ADAPTIVE SPATIAL WAVELETS FOR MOTION-COMPENSATED ORTHOGONAL VIDEO TRANSFORMS

Markus Flierl

ACCESS Linnaeus Center, School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm, Sweden
mflierl@kth.se

ABSTRACT

This paper discusses adaptive spatial wavelets for the class of motion-compensated orthogonal video transforms. Motion-compensated orthogonal transforms (MCOT) are temporal transforms for video sequences that maintain orthonormality while permitting flexible motion compensation. Orthogonality is maintained for arbitrary integer-pixel or sub-pixel motion compensation by cascading a sequence of incremental orthogonal transforms and updating so-called scale counters for each pixel. The energy of the input pictures is accumulated in a temporal low-band while the temporal high-bands are zero if the input pictures are identical after motion compensation. For efficient coding, the temporal subbands should be further spatially decomposed to exploit the spatial correlation within each temporal subband. In this paper, we discuss adaptive spatial wavelets that maintain the orthogonal representation of the temporal transforms. Similar to the temporal transforms, they update scale counters for efficient energy concentration. The type-1 adaptive wavelet is a Haar-like wavelet. The type-2 considers three pixels at a time and achieves better energy compaction than the type-1.

Index Terms— Adaptive wavelet, orthogonal transform, motion-compensated orthogonal video transform, video processing, video coding.

1. INTRODUCTION

In recent years, there has been research to incorporate motion compensation into temporal subband coding schemes [1, 2, 3] by approaching problems arising from multi-connected pixels. In [4], we propose a unidirectionally motion-compensated orthogonal transform that 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.

The work in [5] extends this approach to bidirectional motion compensation. The so-called bidirectionally motion-compensated orthogonal transform is able to consider up to two motion fields per frame. Similar to our work in [4], we factor the transform into a sequence of incremental transforms which are strictly orthogonal. The incremental transforms maintain scale counters that are compatible with the scale counters in [4]. The decorrelation factors of each incremental transform are determined such that an energy-concentration constraint is met for bidirectional motion compensation.

The works in [4] and [5] utilize integer-pel motion compensation only. We have shown that this concept can be combined with sub-pixel motion compensation. A half-pixel accurate motion-compensated orthogonal transform as well as a more general double motion-compensated orthogonal transform has been introduced subsequently.

The mentioned work focuses on the temporal orthogonal transform. To achieve an efficient orthogonal subband representation for a group of pictures, the temporal subbands need to be further decomposed spatially. In the following, we discuss adaptive spatial wavelets that maintain the orthogonal representation of the temporal transforms. Similar to the temporal transforms, they update scale counters for efficient energy concentration. The type-1 adaptive wavelet processes two pixels at a time and the type-2 adaptive wavelet three pixels at a time. The type-2 adaptive wavelet achieves better energy compaction than the type-1.

The paper is organized as follows: Section 2 summarizes MCOT. Section 3 discusses type-1 and Section 4 type-2. Section 5 presents experimental results on the energy compaction of the adaptive orthogonal wavelets when decomposing a group of pictures.

2. MC ORTHOGONAL TRANSFORMS

Motion-compensated orthogonal transforms maintain strict orthogonality with arbitrary motion compensation. They transform a group of K temporally successive pictures into a set of K temporal subbands. The transform is factored into a sequence of incremental transforms that are orthogonal by themselves. The most basic incremental transform can handle an individual pixel with its associated motion information.

MCOTs have energy concentration constraints. The idea is to have a resulting zero high-band pixel if a pixel in the video is subject to the same motion model as used by the corresponding incremental transform. As general motion fields result in multiconnected pixels, MCOT uses the concept of a scale counter to count how often each pixel is used for reference. To achieve the best energy concentration for each incremental transform, scale factors have to be considered that are directly related to the scale counters.

Note, the scale counters need not to be transmitted separately as they result directly from the used motion fields. Moreover, scale counters need not to be integer values. Scale counter update rules exist for each type of incremental transform. As each incremental transform is orthogonal, signal energy is conserved.

3. TYPE-1 SPATIAL TRANSFORM

We consider the case where a spatial decomposition follows the temporal decomposition. The spatial transform that further decomposes the temporal low-band has to consider the scale factors that have been used during the temporal decomposition. In the following, we outline the type-1 spatial transform. It is a separable transform and we apply it first horizontally and then vertically.

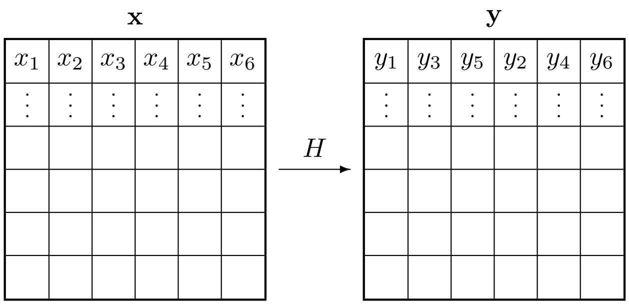
$\mathbf{x}$
$\mathbf{y}$
x_1 x_2 x_3 x_4 x_5 x_6
y_1 y_3 y_5 y_2 y_4 y_6
H

Fig. 1. The adaptive horizontal transform H for a temporal low-band $\mathbf{x}$.

Fig. 1 depicts the temporal low-band $\mathbf{x}$ and its horizontal decomposition $\mathbf{y}$. Let x_{2r+1} and x_{2r+2} be the odd and even horizontal samples of the the temporal low-band $\mathbf{x}$. The adaptive spatial transform H maps these pixels according to

$$\begin{pmatrix} y_{2r+1} \\ y_{2r+2} \end{pmatrix} = H \begin{pmatrix} x_{2r+1} \\ x_{2r+2} \end{pmatrix} \tag{1}$$

into spatial low- and high-band coefficients y_{2r+1} and y_{2r+2}, respectively. The transform matrix H is the orthogonal matrix

$$H = \frac{1}{\sqrt{1+a_n^2}} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix}. \tag{2}$$

The decorrelation factor a is determined by the constraint of energy concentration.

Ideally, if two neighboring pixel values in an image are identical, the resulting high-band pixel is zero. This is achieved, for example, with a standard Haar transform. Due to the MCOT scale counters, corresponding scale factors are introduced. In the case that the original pixel values are identical, i.e., $x_1 = x_2$, we obtain for the adaptive transform

$$\begin{pmatrix} u_1 x_1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix} \begin{pmatrix} v_1 x_1 \\ v_2 x_1 \end{pmatrix}, \tag{3}$$

were v_1 and v_2 are the scale factors of the input pixel and u_1 is the scale factor of the output low-band pixel. The corresponding high-band pixel has no scale factor. The condition of energy concentration is satisfied if

$$a = \frac{v_2}{v_1} \quad \text{and} \tag{4}$$

$$u_1 = \sqrt{v_1^2 + v_2^2}. \tag{5}$$

With the definition of the scale counter in [4]

$$v = \sqrt{n+1}, \tag{6}$$

we can write the scale factors before the transform as $v_1 = \sqrt{n_1+1}$ and $v_2 = \sqrt{n_2+1}$. After the transform, the condition of energy concentration in (5) requires the scale factors to satisfy $u_1 = \sqrt{n_1+1+n_2+1}$. This result allows us to define the *scale counter update rule* for the type-1 transform

$$m_1 = n_1 + n_2 + 1, \tag{7}$$

where the scale factor after the transform is $u_1 = \sqrt{m_1+1}$. This allows us to state the condition of energy concentration in terms of scale counter values. The decorrelation factor is

$$a = \frac{\sqrt{n_2+1}}{\sqrt{n_1+1}} \tag{8}$$

with the scale counters n_1 and n_2 which are maintained according to (7).

After updating the scale counter for the odd pixels, the scale counter for the even pixels in picture $\mathbf{y}$ are set to zero. Consequently, a standard Haar transform ($a = 1$) is applied to obtain the horizontal and diagonal subbands. But an adaptive vertical transform is used to obtain the spatial low-band and the vertical subband. The adaptive vertical transform is analogous to the adaptive horizontal transform.

4. TYPE-2 SPATIAL TRANSFORM

The type-2 spatial transform processes three pixels at a time. It outputs two low-band pixels and one high-band pixel. For example, to process a row of eight pixels, we use the decomposition as depicted in Fig. 2.

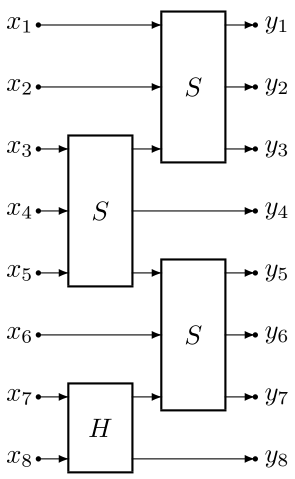

Fig. 2. Decomposition of a row of 8 pixels with the orthogonal type-2 wavelet.

The type-2 incremental transform is defined for three input pixels but generates two low-band pixels. In combination with the type-1 transform, we are able to define an orthogonal transform where the number of low-band and high-band pixels is the same after the transform. In Fig. 2, a row of 8 pixels x_i is decomposed into the low-band pixels y_1, y_3, y_5, y_7 and the high-band pixels y_2, y_4, y_6, y_8. The pixel locations are depicted in Fig. 1. H denotes the type-1 transform and S denotes the type-2 transform.

We construct an orthogonal S 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. $S = S_3S_2S_1$, where S_r denotes a rotation about one axes. We choose the composition

$$S = \begin{pmatrix} \cos(\psi) & 0 & \sin(\psi) \\ 0 & 1 & 0 \\ -\sin(\psi) & 0 & \cos(\psi) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} \tag{9}$$

with the Euler angles ψ, θ, and ϕ. Euler's theorem implies that three rotation angles are sufficient to capture all possible 3×3 orthogonal transforms. The Euler angles will be determined via the constraint of energy concentration.

The three Euler angles for each pixel touched by the incremental transform have to be chosen such that the energy in pixel y_2 is minimized. Consider the pixel triplet x_1, x_2, and x_3 to be processed by the incremental transform S. To determine the Euler angles for the pixel x_2, we assume that the pixel x_2 is perfect copy of the pixels x_1 and x_3 such that $x_2 = x_1 = x_3$. Consequently, the resulting high-band pixel y_2 shall be zero. Note that the pixels x_1 and x_3 may have been processed previously. Therefore, let v_1 and v_3 be the *scale factors* for the pixels x_1 and x_3, respectively, such that $x'_1 = v_1x_1$ and $x'_3 = v_3x_3$. The pixel x_2 is used only once during the transform process and no scale factor needs to be considered. But in general, when considering subsequent dyadic decompositions, scale factors are passed on to higher decomposition levels and, consequently, they need to be considered, i.e., $x'_2 = v_2x_2$. Obviously, for the first decomposition level, $v_2 = 1$. Let u_1 and u_3 be the scale factors for the pixels y_1 and y_3, respectively, after they have been processed. Now, the pixels x'_1, x'_2, and x'_3 are processed as follows:

$$\begin{pmatrix} u_1x_1 \\ 0 \\ u_3x_1 \end{pmatrix} = S_3S_2S_1 \begin{pmatrix} v_1x_1 \\ v_2x_1 \\ v_3x_1 \end{pmatrix} \tag{10}$$

Energy conservation requires that

$$u_1^2 + u_3^2 = v_1^2 + v_2^2 + v_3^2. \tag{11}$$

The Euler angle ϕ in S_1 is chosen such that the two hypotheses x'_1 and x'_3 are weighted equally after being attenuated by their scale factors v_1 and v_3.

$$\tan(\phi) = -\frac{v_1}{v_3} \tag{12}$$

The Euler angle θ in S_2 is chosen such that it meets the zero-energy constraint for the high-band in (10).

$$\tan(\theta) = \frac{v_2}{\sqrt{v_1^2 + v_3^2}} \tag{13}$$

Finally, the Euler angle ψ in S_3 is chosen such that the pixels x_1 and x_3, after the incremental transform, have scalar weights u_1 and u_3, respectively.

$$\tan(\psi) = \frac{u_1}{u_3} \tag{14}$$

But note that we are free to choose this ratio. We have chosen the Euler angle ϕ such that left and right pixels have equal contribution after rescaling with v_1 and v_3. Consequently, we choose the scale factors u_1 and u_3 such that they increase equally.

$$u_1 = \sqrt{v_1^2 + \frac{v_2^2}{2}} \quad \text{and} \quad u_3 = \sqrt{v_3^2 + \frac{v_2^2}{2}} \tag{15}$$

Similar to the type-1 transform, 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 $v = \sqrt{n+1}$ and $u = \sqrt{m+1}$. After applying the incremental transform, the scale counter have to be updated for the modified pixels. For the type-1 transform, the updated scale counter for low-band pixels is given by $m_1 = n_1 + n_2 + 1$, where n_1 and

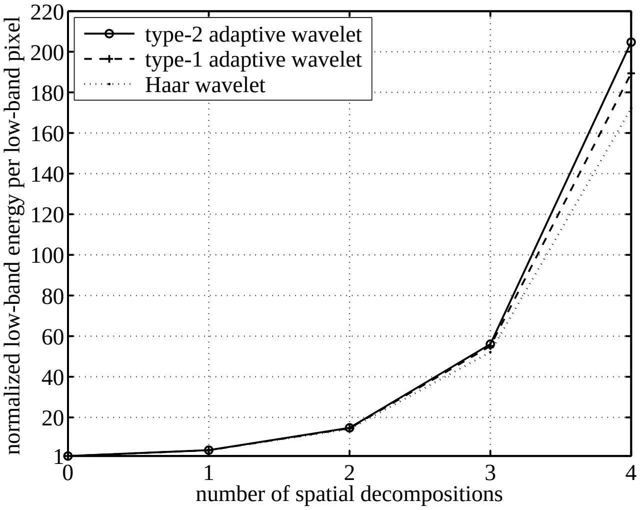

Fig. 3. Normalized spatial low-band energy per low-band pixel over the number of spatial decompositions for the CIF sequence *Soccer* with 3 temporal decompositions ($K = 8$).

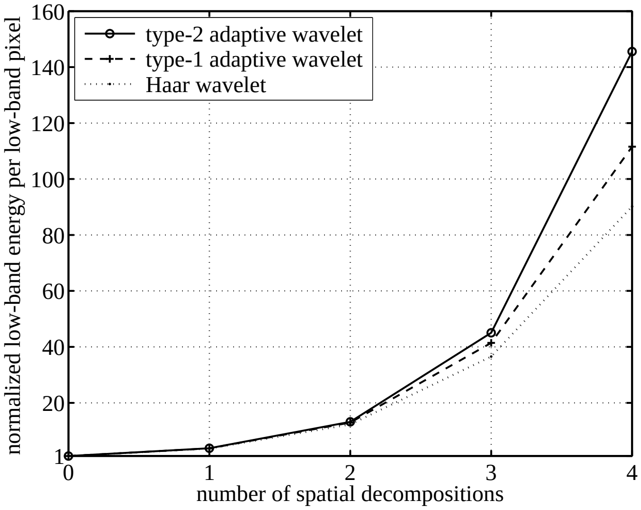

Fig. 4. Normalized spatial low-band energy per low-band pixel over the number of spatial decompositions for the CIF sequence *Bus* with 3 temporal decompositions ($K = 8$).

n_2 are the scale counters of the utilized input pixel pairs. For the type-2 transform, the updated scale counters for low-band pixels result from (15) as follows:

$$m_1 = n_1 + \frac{n_2 + 1}{2} \quad \text{and} \quad m_3 = n_3 + \frac{n_2 + 1}{2} \tag{16}$$

For example, consider the transform in the first decomposition level where $n_2 = 0$. The type-1 transform increases the scale counter by 1 for each used reference pixel, whereas the type-2 transform increases the counter by 0.5 for each of the two used reference pixels.

After updating the scale counter for the odd pixels, the scale counter for the even pixels in picture $\mathbf{y}$ are set to zero. The type-2 adaptive vertical transform is analogous to the type-2 adaptive horizontal transform.

5. EXPERIMENTAL RESULTS

We assess the energy concentration of the type-2 spatial wavelet and compare to that of the type-1 spatial wavelet and the standard Haar wavelet. We use the bidirectionally motion-compensated orthogonal transform [5] to generate temporal low-bands from groups of $K = 8$ video images. The temporal low-bands are then further decomposed with the adaptive spatial wavelets.

Figs. 3 and 4 depict the normalized spatial low-band energy per low-band pixel of temporal low-bands over the number of spatial decompositions for the CIF sequences *Soccer* and *Bus*. Obviously, energy concentration increases with the number of spatial decomposition levels. But note that type-2 achieves better energy compaction than type-1, which also outperforms the standard Haar wavelet. The utilization of the scale counter information improves energy compaction.

6. CONCLUSIONS

This paper discusses type-1 and type-2 adaptive spatial wavelets for the class of motion-compensated orthogonal video transforms. Both maintain the orthogonal representation of the temporal transforms and compact energy better than non-adaptive wavelets. The type-1 adaptive wavelet is a Haar-like wavelet. The type-2 considers three pixels at a time and achieves better energy compaction than the type-1.

7. REFERENCES

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

[2] S.-J. Choi and J. Woods, “Motion-compensated 3-d subband coding of video,” *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[3] 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*, vol. 3, Salt Lake City, UT, May 2001, pp. 1793–1796.

[4] 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.

[5] ——, “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.