Video Coding with Motion-Compensated Lifted Wavelet Transforms

Markus Flierl [1] and Bernd Girod

Information Systems Laboratory, Stanford University

Stanford, CA 94305-9510, USA

Abstract

This article explores the efficiency of motion-compensated three-dimensional transform coding, a compression scheme that employs a motion-compensated transform for a group of pictures. We investigate this coding scheme experimentally and theoretically. The practical coding scheme employs in temporal direction a wavelet decomposition with motion-compensated lifting steps. Further, we compare the experimental results to that of a predictive video codec with single-hypothesis motion compensation and comparable computational complexity. The experiments show that the 5/3 wavelet kernel outperforms both the Haar kernel and, in many cases, the reference scheme utilizing single-hypothesis motion-compensated predictive coding. The theoretical investigation models this motion-compensated subband coding scheme for a group of K pictures with a signal model for K motion-compensated pictures that are decorrelated by a linear transform. We utilize the Karhunen-Loeve Transform to obtain theoretical performance bounds at high bit-rates and compare to both optimum intra-frame coding of individual motion-compensated pictures and single-hypothesis motion-compensated predictive coding. The investigation shows that motion-compensated three-dimensional transform coding can outperform predictive coding with single-hypothesis motion compensation by up to 0.5 bits/sample.

Preprint submitted to Elsevier Science *30 April 2004*

Key words: Video Coding, Motion Compensation, Adaptive Wavelets, Lifting, Three-Dimensional Subband Coding of Video

1 Introduction

For still image compression, wavelet coding schemes are widely favored today, as they combine excellent compression efficiency [1,2] with the possibility of an embedded representation [3,4]. Temporal wavelet coding for video sequences, however, has not had such success. This can be attributed to the difficulty of incorporating motion compensation into the temporal subband decomposition. The recent discovery of motion-compensated lifted wavelet transforms had lead to renewed interest and the hope that temporal subband coding might ultimately outperform predictive hybrid coding, predominant in all current video coding standards.

Early attempts of three-dimensional subband coding did not employ motion compensation. Knauer applied in [5] the Hadamard transform for real-time television compression by considering the sequence of frames as a volume. Further research is published by Vetterli et al. [6], Biemond et al. [7], and Jayant et al. [8]. Zhang et al. investigate memory-constrained 3-D wavelet

Email addresses: mflierl@ieee.org (Markus Flierl), bgirod@stanford.edu (Bernd Girod).

[1] M. Flierl visited the Information Systems Laboratory, Stanford University, and is now with the Signal Processing Institute, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. This work was supported, in part, by the Graduate Research Center "3-D Image Analysis and Synthesis" at the University of Erlangen-Nuremberg, Germany.

transforms without boundary effects and use the lifting structure [9,10]. Barlaud et al. suggest a 3-D scan-based wavelet transform with reduced memory requirements [11]. 3-D wavelet coding with arbitrary regions of support is discussed by Minami et al. [12].

Three-dimensional coding with motion compensation has first been suggested by Kronander [13,14]. Akiyama et al. use global motion compensation and three-dimensional transform coding [15]. Global motion compensation with 3-D wavelet coding is also investigated by Chou et al. in [16]. Zhang and Zafar present a motion-compensated wavelet transform coder for color video compression [17]. Ohm starts his investigation of 3-D subband video coding [18,19] with integer-pel accurate motion compensation and first order quadrature mirror filters in the temporal domain [20,21]. Woods et al. discuss a scheme for object-based spatio-temporal subband coding [22]. They optimize the trade-off of the rate between motion vectors and 3-D subbands [23] and consider digital cinema applications [24]. A bit allocation scheme for subband compression of HDTV is published in [25]. For video communication over wireless channels, Chou and Chen propose a perceptually optimized 3-D subband codec [26].

Mallat [27] discusses the wavelet representation as a suitable tool for multiresolution signal decomposition. Such a decomposition of video signals may permit temporal and spatial scalability. This important feature is investigated by many researchers. Early work by Lippman and Butera discusses interactive retrieval of image sequences [28]. Uz et al. suggest a scheme for interpolative multiresolution coding of digital HDTV [29,30]. Later, Girod and Horn investigate this scheme for Internet video streaming [31]. In addition to the multiresolution representation of the motion-compensated three-dimensional signal, researchers investigate also the multiresolution representation of mo-

tion, like Ohm [32] and Zhang et al. [33]. Taubman and Zakhor propose a common framework for rate and distortion based scaling of highly scalable compressed video [34–38]. Pearlman et al. introduce in [39,40] an embedded wavelet video codec using three-dimensional set partitioning in hierarchical trees [41]. Woods et al. discuss rate-constrained multiresolution transmission of video [42], and present a resolution and frame-rate scalable subband video coder [43]. Ranganath et al. outline a highly scalable wavelet-based codec for very low bit-rate environments and introduces tri-zerotrees [44]. Pesquet-Popescu et al. suggest a method for context modeling in the spatio-temporal trees of wavelet coefficients [45] and propose the strategy of fully scalable zerotree coding [46]. Zhang et al. investigate three-dimensional embedded block coding with optimized truncation [47].

Daubechies and Sweldens factor wavelet transforms into lifting steps [48–50]. This construction scheme permits wavelet transforms that map integers to integers [51], a desirable property for any practical coding scheme. Moreover, the lifting scheme permits also adaptive wavelet transforms [52,53]. Adaptive techniques, such as motion compensation, can be incorporated into the lifting scheme. For the temporal subband decomposition of video signals, Pesquet-Popescu and Bottreau improve the compression efficiency by using a lifting scheme with motion compensation [54]. Zhang et al. also incorporate motion compensation into the lifting scheme [55]. Secker and Taubman investigate the schemes with block motion compensation [56] and deformable mesh motion compensation [57]. In [58], Ohm reviews recent progress and discusses novel aspects with respect to irregular sampled signals, shift variance of wavelet transforms and non-dyadic wavelet processing. Pateux et al. investigate several motion-compensated lifting implementations and compare to standard-

ized hybrid codecs [59]. Barlaud et al. extend their 3-D scan-based wavelet transform codec and use the motion-compensated lifting scheme [60]. The recent publications on the lifting scheme with motion compensation show that this scheme represents an efficient approach for a subband decomposition in temporal direction. The decomposition is invertible, even for motion compensation with sub-pel accuracy. Furthermore, the wavelet decomposition permits full scalability, a desirable feature for heterogeneous networks.

The outline of this article is as follows: Section 2 discusses the temporal subband coding scheme with a motion-compensated lifted Haar wavelet and a motion-compensated lifted 5/3 wavelet. Experimental results and comparisons for several test sequences are presented. A mathematical model of the scheme's performance is developed in Section 3. The motion-compensated wavelet kernels of the practical coding schemes are generalized to obtain theoretical performance bounds.

2 Coding Scheme

Applying a linear transform in temporal direction of a video sequence is not very efficient if significant motion is prevalent. However, a combination of a linear transform and motion compensation seems promising for efficient compression. For wavelet transforms, the so called *Lifting Scheme* [50] can be used to construct the kernels. A two-channel decomposition can be achieved with a sequence of prediction and update steps that form a ladder structure. The advantage is that this lifting structure is able to map integers to integers without requiring invertible lifting steps. Further, motion compensation can be incorporated into the prediction and update steps as proposed in [56]. The fact

that the lifting structure is invertible without requiring invertible lifting steps makes this approach feasible. We cannot count on motion compensation to be invertible in general. If it is invertible, this motion-compensated wavelet transform based on lifting permits a linear transform along the motion trajectories in a video sequence.

In the following, we investigate coding schemes that process video sequences in groups of K pictures (GOP). First, we decompose each GOP in temporal direction. The dyadic decomposition utilizes a motion-compensated wavelet which will be discussed later in more detail. The temporal transform provides K output pictures that are intra-frame encoded. In order to allow a comparison to a basic predictive coder with motion compensation, we utilize for the intra-frame coder a 8×8 DCT with run-length coding. If we employ a Haar wavelet and set the motion vectors to zero, the dyadic decomposition will be an orthonormal transform. Therefore, we select the same quantizer step-size for all K intra-frame encoder. The motion information that is required for the motion-compensated wavelet transform is estimated in each decomposition level depending on the results of the lower level. Further, we employ half-pel accurate motion compensation with bi-linear interpolation. The block-based motion vectors are predicted from spatial neighbors and the motion vector differences are encoded with Huffman codes.

2.1 Motion-Compensated Lifted Haar Wavelet

First, we discuss the lifting scheme with motion compensation for the Haar wavelet [56].

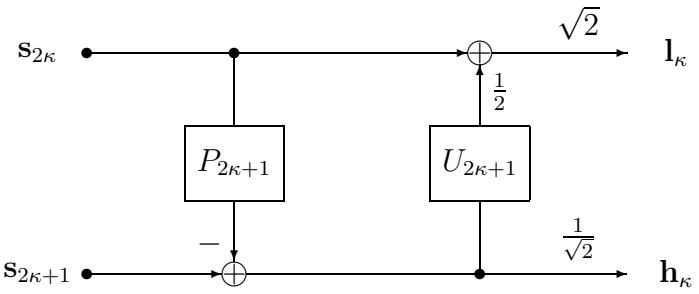

Fig. 1. Haar transform with motion-compensated lifting steps. Both steps, prediction P and update U, utilize block-based motion compensation.

Fig. 1 depicts a Haar transform with motion-compensated lifting steps. The even frames of the video sequence $\mathbf{s}_{2\kappa}$ are used to predict the odd frames $\mathbf{s}_{2\kappa+1}$ (*prediction step P*). The prediction step is followed by an *update step U*. If the motion field between the even and odd frames is invertible, the corresponding motion vectors in the update and prediction steps sum to zero. We use a block-size of 16×16 and half-pel accurate motion compensation in the prediction step and select the motion vectors such that they minimize the squared error in the high-band $\mathbf{h}_\kappa$. In general, this block-motion field is not invertible but we still utilize the negative motion vectors for the update step as an approximation. Additional scaling factors in low- and high-band are necessary to normalize the transform.

2.2 Motion-Compensated Lifted 5/3 Wavelet

The Haar wavelet is a short filter and provides limited coding gain. We expect better coding efficiency with longer wavelet kernels. In the following, we discuss the lifted 5/3 wavelet kernel with motion compensation [56].

Fig. 2 depicts the 5/3 transform with motion-compensated lifting steps. Similar to the Haar transform, the update steps $U_{2\kappa+1}$ use the negative motion

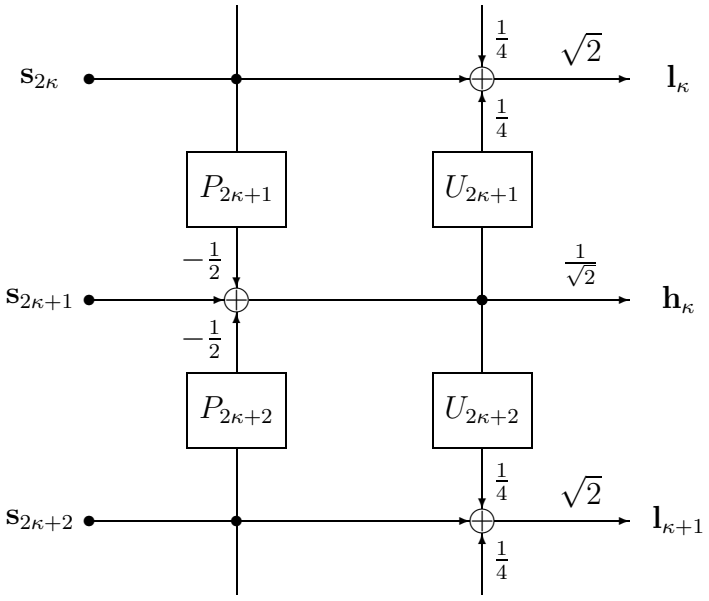

Fig. 2. Lifted 5/3 wavelet with motion compensation.

vectors of the corresponding prediction steps. But for this transform, the odd frames are predicted by a linear combination of two displaced neighboring even frames. Again, we use a block-size of 16×16 and half-pel accurate motion compensation in the prediction steps and choose the motion vectors for $P_{2\kappa+1}$ and $P_{2\kappa+2}$ such that they minimize the squared error in the high-band $\mathbf{h}_{\kappa}$. The corresponding update steps $U_{2\kappa+2}$ use also the negative motion vectors of the corresponding prediction steps.

2.3 Experimental Results

For the experiments, we subdivide the QCIF sequences *Mother & Daughter*, *Container Ship*, *Salesman*, *Mobile & Calendar*, *Foreman*, *News*, and *Car Phone*, each with 288 frames at 30 fps, into groups of K pictures. We decompose the GOPs independently and in the case of the 5/3 wavelet, we refer back to the first picture in the GOP when the GOP terminates. We will justify later, why we choose this cyclic extension to handle the GOP boundaries.

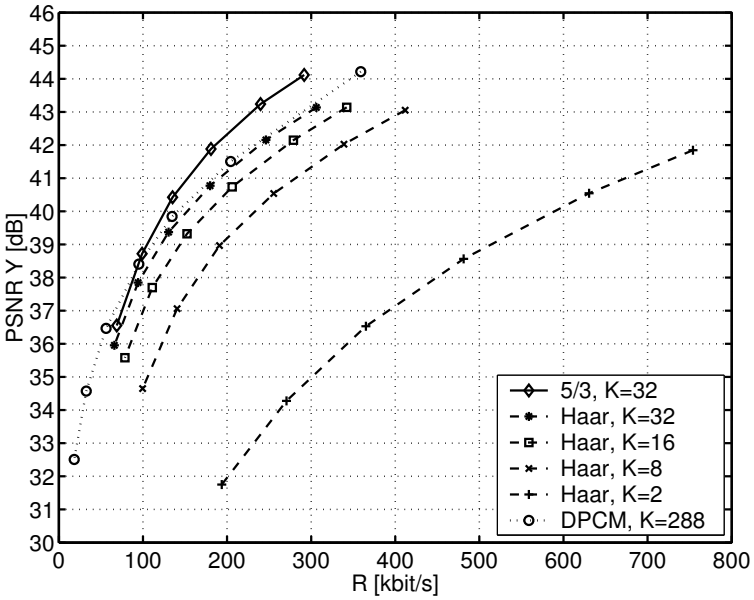

Fig. 3. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mother & Daughter* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

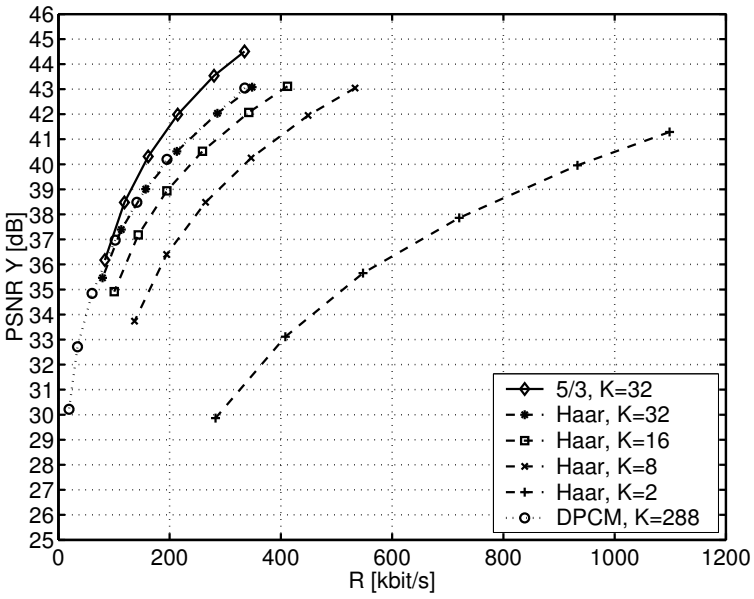

Fig. 4. Luminance PSNR vs. total bit-rate for the QCIF sequence *Container Ship* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

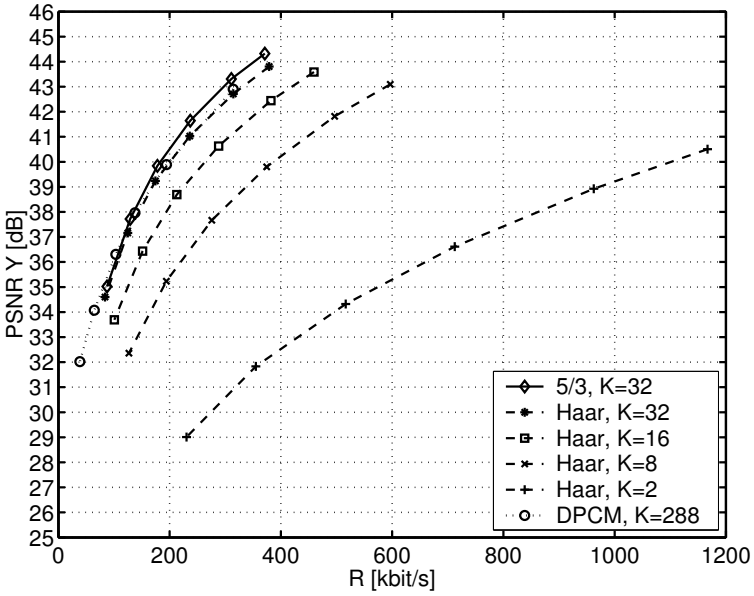

Fig. 5. Luminance PSNR vs. total bit-rate for the QCIF sequence *Salesman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

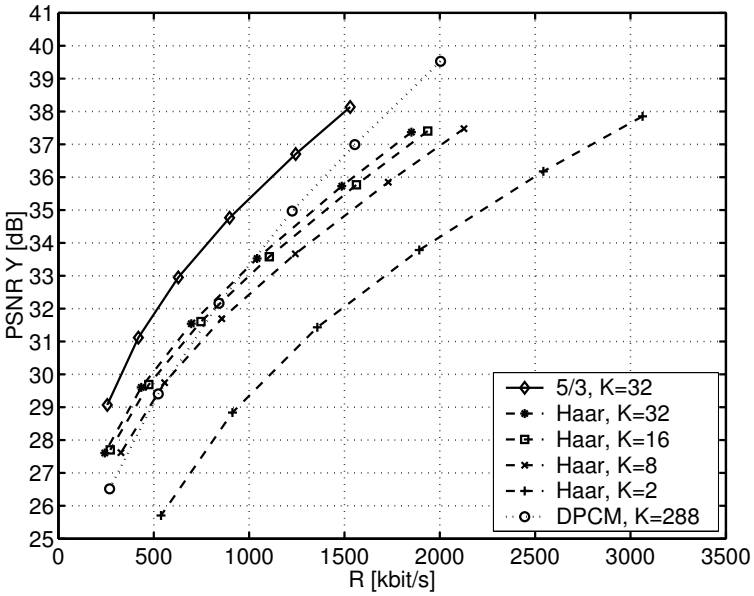

Fig. 6. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mobile & Calendar* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

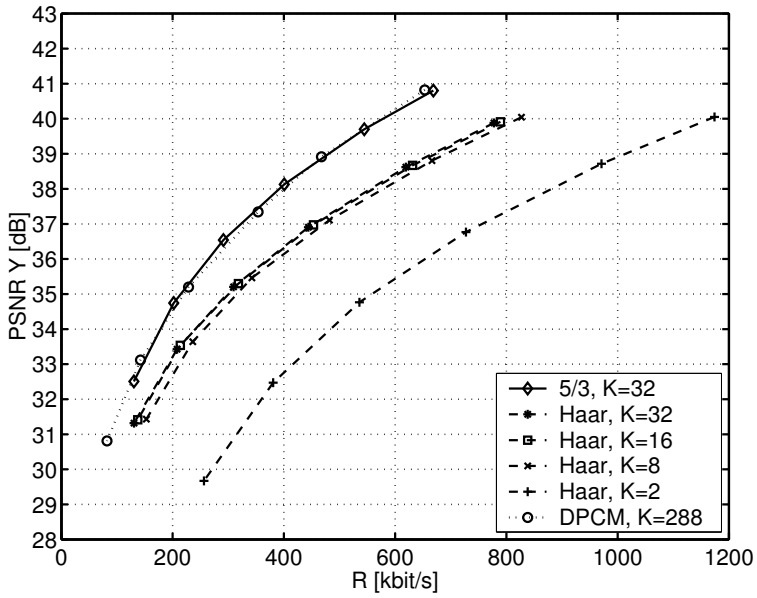

Fig. 7. Luminance PSNR vs. total bit-rate for the QCIF sequence *Foreman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

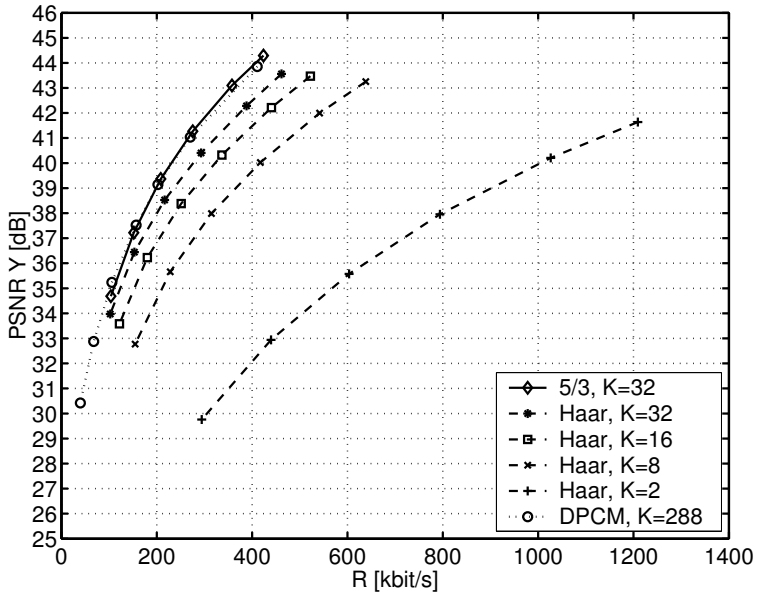

Fig. 8. Luminance PSNR vs. total bit-rate for the QCIF sequence *News* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

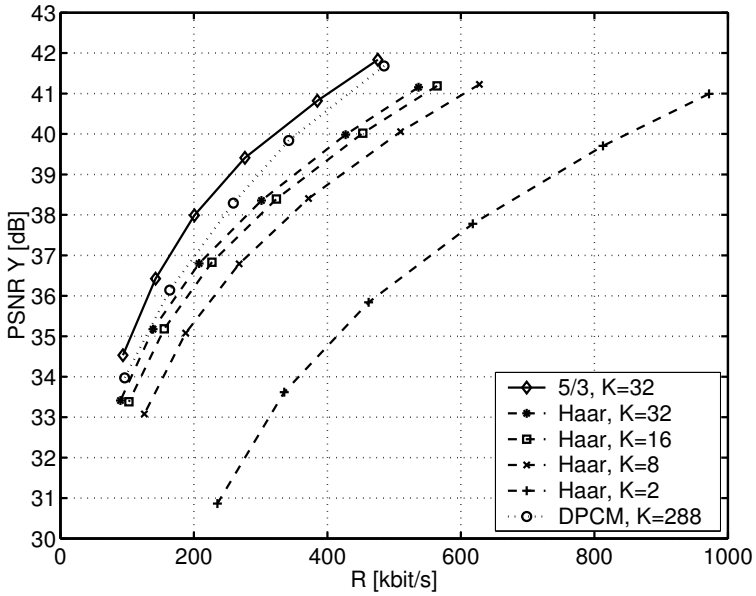

Fig. 9. Luminance PSNR vs. total bit-rate for the QCIF sequence *Car Phone* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

Figs. 3 - 9 show luminance PSNR over the total bit-rate for the test sequences encoded for groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel as well as for groups of $K = 32$ with the 5/3 kernel. The plots depict also coding results for a basic predictive video codec (DPCM) with one intra-frame and 287 inter-frames ($K = 288$).

The basic predictive video codec utilizes single-hypothesis 16×16 block-motion compensation, half-pel accurate motion vectors, one previous reference picture for prediction, and 8×8 DCT. We use the same building blocks for the temporal wavelet codec and the predictive codec such that their computational complexity is comparable. We just replace the temporal wavelet architecture by the temporal predictive architecture where both schemes use the same components (i.e. block-based motion field coding, 8×8 DCT transform coding). The update step in the wavelet architecture increases the computational

complexity only slightly as the motion information of the prediction steps is re-used.

We observe that the bit-rate savings with the Haar kernel diminish very quickly as the GOP size approaches 32 pictures. Note also that the 5/3 decomposition with a GOP size of 32 outperforms the Haar decomposition with a GOP size of 32. For the sequences *Mother & Daughter*, *Container Ship*, and *Salesman*, the Haar wavelet coding scheme with $K = 32$ performs similar to the basic predictive video codec with a very large GOP size. Please note that for *Mobile & Calendar* at lower bit-rates the Haar wavelet coding scheme outperforms the predictive video codec. The 5/3 decomposition with a GOP size of 32 outperforms not only the corresponding Haar decomposition but also the basic predictive video coding scheme with a GOP size of $K = 288$. For the sequences *Foreman*, *News*, and *Car Phone*, the 5/3 wavelet coding scheme performs comparable or slightly better than the predictive video codec. These sequences contain inhomogeneous motion and we suspect that the use of negative motion vectors in the update step permits only an insufficient approximation of the reverse motion field.

Figs. 10 and 11 show luminance PSNR over the total bit-rate for the CIF sequences *Container Ship* and *Mobile & Calendar*, each with 288 frames at 30 fps, encoded for groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel as well as for groups of $K = 32$ with the 5/3 kernel. The plots depict also coding results for the basic predictive video codec (DPCM) with one intra-frame and 287 inter-frames. When we compare the results to that of the corresponding QCIF sequences in Figs. 4 and 6, we observe a similar relative behavior for the Haar and 5/3 kernel as well as the basic predictive video codec. Please note that our coding schemes are block-based. The macroblock size is 16×16,

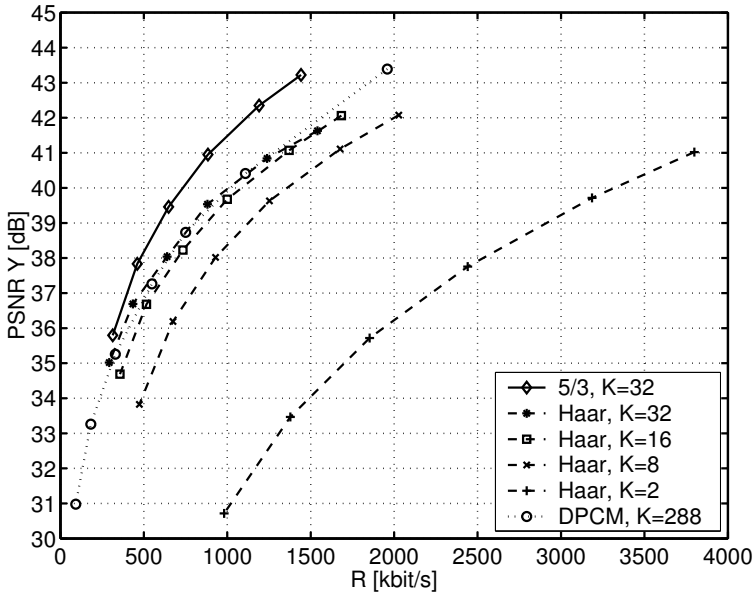

Fig. 10. Luminance PSNR vs. total bit-rate for the CIF sequence *Container Ship* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

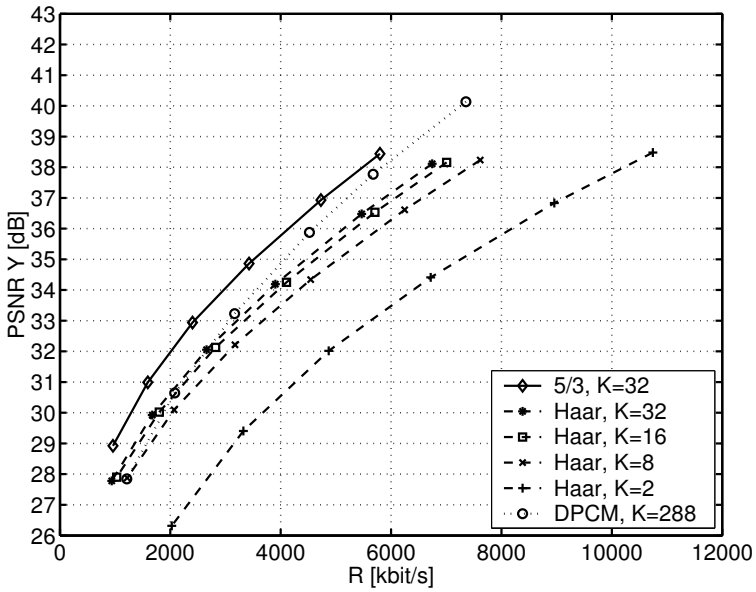

Fig. 11. Luminance PSNR vs. total bit-rate for the CIF sequence *Mobile & Calendar* at 30 fps. A dyadic decomposition is used to encode groups of $K = 2$, 8, 16, and 32 pictures with the Haar kernel, and $K = 32$ with the 5/3 kernel. Results for a basic predictive video codec with 287 inter-frames are given for reference.

independent of the spatial resolution of the coded sequences.

Further, we investigate the behavior of the coding scheme for the case that the frames are degraded by additive noise. This case will be of interest for explaining some of the findings of our theoretical analysis in Section 3. For that, we generate the sequence *Noisy Foreman* by repeating 32 times the first frame of the sequence *Foreman* and adding statistically independent white Gaussian noise of variance 25. As we investigate the residual noise only, this sequence contains no motion. Motion-compensated prediction is not capable of predicting the additive noise in the current frame. In fact, prediction doubles the noise variance in the residual signal and we expect that predictive coding performs inferior. Note that transform coding systems handle this noisy components differently. If the input pictures are degraded by statistically independent additive noise with identical variances, an orthonormal transform like the Haar wavelet preserves the total energy of the noise and distributes it into all subbands equally. This suggests that the transform coding system will perform superior.

Fig. 12 shows luminance PSNR over the total bit-rate for the sequence *Noisy Foreman*. The coding scheme with the Haar wavelet kernel and a dyadic decomposition of 32 pictures is compared to the predictive coding scheme. We observe that the wavelet coding scheme outperforms the predictive coding scheme by approximately 2 dB. The predictive coding scheme is inferior as the statistically independent noise in the current frame cannot be predicted from the previous frame. The maximum gain of 3 dB is not observed due to the intra-picture in the beginning of the GOP and the capability of the predictive codec to select an intra-macroblock if prediction is rate-distortion inefficient.

Finally, we discuss briefly our GOP boundary handling for the 5/3 wavelet. As we encode the GOPs in the sequence independently, we have to solve the

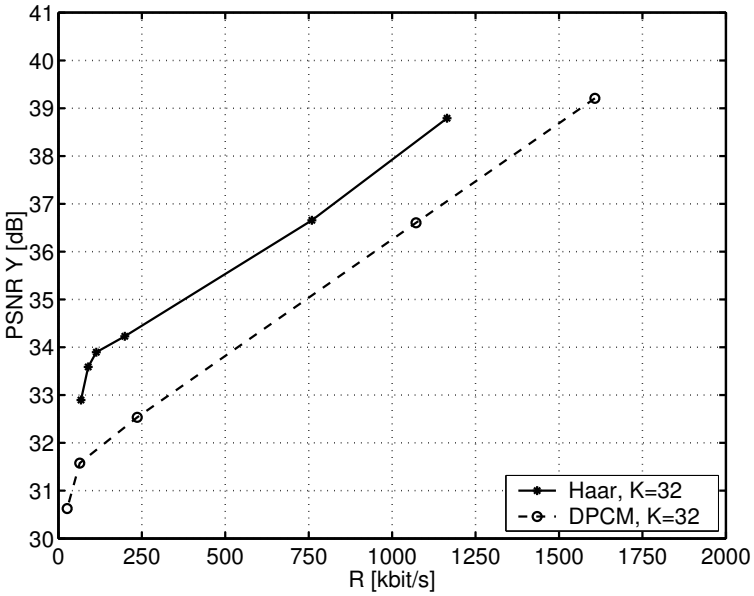

Fig. 12. Luminance PSNR vs. total bit-rate for the QCIF sequence *Noisy Foreman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the motion-compensated Haar wavelet. Results for a basic predictive video codec with 31 inter-frames are given for reference.

boundary problem for the 5/3 wavelet. For the discussion, we consider cyclic and symmetric extensions at the GOP boundary. Note that a GOP begins with the even picture $\mathbf{s}_0$ and terminates always with an odd picture. When the GOP terminates, the cyclic extension refers back to the first picture $\mathbf{s}_0$, and the symmetric extension uses the last even picture in the GOP twice as a reference. In the case of the cyclic extension, the terminal update step modifies the first picture $\mathbf{s}_0$, and in the case of the symmetric extension, the last even picture in the GOP is updated twice. We implemented both extensions for an experimental comparison of the rate-distortion performance. Other approaches are known to deal with the temporal boundary effects, for example the memory-constrained 3-D wavelet transform in [9,10] or the 3-D scan-based wavelet transform in [11,60], but we restrict the discussion to the above mentioned approaches.

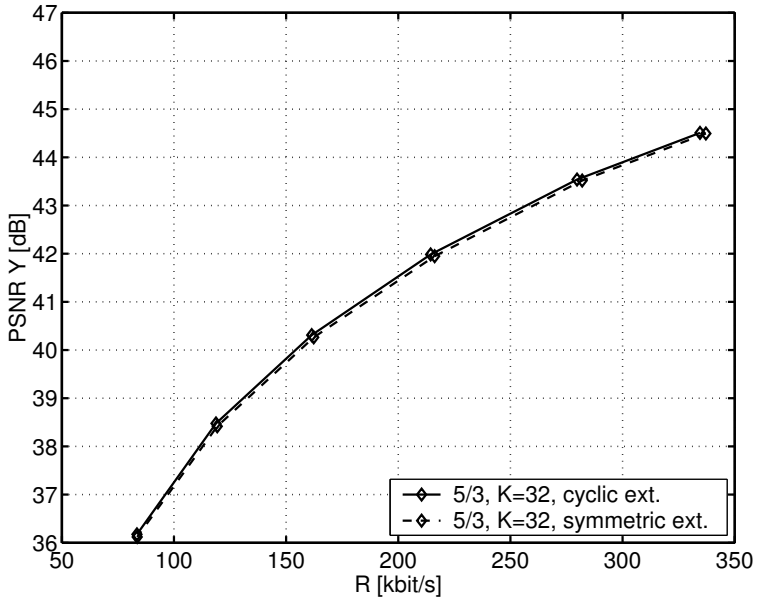

Fig. 13. Luminance PSNR vs. total bit-rate for the QCIF sequence *Container Ship* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the motion-compensated 5/3 wavelet. The compression efficiency of cyclic and symmetric extension is depicted.

Fig. 13 depicts luminance PSNR over the total bit-rate for the sequence *Container Ship*. We observe that both extensions show similar rate-distortion performance. We expect better efficiency for the cyclic extension as it is a multi-frame approach. But it seems that the large temporal distance is disadvantageous. As we observe a small advantage in terms of compression efficiency, also for other sequences like *Foreman*, we use the cyclic extension for our experiments.

3 Mathematical Model of Motion-Compensated Subband Coding

The experimental results show that the temporal subband coding scheme can provide superior compression efficiency when compared to the predictive coding scheme. In the following, we outline a mathematical model to study motion-compensated subband coding in more detail. With that, we derive performance bounds for motion-compensated three-dimensional transform coding and compare to bounds known for motion-compensated predictive coding.

Let $\mathbf{s}_k = \{\mathbf{s}_k[l], l \in \Pi\}$ be scalar random fields over a two-dimensional orthogonal grid Π with horizontal and vertical spacing of 1. The vector $l = (x, y)^T$ denotes a particular location in the lattice Π. We interpret $\mathbf{s}_k$ as the k-th of K pictures to be encoded. Further, the signal $\mathbf{s}_k[l]$ is thought of as samples of a space-continuous, spatially band-limited signal and we obtain a displaced version of it as follows: We shift the ideal reconstruction of the band-limited signal by the continuous-valued displacement vector d and re-sample it on the original grid. With this signal model, a spatially constant displacement operation is invertible.

3.1 Motion-Compensated Lifted Haar Wavelet

With the above signal model, we revisit the motion-compensated lifted Haar wavelet in Fig. 1 and remove the displacement operators in the lifting steps such that we can isolate a lifted Haar wavelet without displacement operators.

Fig. 14 shows the equivalent Haar wavelet where the displacement operators are pre- and post-processing operators with respect to the original Haar trans-

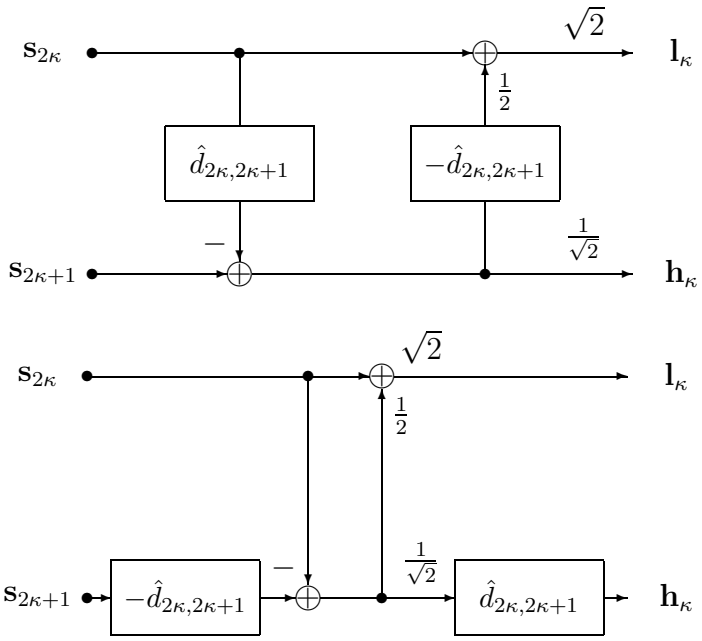

Fig. 14. Haar transform with lifting steps that shift the signal (top). As the shift operation is invertible, an equivalent system without shifts in the lifting steps is possible (bottom).

form. The schemes in Fig. 14 are equivalent, if the displacement operators are linear and invertible.

We continue and perform the dyadic decomposition of a GOP with the equivalent Haar wavelet. For that, the displacements of the equivalent Haar blocks have to be added. We assume that the estimated displacements between pairs of frames are additive such that, e.g., $\hat{d}_{02} + \hat{d}_{23} = \hat{d}_{03}$. As the true displacements are also additive, e.g. $d_{02} + d_{23} = d_{03}$, and differ from the estimated displacement by the displacement error, i.e. $d_{ij} = \hat{d}_{ij} + \Delta_{ij}$, we conclude that the displacement errors are also additive, e.g. $\Delta_{02} + \Delta_{23} = \Delta_{03}$, [61]. To motivate the above assumptions, consider an object in a video sequence that moves on a reversible trajectory. It is desirable to move on the trajectory in any direction and with an arbitrary number of intermediate steps in order to build a path between two given points on the trajectory.

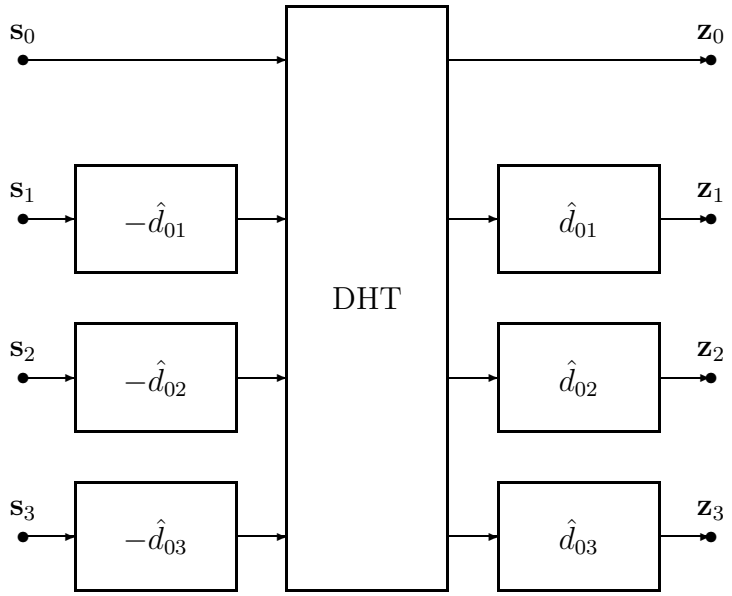

Fig. 15. Dyadic Haar Transform (DHT) without shifts in the lifting steps for $K = 4$ pictures.

Fig. 15 depicts a dyadic decomposition for $K = 4$ pictures based on the equivalent Haar wavelet in Fig. 14. The dyadic Haar transform without displacements in the lifting steps is labeled by DHT. The displacements $\hat{d}_{0k}$ are pre- and post-processing operators with respect to the original dyadic Haar decomposition DHT.

3.2 Motion-Compensated Lifted 5/3 Wavelet

We also apply the invertible displacement operator to the motion-compensated lifted 5/3 wavelet in Fig. 2 and obtain the equivalent 5/3 wavelet in Fig. 16. Due to the structure of the 5/3 wavelet, we have displacements between the frames 2κ & $2\kappa+1$, $2\kappa+2$ & $2\kappa+1$, and 2κ & $2\kappa+2$ (in the next decomposition level). Again, we assume that the estimated displacements are additive such that, e.g., $\hat{d}_{01} - \hat{d}_{21} = \hat{d}_{02}$. With this assumption, the displacement operators

between the levels cancel out and several decomposition levels are possible without displacements between the levels.

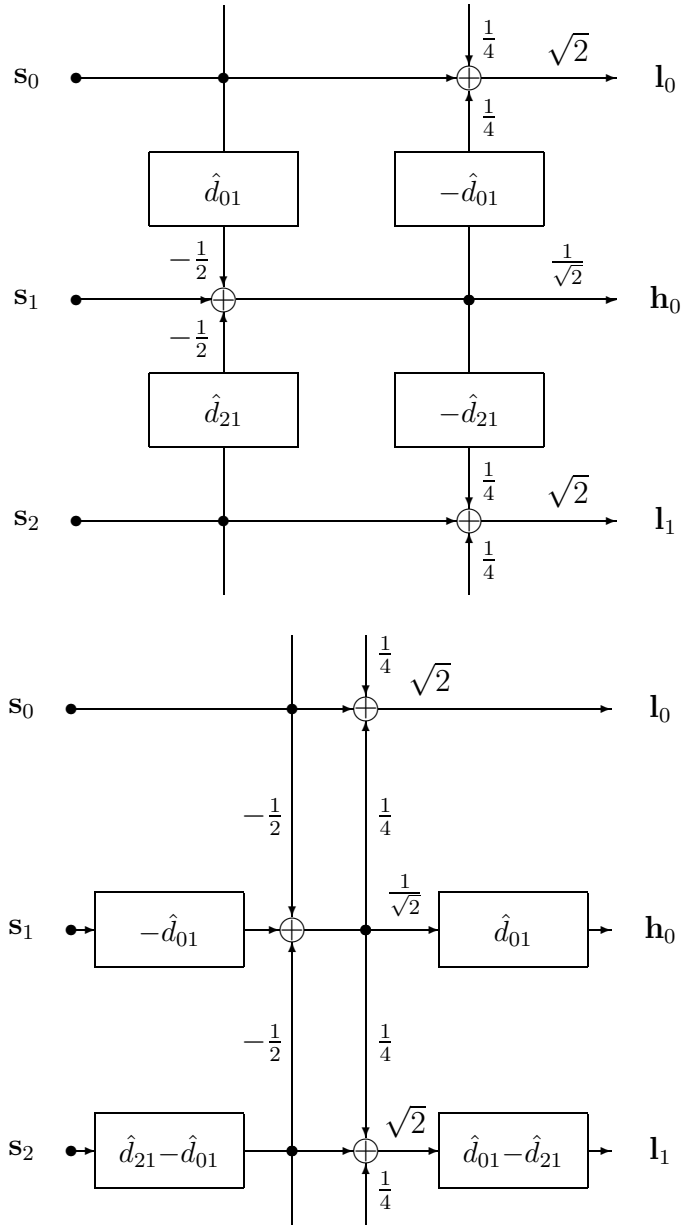

Fig. 16. 5/3 wavelet with lifting steps that shift the signal (top). As the shift operation is invertible, an equivalent system without shifts in the lifting steps is possible (bottom).

The equivalent dyadic 5/3 transform has the same pre- and post-processing displacement operators as the equivalent dyadic Haar transform in Fig. 15 but the DHT is replaced by the original dyadic 5/3 decomposition as depicted in

Fig. 17.

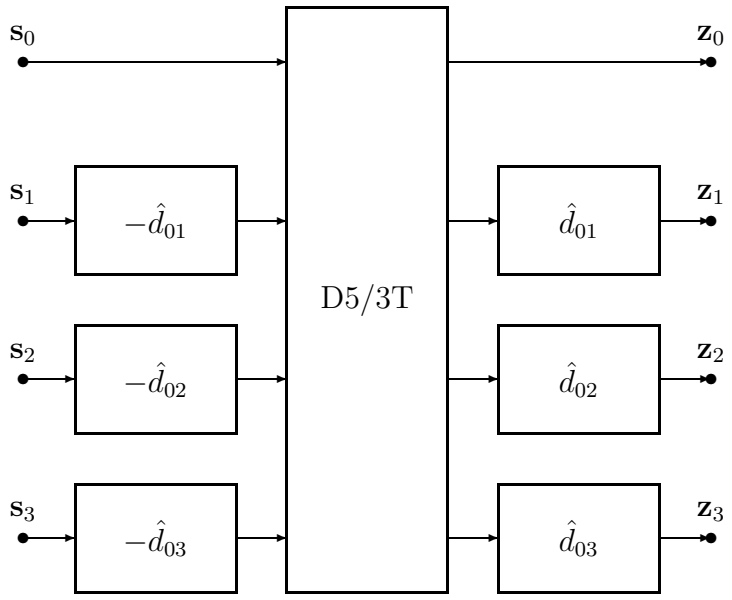

Fig. 17. Dyadic 5/3 Transform (D5/3T) without shifts in the lifting steps for $K = 4$ pictures.

3.3 Signal Model

Now, we assume that the pictures $\mathbf{s}_k$ are shifted versions of a "clean" video signal $\mathbf{v}$ with the true displacements d_{0k} and degraded by independent additive white Gaussian noise $\mathbf{n}_k$. Combining this signal model with the equivalent dyadic decomposition, we can eliminate the absolute displacements and restrict ourselves to the displacement error Δ_{0k} in the k-th picture. In the following, we do not consider particular displacement errors Δ_{0k}. We rather specify statistical properties and consider them as random variables $\mathbf{\Delta}_k$, statistically independent from the "clean" signal $\mathbf{v}$ and the noise $\mathbf{n}_k$. The noise signals $\mathbf{n}_\mu$ and $\mathbf{n}_\nu$ are also mutually statistically independent.

Fig. 18 depicts the model with the displacement-free and linear transform T

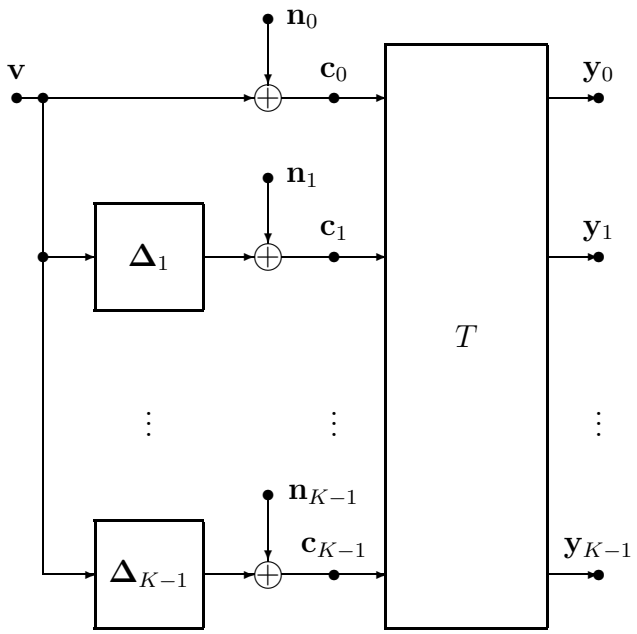

Fig. 18. Motion compensation for a group of K pictures.

for a group of K pictures. The motion-compensated pictures $\mathbf{c}_1, \ldots, \mathbf{c}_{K-1}$ are aligned with respect to the first picture $\mathbf{c}_0$. According to Fig. 15, the signals $\mathbf{z}_k$ are independently intra-frame encoded. As the absolute displacements have no influence on the performance of the intra-frame encoder, we omit them and consider only the direct output signals $\mathbf{y}_k$ of T.

Now, assume that the random fields $\mathbf{v}$ and $\mathbf{c}_k$ are jointly wide-sense stationary with the real-valued scalar two-dimensional power spectral densities $\Phi_{\mathbf{vv}}(\omega)$ and $\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}(\omega)$.

The power spectral densities $\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}(\omega)$ are elements in the power spectral density matrix of the motion-compensated pictures $\Phi_{\mathbf{cc}}$. The power spectral density matrix of the decorrelated signal $\Phi_{\mathbf{yy}}$ is given by $\Phi_{\mathbf{cc}}$ and the transform T,

$$\Phi_{\mathbf{yy}}(\omega) = T(\omega)\Phi_{\mathbf{cc}}(\omega)T^H(\omega), \tag{1}$$

where T^H denotes the Hermitian conjugate of T and $\omega = (\omega_x, \omega_y)^T$ the vector-valued frequency.

We adopt the expressions for the cross spectral densities $\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}$ from [62]

$$\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}(\omega) = E\left\{e^{-j\omega^T(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)}\right\}\Phi_{\mathbf{vv}}(\omega) + \Phi_{\mathbf{n}_\mu \mathbf{n}_\nu}(\omega) \tag{2}$$

and assume a normalized power spectrum $\Phi_{\mathbf{vv}}$ with $\sigma_{\mathbf{v}}^2 = 1$ that corresponds to an exponentially decaying isotropic autocorrelation function with a correlation coefficient between horizontally and vertically adjacent pixels of $\rho_{\mathbf{v}} = 0.93$.

For the k-th displacement error $\mathbf{\Delta}_k$, a 2-D normal distribution with variance $\sigma_{\mathbf{\Delta}}^2$ and zero mean is assumed where the x- and y-components are statistically independent. The expected value in (2) depends on the variance of the displacement error with respect to the reference picture $\mathbf{c}_0$ (*absolute displacement accuracy*) and the variance of the difference displacement error between pairs of non-reference pictures (*relative displacement accuracy*). We assume that each picture in a GOP can be the reference picture $\mathbf{c}_0$. That is, there is no preference among the pictures in a GOP and the variances of the absolute displacement error are the same for all $K - 1$ motion-compensated pictures. Based on the dyadic decomposition with motion-compensated lifted wavelets and the assumption that there is no preference among the pictures in a GOP, we assume that absolute and relative displacement accuracy are identical. The differences of absolute displacement errors are related to the relative displacement errors as we assume in Sections 3.1 and 3.2 additive estimated displacements.

$$\Delta_{0j} - \Delta_{0i} = \Delta_{ij} \tag{3}$$

With that, we obtain for the variances of the absolute and relative displacement error components:

$$E\left\{(\boldsymbol{\Delta}_{\mathbf{x}0j} - \boldsymbol{\Delta}_{\mathbf{x}0i})^2\right\} = E\left\{\boldsymbol{\Delta}_{\mathbf{x}ij}^2\right\} \tag{4}$$

$$2\sigma_{\boldsymbol{\Delta}}^2(1 - \rho_{\boldsymbol{\Delta}}) = \sigma_{\boldsymbol{\Delta}}^2 \tag{5}$$

This is only possible with correlated displacement errors such that $\rho_{\boldsymbol{\Delta}} = 0.5$ [63]. Finally, we abbreviate the expected value in (2) with $P(\omega, \sigma_{\boldsymbol{\Delta}}^2)$ which is the characteristic function of the continuous 2-D Gaussian displacement error.

$$E\left\{e^{-j\omega^T \boldsymbol{\Delta}_{\nu\mu}}\right\} := P(\omega, \sigma_{\boldsymbol{\Delta}}^2) = e^{-\frac{1}{2}\omega^T \omega \sigma_{\boldsymbol{\Delta}}^2} \tag{6}$$

With that, we obtain for the power spectral density matrix of the motion-compensated pictures

$$\frac{\Phi_{\mathbf{cc}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} 1+\alpha(\omega) & P(\omega) & \cdots & P(\omega) \\ P(\omega) & 1+\alpha(\omega) & \cdots & P(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ P(\omega) & P(\omega) & \cdots & 1+\alpha(\omega) \end{pmatrix}. \tag{7}$$

$\alpha = \alpha(\omega)$ is the normalized spectral density of the noise $\Phi_{\mathbf{n}_k\mathbf{n}_k}(\omega)$ with respect to the spectral density of the "clean" video signal.

$$\alpha(\omega) = \frac{\Phi_{\mathbf{n}_k\mathbf{n}_k}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \text{for} \quad k = 0, 1, \ldots, K-1 \tag{8}$$

T represents the dyadic Haar transform or the dyadic 5/3 transform. In terms of decorrelation and coding gain, the 5/3 wavelet performs better than the Haar wavelet as shown in Figs. 3 - 9. In the following, we are interested in

theoretical performance bounds and choose the Karhunen-Loeve Transform (KLT). The normalized eigenvalues of the power spectral density matrix $\Phi_{\mathbf{cc}}$ are $\lambda_1(\omega) = 1 + \alpha(\omega) + (K-1)P(\omega)$ and $\lambda_{2,3,\ldots,K}(\omega) = 1 + \alpha(\omega) - P(\omega)$. The power spectral density matrix of the transformed signals $\Phi_{\mathbf{yy}}$ is diagonal.

$$\frac{\Phi_{\mathbf{yy}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} \lambda_1(\omega) & 0 & \cdots & 0 \\ 0 & \lambda_2(\omega) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K(\omega) \end{pmatrix} \tag{9}$$

The first eigenvector just adds all components and scales with $1/\sqrt{K}$. For the remaining eigenvectors, any orthonormal basis can be used that is orthogonal to the first eigenvector. That is, the KLT for our signal model is not dependent on ω. Note that for this simple signal model, the Haar transform is also a KLT.

3.4 Transform Coding Gain

The rate difference [62] is used to measure the improved compression efficiency for each picture k.

$$\Delta R_k = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(\frac{\Phi_{\mathbf{y}_k\mathbf{y}_k}(\omega)}{\Phi_{\mathbf{c}_k\mathbf{c}_k}(\omega)} \right) d\omega \tag{10}$$

It represents the maximum bit-rate reduction (in bit per sample) possible by optimum encoding of the transformed signal $\mathbf{y}_k$, compared to optimum intra-frame encoding of the signal $\mathbf{c}_k$ for Gaussian wide-sense stationary signals for the same mean square reconstruction error. A negative ΔR_k corresponds

to a reduced bit-rate compared to optimum intra-frame coding. The overall rate difference ΔR is the average over all pictures and is used to evaluate the efficiency of motion-compensated transform coding. Assuming the KLT, we obtain for the overall rate difference

$$\Delta R = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{K-1}{2K} \log_2 \left(1 - \frac{P(\omega, \sigma_{\mathbf{\Delta}}^2)}{1+\alpha(\omega)} \right) + \frac{1}{2K} \log_2 \left(1 + (K-1) \frac{P(\omega, \sigma_{\mathbf{\Delta}}^2)}{1+\alpha(\omega)} \right) d\omega. \tag{11}$$

The case of a very large number of motion-compensated pictures is of special interest for the comparison to video coding with motion-compensated prediction.

$$\Delta R_{K \to \infty} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(1 - \frac{P(\omega, \sigma_{\mathbf{\Delta}}^2)}{1+\alpha(\omega)} \right) d\omega \tag{12}$$

The performance of single-hypothesis motion-compensated prediction with optimum Wiener filter achieves a rate difference of

$$\Delta R_{\mathrm{MCP}} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(1 - \frac{P^2(\omega, \sigma_{\mathbf{\Delta}}^2)}{[1+\alpha(\omega)]^2} \right) d\omega. \tag{13}$$

We obtain this result from [62], Eqn. 21 with $N = 1$ hypothesis per picture and $\alpha_0 = \alpha_1 = \alpha$. This comparison is chosen as both schemes employ single-hypothesis motion compensation per picture where the GOP size is very large, i.e. $K \to \infty$.

Figs. 19 and 20 depict the rate difference according to (11) and (13) over the displacement inaccuracy $\beta = \log_2(\sqrt{12}\sigma_{\mathbf{\Delta}})$ for a residual noise level RNL $= 10 \log_{10}(\sigma_{\mathbf{n}}^2)$ of -100 dB and -30 dB, respectively. Note that the variance of the "clean" video signal $\mathbf{v}$ is normalized to $\sigma_{\mathbf{v}}^2 = 1$. We observe that the rate

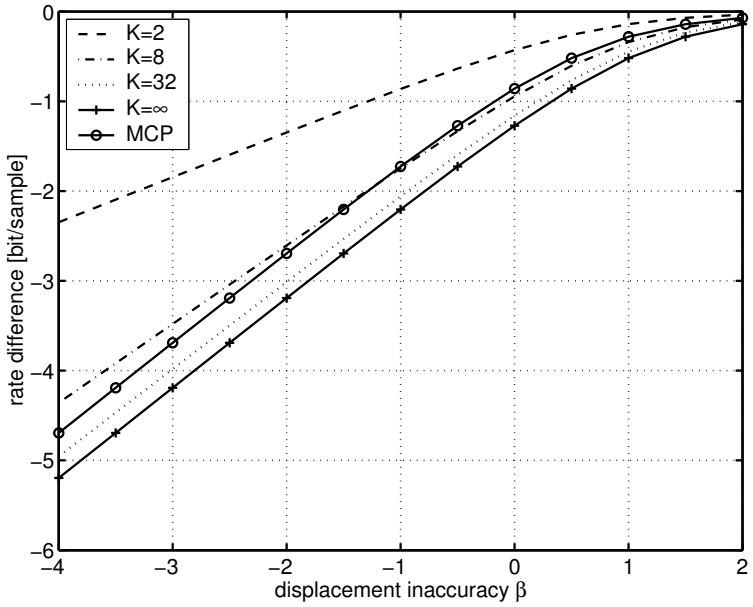

Fig. 19. Rate difference for motion-compensated transform coding with groups of K pictures over the displacement inaccuracy β. The performance of single-hypothesis motion-compensated prediction with Wiener filter is labeled by MCP. The residual noise level is -100 dB.

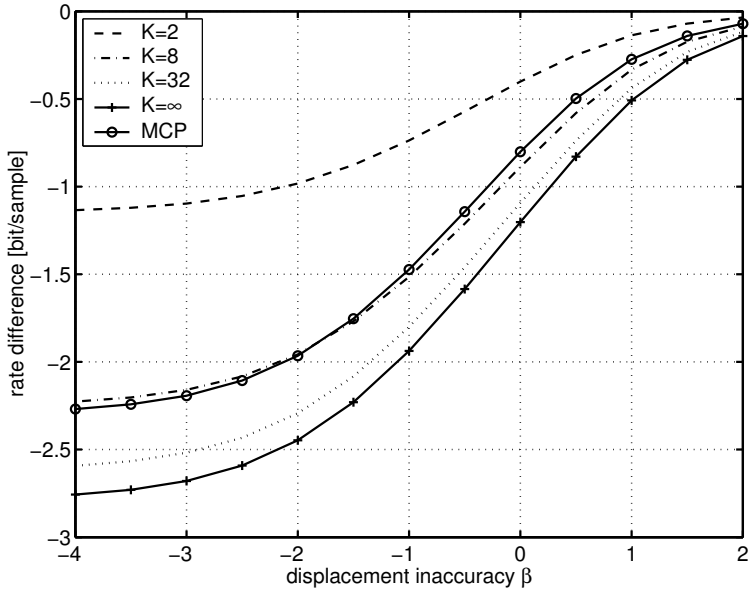

Fig. 20. Rate difference for motion-compensated transform coding with groups of K pictures over the displacement inaccuracy β. The performance of single-hypothesis motion-compensated prediction with Wiener filter is labeled by MCP. The residual noise level is -30 dB.

difference starts to saturate for $K = 32$. This observation is consistent with the experimental results in the previous section. For a very large group of pic-

tures and negligible residual noise, the slope of the rate difference is limited by 1 bit per sample per inaccuracy step, similar to that of single-hypothesis motion-compensated prediction. Further, motion-compensated transform coding outperforms single-hypothesis motion-compensated prediction by at most 0.5 bits per sample. For example, if we encode frames with statistically independent additive noise, motion-compensated prediction is not capable of predicting the additive noise in the current frame. In this case, prediction actually doubles the noise variance in the residual signal and predictive coding performs suboptimally.

4 Conclusions

This article discusses motion-compensated three-dimensional transform coding for groups of K pictures. We investigate experimentally and theoretically motion-compensated lifted wavelet transforms. The experiments show that the 5/3 wavelet kernel outperforms both the Haar kernel and, in many cases, the reference scheme utilizing single-hypothesis motion-compensated predictive coding. The motion-compensated lifted wavelet kernels re-use the motion vectors in the prediction step for the update step by assuming an invertible block-motion field. This assumption seems to be inadequate for sequences with inhomogeneous motion as their rate-distortion performance is weaker than expected.

The theoretical discussion is based on a signal model for K motion-compensated pictures that are decorrelated by a linear transform. The dyadic decomposition of K pictures with motion-compensated lifted wavelets is replaced by an equivalent coding scheme with K motion-compensated pictures

and a dyadic wavelet decomposition without motion compensation. That is, we remove the displacement operators in the lifting steps and generate a set of motion-compensated pictures with an additional constraint on the displacement errors. We generalize the model and employ the Karhunen-Loeve Transform to obtain theoretical performance bounds at high bit-rates for motion-compensated 3-D transform coding.

The analysis of this model gives the following insights: The coding gain for a group of K pictures is limited and saturates with increasing K. For a very large group of pictures and negligible residual noise, the slope of the rate difference is limited by 1 bit per sample per inaccuracy step. The slope of the rate difference for predictive coding with single-hypothesis motion compensation is also limited by 1 bit per sample per inaccuracy step but this coding scheme outperforms predictive coding with single-hypothesis motion compensation by up to 0.5 bits per sample. This is also true for very accurate motion compensation when the residual noise dominates the coding gain.

References

[1] M. Vetterli, J. Kovacevic, Wavelets and Subband Coding, Signal Processing Series, Prentice Hall, Englewood Cliffs, NJ, 1995.

[2] D. Taubman, M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice, Kluwer Academic Publishers, 2001.

[3] J. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, IEEE Transactions on Signal Processing 41 (12) (1993) 3445–3462.

[4] D. Taubman, High performance scalable image compression with EBCOT, IEEE Transactions on Image Processing 9 (7) (2000) 1158–1170.

[5] S. Knauer, Real-time video compression algorithm for Hadamard transform processing, IEEE Transactions on Electromagnetic Compatibility EMC-18 (1976) 28–36.

[6] G. Karlsson, M. Vetterli, Three dimensional sub-band coding of video, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, New York, NY, 1988, pp. 1100–1103.

[7] F. Bosveld, R. Lagendijk, J. Biemond, Hierarchical video coding using a spatio-temporal subband decomposition, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, San Francisco, CA, 1992, pp. 221–224.

[8] C. Podilchuk, N. Jayant, N. Farvardin, Three-dimensional subband coding of video, IEEE Transactions on Image Processing 4 (2) (1995) 125–139.

[9] J. Xu, S. Li, Z. Xiong, Y.-Q. Zhang, Memory-constrained 3D wavelet transforms for video coding without boundary effects, in: Proceedings of the IEEE International Symposium on Intelligent Signal Processing and Communication Systems, Honolulu, HI, 2000.

[10] J. Xu, Z. Xiong, S. Li, Y.-Q. Zhang, Memory-constrained 3D wavelet transform for video coding without boundary effects, IEEE Transactions on Circuits and Systems for Video Technology 12 (9) (2002) 812–818.

[11] C. Parisot, M. Antonini, M. Barlaud, 3D scan based wavelet transform for video coding, in: Proceedings of the IEEE Workshop on Multimedia Signal Processing, Cannes, France, 2001, pp. 403–408.

[12] G. Minami, Z. Xiong, A. Wang, S. Mehrotra, 3-D wavelet coding of video with arbitrary regions of support, IEEE Transactions on Circuits and Systems for Video Technology 11 (9) (2001) 1063–1068.

[13] T. Kronander, Motion compensated 3-dimensional wave-form image coding, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, Glasgow, Scotland, 1989, pp. 1921–1924.

[14] T. Kronander, New results on 3-dimensional motion compensated subband coding, in: Proceedings of the Picture Coding Symposium, Cambridge, MA, 1990, pp. 8.5–1.

[15] T. Akiyama, T. Takahashi, K. Takahashi, Adaptive three-dimensional transform coding for moving pictures, in: Proceedings of the Picture Coding Symposium, Cambridge, MA, 1990, pp. 8.2–1 – 8.2–2.

[16] A. Wang, Z. Xiong, P. Chou, S. Mehrotra, Three-dimensional wavelet coding of video with global motion compensation, in: Proceedings of the Data Compression Conference, Snowbird, UT, 1999, pp. 404–413.

[17] Y.-Q. Zhang, S. Zafar, Motion-compensated wavelet transform coding for color video compression, IEEE Transactions on Circuits and Systems for Video Technology 2 (3) (1992) 285–296.

[18] J.-R. Ohm, Three-dimensional subband coding with motion compensation, IEEE Transactions on Circuits and Systems for Video Technology 3 (5) (1994) 559–571.

[19] J.-R. Ohm, K. Rümmler, Variable-raster multiresolution video processing with motion compensation techniques, in: Proceedings of the IEEE International Conference on Image Processing, Santa Barbara, CA, USA, 1997, pp. 759–762.

[20] J.-R. Ohm, Temporal domain sub-band video coding with motion compensation, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, San Francisco, CA, 1992, pp. 229–232.

[21] J.-R. Ohm, Advanced packet-video coding based on layered VQ and SBC tech-

niques, IEEE Transactions on Circuits and Systems for Video Technology 3 (3) (1993) 208–221.

[22] S.-C. Han, J. Woods, Spatiotemporal subband/wavelet coding of video with object-based motion information, in: Proceedings of the IEEE International Conference on Image Processing, Santa Barbara, CA, 1997, pp. 629–632.

[23] S.-J. Choi, J. Woods, Motion-compensated 3-d subband coding of video, IEEE Transactions on Image Processing 8 (2) (1999) 155–167.

[24] P. Chen, J. Woods, Video coding for digital cinema, in: Proceedings of the IEEE International Conference on Image Processing, Vol. 1, Rochester, NY, 2002, pp. 749–752.

[25] J. Woods, T. Naveen, A filter based bit allocation scheme for subband compression of HDTV, IEEE Transactions on Image Processing 1 (3) (1992) 436–440.

[26] C.-H. Chou, C.-W. Chen, A perceptually optimized 3-D subband codec for video communication over wireless channels, IEEE Transactions on Circuits and Systems for Video Technology 6 (2) (1996) 143–156.

[27] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7) (1989) 674–693.

[28] A. Lippman, W. Butera, Coding image sequences for interactive retrieval, Communications of the ACM 32 (7) (1989) 852–860.

[29] K. Uz, M. Vetterli, D. LeGall, A multiresolution approach to motion estimation and interpolation with application to coding of digital HDTV, in: Proceedings of the IEEE International Symposium on Circuits and Systems, Vol. 2, New Orleans, LA, 1990, pp. 1298–1301.

[30] K. Uz, M. Vetterli, D. LeGall, Interpolative multiresolution coding of advanced

television with compatible subchannels, IEEE Transactions on Circuits and Systems for Video Technology 1 (1) (1991) 86–99.

[31] B. Girod, U. Horn, A scalable codec for Internet video streaming, in: Proceedings of the International Conference on Digital Signal Processing, Vol. 1, Santorini, Greece, 1997, pp. 221–224.

[32] J.-R. Ohm, Motion-compensated 3-d subband coding with multiresolution representation of motion parameters, in: Proceedings of the IEEE International Conference on Image Processing, Vol. 3, Austin, TX, 1994, pp. 250–254.

[33] S. Zafar, Y.-Q. Zhang, B. Jabbari, Multiscale video representation using multiresolution motion compensation and wavelet decomposition, IEEE Journal on Selected Areas in Communications 11 (1) (1993) 24–35.

[34] D. Taubman, A. Zakhor, Rate and resolution scalable subband coding of video, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, Adelaide, Australia, 1994, pp. 493–496.

[35] D. Taubman, A. Zakhor, Highly scalable, low-delay video compression, in: Proceedings of the IEEE International Conference on Image Processing, Austin, TX, 1994, pp. 740–744.

[36] D. Taubman, A. Zakhor, Multirate 3-d subband coding of video, IEEE Transactions on Image Processing 3 (5) (1994) 572–588.

[37] D. Taubman, Directionality and scalability in image and video compression, Ph.D. thesis, University of California, Berkeley (1994).

[38] D. Taubman, A. Zakhor, A common framework for rate and distortion based scaling of highly scalable compressed video, IEEE Transactions on Circuits and Systems for Video Technology 6 (4) (1996) 329–354.

[39] B.-J. Kim, W. Pearlman, An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT), in: Proceedings of the Data Compression Conference, Snowbird, UT, 1997, pp. 251–260.

[40] B.-J. Kim, Z. Xiong, W. Pearlman, Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT), IEEE Transactions on Circuits and Systems for Video Technology 10 (8) (2000) 1374–1387.

[41] A. Said, W. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, IEEE Transactions on Circuits and Systems for Video Technology 6 (3) (1996) 243–250.

[42] T. Naveen, F. Bosveld, J. Woods, R. Lagendijk, Rate constrained multiresolution transmission of video, IEEE Transactions on Circuits and Systems for Video Technology 5 (3) (1995) 193–206.

[43] J. Woods, G. Lilienfield, A resolution and frame-rate scalable subband/wavelet video coder, IEEE Transactions on Circuits and Systems for Video Technology 11 (9) (2001) 1035–1044.

[44] J. Tham, S. Ranganath, A. Kassim, Highly scalable wavelet-based video codec for very low bit-rate environment, IEEE Journal on Selected Areas in Communications 16 (1) (1998) 12–27.

[45] B. Felts, B. Pesquet-Popescu, Efficient context modeling in scalable 3D wavelet-based video compression, in: Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, 2000, pp. 1004–1007.

[46] V. Bottreau, M. Benetiere, B. Felts, B. Pesquet-Popescu, A fully scalable 3D subband video codec, in: Proceedings of the IEEE International Conference on Image Processing, Thessaloniki, Greece, 2001, pp. 1017–1020.

[47] J. Xu, Z. Xiong, S. Li, Y.-Q. Zhang, Three-dimensional embedded subband

coding with optimal truncation (3D ESCOT), Applied and Computational Harmonic Analysis 10 (2001) 290–315.

[48] I. Daubechies, W. Sweldens, Factoring wavelet transforms into lifting steps, J. Fourier Anal. Appl. 4 (1998) 247–269.

[49] W. Sweldens, The lifting scheme: A new philosophy in biorthogonal wavelet constructions, in: Wavelet Applications in Signal and Image Processing III, SPIE 2569, 1995, pp. 68–79.

[50] W. Sweldens, The lifting scheme: A construction of second generation wavelets, SIAM Journal on Mathematical Analysis 29 (2) (1998) 511–546.

[51] R. Calderbank, I. Daubechies, W. Sweldens, B.-L. Yeo, Wavelet transforms that map integers to integers, Appl. Comput. Harmon. Anal. 5 (3) (1998) 332–369.

[52] R. Claypoole, G. Davis, W. Sweldens, R. Baraniuk, Nonlinear wavelet transforms for image coding, in: Proceedings of the 31st IEEE Asilomar Conference on Signals, Systems, and Computers, Vol. 1, Pacific Grove, CA, 1997, pp. 662–667.

[53] R. Claypoole, R. Baraniuk, R. Nowak, Adaptive wavelet transforms via lifting, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, Washington, 1998, pp. 1513–1516.

[54] B. Pesquet-Popescu, V. Bottreau, Three-dimensional lifting schemes for motion compensated video compression, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, Salt Lake City, UT, 2001, pp. 1793–1796.

[55] L. Luo, J. Li, S. Li, Z. Zhuang, Y.-Q. Zhang, Motion-compensated lifting wavelet and its application in video coding, in: Proceedings of the IEEE International Conference on Multimedia and Expo, Tokyo, Japan, 2001, pp. 481–484.

[56] A. Secker, D. Taubman, Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting, in: Proceedings of the IEEE International Conference on Image Processing, Vol. 2, Thessaloniki, Greece, 2001, pp. 1029–1032.

[57] A. Secker, D. Taubman, Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation, in: Proceedings of the IEEE International Conference on Image Processing, Vol. 3, Rochester, NY, 2002, pp. 749–752.

[58] J.-R. Ohm, Motion-compensated wavelet lifting filters with flexible adaptation, in: Proceedings of the International Workshop on Digital Communications, Capri, Italy, 2002, pp. 113–120.

[59] J. Viéron, C. Guillemot, S. Pateux, Motion compensated 2D+t wavelet analysis for low rate FGS video compression, in: Proceedings of the International Workshop on Digital Communications, Capri, Italy, 2002, pp. 129–135.

[60] C. Parisot, M. Antonini, M. Barlaud, Motion-compensated scan based wavelet transform for video coding, in: Proceedings of the International Workshop on Digital Communications, Capri, Italy, 2002, pp. 121–127.

[61] M. Flierl, B. Girod, Investigation of motion-compensated lifted wavelet transforms, in: Proceedings of the Picture Coding Symposium, Saint-Malo, France, 2003, pp. 59–62.

[62] B. Girod, Efficiency analysis of multihypothesis motion-compensated prediction for video coding, IEEE Transactions on Image Processing 9 (2) (2000) 173–183.

[63] M. Flierl, B. Girod, Video coding with motion compensation for groups of pictures, in: Proceedings of the IEEE International Conference on Image Processing, Vol. 1, Rochester, NY, 2002, pp. 69–72.