MOTION-COMPENSATED ORTHOGONAL TRANSFORMS FOR MULTIVIEW VIDEO CODING

Markus Flierl

Max Planck Center for Visual Computing and Communication
Stanford University, California
mflierl@stanford.edu

Invited Paper

ABSTRACT

This paper discusses coding of multiview video with motion-compensated orthogonal transforms. These transforms have recently been introduced to generate strictly orthonormal decompositions of image sequences. Orthonormality is maintained for arbitrary motion compensation among the images. This is of particular interest for coding applications as motion compensation is inaccurate due to quantization of motion information. This is in contrast to the well known motion-compensated lifted wavelets where the properties of the decomposition are motion-dependent. This disadvantage affects compression efficiency, and in particular that of multiview video signals. Multiview video coding schemes utilize motion and disparity compensation when decomposing the multiview imagery into view-temporal subbands. Usually, decompositions in time and view direction are cascaded. Motion-compensated lifted wavelets suffer from their motion-dependent decomposition, and in particular if cascaded for multiview video coding. On the other hand, motion and disparity compensated orthogonal transforms offer the advantage that the overall decomposition is still strictly orthonormal. This paper investigates the advantages of strict orthonormality that is offered by motion-compensated orthogonal transforms.

1. INTRODUCTION

Today's advances in display and camera technology enable new applications utilizing multiple video cameras for communicating dynamic 3-d scenes. Well known examples are free viewpoint video [1] and free viewpoint television (FTV) [2]. For all these applications, efficient coding of multiview imagery is challenging.

For efficient coding of multiview video, statistical dependencies among all images have to be exploited. Disparities between views and motion between temporally successive frames are the most important parameters. To achieve a good trade-off between image quality and bit-rate, the correlation among all pictures has to be exploited efficiently. Usually, this is accomplished with either predictive or subband coding schemes.

Based on the video coding standard H.264/AVC [3], the Joint Video Team (JVT) is developing a Joint Multiview Video Model (JMVM) [4] for multiview video coding. It utilizes motion and disparity compensated prediction to exploit the correlation in temporal and view direction. The JMVM is a predictive coding scheme.

For subband coding of multiview video, an efficiency analysis of motion and disparity compensated coding [5] has been developed. It has been found that high-rate performance bounds may be achieved with motion and disparity compensated orthogonal transforms. For video coding, unidirectionally [6] and bidirectionally [7] motion-compensated orthogonal video transforms have been proposed. In this paper, we investigate multiview video coding with motion-compensated orthogonal transforms applied in temporal as well as in view direction.

Motion-compensated orthogonal transforms decompose image sequences into strictly orthonormal representations. Orthonormality is maintained for arbitrary motion compensation among the input images. Given the quantized motion information in coding applications, motion-compensated orthogonal transforms still generate an orthonormal decomposition of the input image sequence. This is in contrast to the well known motion-compensated lifted wavelets where the properties of the decomposition are motion-dependent. For example, the motion-compensated lifted Haar wavelet is orthonormal if the motion field is single-connecting, e.g. zero motion. But orthonormality is lost if there are multi-connected and unconnected pixels. This loss affects compression efficiency, and in particular that of multiview video signals.

For subband coding of multiview video, decompositions in time and view direction are usually cascaded. This is the case if, for example, temporal subbands are further decomposed in view direction. Motion-compensated lifted wavelets suffer from their motion-dependent decomposition, and in particular if cascaded for multiview video coding. This paper investigates the advantages of strict orthonormality that is offered by motion-compensated orthogonal transforms.

The paper is organized as follows: Section 2 summarizes the class of motion-compensated orthogonal transforms. Section 3 outlines our coding scheme for multiview video based on motion and disparity adaptive orthogonal transforms. Finally, Section 4 reports on the experimental results that we have obtained with our coding scheme.

2. THE CLASS OF MOTION-COMPENSATED ORTHOGONAL TRANSFORMS

The class of motion-compensated orthogonal transforms (MCOT) includes unidirectionally [6] and bidirectionally [7] motion-compensated orthogonal transforms. There are also half-pel accurate orthogonal transforms [8] as well as the general double motion-compensated orthogonal transforms [9]. In the following, we summarize the two main principles that define this class. The first principle factors the orthogonal transform into a sequence of incremental transforms where each is orthogonal by itself. The second principle establishes an energy concentration constraint for each incremental transform such that energy in high-bands is removed efficiently.

2.1 Incremental Transform

We begin with a bidirectionally motion-compensated orthogonal transform. Let $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ be three vectors representing consecutive pictures of an image sequence. The transform T maps these vectors according to

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix} = T \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix} \tag{1}$$

into three vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ which represent the first temporal low-band, the high-band, and the second temporal low-band, respectively. We factor the transform T into a sequence of k incremental transforms T_κ such that

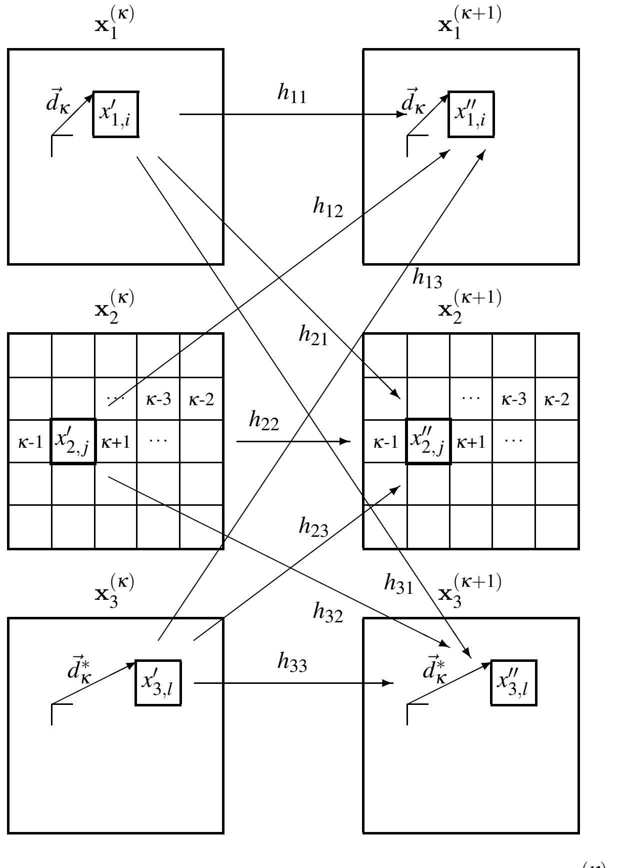
$$T = T_k T_{k-1} \cdots T_\kappa \cdots T_2 T_1, \tag{2}$$

where each incremental transform T_κ is orthogonal by itself, i.e., $T_\kappa T_\kappa^T = I$ holds for all $\kappa = 1, 2, \cdots, k$. 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 κ is represented by the incremental transform T_κ.

Let $\mathbf{x}_1^{(\kappa)}$, $\mathbf{x}_2^{(\kappa)}$, and $\mathbf{x}_3^{(\kappa)}$ be three vectors representing consecutive pictures of an image sequence if $\kappa = 1$, or three output vectors of the incremental transform $T_{\kappa-1}$ if $\kappa > 1$. The incremental transform T_κ maps these vectors according to

$$\begin{pmatrix} \mathbf{x}_1^{(\kappa+1)} \\ \mathbf{x}_2^{(\kappa+1)} \\ \mathbf{x}_3^{(\kappa+1)} \end{pmatrix} = T_\kappa \begin{pmatrix} \mathbf{x}_1^{(\kappa)} \\ \mathbf{x}_2^{(\kappa)} \\ \mathbf{x}_3^{(\kappa)} \end{pmatrix} \tag{3}$$

into three vectors $\mathbf{x}_1^{(\kappa+1)}$, $\mathbf{x}_2^{(\kappa+1)}$, and $\mathbf{x}_3^{(\kappa+1)}$ which will be further transformed into the first temporal low-band, high-band, and second temporal low-band, respectively.



$\mathbf{x}_1^{(\kappa)}$
$\mathbf{x}_1^{(\kappa+1)}$
$\vec{d}_\kappa$
$x'_{1,i}$
h_{11}
$\vec{d}_\kappa$
$x''_{1,i}$
h_{12}
h_{13}
$\mathbf{x}_2^{(\kappa)}$
$\mathbf{x}_2^{(\kappa+1)}$
h_{21}
$\cdots$ κ-3 κ-2
κ-1 $x'_{2,j}$ κ+1 $\cdots$
h_{22}
$\cdots$ κ-3 κ-2
κ-1 $x''_{2,j}$ κ+1 $\cdots$
h_{23}
$\mathbf{x}_3^{(\kappa)}$
h_{31}
$\mathbf{x}_3^{(\kappa+1)}$
h_{32}
$\vec{d}^*_\kappa$
$x'_{3,l}$
h_{33}
$\vec{d}^*_\kappa$
$x''_{3,l}$

Figure 1: The incremental transform T_κ for the three frames $\mathbf{x}_1^{(\kappa)}$, $\mathbf{x}_2^{(\kappa)}$, and $\mathbf{x}_3^{(\kappa)}$ which strictly maintains orthogonality for any bidirectional motion field $(\vec{d}_\kappa, \vec{d}^*_\kappa)$. T_κ minimizes the energy in $x_{2,j}$.

Fig. 1 depicts the process accomplished by the 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 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 l-th pixel $x'_{3,l}$ in the image $\mathbf{x}_3^{(\kappa)}$ which is linked by the motion vector $\vec{d}^*_\kappa$ (or the j-th block with the help of both the i-th and the l-th block if all the pixels of the i-th and l-th block have the motion vectors $\vec{d}_\kappa$ and $\vec{d}^*_\kappa$, respectively). The energy-removed pixel value $x''_{2,j}$ is obtained by a linear combination of the pixel values $x'_{1,i}$, $x'_{2,j}$, and $x'_{3,l}$ with scalar weights h_{21}, h_{22}, and h_{23}. The energy-concentrated pixel value $x''_{1,i}$ is also obtained by a linear combination of the pixel values $x'_{1,i}$, $x'_{2,j}$, and $x'_{3,l}$ but with scalar weights h_{11}, h_{12}, and h_{13}. The energy-concentrated pixel value $x''_{3,l}$ is calculated accordingly. All other pixels are simply kept untouched.

To summarize, the incremental transform T_κ touches only pixels that are linked by the same motion vector pair $(\vec{d}_\kappa, \vec{d}^*_\kappa)$. Of these, T_κ 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:

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

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.

If bidirectional motion compensation is used in step κ, incremental transforms $T_\kappa^{(2)}$ with the general form of (4) are required. For unidirectional motion compensation, the incremental transforms simplifies. For example, unidirectional motion compensation from $\mathbf{x}_1$ does not require the picture $\mathbf{x}_3$. Hence, the pixels in $\mathbf{x}_3$ are not altered in step κ and the submatrix of T_κ which modifies the pixels in $\mathbf{x}_3$ is simply an identity matrix. An incremental transform that accomplishes unidirectional motion compensation in step κ has the following matrix notation:

$$T_\kappa^{(1)} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \cdots & 0 & h_{11} & \cdots & 0 & h_{12} & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots \\ \cdots & 0 & h_{21} & \cdots & 0 & h_{22} & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots \\ \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{5}$$

Further, if any type of motion compensation is not suitable for a pixel or block in $\mathbf{x}_2$, the corresponding incremental transform in step κ is set to

$$T_\kappa^{(0)} = I, \tag{6}$$

where I denotes the identity matrix. We call this the **intra mode** for a pixel or block in the picture $\mathbf{x}_2$. To summarize, the type of incremental transform can be chosen freely in each step κ to match the motion of the affected pixels in $\mathbf{x}_2$ without destroying the property of orthonormality.

In each step κ, the scalar weights $h_{\mu\nu}$ are arranged into the matrix H_κ. The incremental transform T_κ is orthogonal if H_κ is also orthogonal. We accomplish unidirectional motion compensation with a 2×2 matrix $H_\kappa^{(1)}$, and bidirectional motion compensation with a 3×3 matrix $H_\kappa^{(2)}$. In general, p-hypothesis motion requires a $(p+1) \times (p+1)$ matrix $H_\kappa^{(p)}$. The coefficients of H_κ in each step

κ 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 $\mathbf{x}_1$ and $\mathbf{x}_3$ may be touched multiple times or never. Further, the order in which the incremental transforms T_κ 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 coefficients $h_{\mu\nu}$ of H_κ have to be chosen such that the energy in image $\mathbf{x}_2$ is minimized. [6] and [7] discuss a method that reduces the energy in the high-band to zero for any motion vector field if the input pictures are identical and of constant intensity. For that, so called **scale factors** u and v are introduced to capture the effect of previous incremental transforms on the intensity of each pixel.

For example, let us consider unidirectional motion compensation [6]. With the notation in Fig. 1 and above assumption, we have $x''_{1,i} = u_1 x_{1,i}$ and $x'_{1,i} = v_1 x_{1,i}$. Now, energy concentration is accomplished if

$$\begin{pmatrix} u_1 x_{1,i} \\ 0 \end{pmatrix} = H_\kappa^{(1)} \begin{pmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \end{pmatrix} \qquad (7)$$

is satisfied, i.e., the high-band coefficient is zero. Obviously, energy conservation requires that the scale factors satisfy

$$u_1^2 = v_1^2 + v_2^2. \qquad (8)$$

Further, energy concentration determines also the decorrelation factor that is the sole degree of freedom for the 2×2 matrix $H_\kappa^{(1)}$. Details are given in [6].

Interestingly, this decorrelation factor is determined only by the scale factors v_1 and v_2. Moreover, the scale factors are linked to so called **scale counters** m and n such that

$$u = \sqrt{m+1} \quad \text{and} \quad v = \sqrt{n+1}. \qquad (9)$$

The scale counters simply count for each pixel how often it is used as reference for motion compensation. Processing starts with scale counters set to zero for all pixels. For unidirectional motion compensation, the **scale counter update rule** is simply

$$m_1 = n_1 + n_2 + 1. \qquad (10)$$

Bidirectional motion compensation uses two reference pixels at the same time. Hence, two scale counters have to be updated. In [7], the scale counter update rule

$$m_1 = n_1 + \frac{n_2+1}{2} \quad \text{and} \quad m_3 = n_3 + \frac{n_2+1}{2} \qquad (11)$$

is used. Note that for bidirectional motion compensation, the 3×3 matrix $H_\kappa^{(2)}$ can be factored into rotations about three axes with the help of Euler's rotation theorem. This implies that the bidirectionally motion-compensated incremental transform has only three degrees of freedom. For general p-hypothesis motion, the extension of Euler's theorem to the p+1-dimensional space can be utilized.

2.3 Dyadic Transform for Groups of Pictures

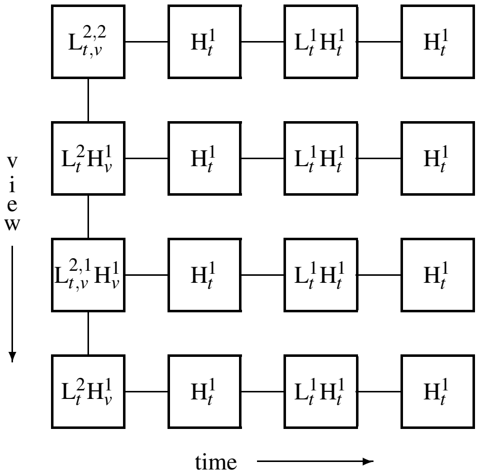
The bidirectional transform in [7] is defined for three input pictures and generates two temporal low-bands. In combination with the unidirectional transform in [6], we are able to define an orthogonal transform with only one temporal low-band for each group of pictures whose number of pictures is larger than two and a power of two. Details are given in [7].

As already discussed above, we are free to choose the type of motion compensation and, if necessary, the intra mode for each incremental transform individually. Hence, the dyadic structure for groups of pictures permits an intra block mode as well as block-wise decisions between unidirectional and bidirectional motion compensation. This adaptivity is used in our multiview video coding scheme.

3. CODING SCHEME

Our coding scheme cascades the decompositions in time and view direction. First, each view is independently decomposed with motion-compensated orthogonal transforms. Second, the resulting temporal low-bands are further decomposed in view direction with disparity-compensated orthogonal transforms.

We arrange the multiview video data into a **Matrix of Pictures** (MOP). Each MOP consists of N image sequences, each with K temporally successive pictures. With that, we consider the correlation among all the pictures within a MOP.



$\mathrm{L}_{t,v}^{2,2}$
H_t^1
$\mathrm{L}_t^1\mathrm{H}_t^1$
H_t^1
$\mathrm{L}_t^2\mathrm{H}_v^1$
H_t^1
$\mathrm{L}_t^1\mathrm{H}_t^1$
H_t^1
$\mathrm{L}_{t,v}^{2,1}\mathrm{H}_v^1$
H_t^1
$\mathrm{L}_t^1\mathrm{H}_t^1$
H_t^1
$\mathrm{L}_t^2\mathrm{H}_v^1$
H_t^1
$\mathrm{L}_t^1\mathrm{H}_t^1$
H_t^1
view
time

Figure 2: Matrix of pictures (MOP) for $N = 4$ image sequences, each comprising of $K = 4$ temporally successive pictures. The coding structure is also shown. The temporal decomposition of each view is followed by one view decomposition only.

We explain our decomposition of the multiview video signal with the example in Fig. 2. It depicts a MOP of $N = 4$ image sequences, each comprising of $K = 4$ temporally successive pictures. Each MOP is encoded with one low-band picture and $NK-1$ high-band pictures. First, a 2-level multiresolution decomposition of each view sequence in temporal direction is accomplished with motion-compensated orthogonal transforms. The first frame of each view is represented by the temporal low-band L_t^2, the remaining frames of each view by temporal high-bands H_t^1. Second, a 2-level multiresolution decomposition of the temporal low-bands L_t^2 in view direction is accomplished with disparity-compensated orthogonal transforms. After the decomposition of N temporal low-bands, we obtain the MOP low-band $\mathrm{L}_t^2\mathrm{L}_v^2$ and the remaining $N-1$ view high-bands H_v^1. We will use only the disparity fields among the views at the first time instant in the MOP. Therefore, we do not further decompose the temporal high-bands H_t^1 in view direction. It is subject to further study whether such a decomposition with additional disparity fields will provide a superior overall performance.

4. EXPERIMENTAL RESULTS

We investigate the efficiency of motion and disparity compensated orthogonal transforms for multiview video coding. For that, we use the two multiview video data sets *Ballet* and *Breakdancers*, each with 4 views, 32 temporal frames, 15 fps, and a spatial resolution of 256×192. For simplicity, we have reduced the original spatial resolution of the data sets with the MPEG downsampling filters.

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 and disparity 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 and disparity vectors.

Motion and disparity compensation is limited to 8×8 blocks

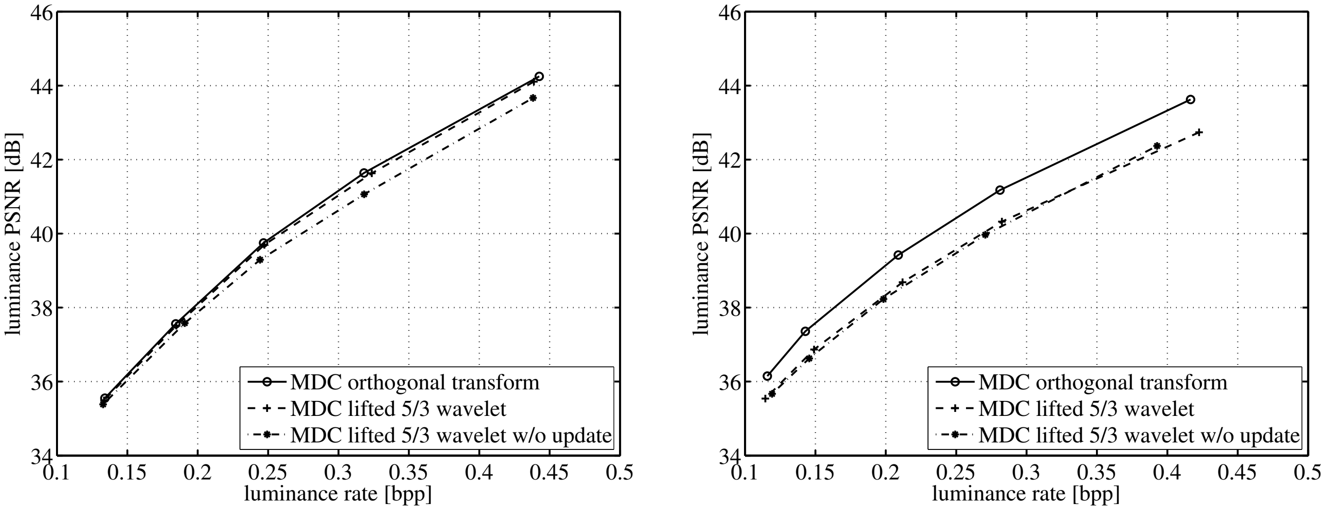


Figure 3: PSNR over bit-rate for the luminance signal of the data set *Ballet*. For a temporal GOP size of $K = 2$, the performance of 3 decompositions is given for view GOP sizes of $N = 1$ **(left)** and $N = 2$ **(right)**. The decompositions are obtained with the motion and disparity compensated orthogonal transform as well as the lifted 5/3 wavelet with and without update step.

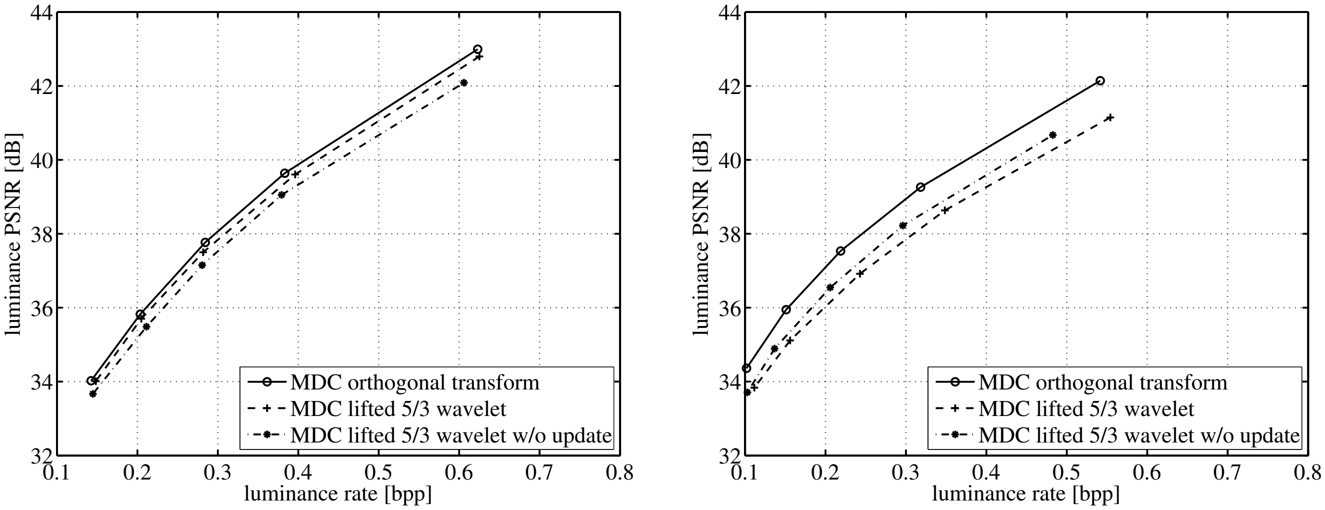


Figure 4: PSNR over bit-rate for the luminance signal of the data set *Breakdancers*. For a temporal GOP size of $K = 2$, the performance of 3 decompositions is given for view GOP sizes of $N = 1$ **(left)** and $N = 4$ **(right)**. The decompositions are obtained with the motion and disparity compensated orthogonal transform as well as the lifted 5/3 wavelet with and without update step.

and integer-pel accuracy. An extension to sub-pel accuracy is possible [8] [9]. Conditional motion and disparity estimation is used for the bidirectional type. The distortion for the bidirectional type is determined and compared to that of the unidirectional type and the intra type. Currently, the type with the smallest distortion is chosen. We investigate energy compaction of transforms and do not consider bit-rate for motion and disparity. For simplicity, the resulting temporal subbands are coded with JPEG 2000. If no intra mode is chosen, the high-bands are coded directly. The low-band of the MOP is always rescaled by (9) before encoding. Lagrangian costs are used for optimal rate allocation. Note that the scale factors of the low- and high-bands are considered in the distortion term.

4.1 Cascaded Orthogonal Transforms

Figs. 3 and 4 depict the rate-distortion performance of the luminance signal for *Ballet* and *Breakdancers*, respectively. Results for the motion and disparity compensated (MDC) orthogonal transform as well as the MDC lifted 5/3 wavelet with and without update step are given. The same block motion/disparity fields are used for both orthogonal transform and 5/3 wavelet. Therefore, the rate of the motion and disparity fields is not included. To assess the advantage of strict orthogonality for cascaded decompositions, we choose a common temporal GOP size $K = 2$. The left plots in both figures show the case of view GOP size $N = 1$, i.e., without cascaded decompositions. The views are simply coded independently. The right plots in both figures show cases with cascaded decompositions. In any case, the MDC orthogonal transform compares favorably with the MDC lifted 5/3 wavelet with and without update step. Moreover, the relative advantage over the MDC lifted 5/3 wavelet is larger for cascaded decompositions.

4.2 Performance of View-Temporal Decompositions

Fig. 5 depicts energy compaction for view GOP sizes of $N = 1,2,4$. The left plot of Fig. 6 depicts energy compaction for temporal GOP sizes of $K = 1,2,4,8$. Energy compaction improves with growing GOP size. The overall performance will improve with rate-distortion optimal type decisions and block-based intra coding. The right plot of Fig. 6 depicts the rate difference to the reference schemes with view GOP size $N = 1$ (left plot) for schemes with view GOP size of $N = 2,4$ and temporal GOP size of $K = 1,2,4,8$ at 35 dB. Please note the similarity to the theoretical results in [5].

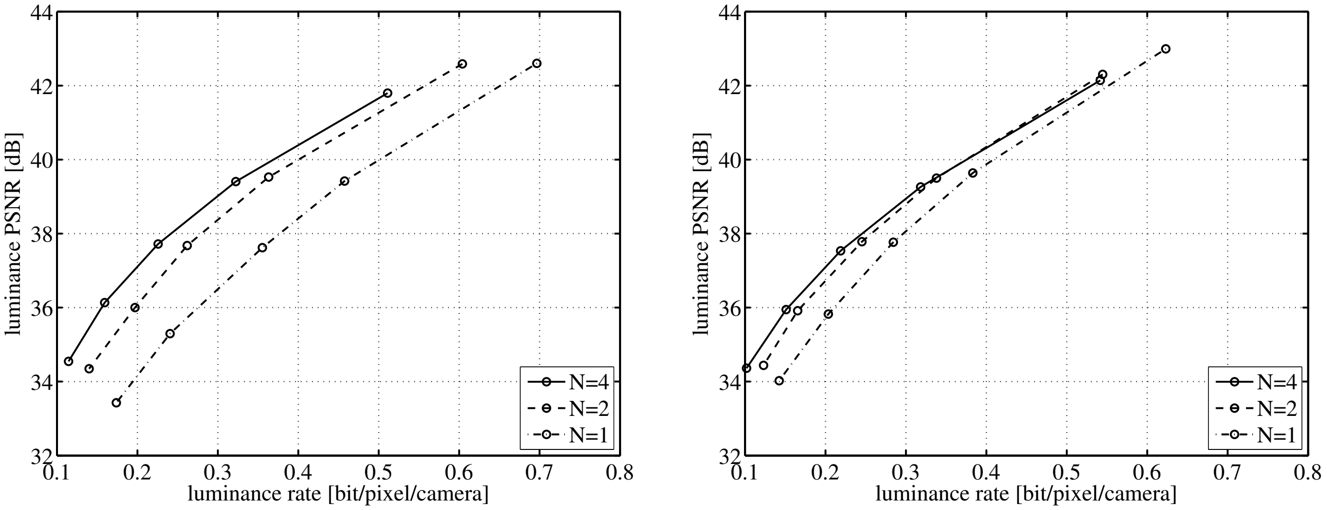


Figure 5: PSNR over bit-rate for the luminance signal of the data set *Breakdancers*. The performance of view GOP sizes of $N = 1, 2, 4$ are compared for a temporal GOP size of $K = 1$ **(left)** and $K = 2$ **(right)**.

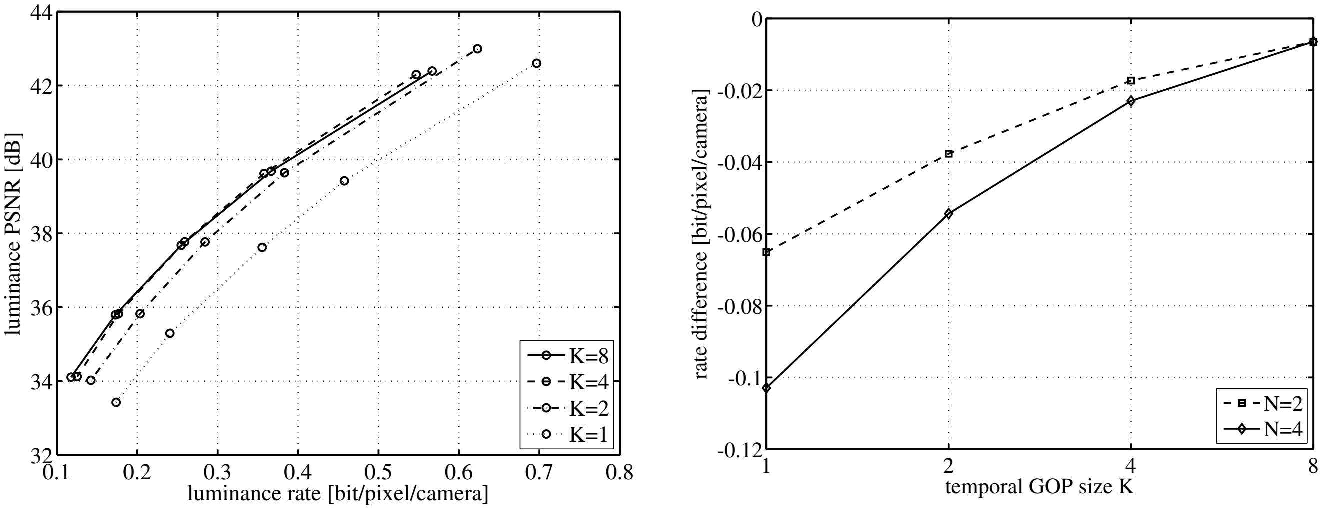


Figure 6: **Left:** PSNR over bit-rate for the luminance signal of the data set *Breakdancers*. The performance of temporal GOP sizes of $K = 1, 2, 4, 8$ is given for a view GOP size of $N = 1$. **Right:** For schemes with view GOP size of $N = 2, 4$, the rate difference to the reference schemes with view GOP size $N = 1$ is given for temporal GOP sizes of $K = 1, 2, 4, 8$ at 35 dB.

5. CONCLUSIONS

This paper investigates motion and disparity compensated orthogonal transforms for multiview video coding. It demonstrates the advantage of strict orthogonality for cascaded decompositions. Future work will develop a complete coding scheme and provide comparisons to standardized schemes.

REFERENCES

[1] A. Smolic and P. Kauff, “Interactive 3-D video representation and coding technologies,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 98–110, Jan. 2005.

[2] M. Tanimoto, “FTV (free viewpoint television) creating ray-based image engineering,” in *Proceedings of the IEEE International Conference on Image Processing*, Genova, Italy, Sept. 2005.

[3] *ITU-T Rec. H.264 – ISO/IEC 14496-10 AVC : Advanced Video Coding for Generic Audiovisual Services*, ITU-T and ISO/IEC Joint Video Team, 2005.

[4] A. Vetro, Y. Su, H. Kimata, and A. Smolic, “Joint multiview video model JMVM 2.0,” ITU-T and ISO/IEC Joint Video Team, Document JVT-U207, Nov. 2006, http:// ftp3. itu. int/ av-arch/ jvt-site/ 2006_10_Hangzhou/ JVT-U207.zip.

[5] M. Flierl, A. Mavlankar, and B. Girod, “Motion and disparity compensated coding for video camera arrays,” in *Proceedings of the Picture Coding Symposium*, Beijing, China, Apr. 2006.

[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] ——, “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.

[8] ——, “Half-pel accurate motion-compensated orthogonal video transforms,” in *Proceedings of the Data Compression Conference*, Snowbird, UT, Mar. 2007.

[9] ——, “A double motion-compensated orthogonal transform with energy concentration constraint,” in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 6508, San Jose, CA, Jan. 2007.