

# A Double Motion-Compensated Orthogonal Transform with Energy Concentration Constraint

Markus Flierl and Bernd Girod

Max Planck Center for Visual Computing and Communication  
Stanford University, Stanford, CA 94305

## ABSTRACT

This paper discusses a transform for successive pictures of an image sequence which strictly maintains orthogonality while permitting 2-hypothesis motion compensation between pairs of pictures. We extend previous work on an orthogonal transform for image sequences which uses only 1-hypothesis motion compensation between pairs of pictures. This work is motivated by the well known fact that the motion-compensated lifted Haar wavelet maintains orthogonality only approximately. In the case of zero motion fields, the motion-compensated lifted Haar wavelet is known to be orthogonal. But for complex motion fields with many multi-connected and unconnected pixels, the motion-compensated lifted Haar wavelet cannot accurately maintain its transform property and, hence, suffers a performance degradation. The presented double motion-compensated orthogonal transform strictly maintains orthogonality for any motion field. For the transform, each pixel in the high-band is compensated by a linear combination of two motion-compensated pixels chosen from the corresponding low-band. That is, each pixel in the high-band is associated with two motion vectors. This is in contrast to the previously presented single motion-compensated orthogonal transform where each pixel in the high-band is compensated by only one motion-compensated pixel chosen from the corresponding low-band. In terms of energy concentration, the double motion-compensated orthogonal transform outperforms the single motion-compensated orthogonal transform and compares favorably with the double motion-compensated lifted Haar wavelet.

**Keywords:** Temporal subband coding of video, motion-compensated orthogonal video transforms, complementary motion-compensated signals.

## 1. INTRODUCTION

We address the problem of representing image sequences for coding and communication applications. Well known methods are standard hybrid video coding techniques as well as subband coding schemes. To achieve high compression efficiency, standard hybrid video encoders operate in a closed-loop fashion such that the total distortion across the reconstructed pictures equals the total distortion in the corresponding intra picture and encoded displaced frame differences. In case of channel errors, decoded reference frames differ from the optimized reference frames at the encoder and error propagation is observed. On the other hand, transform coding schemes operate in an open-loop fashion. Consider high-rate transform coding schemes in which the analysis transform produces independent transform coefficients. With uniform quantization, these schemes are optimal when utilizing an orthogonal transform.<sup>1</sup> Further, energy conservation holds for orthogonal transforms such that the total quantization distortion in the coefficient domain equals that in the image domain. In case of channel errors, the error energy in the image domain equals that in the coefficient domain. Hence, the error energy is preserved in the image domain and is not amplified by the decoder, as is the case, e.g., for predictive decoders.

During the last decade, there have been attempts to incorporate motion compensation into temporal subband coding schemes by approaching problems arising from multi-connected pixels. Ref. 2 proposes to distinguish between connected, covered, and uncovered pixels when incorporating motion compensation for filtering in temporal direction. Motion-compensated filtering in Ref. 3 addresses the problem of double-connected pixels and proposes a method to resolve the ambiguity. Refs. 4–6 choose a lifting implementation for the temporal filter and incorporate motion compensation into the lifting steps. In contrast to the early work, the lifting implementation permits a reversible filter structure, but still, it struggles with unconnected, connected, and

---

Contact e-mail: [mflerl@stanford.edu](mailto:mflerl@stanford.edu)

multi-connected pixels when performing the update step. To minimize the impact on the reconstructed images, Ref. 7 proposes an optimum update step that minimizes the mean-squared reconstruction error. But note that the motion-compensated lifted Haar wavelet maintains orthogonality only approximately. For a zero motion field, the motion-compensated lifted Haar wavelet is known to be orthogonal. But for complex motion fields with many multi-connected and unconnected pixels, the motion-compensated lifted Haar wavelet cannot accurately maintain its transform property.

We present in Ref. 8 a motion-compensated orthogonal transform which strictly maintains orthogonality for any motion field. The transform is factored into a sequence of incremental transforms that are strictly orthogonal. The incremental transforms maintain scale counters to keep track of the scale factors that are introduced to ensure orthogonality. The decorrelation factor of each incremental transform is determined by the scale counters and is chosen such that the transform meets an energy concentration constraint. The experiments show that this orthogonal transform offers an improved energy compaction when compared to motion-compensated lifted Haar wavelets and closed-loop hierarchical P pictures. But note that the transform in Ref. 8 compensates each pixel in the high-band by only one motion-compensated pixel chosen from the corresponding low-band. That is, each pixel in the high-band is associated with only one motion vector. Hence, it will be called a single motion-compensated orthogonal transform.

It is well known that a linear combination of multiple motion-compensated signals improves the accuracy of motion compensation.<sup>9</sup> For example, 2-hypothesis motion compensation is widely used and the bi-predictive mode of H.264/AVC is a popular example. This paper discusses an extension of the work in Ref. 8 by permitting up to two hypotheses per high-band pixel. In particular, each pixel in the high-band is compensated by a linear combination of two motion-compensated pixels chosen from the corresponding low-band. That is, each pixel in the high-band is associated with two motion vectors. Hence, it will be called a double motion-compensated orthogonal transform. If we limit the pair of motion vectors to address directly neighboring pixels in the corresponding low-band, we achieve sub-pixel accurate motion compensation. But if we choose optimal pairs of complementary motion vectors, we perform the generic 2-hypothesis motion compensation.

In contrast to our previous work in Ref. 8 where only single motion compensation between two pictures is considered, the here presented double motion-compensated orthogonal transform is able to consider up to two motion fields per picture. Similar to our previous work, we factor the transform into a sequence of incremental transforms that are strictly orthogonal and that maintain scale counters that are compatible with the scale counters in Ref. 8. The decorrelation factors of each incremental transform are determined such that an energy concentration constraint for double motion compensation is met. In the experiments, we compare the double motion-compensated orthogonal transform to the double motion-compensated lifted Haar wavelet in Ref. 10. We choose this adaptive lifted Haar wavelet as it permits also generic 2-hypothesis motion compensation. This enables us to compare the two transforms while utilizing the same motion fields.

The paper is organized as follows: Section 2 introduces the double motion-compensated orthogonal transform and discusses its incremental transform as well as the energy concentration constraint. Section 3 presents the experimental results with the double motion-compensated orthogonal transform and lifted Haar wavelet.

## 2. DOUBLE MOTION-COMPENSATED ORTHOGONAL TRANSFORM

This section discusses how the transform is factored into incremental transforms. We outline the construction of the incremental transform and the incorporation of the energy concentration constraint.

Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be two vectors representing consecutive pictures of an image sequence. The transform  $T$  maps these vectors according to

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = T \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \quad (1)$$

into two vectors  $\mathbf{y}_1$  and  $\mathbf{y}_2$  which represent the temporal low- and high-band, respectively. Now, we factor the transform  $T$  into a sequence of  $k$  incremental transforms  $T_\kappa$  such that

$$T = T_k T_{k-1} \cdots T_\kappa \cdots T_2 T_1, \quad (2)$$

where each incremental transform  $T_\kappa$  is orthogonal by itself, i.e.,  $T_\kappa T_\kappa^T = I$  holds for all  $\kappa = 1, 2, \dots, k$ , where  $I$  denotes the identity matrix. This guarantees that the transform  $T$  is also orthogonal.

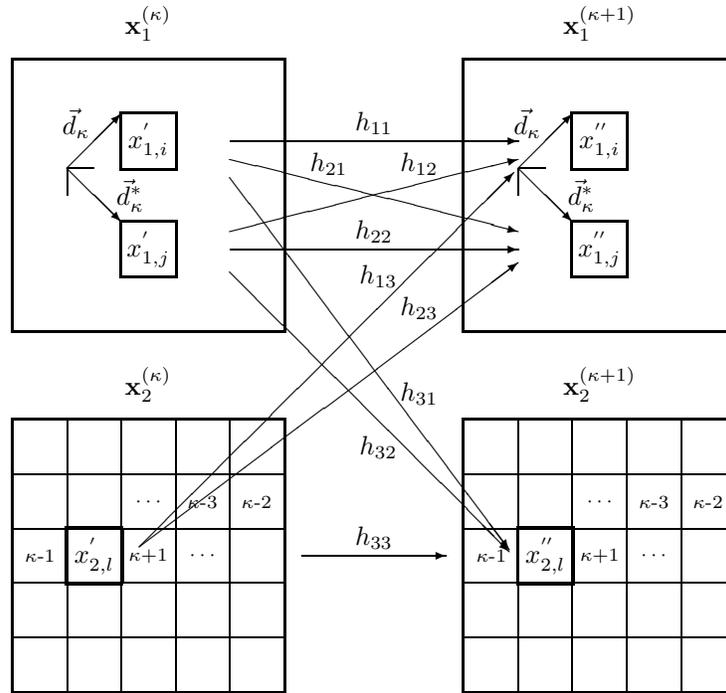
It can be imagined that the pixels of the image  $\mathbf{x}_2$  are processed from top-left to bottom-right in  $k$  steps where each step  $\kappa$  is represented by the incremental transform  $T_\kappa$ .

### 2.1. Incremental Transform

Let  $\mathbf{x}_1^{(\kappa)}$  and  $\mathbf{x}_2^{(\kappa)}$  be two vectors representing consecutive pictures of an image sequence if  $\kappa = 1$ , or two output vectors of the incremental transform  $T_{\kappa-1}$  if  $\kappa > 1$ . The incremental transform  $T_\kappa$  maps these vectors according to

$$\begin{pmatrix} \mathbf{x}_1^{(\kappa+1)} \\ \mathbf{x}_2^{(\kappa+1)} \end{pmatrix} = T_\kappa \begin{pmatrix} \mathbf{x}_1^{(\kappa)} \\ \mathbf{x}_2^{(\kappa)} \end{pmatrix} \quad (3)$$

into two vectors  $\mathbf{x}_1^{(\kappa+1)}$  and  $\mathbf{x}_2^{(\kappa+1)}$  which will be further transformed into the temporal low- and high-band, respectively.



**Figure 1.** The incremental transform  $T_\kappa$  for two frames  $\mathbf{x}_1^{(\kappa)}$  and  $\mathbf{x}_2^{(\kappa)}$  which strictly maintains orthogonality for any 2-motion field ( $\vec{d}_\kappa, \vec{d}_\kappa^*$ ) between the two frames.

Fig. 1 depicts the process accomplished by the incremental transform  $T_\kappa$  with its input and output images as defined above. The incremental transform removes the energy of the  $l$ -th pixel  $x'_{2,l}$  in the image  $\mathbf{x}_2^{(\kappa)}$  with the help of both the  $i$ -th pixel  $x'_{1,i}$  in the image  $\mathbf{x}_1^{(\kappa)}$  which is linked by the motion vector  $\vec{d}_\kappa$  and the  $j$ -th pixel  $x'_{1,j}$  in the image  $\mathbf{x}_1^{(\kappa)}$  which is linked by the motion vector  $\vec{d}_\kappa^*$  (or the  $l$ -th block with the help of both the  $i$ -th and the  $j$ -th block if the blocks in image  $\mathbf{x}_1^{(\kappa)}$  do not overlap and if all the pixels of the  $i$ -th and  $j$ -th block have the motion vectors  $\vec{d}_\kappa$  and  $\vec{d}_\kappa^*$ , respectively). The energy-removed pixel value  $x''_{2,l}$  is obtained by a linear combination of the pixel values  $x'_{1,i}$ ,  $x'_{1,j}$ , and  $x'_{2,l}$  with scalar weights  $h_{31}$ ,  $h_{32}$ , and  $h_{33}$ . The energy concentrated pixel value  $x''_{1,i}$  is also obtained by a linear combination of the pixel values  $x'_{1,i}$ ,  $x'_{1,j}$ , and  $x'_{2,l}$  but with scalar weights  $h_{11}$ ,  $h_{12}$ , and  $h_{13}$ . The energy concentrated pixel value  $x''_{1,j}$  is calculated accordingly. All other pixels are simply kept untouched.

To summarize, the incremental transform  $T_\kappa$  touches only pixels that are linked by the same motion vector pair  $(\vec{d}_\kappa, \vec{d}_\kappa^*)$ . Of these,  $T_\kappa$  performs only a linear combination with three pixels that are connected by this motion vector pair. All other pixels are kept untouched. This is reflected in the following matrix notation:

$$T_\kappa = \begin{pmatrix} \ddots & \vdots \\ \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & h_{11} & 0 & \cdots & 0 & h_{12} & 0 & \cdots & 0 & h_{13} & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & h_{21} & 0 & \cdots & 0 & h_{22} & 0 & \cdots & 0 & h_{23} & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & h_{31} & 0 & \cdots & 0 & h_{32} & 0 & \cdots & 0 & h_{33} & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots \\ \vdots & \ddots \end{pmatrix} \quad (4)$$

The diagonal elements equal to 1 represent the untouched pixels and the elements  $h_{\mu\nu}$  represent the pixels subject to linear operations. All other entries are zero.

Now, the scalar weights  $h_{\mu\nu}$  are arranged into the  $3 \times 3$  matrix  $H$ . The incremental transform  $T_\kappa$  is orthogonal if  $H$  is also orthogonal.

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (5)$$

We construct an orthogonal  $H$  with the help of Euler's rotation theorem which states that any rotation can be given as a composition of rotations about three axes, i.e.  $H = H_3 H_2 H_1$ , where  $H_r$  denotes a rotation about one axes. We choose the composition

$$H = \begin{pmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

with the Euler angles  $\psi$ ,  $\theta$ , and  $\phi$ . The Euler angles will be determined in the next subsection which discusses the energy concentration constraint.

Note that, to carry out the full transform  $T$ , each pixel in  $\mathbf{x}_2$  is touched only once whereas the pixels in  $\mathbf{x}_1$  may be touched multiple times or never. Further, the order in which the incremental transforms  $T_\kappa$  are applied does not affect the orthogonality of  $T$ , but it may affect the energy concentration of the transform  $T$ .

## 2.2. Energy Concentration Constraint

The three Euler angles for each pixel touched by the incremental transform have to be chosen such that the energy in image  $\mathbf{x}_2$  is minimized. Consider the pixel triplet  $x_{1,i}$ ,  $x_{1,j}$ , and  $x_{2,l}$  to be processed by the incremental transform  $T_\kappa$ . To determine the Euler angles for the pixel  $x_{2,l}$ , we assume that the pixel  $x_{2,l}$  is connected to the pixels  $x_{1,i}$  and  $x_{1,j}$  such that  $x_{2,l} = x_{1,i} = x_{1,j}$ . Consequently, the resulting high-band pixel  $x''_{2,l}$  shall be zero. Note that the pixels  $x_{1,i}$  and  $x_{1,j}$  may have been processed previously by  $T_\tau$ , where  $\tau < \kappa$ . Therefore, let  $v_1$  and  $v_2$  be the *scale factors* for the pixels  $x_{1,i}$  and  $x_{1,j}$ , respectively, such that  $x'_{1,i} = v_1 x_{1,i}$  and  $x'_{1,j} = v_2 x_{1,j}$ . The pixel  $x_{2,l}$  is used only once during the transform process  $T$  and no scale factor needs to be considered. But in general, when considering subsequent dyadic decompositions with  $T$ , scale factors are passed on to higher decomposition levels and, consequently, they need to be considered, i.e.,  $x'_{2,l} = v_3 x_{2,l}$ . Obviously, for the first

decomposition level,  $v_3 = 1$ . Let  $u_1$  and  $u_2$  be the scale factors for the pixels  $x_{1,i}$  and  $x_{1,j}$ , respectively, after they have been processed by  $T_\kappa$ . Now, the pixels  $x'_{1,i}$ ,  $x'_{1,j}$ , and  $x'_{2,l}$  are processed by  $T_\kappa$  as follows:

$$\begin{pmatrix} u_1 x_{1,i} \\ u_2 x_{1,i} \\ 0 \end{pmatrix} = H_3 H_2 H_1 \begin{pmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \\ v_3 x_{1,i} \end{pmatrix} \quad (7)$$

Energy conservation requires that

$$u_1^2 + u_2^2 = v_1^2 + v_2^2 + v_3^2. \quad (8)$$

The Euler angle  $\phi$  in  $H_1$  is chosen such that the two hypotheses  $x'_{1,i}$  and  $x'_{1,j}$  are weighted equally after being attenuated by their scale factors  $v_1$  and  $v_2$ .

$$\tan(\phi) = -\frac{v_1}{v_2} \quad (9)$$

The Euler angle  $\theta$  in  $H_2$  is chosen such that it meets the zero-energy constraint for the high-band in (7).

$$\tan(\theta) = \frac{v_3}{\sqrt{v_1^2 + v_2^2}} \quad (10)$$

Finally, the Euler angle  $\psi$  in  $H_3$  is chosen such that the pixels  $x_{1,i}$  and  $x_{1,j}$ , after the incremental transform  $T_\kappa$ , have scalar weights  $u_1$  and  $u_2$ , respectively.

$$\tan(\psi) = \frac{u_1}{u_2} \quad (11)$$

But note that we are free to choose this ratio. We have chosen the Euler angle  $\phi$  such that the  $i$ -th pixel  $x_{1,i}$  and the  $j$ -th pixel  $x_{1,j}$  have equal contribution after rescaling with  $v_1$  and  $v_2$ . Consequently, we choose the scale factors  $u_1$  and  $u_2$  such that they increase equally.

$$u_1 = \sqrt{v_1^2 + \frac{v_3^2}{2}} \quad \text{and} \quad u_2 = \sqrt{v_2^2 + \frac{v_3^2}{2}} \quad (12)$$

Similar to the work in Ref. 8, we utilize *scale counters* to keep track of the scale factors. Scale counters simply count how often a pixel is used as reference for motion compensation. Before any transform is applied, the scale counter for each pixel is  $n = 0$  and the scale factor is  $v = 1$ . For arbitrary scale counter  $n$  and  $m$ , the scale factors are

$$v = \sqrt{n+1} \quad \text{and} \quad u = \sqrt{m+1}. \quad (13)$$

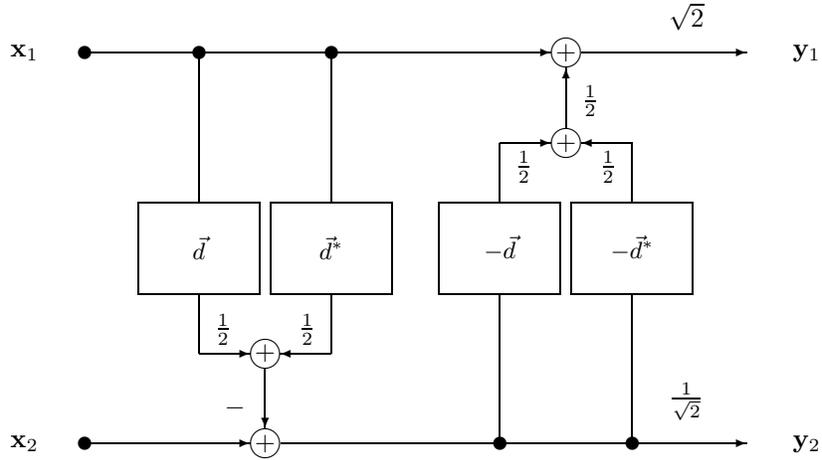
After applying the incremental transform, the scale counter have to be updated for the modified pixels. For the single motion-compensated orthogonal transform in Ref. 8, the updated scale counter for low-band pixels is given by  $m_1 = n_1 + n_2 + 1$ , where  $n_1$  and  $n_2$  are the scale counters of the utilized input pixel pairs. For the double motion-compensated orthogonal transform, the updated scale counters for low-band pixels result from (12) as follows:

$$m_1 = n_1 + \frac{n_3 + 1}{2} \quad \text{and} \quad m_2 = n_2 + \frac{n_3 + 1}{2} \quad (14)$$

For example, consider the transform in the first decomposition level where  $n_3 = 0$ . The single motion-compensated transform increases the scale counter by 1 for each used reference pixel, whereas the double motion-compensated transform increases the counter by 0.5 for each of the two used reference pixels.

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

Experimental results assessing the energy compaction are obtained for the QCIF sequences *Bus*, *Crew*, *Soccer*, *Mother & Daughter*, and *City*. Our scheme with the double motion-compensated orthogonal transform is compared to schemes which use a double motion-compensated lifted Haar wavelet<sup>10</sup> with and without update step. In addition, the performance of the single motion-compensated orthogonal transform<sup>8</sup> is reported. Note that we choose the double motion-compensated lifted Haar wavelet as depicted in Fig. 2 as it permits also generic



**Figure 2.** Double motion-compensated lifted Haar wavelet<sup>10</sup> with input pictures  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . Each lifting step averages two motion-compensated signals which are generated by the 2-motion field  $(\vec{d}, \vec{d}^*)$ .

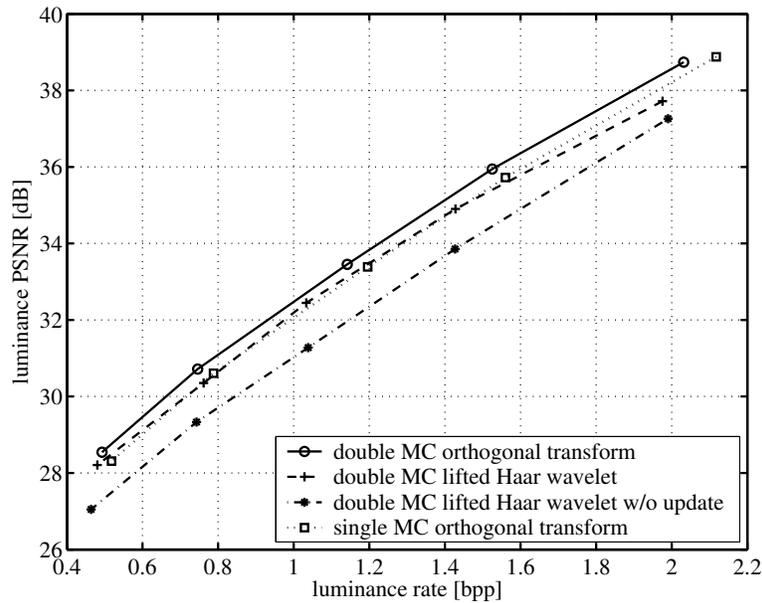
2-hypothesis motion compensation. This enables us to compare the two double motion-compensated transforms while utilizing the same motion fields.

For the lifted Haar wavelet in Fig. 2, the prediction step averages two motion-compensated signals generated by the integer-pixel accurate 2-motion field  $(\vec{d}, \vec{d}^*)$ . The update step performs the same operations with the inverse motion field. Hence, each pixel that is used as reference pixel in the prediction step receives its corresponding update signal in the update step. We achieve sub-pixel accurate motion compensation by averaging the two integer-pixel accurate signals. For conventional sub-pixel accurate motion compensation, spatial filtering is used to determine sub-pixel intensity values. If spatial filtering is used in the prediction step, the required corresponding update step is difficult to achieve. The scheme in Fig. 2 does not require conventional spatial filtering to achieve sub-pixel accurate motion compensation.

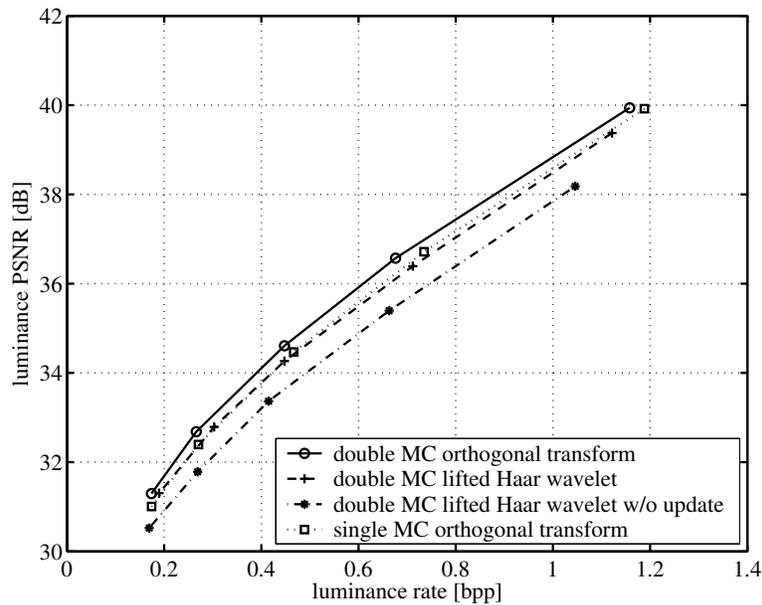
For the coding process with the orthogonal transforms, a scale counter  $n$  is maintained for every pixel of each picture. The scale counters are an immediate results of the utilized motion vectors and are only required for the processing at encoder and decoder. The scale counters do not have to be encoded as they can be recovered from the motion vectors.

All schemes operate with a GOP size of 16 frames. Sub-pixel accurate motion compensation is achieved by averaging two motion-compensated signals resulting from integer-pixel accurate 2-motion fields. The block size for motion compensation is limited to  $8 \times 8$ . Conditional motion estimation is used for 2-motion estimation. In particular, the first motion vector is obtained by block matching. Given the first motion vector, conditional motion estimation optimizes only the second motion vector within a conditional search window of  $\pm 5$ , centered at the location of the first motion vector. For simplicity, the first motion vector is not refined conditionally. If a 2-motion is not efficient for a block, the single motion-compensated orthogonal transform and lifted Haar wavelet are used. Further, the resulting temporal subbands are simply coded with JPEG 2000. The temporal high-bands are coded directly, whereas the temporal low-band is rescaled with (13) before encoding. For optimal rate allocation, Lagrangian costs are determined where the distortion term considers the scale factors applied to the temporal low-band.

Figs. 3, 4, 5, 6, and 7 depict the rate distortion performances for the luminance signals of the test sequences. Results for the double motion-compensated (MC) orthogonal transform as well as the double MC lifted Haar wavelet with and without update step<sup>10</sup> are given. Note that all three schemes utilize the same motion vectors. In order to assess the energy compaction, no intra modes have been used for all temporal coding schemes. For *Bus*, *Crew*, and *Soccer*, the double MC orthogonal transform outperforms the double MC lifted Haar wavelet with and without update step. For *Mother & Daughter* and *City*, the double MC orthogonal transform performs similar to the double MC lifted Haar wavelet with update step, but outperforms the double MC lifted Haar wavelet without update step. For the comparison with the single motion-compensated orthogonal transform,<sup>8</sup>



**Figure 3.** PSNR over bit rate for the luminance signal of the QCIF sequence *Bus* at 15 fps with 64 frames.



**Figure 4.** PSNR over bit rate for the luminance signal of the QCIF sequence *Crew* at 15 fps with 64 frames.

the second motion vector of the 2-motion is simply not used. The saved bit rate is not considered. For a direct comparison with the single MC orthogonal transform, a rate-constrained transform selection is required. This is possible, but currently not implemented.

For *Bus*, *Crew*, and *Soccer*, the significant motion in the sequences degrades the performance of the double MC lifted Haar wavelet as the update step introduces additional noise. For *Mother & Daughter* and *City*, the double MC orthogonal transform performs similar to the double MC lifted Haar wavelet with update step as the weak/regular motion in the sequence does not substantially harm the lifted Haar wavelet.

Figs. 8 and 9 show the luminance PSNR over the frame number for the sequences *Bus* and *City*, respectively.

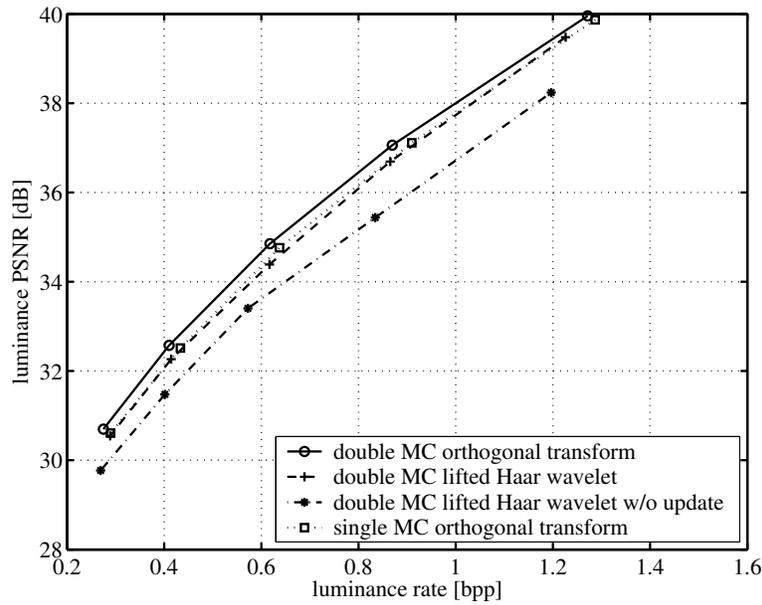


Figure 5. PSNR over bit rate for the luminance signal of the QCIF sequence *Soccer* at 15 fps with 64 frames.

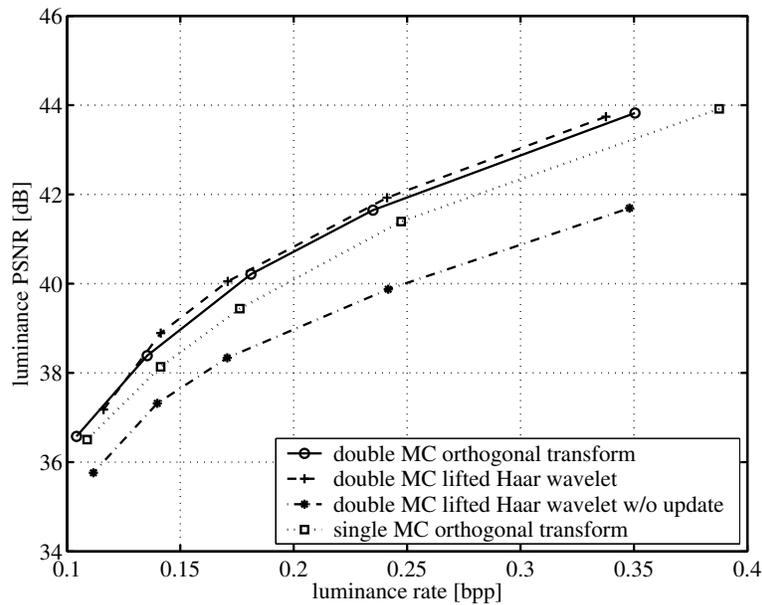
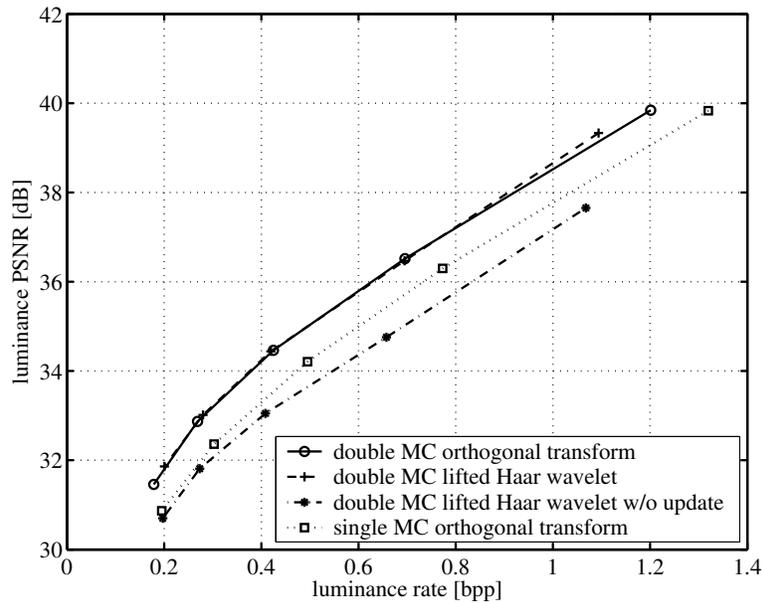


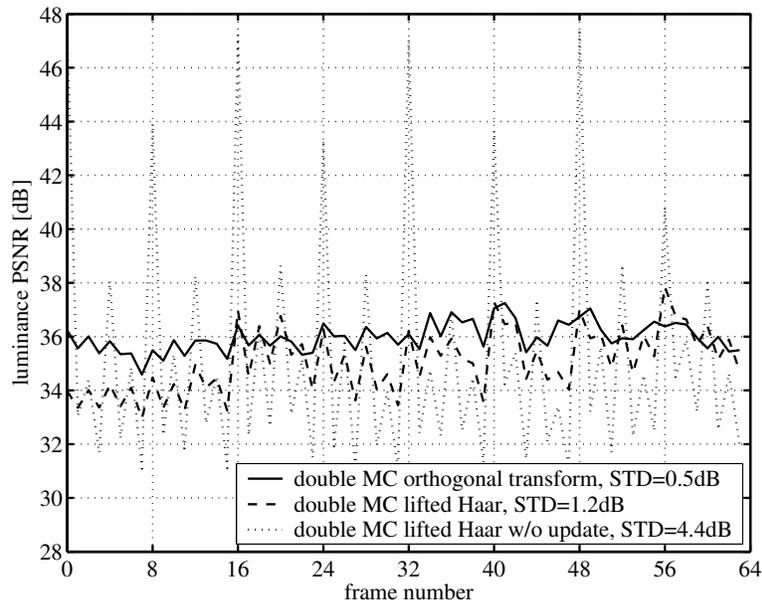
Figure 6. PSNR over bit rate for the luminance signal of the QCIF sequence *Mother & Daughter* at 30 fps with 64 frames.

Compared to the double MC lifted Haar wavelet, the double MC orthogonal transform offers the smallest PSNR fluctuations over time while providing the best rate distortion performance. For *Bus*, the standard deviation is as small as 0.5 dB compared to 1.2 dB and 4.4 dB for the lifted Haar wavelet with and without update step, respectively.

As noted above, limiting the second motion vector to address directly neighboring pixels in the corresponding low-band will achieve sub-pixel accurate motion compensation. This will reduce complexity as well as efficiency of the generic 2-hypothesis motion, but, on the other hand, will save the bit rate of the second motion vector. Future work will investigate this trade-off between complexity and efficiency.



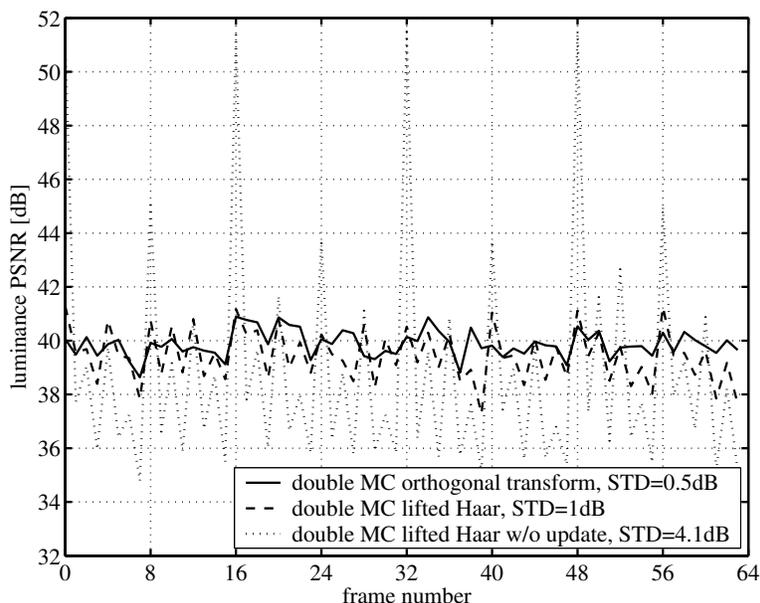
**Figure 7.** PSNR over bit rate for the luminance signal of the QCIF sequence *City* at 15 fps with 64 frames.



**Figure 8.** PSNR of the luminance signal over the frame number for the QCIF sequence *Bus* at 15 fps with 64 frames. The standard deviation of the luminance PSNR over time for the double MC orthogonal transform is with 0.5 dB significantly smaller than that of the double MC lifted Haar wavelet with 1.2 dB and 4.4 dB.

#### 4. CONCLUSIONS

This paper presents a double motion-compensated orthogonal transform which strictly maintains orthogonality for any 2-motion field. In terms of energy compaction, it outperforms the single motion-compensated orthogonal transform and provides benefits over the double motion-compensated lifted Haar wavelet. The incremental orthogonal  $3 \times 3$  transform permits a linear combination of two integer-pixel accurate motion-compensated signals. This achieves sub-pixel accurate motion compensation between two frames. This is in contrast to the single motion-compensated orthogonal transform which uses only an incremental  $2 \times 2$  transform and, hence, is



**Figure 9.** PSNR of the luminance signal over the frame number for the QCIF sequence *City* at 15 fps with 64 frames. The standard deviation of the luminance PSNR over time for the double MC orthogonal transform is with 0.5 dB significantly smaller than that of the double MC lifted Haar wavelet with 1.0 dB and 4.1 dB.

limited to integer-pixel accurate motion compensation. In summary, the orthogonality principle improves energy compaction, provides a highly robust video representation, and permits 2-motion compensation.

## REFERENCES

1. V. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine* **18**, pp. 9–21, Sept. 2001.
2. J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing* **3**, pp. 559–571, Sept. 1994.
3. S.-J. Choi and J. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Transactions on Image Processing* **8**, pp. 155–167, Feb. 1999.
4. 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.
5. 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.
6. M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *Signal Processing: Image Communication* **19**, pp. 561–575, Aug. 2004.
7. B. Girod and S. Han, "Optimum update for motion-compensated lifting," *IEEE Signal Processing Letters* **12**, pp. 150–153, Feb. 2005.
8. M. Flierl and B. Girod, "A motion-compensated orthogonal transform with energy-concentration constraint," in *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, (Victoria, BC), Oct. 2006.
9. 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.
10. M. Flierl, P. Vanderghenst, and B. Girod, "Video coding with lifted wavelet transforms and complementary motion-compensated signals," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, **5308**, pp. 497–508, (San Jose, CA), Jan. 2004.