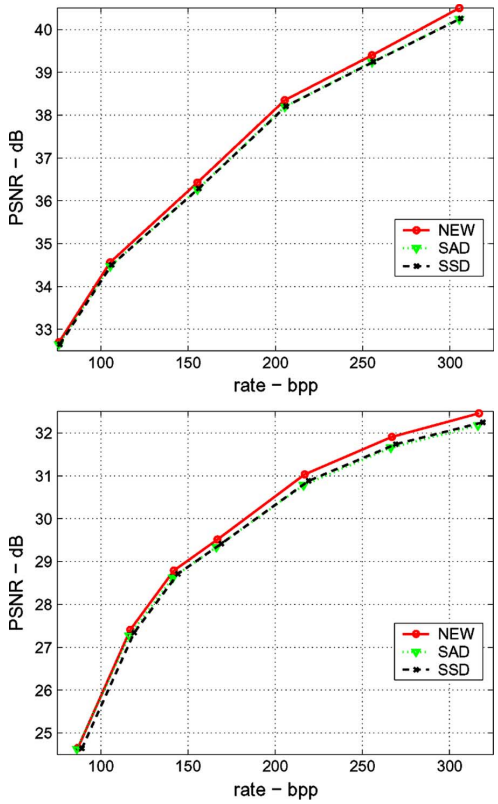

Fig. 5. RD comparison at low bit-rates. Top: "akiyo." Bottom: "foreman."

quite low bit-rates the proposed method is superior to the classical ones but the difference is smaller. In this figure we report only the results for full pixel precision, which proved to be the best choice at rates below 250 kbps.

Finally, we considered the possibility to use the proposed algorithm even for different temporal filter, like the common (2,2) LS. In other words, we repeated the previous experiment but we changed the temporal filter. Actually, in [12] authors already showed that a criterion which minimizes the high frequency subband variances, can improve the RD performance of a (2,2)-based video encoder, even if they used a low-complexity approximation of this criterion. Our test confirmed the performance increases in this case as well.

The optimal technique is far more complex than the classical ones, since it needs the joint estimation of backward and forward vectors, with a quadratic complexity with respect to the number of test vectors. This prevents the use of the proposed criterion in software-only real-time applications. Faster, suboptimal search strategies such as those proposed in [12] can be envisaged in order to reduce its complexity, but a complete analysis of the complexity-performance tradeoff for the proposed method is beyond the scope of this paper.

VI. Conclusion

In this paper, we dealt with the problem of optimal ME for WT video coding. We showed that, by using the coding gain as merit figure, we can define an optimal ME criterion which can be used when the temporal filter is the (2,0) LS. Moreover, our analytical development justifies the good performances of classical ME methods, as they are characterized by a ME criterion very similar to the optimal one. Experiments show a consistent increase of coding gain with respect to classical methods, with a negligible variation of coding cost. This results in a significant improvement of overall performances when this technique is integrated into a complete WT video encoder. Moreover, the proposed criterion can be effectively used even for the common (2,2) lifting scheme.

On the other hand, the proposed method has a remarkable complexity, which means that in order to use it for real time application, suboptimal low-complexity search strategies must be considered. Future works will explore the performance tradeoff related to a low-complexity search strategies. Moreover, we intend to study whether an optimal criterion can be found for the more common (2,2) LS and what the advantages could be in this case.

References

[1] J. Jain and A. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. 29, pp. 1799–1808, Dec. 1981.

[2] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. 5, pp. 1140–1154, Aug. 1987.

[3] J.-R. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[4] S. Choi and J. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[5] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3-D wavelet transform based on lifting," in *Proc. IEEE Int. Conf. Image Process.*, Thessaloniki, Greece, Oct. 2001, pp. 1029–1032.

[6] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2001, pp. 1793–1796.

[7] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[8] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transforms," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[9] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 245–267, 1998.

[10] J. Konrad, "Transversal versus lifting approach to motion-compensated temporal discrete wavelet transform of image sequences: Equivalence and tradeoffs," in *Proc. SPIE Vis. Comm. Image Process.*, Jan. 2004.

[11] T. André, M. Cagnazzo, M. Antonini, M. Barlaud, N. Božinović, and J. Konrad, "(N,0) motion-compensated lifting-based wavelet transform," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Montreal, QC, Canada, May 2004, vol. 3, pp. 121–124.

[12] G. Pau, C. Tillier, and B. Pesquet-Popescu, "Optimization of the predict operator in lifting-based motion compensated temporal filtering," in *Proc. SPIE Vis. Commun. Image Process.*, 2004, pp. 712–720.

[13] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Nowell, MA: Kluwer, Jan. 1988.

[14] B. E. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. Data Comp. Conf.*, Mar. 1996, pp. 387–395.

[15] J. Katto and Y. Yasuda, "Performance evaluation of subband coding and optimization of its filter coefficients," in *Pro. SPIE Vis. Commun. Image Process.*, Nov. 1991, vol. 1605, pp. 95–106.

[16] O. Egger, P. Fleury, T. Ebrahimi, and M. Kunt, "High-performance compression of visual information-A tutorial review—Part I: Still pictures," *Proc. EEE*, vol. 87, no. 6, pp. 976–1011, Jun. 1999.

[17] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[18] B. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[19] M. Cagnazzo, T. André, M. Antonini, and M. Barlaud, "A model-based motion compensated video coder with JPEG2000 compatibility," in *Proc. IEEE Int. Conf. Image Process.*, Singapore, Oct. 2004, pp. 2255–2258.