Coding of Coefficients of two-dimensional non-separable Adaptive Wiener Interpolation Filter

Y. Vatis, B. Edler, I. Wassermann, D. T. Nguyen and J. Ostermann

ABSTRACT

Standard video compression techniques apply motion-compensated prediction combined with transform coding of the prediction error. In the context of prediction with fractional-pel motion vector resolution it was shown, that aliasing components contained in an image signal are limiting the prediction accuracy obtained by motion compensation. In order to consider aliasing, quantisation and motion estimation errors, camera noise, etc., we analytically developed a two-dimensional (2D) non-separable interpolation filter, which is calculated for each frame independently by minimising the prediction error energy. For every fractional-pel position to be interpolated, an individual set of 2D filter coefficients is determined. Since transmitting filter coefficients as side information results in an additional bit rate, which is almost independent for different total bit rates and image resolutions, the overall gain decreases when total bit rates decrease. In this paper we present an algorithm, which regards the non-separable two-dimensional filter as a polyphase filter. For each frame, predicting the interpolation filter impulse response through evaluation of the polyphase filter, we only have to encode the filter coefficients prediction error. This enables bit rate savings, needed for transmitting filter coefficients of up to 75% compared to PCM coding. A coding gain of up to 1,2 dB Y-PSNR at same bit rate or up to 30% reduction of bit rate is obtained for HDTV-sequences compared to the standard H.264/AVC. Up to 0,5 dB (up to 10% bit rate reduction) are achieved for CIF-sequences.

Keywords: video coding, H.264/AVC, adaptive interpolation, adaptive Wiener filter

1. INTRODUCTION

In order to reduce the bit rate of video signals, the ISO and ITU coding standards apply hybrid video coding with motion-compensated prediction combined with transform coding of the prediction error. In the first step the motion-compensated prediction is performed. The temporal redundancy, i.e. the correlation between already transmitted images and the current image is exploited. In a second step, the prediction error is transform coded, thus the spatial redundancy is reduced.

In order to perform the motion-compensated prediction, the current image of a sequence is split into blocks. For each block, a displacement vector $\vec{d_i}$ is estimated and transmitted that refers to the corresponding position in a reference image. The displacement vectors have a fractional-pel resolution. Today's standard H.264/AVC[1] is based on $\frac{1}{4}$ pel displacement resolution. Displacement vectors with fractional resolution may refer to positions in the reference image, which are located between the sampled positions. In order to estimate and compensate the fractional-pel (sub-pel) displacements, the reference image has to be interpolated on the sub-pel positions. H.264/AVC uses a 6-tap Wiener interpolation filter with filter coefficients similar to the proposal of Werner.[2] The interpolation process is depicted in figure 1 and can be subdivided into two steps. In the first step, the half-pel positions aa, bb, b, hh, ii, jj and cc, dd, h, ee, ff, gg are calculated, using a horizontal or vertical 6-tap Wiener filter, respectively. Using the same Wiener filter applied at sub-pel positions aa, bb, b, hh, ii, jj the sub-pel position j is computed. In the second step, the residual quarter-pel positions are obtained using a bilinear filter applied at already calculated half-pel positions and existing full-pel positions.

Wedi proposed an adaptive interpolation filter,[3] which is independently estimated for every image. This approach enables to take into account the alteration of image signal properties, especially aliasing, on the basis of minimisation of the prediction error energy. The filter coefficients, which are used for the calculation of the

---

E-mail:
vatis@tnt.uni-hannover.de, edler@tnt.uni-hannover.de, wasserma@tnt.uni-hannover.de, nguyen@tnt.uni-hannover.de, osterman@tnt.uni-hannover.de

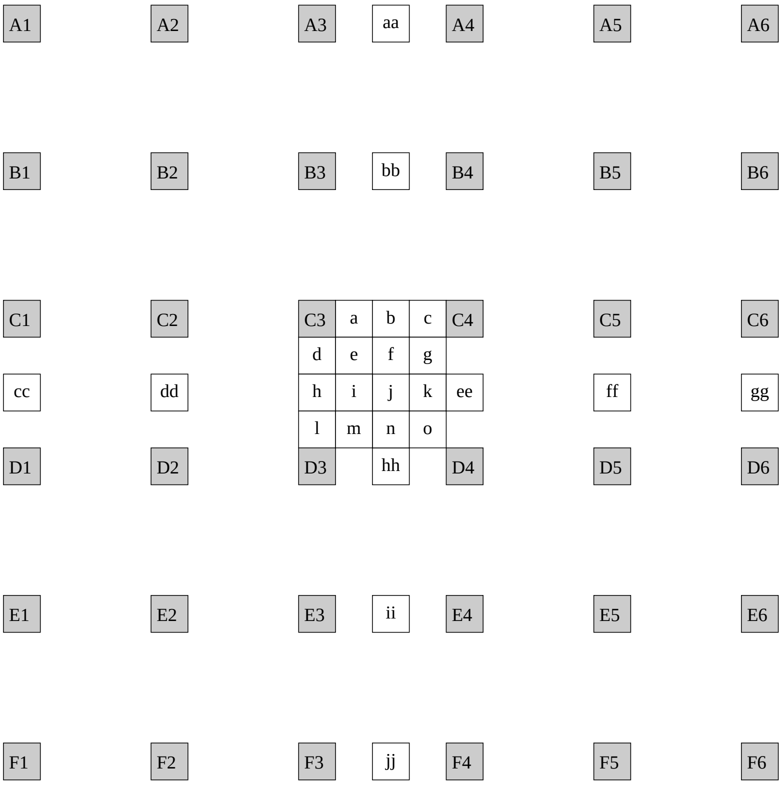

Figure 1. Integer samples (shaded blocks with upper-case letters) and fractional sample positions (white blocks with lower-case letters).

half-pel positions, are estimated iteratively using a numerical approach. The quarter-pel positions are calculated using a bilinear filter. In order to guarantee the convergence of the approach, the displacement vectors, estimated during the first step using the standard filter set, are used in further iterations.

Further, Wedi proposed a 3D filter[4] combining two techniques: a two-dimensional spatial filter with motion compensated interpolation filter (MCIF). MCIF does not only utilise samples of the frame $s'(t-1)$ to be interpolated at time instance $t-1$, but also the samples from the interpolated frame at time instance $t-2$ in order to interpolate frame $s'_i(t-1)$. A disadvantage of MCIF is the sensitivity concerning displacement vector estimation errors. Thus, quality improvement could only be achieved in combination with $\frac{1}{8}$-pel displacement vector resolution.

In this paper, we present a new strategy for calculation of optimal filter coefficients for a 2D non-separable motion compensated interpolation filter[5] and an approach for efficient conding of filter coefficients, required especially at low bit rates and small resolution videos. In the following section, the new scheme of interpolation filter is described. In section 3, an algorithm for coding of the filter coefficients is proposed. Experimental results are given in 4. The paper closes with a summary.

2. NON-SEPARABLE TWO-DIMENSIONAL ADAPTIVE FILTER

In order to reach the practical bound for the gain, achieved by means of adaptive filter, a new adaptive filter scheme has been developed. For every sub-pel position SP $(a \dots o)$, see figure 1 an individual set of coefficients is analytically calculated, such that no bilinear interpolation is used. If the sub-pel position to be interpolated is located at a, b, c, d, h, l, see figure 1, a one-dimensional 6-tap filter is calculated, using the samples $C1 \dots C6$ for the sub-pel positions a, b, c and $A3 \dots F3$ for d, h, l respectively. For each of the remaining sub-pel positions e, f, g, i, j, k, m, n and o, a two-dimensional 6x6-tap filter is calculated. For all sub-pel positions, the filter coefficients are calculated in a way that the prediction error energy is minimised, i.e. the mean squared difference between the original and the predicted image signals. Note, that we here limit the size of the filter to 6x6 and the displacement vector resolution to quarter-pel, but other filter sizes and displacement vector resolutions are also conceivable with our approach.

In the following, we describe the calculation of the filter coefficients more precisely. Let us assume, that h_{00}^{SP}, h_{01}^{SP},... , h_{54}^{SP}, h_{55}^{SP} are the 36 filter coefficients of a 6x6-tap 2D filter used for a particular sub-pel position SP. Then the values p^{SP} $(p^a \dots p^o)$ to be interpolated are computed by a two-dimensional convolution:

$$p^{SP} = \sum_{i=1}^{6}\sum_{j=1}^{6} P_{i,j} h_{i-1,j-1}^{SP} \tag{1}$$

where P_{ij} is an integer sample value $(A1 \dots F6)$. The calculation of the coefficients and the motion compensation are performed in the following steps:

1. Displacement vectors $\vec{d_t} = (mvx, mvy)$ are estimated using the non-adaptive standard interpolation filter for the image to be coded.

2. 2D filter coefficients $h_{i,j}^{SP}$ are calculated for each sub-pel position SP independently by minimisation of the prediction error energy:

$$\left(e^{SP}\right)^2 = \sum_x \sum_y \left(S_{x,y} - \sum_{i=1}^{6}\sum_{j=1}^{6} h_{i-1,j-1}^{SP} P_{\tilde{x}+i,\tilde{y}+j} \right)^2 \tag{2}$$

$$\text{with } \tilde{x} = x + \lfloor mvx \rfloor - FO, \tilde{y} = y + \lfloor mvy \rfloor - FO$$

where $S_{x,y}$ is an original image, $P_{x,y}$ a previously decoded image, mvx and mvy are the estimated displacement vector components, FO - a so called *F*ilter *O*ffset caring for centreing of the filter ($FO = \frac{1}{2} \cdot filter\_size - 1$, in case of a 6-tap filter so $FO = 2$) and $\lfloor \dots \rfloor$-operator is the *floor function*, which maps the estimated displacement vector mv to the next full-pel position smaller than mv. This is a necessary step, since the previously decoded images contain information only at full-pel positions. Note, only the sub-pel positions, which were calculated by motion estimation, are used for the error minimisation. Thus, for each of the sub-pel positions $a \dots o$ an independent set of equations is set up by computing the derivative of $\left(e^{SP}\right)^2$ with respect to the filter coefficient $h_{i,j}^{SP}$. The number of equations is equal to the number of filter coefficients used for current sub-pel position SP.

$$\begin{aligned}
0 &= \frac{\partial \left(e^{SP}\right)^2}{\partial h_{k,l}^{SP}} \\
&= \frac{\partial}{\partial h_{k,l}^{SP}} \left(\sum_x \sum_y \left(S_{x,y} - \sum_{i=1}^{6}\sum_{j=1}^{6} h_{i,j}^{SP} P_{\tilde{x}+i,\tilde{y}+j} \right)^2 \right) \\
&= \sum_x \sum_y \left(S_{x,y} - \sum_{i=1}^{6}\sum_{j=1}^{6} h_{i,j}^{SP} P_{\tilde{x}+i,\tilde{y}+j} \right) P_{\tilde{x}+k,\tilde{y}+l} \\
& \qquad \forall k,l \in \{0;5\}
\end{aligned}$$

For each sub-pel position $e, f, g, i, j, k, m, n, o$ using a 6x6-tap 2D filter, a system of 36 equations with 36 unknowns has to be solved. For the remaining sub-pel positions requiring a 1D filter, systems of 6 equations have to be solved. This results in 360 filter coefficients (9 2D filter sets with 36 coefficients each and 6 1D filter sets with 6 coefficients per set).

3. New displacement vectors are estimated. For the purpose of interpolation, the adaptive filter computed in step 2 is applied. This step enables reducing motion estimation errors, caused by aliasing, camera noise, etc. on the one hand and treating the problem in the rate-distortion sense on the other hand.

The steps 2 and 3 can be repeated, until a particular quality improvement threshold is achieved. Since some of the displacement vectors are different after the 3. step, it is conceivable to estimate new filter coefficients, adapted to the new displacement vectors. However, this would result in a higher encoder complexity.

Since transmitting 360 filter coefficients may result in a high additional bit rate, the overall coding gain can be drastically reduced, especially for video sequences with small spatial resolution. In order to reduce the side information, we assume, that statistical properties of an image signal are symmetric. Thus, the filter coefficients are assumed to be equal, in case the distance of the corresponding full-pel positions to the current sub-pel position are equal (the distance equality between the pixels in x- and y-direction is also assumed, i.e. if the image signal is interlaced, a scaling factor should be considered etc.).

Let us denote h^a_{C1} as a filter coefficient used for computing the interpolated pixel at sub-pel position a from the integer position $C1$ as depicted in figure 1. The remaining filter coefficients are indexed in the same manner. Then, based on symmetry assumptions, only 5 independent 1D or 2D filter sets, consisting of different numbers of coefficients are required. Thus, for the sub-pel positions a, c, d, l only one filter with 6 coefficients is estimated, since:

$$\begin{gathered} h^a_{C1} = h^d_{A3} = h^c_{C6} = h^l_{F3};\ h^a_{C2} = h^d_{B3} = h^c_{C5} = h^l_{E3} \\ h^a_{C3} = h^d_{C3} = h^c_{C4} = h^l_{D3};\ h^a_{C4} = h^d_{D3} = h^c_{C3} = h^l_{C3} \\ h^a_{C5} = h^d_{E3} = h^c_{C2} = h^l_{B3};\ h^a_{C6} = h^d_{F3} = h^c_{C1} = h^l_{A3} \end{gathered}$$

The same assumptions, applied at sub-pel positions b and h result in 3 coefficients for these sub-pel positions. In the same way, we get 21 filter coefficients for sub-pel positions e, g, m, o, 18 filter coefficients for sub-pel positions f, i, k, n and 6 filter coefficients for the sub-pel position j. In total, this reduces the number of needed filter coefficients from 360 to 54, exploiting the assumption, that statistical properties of an image signal are symmetric.

3. PREDICTION AND CODING OF THE FILTER COEFFICIENTS

The process of coding of the filter coefficients can be subdivided into 4 steps:

1. Quantisation. In this step, it is important to find a good compromise between the accuracy required for filter coefficients and the size of the side information needed for the transmission of the filter coefficients. Therefore, we perform a coarse quantisation (Q1) to achieve representative values, which are later mapped to more accurate ones (Q2). The accurate values are used only for the actual interpolation. In our experiments (see section 4), we set Q1 to 6 bits and Q2 to 10 bits, accuracy which provides a quality almost identical to the one provided by unquantised filter coefficients.
2. Temporal (inter) prediction. This type of prediction is used for the filter coefficients, applied to the sub-pel position b. Only differences of the current filter set to the filter set, obtained by the prediction of the previous frame, need to be transmitted.
3. Spatial (intra) prediction. This type of prediction is used for other sup-pel positions. Exploiting the symmetry of statistical properties of an image signal and knowing that no bilinear interpolation is used, coefficients of 2D filters for the different sub-pel positions can be regarded as samples of a common 2D filter, also called *polyphase filter*.[6] So, knowing the impulse response of the common filter at particular positions, we can predict its impulse response at other positions by interpolation. This process is depicted in figure 2 for 1D case. Knowing impulse response of the filter at sub-pel position b, obtained by means of inter prediction, impulse response of the filter at sub-pel position a is predicted by interpolation. For the purpose of interpolation, the standard Wiener interpolation filter is used.

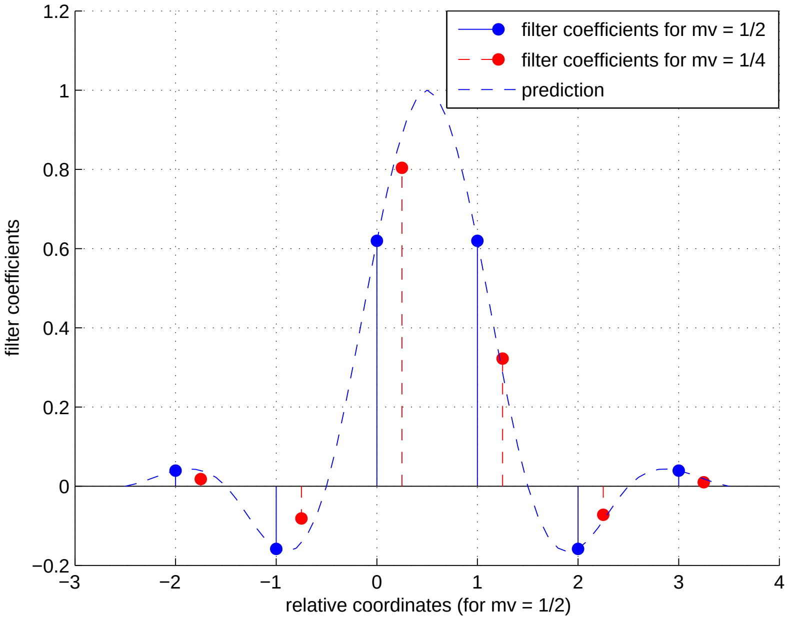

Figure 2. Intra prediction of the impulse response of a 6-tap 1D Wiener filter at sub-pel position a (displacement vector $\frac{1}{4}$) from the impulse response at sub-pel position b (displacement vector $\frac{1}{2}$).

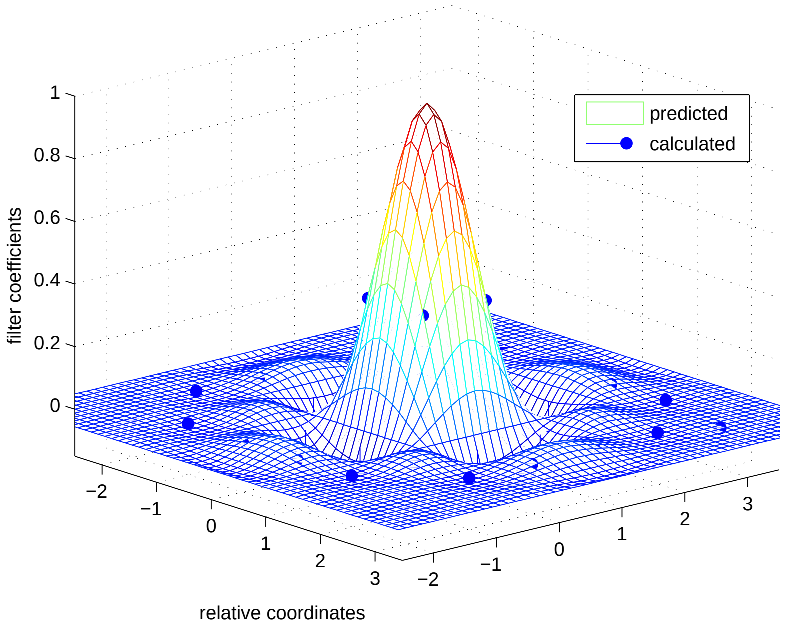

Figure 3. Intra prediction of impulse response of a 6-tap 2D Wiener filter at sub-pel position j (displacement vector $\left[\frac{1}{2}, \frac{1}{2}\right]$) and actually calculated filter coefficients.

So, with h^a and h^b, and accordingly h^c, h^d, h^h and h^l, we can predict 2D filter coefficients by multiplication:

$$h^e = h^d h^a; h^f = h^d h^b; h^j = h^h h^b$$

Figure 3 illustrates an example with interpolated impulse response of a predicted filter at sub-pel position j and actually calculated filter coefficients.

4. Entropy coding. The differences between calculated and predicted filter coefficients are coded with exp-Golomb code.[1] Exp-Golomb code is well-suitable for the kind of distributions, achieved by described approach and is implemented in the standard H.264/AVC. Thus, no additional look up tables or calculations are needed.

4. EXPERIMENTAL RESULTS

In our experiments for evaluating the coding efficiency of the adaptive Wiener interpolation filter, we coded several HDTV- and CIF-sequences. All simulations were performed using the baseline profile of H.264/AVC. Since the side information needed for the transmission of the filter coefficients is negligibly low for HDTV-sequences, we did not distinguish between PCM-coding of the filter coefficients and appliying the proposed algorithm for the coding of the filter coefficients. Nevertheless, we show some results (applying PCM-coding of filter coefficients) in order to point up quality improvements achieved by applying new adaptive filter scheme. In figure 4 four different curves are depicted, representing the H.264/AVC standard and enhancement with adaptive filter coefficients for 1 and 5 reference frames, applied to the HDTV-sequences *Raven*, *Crew*, *ShuttleStart* and *City.* Applying adaptive 2D interpolation filter outperforms the standard H.264/AVC for all bit rates. Performance gains of up to 1,2 dB for the same bit rates are achieved.

The main goal of our approach are sequences with low spatial resolution. In figure 5 three different curves are depicted, representing the enhancement with adaptive filter scheme with PCM-coded filter coefficients, with proposed algorithm for the coding of filter coefficients and standard H.264/AVC, applied to the CIF-sequences *Foreman*, *Mobile&Calendar*, *Bus* and *Basketball.* Since sequences with small spatial resolution generally have low-pass characteristics, the optimal interpolation filter does not differ very much from the standard Wiener filter. Therefore, overall gains as high as those for HDTV-sequences cannot be achieved. Furthermore, the absolute bit rates for CIF-sequences are lower than the ones for HDTV-sequences. Thus, transmitting the side-information for the filter coefficients can reduce overall gains significantly, especially for low bit rates. Nevertheless, applying the proposed algorithm for coding of the filter coefficients, enables bit rate savings for the side information of up to 75% compared to PCM coding. Thus, total gains of up to 0.5 dB can be kept even for very low bit rates.

5. CONCLUSIONS

A two-dimensional non-separable adaptive interpolation filter for motion and aliasing compensated prediction is presented. The motion compensated filter is based on coefficients that are adapted to the non-stationary statistical features of the image signal once per frame. The coefficient estimation is carried out analytically by minimising the prediction error energy of the current frame. Thus, aliasing, quantisation and displacement estimation errors are considered. An efficient algorithm for coding of the filter coefficients is presented. A coding gain of up to 1,2 dB or up to 30% bit rate reduction for the same quality for HDTV-sequences and up to 0,5 dB (up to 10% bit rate reduction) for CIF-sequences are achieved even for low bit rates compared to the standard H.264/AVC.

An additional advantage of the non-separable interpolating scheme is avoiding of double quantisation, which is inevitable for quarter-pel positions, if separable filter scheme is used. Disadvantages are slightly increased decoder complexity (number of operations needed for interpolation increases up to 10%) and lower suitability for 16-bit arithmetic.

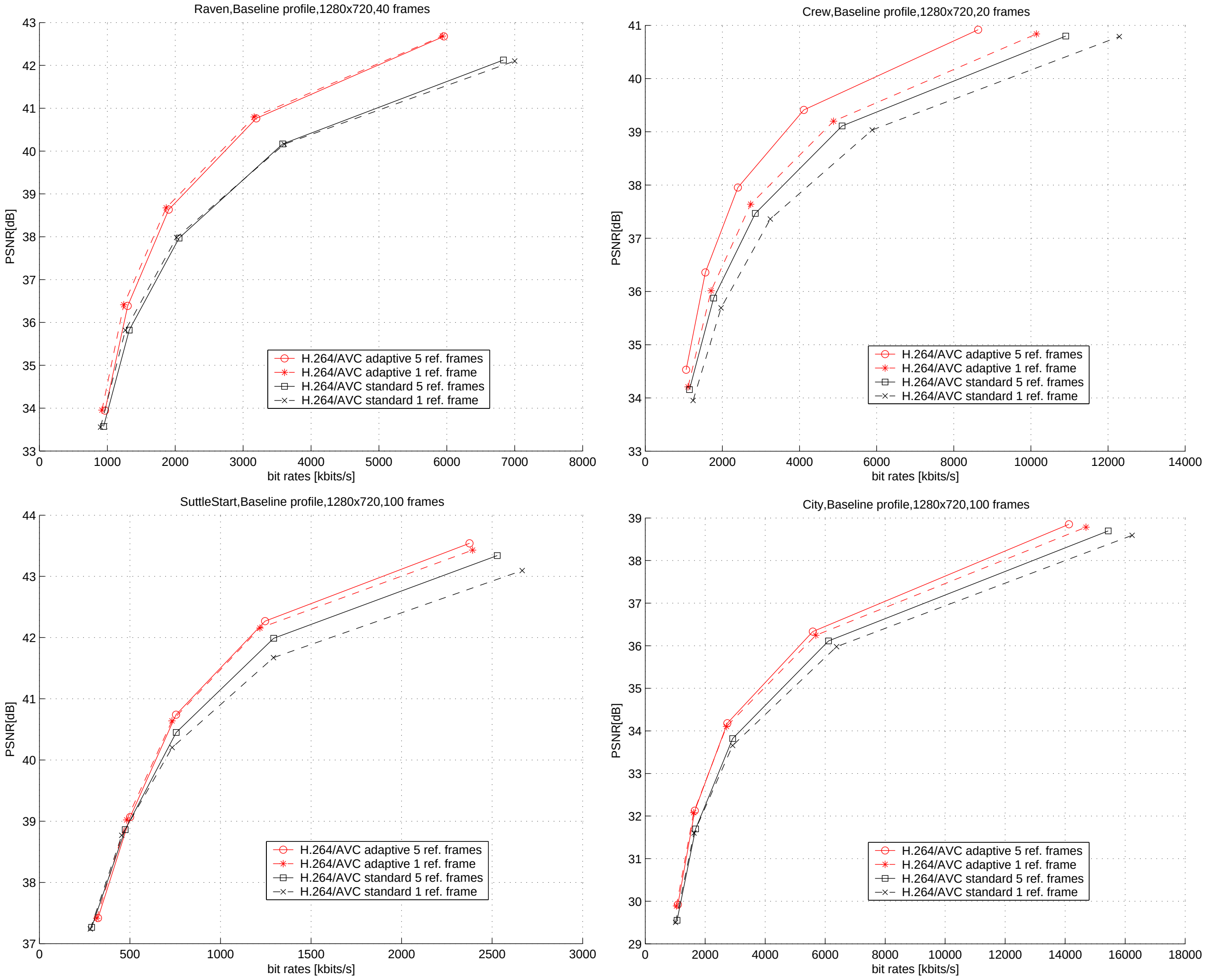

Figure 4. Bit rate, provided by adaptive interpolation filter and by the standard interpolation filter of H.264/AVC for HDTV-sequences Raven (top left), Crew (top right), ShuttleStart (down left) and City (down right) for 1 and 5 reference frames.

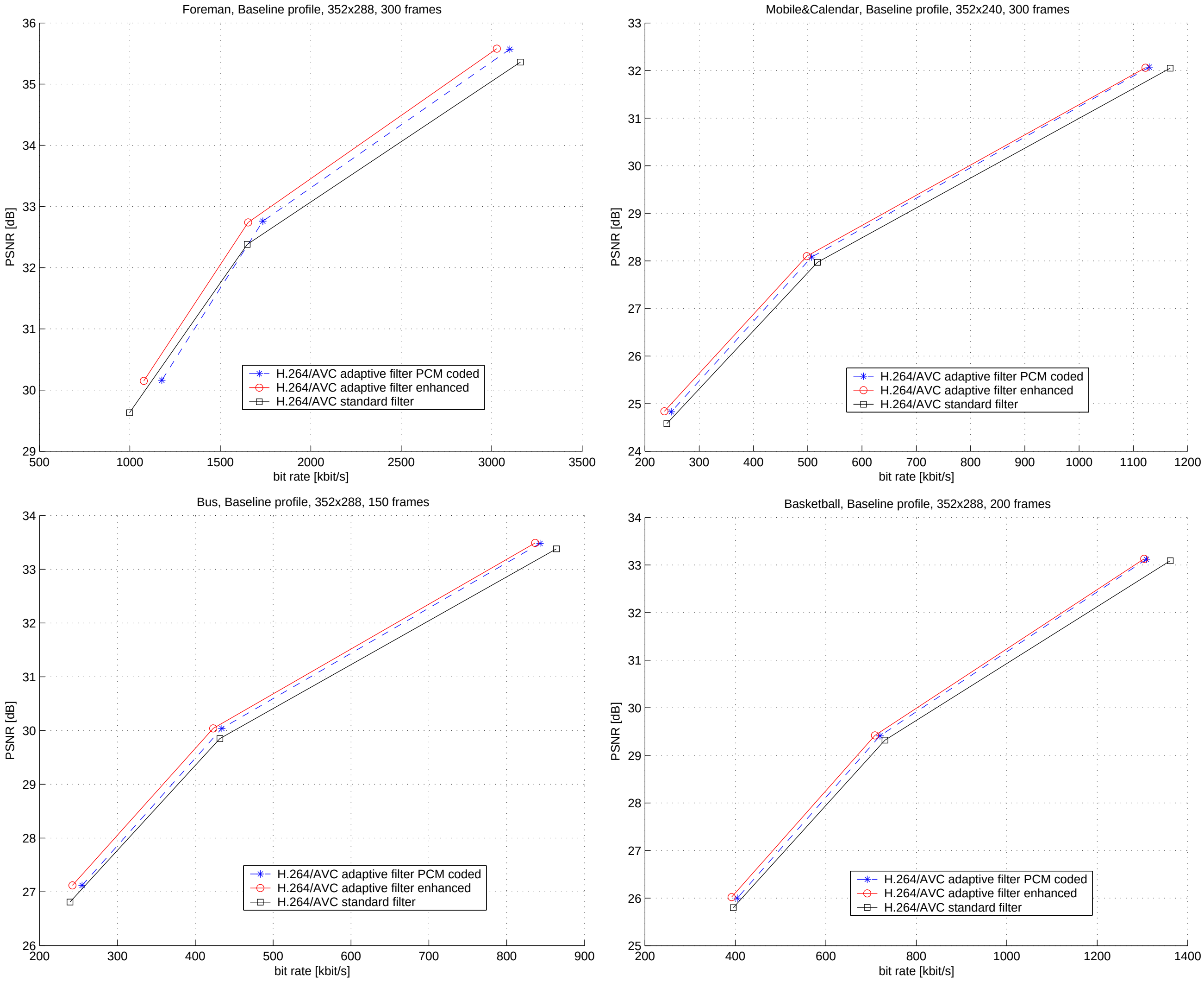

Figure 5. Bit rate, provided by adaptive interpolation filter using PCM coding scheme (dashed), enhanced entropy coding algorithm (red) and by the standard interpolation filter of H.264/AVC (black) for CIF-sequences foreman (top left), mobile&calendar (top right), bus (down left) and basketball (down right), 5 reference frames.

REFERENCES

1. JVT of ISO/IEC & ITU-T, Draft ITU-T Recommendation H.264 and Draft ISO/IEC 14496-10 AVC,Doc JVT-Go50, Pattaya, Thailand, March2003
2. O. Werner, "Drift analysis and drift reduction for multiresolution hybrid video coding", *Signal Processing: Image Commun.*, vol. 8, no. 5, July 1996
3. T. Wedi,"Adaptive Interpolation Filter for Motion Compensated Hybrid Video Coding", *Proc. Picture Coding Symposium (PCS 2001)*, Seoul, Korea, April 2001
4. T. Wedi, "Adaptive Interpolation Filter for Motion and Aliasing Compensated Prediction", *Proc. Electronic Imaging 2002: Visual Communications and Image Processing (VCIP 2002)*, San Jose, California USA, January 2002.
5. Y. Vatis, B. Edler, D. T. Nguyen and J. Ostermann, "Motion- and Aliasing-compensated Prediction using a two-dimensional non-separable Adaptive Wiener Interpolation Filter", submitted to *Image Communications and Image Processing (ICIP)*, Genova, Italy, September, 2005
6. Crochiere, R. E., and L. R. Rabiner, *Multi-Rate Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1983, Pgs. 88-91