Video Coding with Lifted Wavelet Transforms and Complementary Motion-Compensated Signals

Markus Flierl[a], Pierre Vandergheynst[a], and Bernd Girod[b]
[a]Signal Processing Institute, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland
[b]Information Systems Laboratory, Stanford University, Stanford, CA 94305, USA
markus.flierl@epfl.ch, pierre.vandergheynst@epfl.ch, bgirod@stanford.edu

ABSTRACT

This paper investigates video coding with wavelet transforms applied in the temporal direction of a video sequence. The wavelets are implemented with the lifting scheme in order to permit motion compensation between successive pictures. We improve motion compensation in the lifting steps and utilize complementary motion-compensated signals. Similar to superimposed predictive coding with complementary signals, this approach improves compression efficiency. We investigate experimentally and theoretically complementary motion-compensated signals for lifted wavelet transforms. Experimental results with the complementary motion-compensated Haar wavelet and frame-adaptive motion compensation show improvements in coding efficiency of up to 3 dB. The theoretical results demonstrate that the lifted Haar wavelet scheme with complementary motion-compensated signals is able to approach the bound for bit-rate savings of 2 bits per sample and motion-accuracy step when compared to optimum intra-frame coding of the input pictures.

Keywords: Video Coding, Adaptive Wavelets, Lifting, Three-Dimensional Subband Coding of Video, Complementary Motion-Compensated Signals, Superimposed Prediction

1. INTRODUCTION

Applying a linear transform in temporal direction of a video sequence may not yield high compression efficiency if significant motion is prevalent. A linear transform along motion trajectories seems more suitable but requires a motion-adaptive transform for the input pictures. For wavelet transforms, this adaptivity can be achieved by constructing the kernel with the so called lifting scheme[1]: A two-channel decomposition is realized by a sequence of prediction and update steps that form a ladder structure. Adaptivity is permitted by incorporating motion compensation into prediction and update steps as proposed in Ref. 2, 3. The fact that the lifting structure is able to map integers to integers without requiring invertible lifting steps makes this approach feasible.

The theoretical investigation in Ref. 4 models a 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. The Karhunen-Loeve Transform is utilized to obtain theoretical performance bounds at high bit-rates. A comparison to both optimum intra-frame coding of the input pictures and motion-compensated predictive coding is given. Further, it is shown that the motion-compensated subband coding scheme can achieve bit-rate savings up to 1 bit per sample and motion-accuracy step when compared to optimum intra-frame coding. Note that a motion-accuracy step corresponds to an improvement from, e.g., integer-pel to half-pel accuracy or half-pel to quarter-pel accuracy. Moreover, Ref. 4 demonstrates that this scheme can outperform predictive coding with motion compensation by at most 0.5 bits. Predictive coding fails for statistically independent signal components. In the worst case, the prediction error variance is two times the signal variance which corresponds to a degradation of 0.5 bits per sample when assuming Gaussian signals.

It is known that the efficiency of motion-compensated prediction can be improved by utilizing superimposed motion-compensated signals as employed in MPEG's B-pictures. Prediction with linear combinations of motion-compensated signals is also called multihypothesis motion-compensated prediction.[5,6] B-pictures and overlapped block motion compensation are well known examples. The advantage of averaging multiple motion-compensated signals roots in the suppression of statistically independent noise components and, consequently, the improvement in prediction efficiency. Ref. 7 investigates superimposed prediction with complementary motion-compensated signals. The multiple motion-compensated signals with their associated displacement errors are chosen such that the superposition of the motion-compensated signals minimizes the degradation of the prediction signal due to

the displacement errors and, consequently, improves prediction performance. Motion-compensated signals chosen according to this criterion are called complementary. The investigation shows that already two complementary motion-compensated signals provide a large portion of the theoretically possible gain obtained with a very large number of complementary signals. In addition, the superposition of complementary motion-compensated signals benefits also from the suppression of statistically independent noise components. It is observed that complementary motion-compensated signals achieve bit-rate savings of up to 2 bits per sample and motion-accuracy step when compared to optimum intra-frame coding. Note that the bit-rate savings for motion-compensated prediction are limited to 1 bit per sample and motion-accuracy step.

The paper investigates experimentally and theoretically how complementary motion-compensated signals improve the efficiency of inter-frame wavelet coding of video signals. In Sec. 2, we modify the motion-compensated Haar wavelet as presented in Ref. 2, 3 and utilize complementary motion-compensated signals for the transform. Of particular interest are pairs of complementary signals as they provide a good trade-off between efficiency and algorithmic complexity. Sec. 2.1 outlines the coding scheme that uses also frame-adaptive motion compensation. The obtained experimental results are discussed in Sec. 2.2.

The theoretical investigation in Sec. 3 models complementary motion-compensated signals and utilizes them for the lifted Haar wavelet scheme. Again, of particular interest are pairs of complementary signals. Sec. 3.2 determines performance bounds for temporal wavelet coding with the complementary motion-compensated Haar wavelet and compares them to that of motion-compensated lifted wavelet coding. This comparison supports the experimental results that we obtain with our inter-frame wavelet coding scheme utilizing complementary motion-compensated signals. Sec. 3.3 relates the obtained performance bounds to that of superimposed predictive coding with complementary motion-compensated signals. We discuss the relation between both bounds that are obtained with the same number of complementary motion-compensated signals for a given number of input pictures. With this study, we are able to provide a theoretical comparison between predictive and wavelet coding with complementary motion-compensated signals.

2. HAAR WAVELET WITH COMPLEMENTARY MOTION-COMPENSATED SIGNALS

The classic motion-compensated Haar wavelet as proposed in Ref. 2, 3 permits motion compensation in the prediction and update steps of the lifting structure. The motivation for motion compensation in the lifting steps is to perform a wavelet transform along the motion trajectories in a video sequence for more efficient decorrelation of successive pictures. As the true motion in a video sequence is not known a priori, the encoder is bound to utilize only an estimate of the motion for compensation in the lifting steps. Efficient motion compensation relies on accurate motion estimates. But any practical coding scheme has to deal with inaccurate motion compensation due to quantization of motion information. One approach to encounter the degradation due to inaccurate motion compensation is to utilize complementary motion-compensated signals.[7] The rationale for this approach is to accept the degradation of one inaccurate motion-compensated signal but to combine it with at least another inaccurate motion-compensated signal such that the superimposed signal causes less degradation than each individual signal will inflict. In the following, we extend the motion compensation in the prediction and update steps of the Haar wavelet such that we are able to utilize complementary motion-compensated signals in the lifting steps.

Consider two pictures $\mathbf{s}_\mu$ and $\mathbf{s}_\nu$ as well as the associated true displacement vector $d_{\mu\nu} = (d_{\mu\nu,x}, d_{\mu\nu,y})^T$ that captures the true motion information between the two pictures. For coding purposes, we estimate the motion and obtain the estimated displacement vector $\hat{d}_{\mu\nu}$. We relate estimated and true displacement vector by adding the displacement error $\Delta_{\mu\nu}$, such that $d_{\mu\nu} = \hat{d}_{\mu\nu} + \Delta_{\mu\nu}$. As outlined before, we extend the lifting steps of the Haar wavelet such that they superimpose complementary motion-compensated signals. Fig. 1 depicts the example of the adaptive Haar wavelet where $N = 2$ motion-compensated signals with estimated displacements $\hat{d}^{(1)}_{2\kappa,2\kappa+1}$ and $\hat{d}^{(2)}_{2\kappa,2\kappa+1}$ are averaged in the prediction step as well as in the update step. Note that we utilize for the update step just the negative vectors $-\hat{d}^{(1)}_{2\kappa,2\kappa+1}$ and $-\hat{d}^{(2)}_{2\kappa,2\kappa+1}$ of the estimated displacement vectors in the prediction step. This is the best choice if the motion field between the two pictures is invertible. Otherwise, we obtain just a suboptimal approximation with low computational complexity.

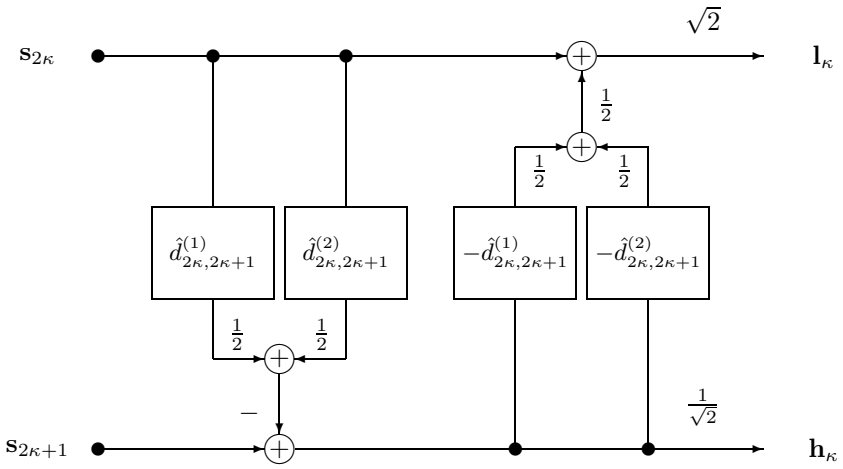

Figure 1. Haar transform of the pictures $\mathbf{s}_{2\kappa}$ and $\mathbf{s}_{2\kappa+1}$ with $N = 2$ superimposed motion-compensated signals in the lifting steps. The motion-compensated signals in the steps are just averaged.

Ref. 7 shows that the superposition of just two complementary motion-compensated signals is very efficient. Combining more than two complementary signals improves further the efficiency but increases significantly the complexity of the estimation algorithm. We obtain the displacement vectors for the complementary signals in the prediction step by minimizing the rate-distortion costs associated with the high band $\mathbf{h}_\kappa$. Ref. 7 shows further that minimizing the mean square prediction error causes the displacement errors of the complementary motion-compensated signals to be maximally negatively correlated. The superposition of these motion-compensated signals leads to lower mean square errors than each individual signal will be able to achieve. We note that simple averaging of complementary motion-compensated signals is optimal if motion compensation is very accurate for all signals.

2.1. Coding Scheme

We process the video sequence in groups of $K = 32$ pictures (GOPs). First, we decompose each GOP in temporal direction with the complementary motion-compensated Haar wavelet. Second, these K output pictures are intra-frame encoded. For simplicity, we utilize a 8×8 DCT with scalar quantization and run-length coding. The even frames of the video sequence $\mathbf{s}_{2\kappa}$ are displaced and superimposed to predict the odd frames $\mathbf{s}_{2\kappa+1}$. The prediction step is followed by an update step with the negative displacements of the prediction step. We use a block-size of 16×16 and half-pel accurate motion compensation with bi-linear interpolation. In general, the block-motion field is not invertible but we still utilize the negative motion vectors for the update step. Further, the coding scheme with the complementary motion-compensated Haar wavelet is adaptive in the number of motion-compensated signals on a block basis. We employ at most $N = 2$ complementary motion-compensated signals but we also permit just one motion-compensated signal ($N = 1$) to obtain the classic motion-compensated Haar wavelet. Depending on the video signal and the bit-rate constraint, complementary motion-compensated signals might not be rate-distortion efficient at low bit-rates. As two complementary signals require a larger bit-rate for the displacement information, the adaptivity in the number of combined motion-compensated signals helps to improve the efficiency at low bit-rates.

So far, we considered the fix coding structure of the lifted Haar wavelet. If we consider one GOP as an independently decodable unit, we are able to utilize all even pictures in a GOP to be used as "references" for block-based motion compensation in the prediction step. Obviously, the corresponding update step has to be modified. In Ref. 8, we investigate the lifted Haar transform with frame-adaptive motion compensation and employ the following rule: Each even picture that is used as "reference" for motion compensation in the prediction step receives also a motion-compensated signal component in the corresponding update step. For this update step, we use again the negative motion vector of the corresponding prediction step. In Ref. 8, we

reference up to M even pictures $\mathbf{s}_{2\tau}$ that are direct neighbors of the picture $\mathbf{s}_{2\kappa+1}$. If we set $M = 1$, we use just the picture $\mathbf{s}_{2\kappa}$ such that we obtain the classic lifted Haar wavelet. Depending on the video signal and the bit-rate constraint, the encoder determines for each block the best picture in the set of M even pictures. In this case, the displacement information is extended by a picture reference parameter.

Also $N = 2$ complementary motion-compensated signals can be determined with this frame-adaptive method. If we set $M = 1$, we choose $N = 2$ complementary signals from the same picture $\mathbf{s}_{2\kappa}$ as depicted in Fig. 1. But if we permit $M > 1$, we chose $N = 2$ complementary signals from up to M pictures $\mathbf{s}_{2\tau}$. Note that in this case, the coding scheme is able to select each complementary signal from different even pictures $\mathbf{s}_{2\tau}$ that belong to the set of pictures of size M. If we deal with the borders of the GOP, we permit also even pictures from the opposite border of the GOP.

2.2. Experimental Results

For the experimental results, we subdivide the test sequences *Foreman*, *Mobile & Calendar*, *Container Ship*, and *Mother & Daughter*, each with 288 frames, into groups of $K = 32$ pictures. We decompose the GOPs independently and the Haar kernel causes no boundary problems. Block-based rate-constrained motion estimation is used to minimize the Lagrangian costs of the blocks in the high bands. If $N = 2$ complementary signals are estimated, the pairs of displacement parameters are estimated by an iterative algorithm such that the total Lagrangian costs are minimized. The costs are determined by the energy of the block in the high band and an additive bit-rate term that is weighted by the Lagrange multiplier λ. The bit-rate term is the sum of the lengths of the codewords that are used to signal one ($N = 1$) or two ($N = 2$) displacements for each prediction step. The quantizer step-size Q is related to the Lagrange multiplier λ such that $\lambda = 0.2Q^2$. Employing the Haar wavelet and setting the displacements to zero, the dyadic decomposition will be an orthonormal transform. Therefore, we choose the same quantizer step-size for all K intra-frame encoder. The motion information that is required for the complementary motion-compensated wavelet transform is estimated in each decomposition level depending on the results of the lower level. Independent of the number of superimposed signals N, the coding scheme permits up to M even pictures in a GOP from which motion-compensated signals can be chosen.

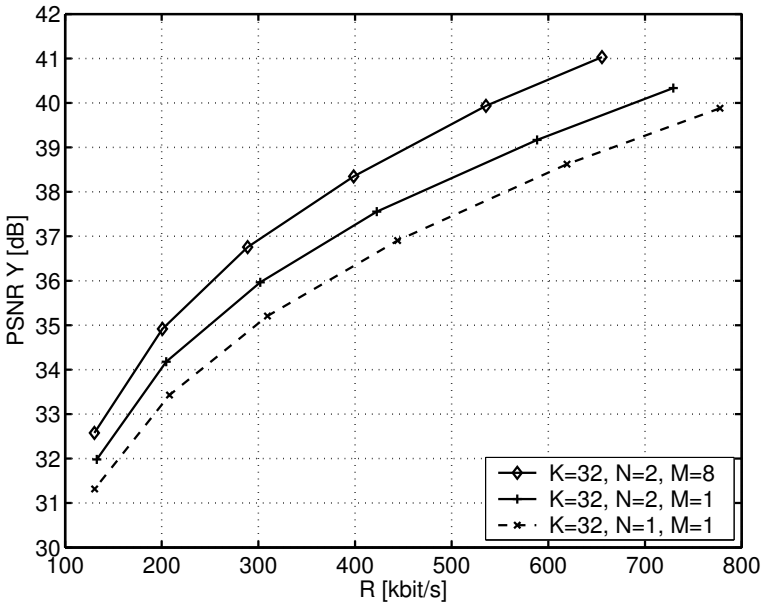

Figure 2. Luminance PSNR vs. total bit-rate for the QCIF sequence *Foreman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures. The motion-compensated Haar wavelet ($N = 1$) is compared to the Haar wavelet with two complementary motion-compensated signals ($N = 2$). In the case of two complementary signals, we select either from $M = 1$ or $M = 8$ even pictures in the GOP.

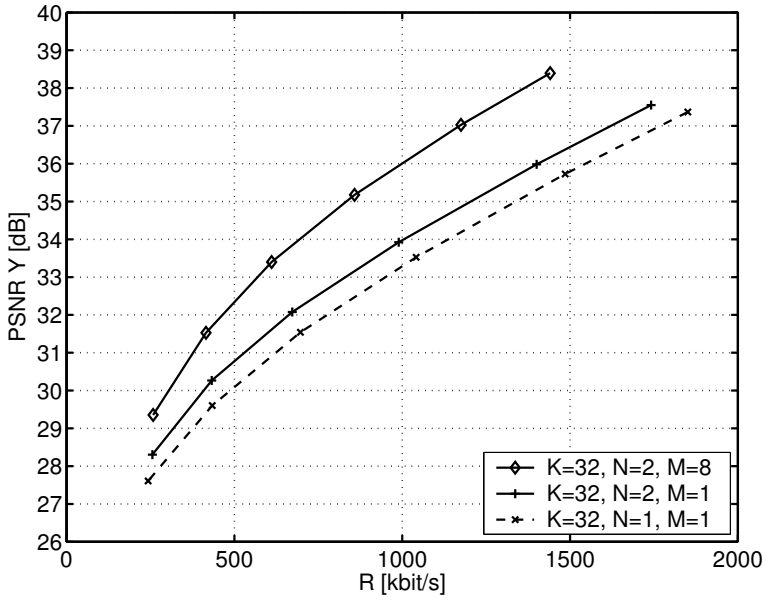

Figure 3. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mobile & Calendar* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures. The motion-compensated Haar wavelet ($N = 1$) is compared to the Haar wavelet with two complementary motion-compensated signals ($N = 2$). In the case of two complementary signals, we select either from $M = 1$ or $M = 8$ even pictures in the GOP.

Figs. 2 and 3 show the luminance PSNR over the total bit-rate for the sequence *Foreman* and *Mobile & Calendar*, respectively. We subdivide the sequences, each with 288 frames, into groups of $K = 32$ pictures and encode them with the complementary motion-compensated Haar wavelet. The classic Haar wavelet with one motion-compensated signal ($N = 1$) and one "reference" picture ($M = 1$) provides the reference performance. Choosing $N = 2$ complementary motion-compensated signals from $M = 1$ picture, as depicted in Fig. 1, provides an improvement in compression efficiency of up to 1 dB. Permitting frame-adaptive motion compensation with up to $M = 8$ even pictures improves the efficiency up to 3 dB for the investigated test sequences. Please note the relation to the 5/3 wavelet in the case $N = 2$ and $M = 2$: If one of the two motion-compensated signals is always chosen from the previous even picture and the other from the subsequent even picture for all blocks in the odd picture, we obtain the classic motion-compensated 5/3 wavelet as presented in Ref. 3.

Finally, we compare the compression efficiency of the complementary motion-compensated Haar wavelet to that of superimposed motion-compensated predictive coding. Ref. 9 presents a predictive codec for video signals that permits linear combinations of N complementary motion-compensated signals chosen from M reference pictures. This codec is used to provide the experimental results on predictive coding. The comparison is reasonable in the sense that both codecs use identical building blocks but differ in their structure, i.e. motion-compensated transform coding vs. motion-compensated predictive coding.

Figs. 4 and 5 depict the experimental comparison for the sequences *Container Ship* and *Mother & Daughter*. The motion-compensated predictive coding scheme and the motion-compensated transform coding scheme are labeled by MCP and MCT, respectively. For transform coding, a dyadic decomposition is used to encode groups of $K = 32$ pictures. For predictive coding, one intra-picture is followed by 287 inter-pictures. Note that both coding schemes use either $N = 1$ motion-compensated signal chosen from $M = 1$ picture or $N = 2$ complementary motion-compensated signals chosen from $M = 8$ pictures. The comparison between the two coding structures for $N = 1$ motion-compensated signal and $M = 1$ "reference" picture is also discussed in Ref. 4. The experimental comparison of the coding schemes with $N = 2$ complementary motion-compensated signals and $M = 8$ "reference" pictures shows that the compression efficiency for both schemes is comparable at high bit-rates.

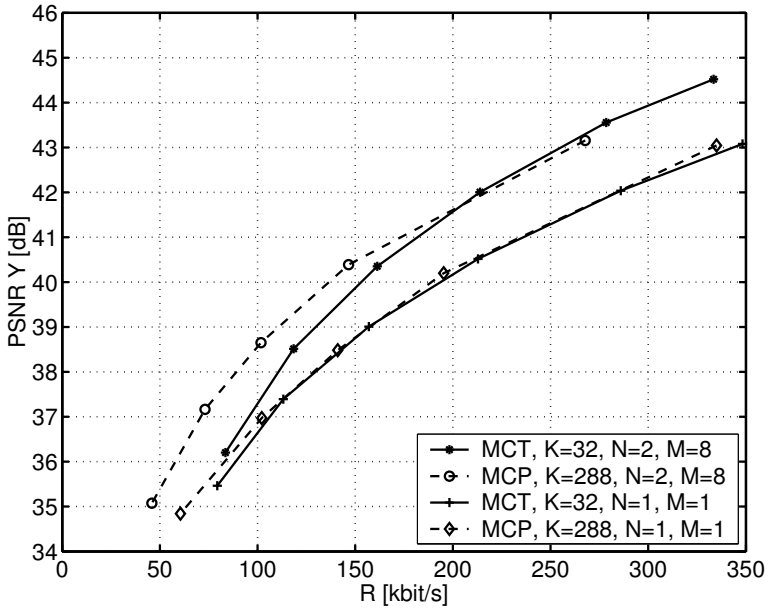

Figure 4. Luminance PSNR vs. total bit-rate for the QCIF sequence *Container Ship* at 30 fps. Transform coding is compared to predictive coding. For transform coding, a dyadic decomposition is used to encode groups of $K = 32$ pictures. For predictive coding, one intra-picture is followed by 287 inter-pictures. Both coding schemes use either $N = 1$ motion-compensated signal chosen from $M = 1$ picture or $N = 2$ complementary motion-compensated signals chosen from $M = 8$ pictures.

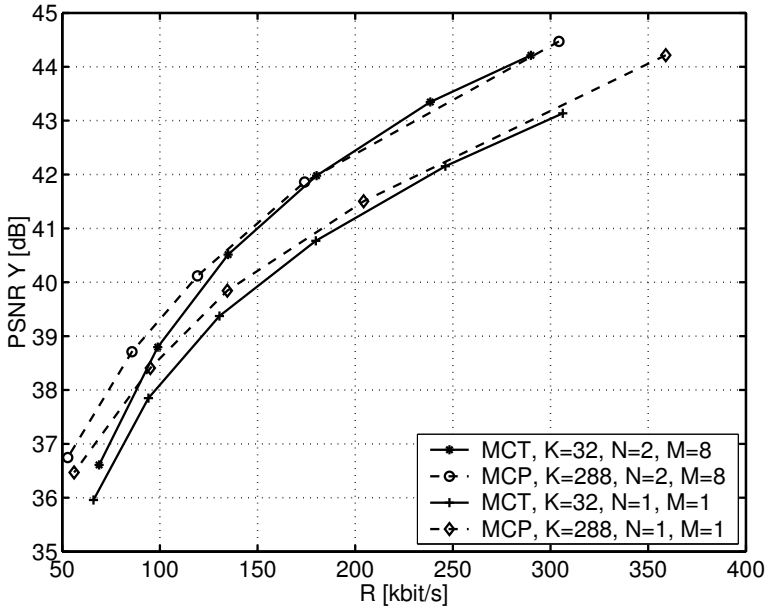

Figure 5. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mother & Daughter* at 30 fps. Transform coding is compared to predictive coding. For transform coding, a dyadic decomposition is used to encode groups of $K = 32$ pictures. For predictive coding, one intra-picture is followed by 287 inter-pictures. Both coding schemes use either $N = 1$ motion-compensated signal chosen from $M = 1$ picture or $N = 2$ complementary motion-compensated signals chosen from $M = 8$ pictures.

3. SIGNAL MODEL FOR HAAR WAVELET WITH COMPLEMENTARY MOTION-COMPENSATED SIGNALS

The theoretical investigation in Ref. 4 models a 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. It is shown that the motion-compensated subband coding scheme can achieve bit-rate savings up to 1 bit per sample and motion-accuracy step when compared to optimum intra-frame coding. But it is also known that predictive coding with linear combinations of complementary motion-compensated signals can achieve bit-rate savings up to 2 bits per sample and motion-accuracy step when compared to optimum intra-frame coding.[7] In the following, we extend the model in Ref. 4 for the special case of the motion-compensated Haar wavelet and utilize superimposed complementary motion-compensated signals.

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 $d = (d_x, d_y)^T$ and re-sample it on the original grid. With this signal model, a spatially constant displacement operation is invertible.

According to Fig. 1, the prediction step averages $N = 2$ signals that are displaced by the vectors $\hat{d}^{(1)}_{2\kappa,2\kappa+1}$ and $\hat{d}^{(2)}_{2\kappa,2\kappa+1}$. Further, we assume that the true displacements are identical for both motion-compensated signals. Given the definition of the true displacement, we obtain for the transfer function of the prediction step $e^{-j\omega^T d_{2\kappa,2\kappa+1}} \frac{1}{2}\left[e^{j\omega^T \Delta^{(1)}_{2\kappa,2\kappa+1}} + e^{j\omega^T \Delta^{(2)}_{2\kappa,2\kappa+1}}\right]$, where $\omega = (\omega_x, \omega_y)^T$ denotes the vector of spatial frequencies. As we assume ideal complementary signals, the variances of both displacement errors are identical and the correlation coefficient is maximal negative $\rho_{\mathbf{\Delta}^{(1)}\mathbf{\Delta}^{(2)}} = -1$.[7] For our model, we achieve this by setting $\Delta^{(2)}_{2\kappa,2\kappa+1} = -\Delta^{(1)}_{2\kappa,2\kappa+1}$ and the transfer function for the prediction step simplifies to $e^{-j\omega^T d_{2\kappa,2\kappa+1}} \cos(\omega^T \Delta_{2\kappa,2\kappa+1})$, as depicted in Fig. 6.

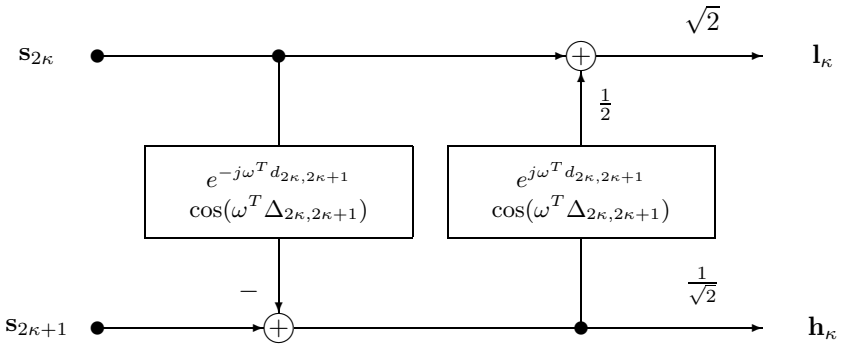

Figure 6. Haar transform with $N = 2$ complementary motion-compensated signals in the lifting steps. The estimated displacements $\hat{d}^{(1)}_{2\kappa,2\kappa+1}$ and $\hat{d}^{(2)}_{2\kappa,2\kappa+1}$ are complementary such that $\hat{d}^{(1)}_{2\kappa,2\kappa+1} = d_{2\kappa,2\kappa+1} - \Delta_{2\kappa,2\kappa+1}$ and $\hat{d}^{(2)}_{2\kappa,2\kappa+1} = d_{2\kappa,2\kappa+1} + \Delta_{2\kappa,2\kappa+1}$. The true displacement between the even and odd frames is $d_{2\kappa,2\kappa+1}$ and the motion-compensated signals are just averaged.

For the signal model, we assume that the pictures $\mathbf{s}_k$ originate from the video signal $\mathbf{v}$. They are degraded by independent additive white Gaussian noise $\mathbf{n}_k$, where the noise signals $\mathbf{n}_\mu$ and $\mathbf{n}_\nu$ are mutually statistically independent. Note that the true displacements in the general signal model and the transform match and cancel out. Therefore, without loss of generality, we assume that the true displacements are zero and investigate the influence of the displacement errors $\Delta_{\mu\nu}$. But we do not consider particular displacement errors $\Delta_{\mu\nu}$. We rather specify statistical properties and consider them to be random variables $\mathbf{\Delta}_{\mu\nu}$, statistically independent from the video signal $\mathbf{v}$ and the noise $\mathbf{n}_k$. Fig. 7 shows the model for the pictures $\mathbf{s}_k$ and the motion-compensated transform $T(\omega, \mathbf{\Delta}_{\mu\nu})$ which is dependent on the displacement errors $\mathbf{\Delta}_{\mu\nu}$. The signals $\mathbf{z}_k$ are independently intra-frame encoded. Note that true displacements which are non-zero do not influence the performance of the optimal intra-frame encoder.

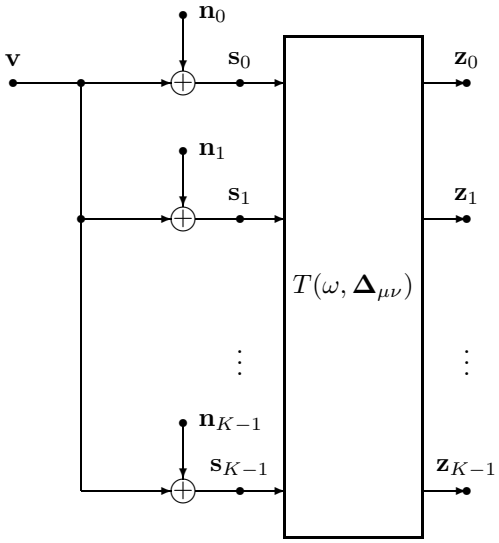

Figure 7. Signal model for motion-compensated transform coding of K pictures.

Now, assume that the random fields $\mathbf{v}$ and $\mathbf{s}_k$ are jointly wide-sense stationary with the real-valued scalar two-dimensional power spectral densities $\Phi_{\mathbf{vv}}(\omega)$ and $\Phi_{\mathbf{s}_\mu \mathbf{s}_\nu}(\omega)$.[4] The power spectral densities of the pictures $\mathbf{s}_k$ are elements in the power spectral density matrix of the input pictures

$$\Phi_{\mathbf{ss}}(\omega) = \begin{pmatrix} 1+\alpha(\omega) & 1 & \cdots & 1 \\ 1 & 1+\alpha(\omega) & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1+\alpha(\omega) \end{pmatrix} \Phi_{\mathbf{vv}}(\omega), \tag{1}$$

where $\alpha(\omega)$ is the normalized spectral density of the noise $\Phi_{\mathbf{n}_k \mathbf{n}_k}$ with respect to the spectral density of the video signal $\mathbf{v}$.[4]

$$\alpha(\omega) = \frac{\Phi_{\mathbf{n}_k \mathbf{n}_k}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \text{for} \quad k = 0, 1, \ldots, K-1 \tag{2}$$

The power spectral density matrix of the decorrelated signal $\Phi_{\mathbf{zz}}$ is given by the transform $T(\omega, \mathbf{\Delta}_{\mu\nu})$,

$$\Phi_{\mathbf{zz}}(\omega) = E\left\{T(\omega, \mathbf{\Delta}_{\mu\nu}) \Phi_{\mathbf{ss}}(\omega) T^H(\omega, \mathbf{\Delta}_{\mu\nu})\right\}, \tag{3}$$

where T^H denotes the Hermitian conjugate of T, and $E\{\cdot\}$ the expectation over the displacement error $\mathbf{\Delta}_{\mu\nu}$. As we intra-frame encode the output pictures $\mathbf{z}_k$ independently, it is sufficient to determine the diagonal elements of (3) with

$$\Phi_{\mathbf{z}_k \mathbf{z}_k}(\omega) = [1+\alpha(\omega)] \sum_{m=0}^{K-1} E\left\{|t_{km}(\omega, \mathbf{\Delta}_{\mu\nu})|^2\right\} + \sum_{\substack{m,n \\ m \neq n}} E\left\{t_{km}(\omega, \mathbf{\Delta}_{\mu\nu}) t^*_{kn}(\omega, \mathbf{\Delta}_{\mu\nu})\right\}, \tag{4}$$

where t_{km} denotes the elements of the transform matrix T. Note that this result is based on the symmetric structure of the power spectral density matrix of the input pictures in (1).

To complete the model, we discuss the properties of the displacements. The true displacements are additive, i.e. $d_{0\mu} + d_{\mu\nu} = d_{0\nu}$, and differ from the estimated displacements by the displacement error, i.e. $d_{\mu\nu} = \hat{d}_{\mu\nu} + \Delta_{\mu\nu}$. We assume that the estimated displacements are also additive such that $\hat{d}_{0\mu} + \hat{d}_{\mu\nu} = \hat{d}_{0\nu}$. Consequently, the displacement errors are additive too.

$$\Delta_{0\mu} + \Delta_{\mu\nu} = \Delta_{0\nu} \tag{5}$$

We assume further, that there is no preference among the input pictures, that is, the order of the input pictures has no influence on the coding efficiency. In addition, the accuracy of motion compensation is the same for all pictures such that the variances of the displacement error are invariant $E\left\{\mathbf{\Delta}_{\mu\nu}^2\right\} = \sigma_{\mathbf{\Delta}}^2$. All displacement errors $\mathbf{\Delta}_{\mu\nu}$ are jointly 2-D normal distributed with variance $\sigma_{\mathbf{\Delta}}^2$ and zero mean, where the x- and y-components are statistically independent. Note that the variance constraint in combination with the additivity of displacement errors in (5) results in correlated displacement errors; for example, $\rho_{\mathbf{\Delta}_{0\mu}\mathbf{\Delta}_{0\nu}} = \frac{1}{2}$.

3.1. Example for $K = 2$

As an example, we determine the power spectral densities $\Phi_{\mathbf{z}_k\mathbf{z}_k}$ of the output pictures for groups of $K = 2$ pictures. We simplify the transform T in Fig. 6 and assume that $|\omega^T\Delta_{\mu\nu}| << \frac{\pi}{2}$ such that we are able to approximate $2-\cos^2(\omega^T\Delta_{\mu\nu}) \approx 1$. This approximation holds in particular for very accurate motion compensation.

$$\begin{pmatrix} Z_0(\omega,\Delta_{01}) \\ Z_1(\omega,\Delta_{01}) \end{pmatrix} \approx \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \cos(\omega^T\Delta_{01}) \\ -\cos(\omega^T\Delta_{01}) & 1 \end{pmatrix} \begin{pmatrix} S_0(\omega) \\ S_1(\omega) \end{pmatrix} \tag{6}$$

We obtain for the power spectral densities of the output pictures the approximations

$$\Phi_{\mathbf{z}_1\mathbf{z}_1}(\omega) \approx [1+\alpha(\omega)]\left[\frac{3}{4}+\frac{1}{4}P(\omega,4\sigma_{\mathbf{\Delta}}^2)\right] + P(\omega,\sigma_{\mathbf{\Delta}}^2) \quad \text{and} \tag{7}$$

$$\Phi_{\mathbf{z}_2\mathbf{z}_2}(\omega) \approx [1+\alpha(\omega)]\left[\frac{3}{4}+\frac{1}{4}P(\omega,4\sigma_{\mathbf{\Delta}}^2)\right] - P(\omega,\sigma_{\mathbf{\Delta}}^2), \tag{8}$$

where $P(\omega,\sigma_{\mathbf{\Delta}}^2) := E\left\{e^{-j\omega^T\mathbf{\Delta}_{\mu\nu}}\right\} = e^{-\frac{1}{2}\omega^T\omega\sigma_{\mathbf{\Delta}}^2}$ denotes the characteristic function of the displacement error with variance $\sigma_{\mathbf{\Delta}}^2$.

3.2. Transform Coding Gain

The rate difference[6] is used to measure the improved compression efficiency for each picture k.

$$\Delta R_k = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log_2\left(\frac{\Phi_{\mathbf{z}_k\mathbf{z}_k}(\omega)}{\Phi_{\mathbf{s}_k\mathbf{s}_k}(\omega)}\right)d\omega \tag{9}$$

It represents the maximum bit-rate reduction (in bit per sample) possible by optimum encoding of the transformed signal $\mathbf{z}_k$, compared to optimum intra-frame encoding of the signal $\mathbf{s}_k$ 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.[6] The overall rate difference ΔR is the average over K pictures and is used to evaluate the efficiency of transform coding. For example, the complementary motion-compensated Haar wavelet with $K = 2$ achieves an overall rate difference of approximately

$$\Delta R_{K=2} \approx \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{4}\log_2\left(\frac{3}{4}+\frac{1}{4}P(\omega,4\sigma_{\mathbf{\Delta}}^2)+\frac{P(\omega,\sigma_{\mathbf{\Delta}}^2)}{1+\alpha(\omega)}\right)+\frac{1}{4}\log_2\left(\frac{3}{4}+\frac{1}{4}P(\omega,4\sigma_{\mathbf{\Delta}}^2)-\frac{P(\omega,\sigma_{\mathbf{\Delta}}^2)}{1+\alpha(\omega)}\right)d\omega. \tag{10}$$

Figs. 8 and 9 depict the overall rate difference for the Haar wavelet with motion compensation ($N = 1$) and $N = 2$ complementary motion-compensated signals 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. The variance of the video signal $\mathbf{v}$ is normalized to $\sigma_{\mathbf{v}}^2 = 1$ and the results for $N = 1$ are taken from Ref. 4. The bounds for the complementary motion-compensated Haar wavelet ($N = 2$) are obtained by evaluating (4) with the approximation in Sec. 3.1 for $K = 2, 4$, and 8. Note that the complementary motion-compensated Haar wavelet ($N = 2$) achieves a rate difference up to 1 bit per sample and motion-accuracy step already for a GOP size of $K = 2$, whereas for $N = 1$, a very large GOP size is required to achieve this slope.[4] On the other hand, the results for $N = 2$ suggest that a rate difference of up to 2 bits per sample and motion-accuracy step is feasible for a very large GOP size. In the presence of residual noise, both approaches with the same GOP size K converge for very accurate motion compensation to the same rate difference as we assume identical residual noise levels.

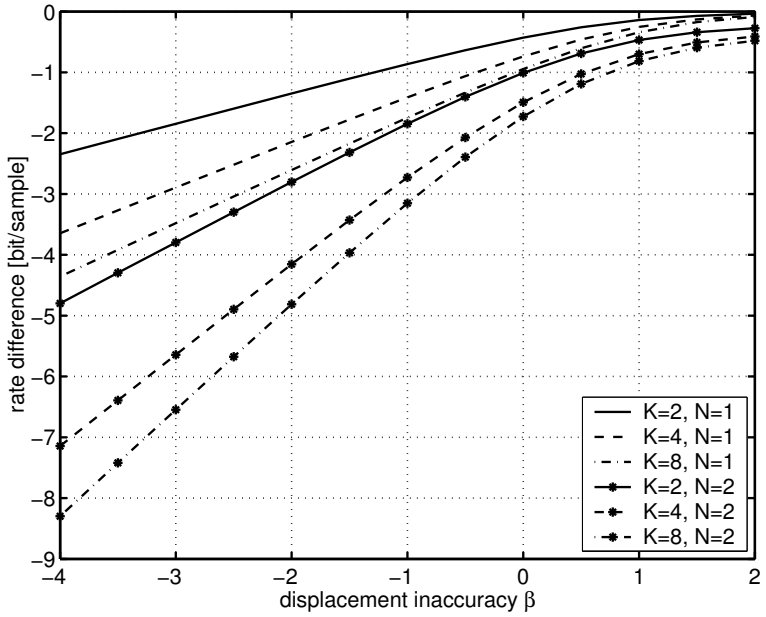

Figure 8. Theoretical bounds for the Haar wavelet with motion compensation ($N = 1$) and $N = 2$ complementary motion-compensated signals. RNL = -100 dB.

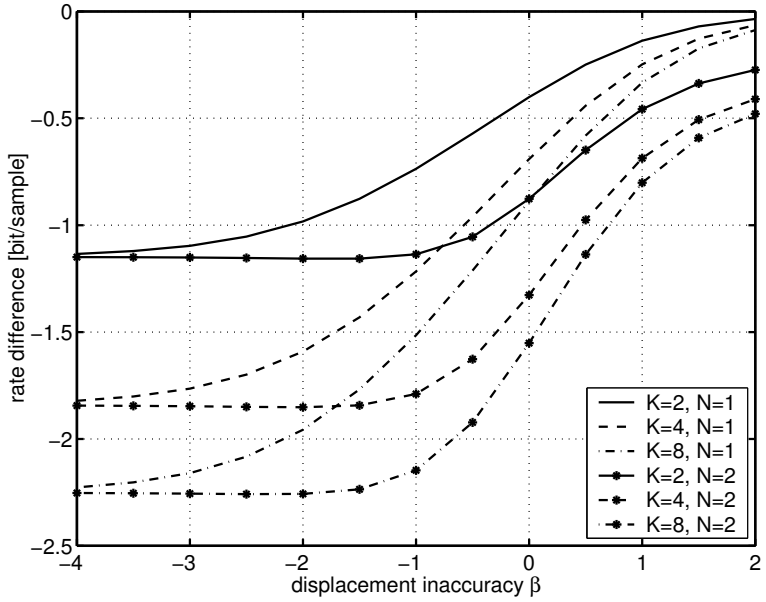

Figure 9. Theoretical bounds for the Haar wavelet with motion compensation ($N = 1$) and $N = 2$ complementary motion-compensated signals. RNL = -30 dB.

3.3. Comparison to Superimposed Predictive Coding

As already pointed out in Ref 4, motion-compensated transform coding without complementary signals ($N = 1$) achieves a rate difference up to 1 bit per sample and motion-accuracy step. A similar result is obtained for motion-compensated predictive coding. But superimposed predictive coding with complementary motion-compensated signals is known to achieve a rate difference up to 2 bits per sample and motion-accuracy step.[7] The following comparison to superimposed predictive coding considers the case of complementary motion-compensated signals.

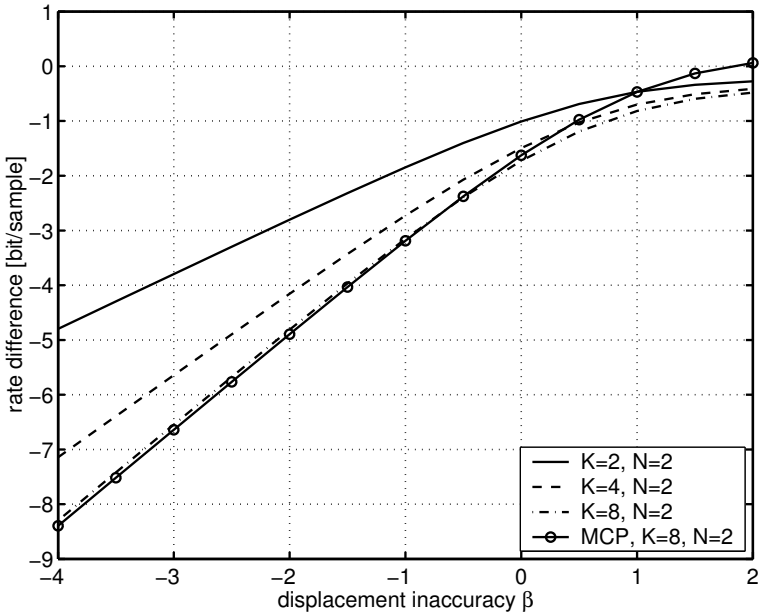

Figure 10. Theoretical bounds for the Haar wavelet and superimposed prediction with $N = 2$ complementary motion-compensated signals. The complementary signals are just averaged. RNL = -100 dB.

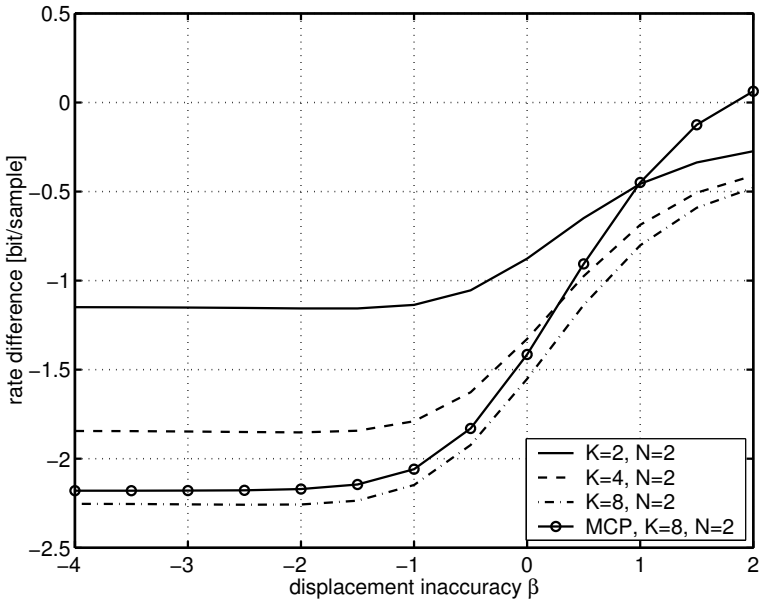

Figure 11. Theoretical bounds for the Haar wavelet and superimposed prediction with $N = 2$ complementary motion-compensated signals. The complementary signals are just averaged. RNL = -30 dB.

Figs. 10 and 11 depict the overall rate difference for the complementary motion-compensated Haar transform ($N = 2$) with $K = 2, 4$, and 8 over the displacement inaccuracy β for a residual noise level of -100 dB and -30 dB, respectively. In addition, the rate difference of superimposed predictive coding with $N = 2$ complementary motion-compensated signals and a GOP size of $K = 8$ is added to the plots.[7] In both cases, the complementary signals are just averaged. In the noise-free case, the complementary motion-compensated Haar transform ($N = 2$) with a GOP size of $K = 8$ achieves a slope up to $\frac{7}{8} \cdot 2$ bits per sample and motion-accuracy step, similar to that of

superimposed predictive coding with a GOP size of $K = 8$. For superimposed predictive coding, the maximum slope of 2 bits per sample and motion-accuracy step is reached for a very large GOP size. If just K pictures are encoded, we have one out of K pictures that is intra-encoded. The overall rate difference is the average over all K pictures and is at most $\frac{K-1}{K} \cdot 2$ bits per sample and motion-accuracy step. Finally, in the presence of residual noise and a GOP size of $K = 8$, the complementary motion-compensated Haar transform ($N = 2$) performs slightly better than superimposed predictive coding with $N = 2$ complementary signals.

4. CONCLUSIONS

We investigate experimentally and theoretically how complementary motion-compensated signals improve the efficiency of inter-frame wavelet coding of video signals. We modify the classic motion-compensated Haar wavelet and incorporate up to $N = 2$ complementary motion-compensated signals that are just averaged. The experimental results show that complementary signals improve the compression efficiency of motion-compensated transform coding. When combined with frame-adaptive motion compensation, gains up to 3 dB can be observed for our coding scheme. Further, we discuss a signal model for the motion-compensated Haar wavelet with complementary signals and determine bounds for the transform coding gain. The theoretical results suggest that the Haar wavelet coding scheme with $N = 2$ complementary motion-compensated signals achieves an overall rate difference up to 2 bits per sample and motion-accuracy step for increasing GOP size. We observe that this theoretical performance is comparable to that of superimposed predictive coding with $N = 2$ complementary motion-compensated signals.

REFERENCES

1. W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis* **29**(2), pp. 511–546, 1998.
2. B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, **3**, pp. 1793–1796, (Salt Lake City, UT), May 2001.
3. A. Secker and 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*, **2**, pp. 1029–1032, (Thessaloniki, Greece), Oct. 2001.
4. M. Flierl and B. Girod, "Investigation of motion-compensated lifted wavelet transforms," in *Proceedings of the Picture Coding Symposium*, pp. 59–62, (Saint-Malo, France), Apr. 2003.
5. G. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, **5**, pp. 437–440, (Minneapolis, MN), Apr. 1993.
6. B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Transactions on Image Processing* **9**, pp. 173–183, Feb. 2000.
7. M. Flierl and B. Girod, "Multihypothesis motion estimation for video coding," in *Proceedings of the Data Compression Conference*, pp. 341–350, (Snowbird, Utah), Mar. 2001.
8. M. Flierl, "Video coding with lifted wavelet transforms and frame-adaptive motion compensation," in *Visual Content Processing and Representation*, N. García, J. Martínez, and L. Salgado, eds., *Lecture Notes in Computer Science* **2849**, pp. 243–251, Springer-Verlag, Berlin, 2003. Proceedings of the 8th International Workshop VLBV, Madrid, Spain, Sept. 2003.
9. M. Flierl, T. Wiegand, and B. Girod, "Rate-constrained multihypothesis prediction for motion compensated video compression," *IEEE Transactions on Circuits and Systems for Video Technology* **12**, pp. 957–969, Nov. 2002.