Half-Pel Accurate Motion-Compensated Orthogonal Video Transforms

Markus Flierl and Bernd Girod
Max Planck Center for Visual Computing and Communication
Stanford University, Stanford, CA 94305
mflierl@stanford.edu

Abstract

Motion-compensated lifted wavelets have received much interest for video compression. While they are biorthogonal, they may substantially deviate from orthonormality due to motion compensation, even if based on an orthogonal or near-orthogonal wavelet. A temporal transform for video sequences that maintains orthonormality while permitting flexible motion compensation would be very desirable. We have recently introduced such a transform for integer-pel accurate motion compensation from one previous frame. In this paper, we extend this idea to half-pel accurate motion compensation. Orthonormality is maintained for arbitrary half-pel motion compensation by cascading a sequence of incremental orthogonal transforms. The half-pel intensity values are obtained by averaging neighboring integer-pel positions. Depending on the number of averaged integer-pel values, we use different types of incremental transforms. The cascade of incremental transforms allows us to choose in each step the optimal type of incremental transform and, hence, the optimal half-pel position. Half-pel motion-compensated blocks of arbitrary shape and size can be used as the granularity of the cascade can be as small as one pixel. The new half-pel accurate motion-compensated orthogonal video transform compares favorably with the integer-pel accurate orthogonal transform.

1 Introduction

Well known methods for representing image sequences for coding and communication applications are standard hybrid video coding techniques as well as motion-compensated 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 [1]. 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 [2, 3, 4, 5] by approaching problems arising from multi-connected pixels. For example, [4, 5] choose a reversible lifting implementation for the temporal filter and incorporate motion compensation into the lifting steps. In particular, the motion-compensated lifted Haar wavelet maintains orthogonality only for single-connecting motion fields. But for complex motion fields with many multi-connected and unconnected pixels, the reversible motion-compensated lifted Haar wavelet loses the property of orthogonality. In [6], we propose an integer-pel accurate motion-compensated orthogonal transform that strictly maintains orthogonality for any motion field. Orthogonality is maintained by factoring the transform into a sequence of incremental transforms that are strictly orthogonal. An energy concentration constrained is used to minimize the energy in the temporal high bands. 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.

In contrast to our previous work in [6] where only integer-pel motion compensation is considered, this paper extends the approach to half-pel accurate motion compensation with averaging filters. With that, we extend also our previous work in [7] where a linear combination of two arbitrarily motion-compensated signals is investigated.

The paper is organized as follows: Section 2 presents the concept of motion-compensated orthogonal transforms with 1-hypothesis, 2-hypothesis, and 4-hypothesis motion compensation. Section 3 outlines the corresponding energy concentration constraints. Section 4 discusses half-pel accurate motion compensation with averaging filters and its realization with orthogonal transforms. Section 5 presents the experimental results.

2 Motion-Compensated Orthogonal Transforms

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

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = T \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \tag{1}$$

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_κ such that

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

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

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_κ maps these vectors according to

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

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.

To picture the sequence of transformed image pairs $(\mathbf{x}_1^{(\kappa)}, \mathbf{x}_2^{(\kappa)})$, 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 κ is represented by the incremental transform T_κ.

2.1 1-Hypothesis Motion Compensation

For 1-hypothesis motion compensation, each pixel in the image $\mathbf{x_2}$ is linked to only one pixel in the image $\mathbf{x_1}$.

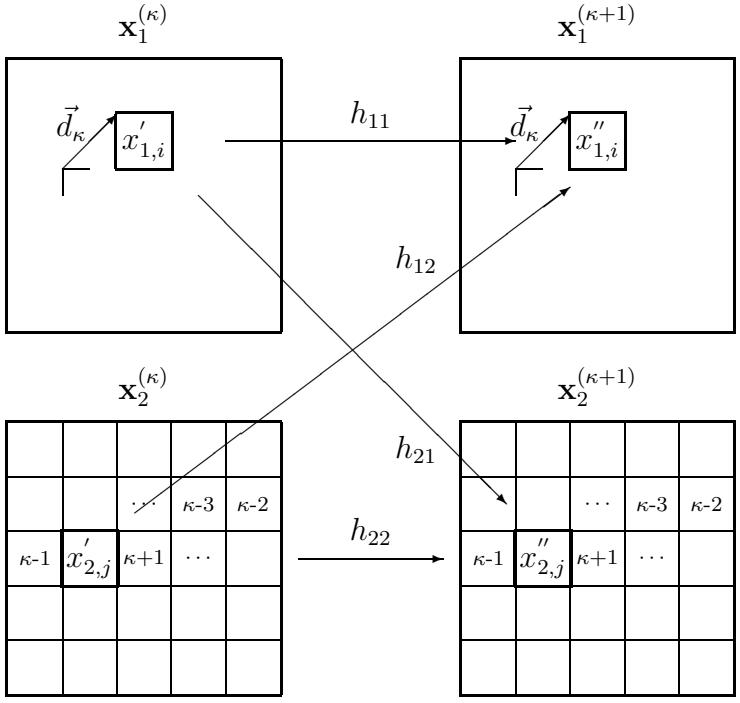

Figure 1: The incremental transform T_κ for two frames $\mathbf{x}_1^{(\kappa)}$ and $\mathbf{x}_2^{(\kappa)}$ which strictly maintains orthogonality for any 1-hypothesis motion field.

Fig. 1 depicts the process accomplished by the 1-hypothesis incremental transform T_κ with its input and output images as defined above. The incremental transform removes the energy of the j-th pixel $x'_{2,j}$ in the image $\mathbf{x}_2^{(\kappa)}$ with the help of the i-th

pixel $x'_{1,i}$ in the image $\mathbf{x}_1^{(\kappa)}$ which is linked by the motion vector $\vec{d}_\kappa$ (or of the j-th block with the help of the i-th block if all the pixels of the block have the same motion vector $\vec{d}_\kappa$). The energy-removed pixel value $x''_{2,j}$ is obtained by a linear combination of the pixel values $x'_{1,i}$ and $x'_{2,j}$ with scalar weights h_{21} and h_{22}. The energy-concentrated pixel value $x''_{1,i}$ is also obtained by a linear combination of the pixel values $x'_{1,i}$ and $x'_{2,j}$ but with scalar weights h_{11} and h_{12}. All other pixels are simply kept untouched.

The scalar weights $h_{\mu\nu}$ are arranged into the matrix

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \tag{4}$$

which is required to be orthogonal. For a 2×2 matrix, one scalar *decorrelation factor* a is sufficient to capture all possible orthogonal transforms. We use the form

$$H = \frac{1}{\sqrt{1+a^2}} \begin{bmatrix} 1 & a \\ -a & 1 \end{bmatrix}, \tag{5}$$

where a is a positive real value to remove the energy in the image $\mathbf{x}_2$ and to concentrate the energy in the image $\mathbf{x}_1$.

T_κ performs only a linear combination with pixel pairs that are connected by the associated motion vector. All other pixels are simply kept untouched. This is reflected with the following matrix notation

$$T_\kappa = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & h_{11} & 0 & \cdots & 0 & h_{12} & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & h_{21} & 0 & \cdots & 0 & h_{22} & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{6}$$

where the diagonal elements equal to 1 represent the untouched pixels and where the elements $h_{\mu\nu}$ represent the pixels subject to linear operations.

2.2 2-Hypothesis Motion Compensation

For 2-hypothesis motion compensation, each pixel in the image $\mathbf{x_2}$ is linked to two pixel in the image $\mathbf{x_1}$.

Fig. 2 depicts the process accomplished by the 2-hypothesis incremental transform T_κ. For this case, we have nine scalar weights that are arranged into the 3×3 orthogonal matrix H. We construct 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.

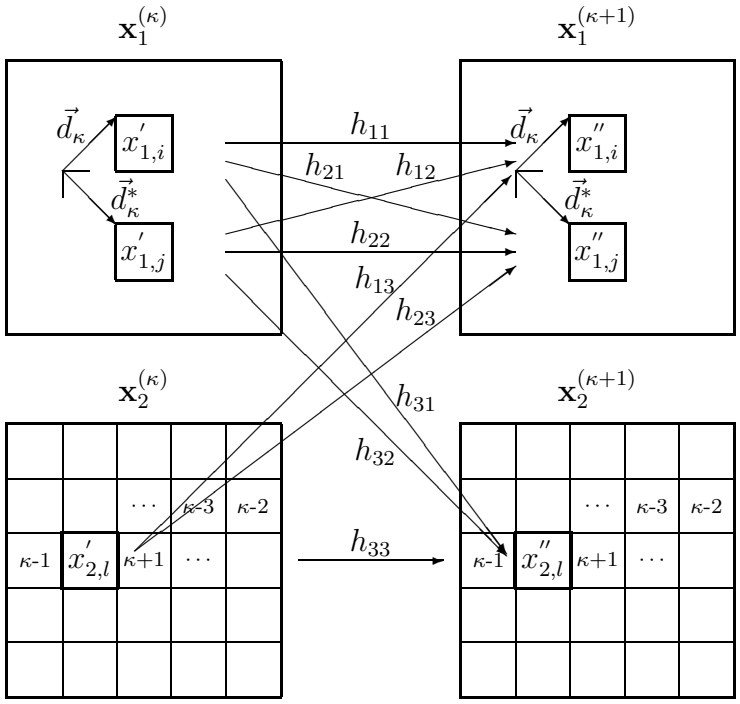

Figure 2: The incremental transform T_κ for two frames $\mathbf{x}_1^{(\kappa)}$ and $\mathbf{x}_2^{(\kappa)}$ which strictly maintains orthogonality for any 2-hypothesis motion field $(\vec{d}_\kappa, \vec{d}^*_\kappa)$.

$H = H_3 H_2 H_1$, where H_r denotes a rotation about one axes. We choose the following composition with the Euler angles ψ, θ, and ϕ:

$$H = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

2.3 4-Hypothesis Motion Compensation

For 4-hypothesis motion compensation, each pixel in the image $\mathbf{x_2}$ is linked to four pixel in the image $\mathbf{x_1}$. Here, 25 scalar weights are arranged into the 5×5 orthogonal matrix H. We construct H by a composition of rotations about 7 axis. We choose the composition

$$H = H_a(\phi_7) H_b(\phi_6) H_c(\phi_5) H_d(\phi_4) H_c(\phi_3) H_b(\phi_2) H_a(\phi_1) \tag{8}$$

with the following individual rotations:

$$H_a(\phi) = \begin{bmatrix} \cos\phi & \sin\phi & 0 & 0 & 0 \\ -\sin\phi & \cos\phi & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad H_b(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \cos\phi & \sin\phi & 0 \\ 0 & 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$H_c(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos\phi & 0 & \sin\phi & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -\sin\phi & 0 & \cos\phi & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad H_d(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \cos\phi & \sin\phi \\ 0 & 0 & 0 & -\sin\phi & \cos\phi \end{bmatrix} \quad (10)$$

3 Energy Concentration Constraint

In the following, the decorrelation factors and angles of the incremental transforms are chosen such that the energy in the temporal low band is concentrated.

3.1 1-Hypothesis Motion Compensation

Consider the pixel pair $x_{1,i}$ and $x_{2,j}$ to be processed by the incremental transform T_κ. To determine the decorrelation factor a for the pixel $x_{2,j}$, we assume that the pixel $x_{2,j}$ is connected to the pixel $x_{1,i}$ such that $x_{2,j} = x_{1,i}$. Consequently, the resulting "high-band to be" pixel $x''_{2,j}$ shall be zero. Note that the pixel $x_{1,i}$ may have been processed previously by T_τ, where $\tau < \kappa$. Therefore, let v_1 be the *scale factor* for the pixel $x_{1,i}$ such that $x'_{1,i} = v_1 x_{1,i}$. At the first wavelet decomposition level, each pixel $x_{2,j}$ of the picture $\mathbf{x}_2$ is used only once during the transform process. Therefore, there is no need to maintain a scale counter for the pixels in the picture $\mathbf{x}_2$. But at the second level of temporal decomposition, both pictures $\mathbf{x}_1$ and $\mathbf{x}_2$ are temporal low-bands resulting from transforms at the first level. Therefore, we need to consider also a scale factor v_2 for the pixel $x_{2,j}$, i.e., $x'_{2,j} = v_2 x_{2,j}$. Let u_1 be the scale factor for the pixel $x_{1,i}$ after it has been processed by T_κ. Now, the pixels $x'_{1,i}$ and $x'_{2,j}$ are processed by T_κ as follows:

$$\begin{bmatrix} u_1 x_{1,i} \\ 0 \end{bmatrix} = \frac{1}{\sqrt{1+a^2}} \begin{bmatrix} 1 & a \\ -a & 1 \end{bmatrix} \begin{bmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \end{bmatrix} \quad (11)$$

The conditions of energy conservation and energy concentration are satisfied if

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

3.2 2-Hypothesis Motion Compensation

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_κ. 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_τ, 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_κ. Now, the pixels $x'_{1,i}$, $x'_{1,j}$, and $x'_{2,l}$ are processed by T_κ as follows:

$$\begin{bmatrix} u_1 x_{1,i} \\ u_2 x_{1,i} \\ 0 \end{bmatrix} = H_3 H_2 H_1 \begin{bmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \\ v_3 x_{1,i} \end{bmatrix} \tag{13}$$

Energy conservation requires that $u_1^2 + u_2^2 = v_1^2 + v_2^2 + v_3^2$. The Euler angle ϕ 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.

$$\tan\phi = -\frac{v_1}{v_2} \tag{14}$$

The Euler angle θ in H_2 is chosen such that it meets the zero-energy constraint for the high-band in (13).

$$\tan\theta = \frac{v_3}{\sqrt{v_1^2 + v_2^2}} \tag{15}$$

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

$$\tan\psi = \frac{u_1}{u_2} \tag{16}$$

But note that we are free to choose this ratio. We have chosen the angle ϕ 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 their energy increases equally.

$$u_1 = \sqrt{v_1^2 + \frac{v_3^2}{2}} \quad \text{and} \quad u_2 = \sqrt{v_2^2 + \frac{v_3^2}{2}} \tag{17}$$

3.3 4-Hypothesis Motion Compensation

For 4-hypothesis motion compensation, we choose 7 angles to minimize the energy of pixels in the image $\mathbf{x}_2$. To determine the angles for the pixel $x_{2,l}$, we assume that it is connected to the pixels $x_{1,i}$, $x_{1,j}$, $x_{1,\mu}$, and $x_{1,\nu}$ such that $x_{2,l} = x_{1,i} = x_{1,j} = x_{1,\mu} = x_{1,\nu}$. Let v_1, v_2, v_3, and v_4 be the scale factors for the four pixels in $\mathbf{x}_1$ and let v_5 be that of the pixel $x_{2,l}$. Let u_1, u_2, u_3, and u_4 be the scale factors for the four pixels in $\mathbf{x}_1$ after they have been processed by T_κ. Now, the four pixels in $\mathbf{x}_1$ as well as the pixel $x'_{2,l}$ are processed by T_κ as follows:

$$\begin{bmatrix} u_1 x_{2,l} \\ u_2 x_{2,l} \\ u_3 x_{2,l} \\ u_4 x_{2,l} \\ 0 \end{bmatrix} = H_a(\phi_7) H_b(\phi_6) H_c(\phi_5) H_d(\phi_4) H_c(\phi_3) H_b(\phi_2) H_a(\phi_1) \begin{bmatrix} v_1 x_{2,l} \\ v_2 x_{2,l} \\ v_3 x_{2,l} \\ v_4 x_{2,l} \\ v_5 x_{2,l} \end{bmatrix} \tag{18}$$

Energy conservation requires that $u_1^2 + u_2^2 + u_3^2 + u_4^2 = v_1^2 + v_2^2 + v_3^2 + v_4^2 + v_5^2$. The angle ϕ_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. The same argument holds for angle ϕ_2. The angle ϕ_3 is chosen such that the combination of the two previous hypothesis pairs is also weighted equally after being attenuated by their combined scale factors. With this, we generally achieve weights that are powers of two, and in this particular case, we achieve a weight of 1/4 for each hypothesis. The angle ϕ_4 is chosen such that it meets the zero-energy constraint for the high-band in (18). Finally, the angles ϕ_5, ϕ_6, and ϕ_7 are chosen such that the pixels in $\mathbf{x}_1$, after the incremental transform T_κ, have scalar weights u_ρ, $\rho = 1, 2, 3, 4$.

$$\tan\phi_1 = -\frac{v_1}{v_2}, \qquad \tan\phi_2 = -\frac{v_3}{v_4}, \qquad \tan\phi_3 = -\frac{\sqrt{v_1^2 + v_2^2}}{\sqrt{v_3^2 + v_4^2}} \tag{19}$$

$$\tan\phi_4 = \frac{v_5}{\sqrt{v_1^2 + v_2^2 + v_3^2 + v_4^2}} \tag{20}$$

$$\tan\phi_5 = \frac{\sqrt{u_1^2 + u_2^2}}{\sqrt{u_3^2 + u_4^2}}, \qquad \tan\phi_6 = \frac{u_3}{u_4}, \qquad \tan\phi_7 = \frac{u_1}{u_2} \tag{21}$$

But note that we are free to choose the ratios among u_ρ. We have chosen the angles such that each hypothesis has equal contribution. Consequently, we choose the scale factors u_ρ such that their energy increases equally.

$$u_\rho = \sqrt{v_\rho^2 + \frac{v_5^2}{4}} \quad \text{for} \quad \rho = 1, 2, 3, 4 \tag{22}$$

4 Half-Pel Accurate Motion Compensation

With these three types of incremental transforms, we are able to achieve half-pel accurate motion compensation where half-pel intensity values are obtained by averaging neighboring integer-pel positions.

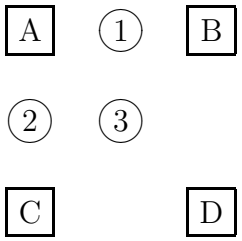

Figure 3: Interpolation for half-pel accurate motion compensation.

Fig. 3 depicts the neighboring integer-pel positions labeled by A to D. Half-pel positions are labeled by 1 to 3. The half-pel intensity value of 1 is obtained by averaging integer positions A and B. For the value of 2, positions A and C are used. In both cases, the 2-hypothesis incremental transform is used. The value of 3 is obtained by

averaging integer values at positions A to D. In this case, the 4-hypothesis incremental transform is utilized. The granularity of the cascade of incremental transforms allows half-pel motion-compensated blocks of arbitrary shape and size.

5 Experimental Results

Experimental results assessing the energy compaction are obtained for the CIF sequence *City*. We compare half-pel and integer-pel accurate motion-compensated orthogonal transforms. For the coding process with the orthogonal transforms, a scale counter $n = v^2 - 1$ 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.

Both schemes operate with a GOP size of 16 frames. The block size for motion compensation is limited to 8×8. The bit rate to signal the half-pel position is included in the total rate. For each block, the half-pel position with the smallest Lagrangian costs is chosen. 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 the scale factor $v = \sqrt{n+1}$ before encoding. For optimal rate allocation, Lagrangian costs are determined where the distortion term considers the scale factors applied to the temporal low-band.

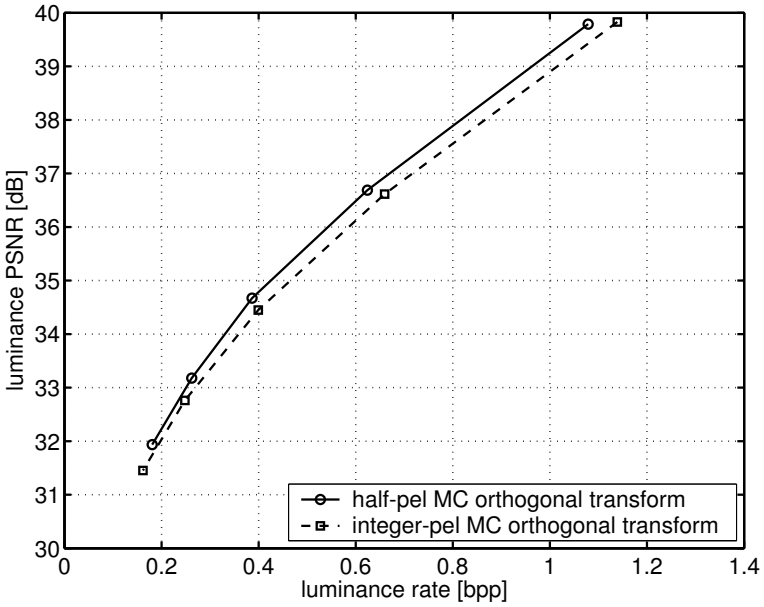

Figure 4: PSNR over bit rate for the luminance signal of the CIF sequence *City* at 30 fps with 64 frames.

Fig. 4 depicts the rate distortion performance for the luminance signal of the test sequence. Results for both half-pel and integer-pel accurate motion-compensated orthogonal transforms are given. In order to assess the energy compaction, no intra

modes have been used for both temporal coding schemes. We observe that the half-pel accurate transform outperforms the integer-pel accurate transform, particularly at high rates.

6 Conclusions

The paper presents a half-pel accurate motion-compensated orthogonal video transform. Orthonormality is maintained for arbitrary half-pel motion compensation by cascading a sequence of incremental orthogonal transforms. The half-pel intensity values are obtained by averaging neighboring integer-pel positions. Depending on the number of averaged integer-pel values, we use different types of incremental transforms. The cascade of incremental transforms allows us to choose in each step the optimal type of incremental transform and, hence, the optimal half-pel position. Bidirectionally motion-compensated orthogonal transforms will be introduced in [8].

References

[1] V.K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 9–21, Sept. 2001.

[2] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–571, Sept. 1994.

[3] S.-J. Choi and J.W. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, no. 2, 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*, Salt Lake City, UT, May 2001, vol. 3, pp. 1793–1796.

[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*, Thessaloniki, Greece, Oct. 2001, vol. 2, pp. 1029–1032.

[6] 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.

[7] M. Flierl and B. Girod, "A double motion-compensated orthogonal transform with energy concentration constraint," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, San Jose, CA, Jan. 2007, vol. 6508.

[8] M. Flierl and B. Girod, "A new bidirectionally motion-compensated orthogonal transform for video coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, Apr. 2007, to appear.