LOCALLY ADAPTIVE NON-SEPARABLE INTERPOLATION FILTER FOR H.264/AVC

Yuri Vatis, Jörn Ostermann

Institut für Informationsverarbeitung
Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany

ABSTRACT

In order to reduce the bit rate of video signals, current coding standards apply hybrid coding with motion-compensated prediction and transform coding of the prediction error. In former publications, it was shown that aliasing components contained in an image signal, as well as inaccuracies in description of motion, are limiting the prediction efficiency obtained by motion compensation. For H.264/AVC, we showed that the analytical calculation of an optimal interpolation filter at given constraints (6-tap, frame-based) is possible, resulting in total coding improvements of up to 0.9 dB for HDTV sequences and up to 0.5 dB for CIF sequences and Main Profile[1]. However, applying an adaptive interpolation filter to the entire image enables only an average prediction improvement. Further improvements are possbile if the filter is also adapted to local image characteristics. Thus, an optimal adaptive interpolation filter is calculated only for the macroblocks, which prediction is distorted by aliasing or by inaccurately estimated motion. This enables further improvements of up to 0.2 dB, compared to globally adaptive filter or up to 0.6 dB, compared to the standard H.264/AVC, for CIF sequences.

Index Terms— Video coding, Motion compensation, Wiener filtering

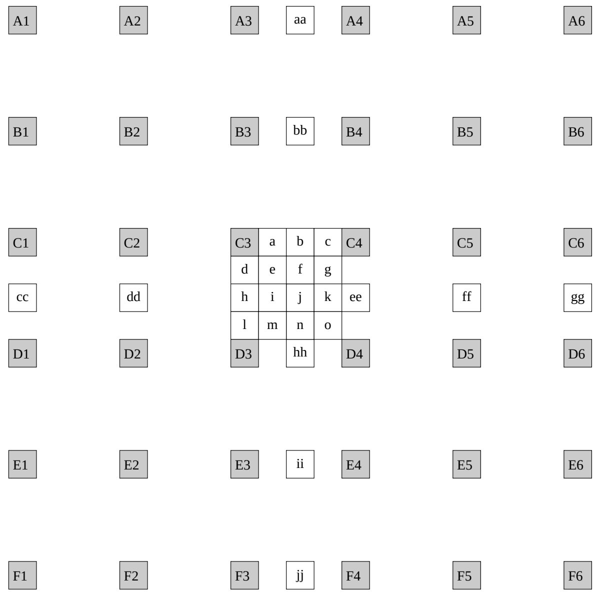

Fig. 1. Integer samples (shaded blocks with upper-case letters) and fractional sample positions (unshaded blocks with lower-case letters).

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 consecutive images, is exploited for the prediction of the current image from already transmitted images. In a second step, the residual error is transform coded, and thus the spatial redundancy is reduced.

In order to perform motion-compensated prediction, for each block of the current image a displacement vector $\vec{d_i}$ is estimated and transmitted, which refers to the corresponding position in a reference image. The displacement vectors may have fractional-pel resolution. Today's standard H.264/AVC uses vectors with $\frac{1}{4}$-pel displacement resolution [3]. Thus, displacement vectors may refer to positions in the reference image, which are located between the sampled positions. In order to estimate and compensate for the fractional-pel displacements, the reference images are interpolated to compute the sub-pel positions. H.264/AVC uses an invariant 6-tap Wiener interpolation filter [3]. The interpolation process is depicted in Figure 1 and can be subdivided into two steps. First, the half-pel positions aa, bb, b, hh, ii, jj and cc, dd, h, ee, ff, gg are calculated, using a 6-tap Wiener filter. 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.

Due to non-ideal low-pass filters in the image acquisition process, the Nyquist Sampling Theorem is not fulfilled and aliasing disturbs the motion-compensated prediction [4]. This

[1]If case of Baseline profile, the improvements are increased at up to 0.2 dB

leads to an additional prediction error, which has to be coded. The amount of aliasing is scene and camera dependent. Furthermore, an inaccurate motion description, caused by a simplified motion model assumption (translative, 2D-blocks, $\frac{1}{4}$-pel accuracy), leads to further error increase. In our previous publications, we showed [2] that an analytical calculation of an optimal filter at given constraints (6-tap, frame-based) is possible, resulting in total quality improvements of up to 1.1 (0.9) dB for HDTV sequences and up to 0.6 (0.5) dB for CIF sequences in the case of the Baseline (Main) Profile. However, further improvements are possible if the filter is also adapted to local image characteristics.

In the following Section 2, the globally adaptive interpolation filter scheme is described. In Section 3, the locally adaptive filter is presented in detail. Experimental results are given in Section 4. The paper closes with conclusions.

2. GLOBALLY ADAPTIVE INTERPOLATION FILTER

In this Section, the globally adaptive interpolation filter scheme presented in [2], is described in detail. For every different sub-pel position, an independent set of coefficients is analytically calculated, such that there is no more need for bilinear interpolation. If the sub-pel position to be interpolated is located on a column or on a row (i.e. the sub-pel positions a, b, c, d, h, l, see Figure 1), for each position a one-dimensional 6-tap filter is calculated, using the samples $C1 \ldots C6$ for the sub-pel positions a, b, c and $A3 \ldots F3$ 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 estimated such that they minimise the prediction error energy, i.e. the square difference between the original and the predicted image signals. Note, that in this proposal, we limit ourselves to a 6x6-tap filter and a quarter-pel displacement vector resolution, but other filter sizes and displacement vector resolutions are conceivable.

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

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

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

1. Displacement vectors $\vec{d_t}$ are estimated using the fixed standard interpolation filter for the image to be coded.

2. 2D filter coefficients $h_{i,j}$ are estimated for each sub-pel position SP, independently, by the minimisation of the prediction error according to equation (2):

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

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

where $S_{x,y}$ is an original image, $P_{x+\lfloor mvx \rfloor, y+\lfloor mvy \rfloor}$ a previously decoded image, mvx and mvy are the estimated displacement vectors, FO - a so called *F*ilter *O*ffset caring for the centring of the filter and $\lfloor \ldots \rfloor$ - 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, for the error minimisation, only the sub-pel positions are used, which are actually used by motion compensation. Thus, for each of the sub-pel positions $a \ldots o$ an independent set of equations is set up by computing the derivative of $\left(e^{sp}\right)^2$ with respect to the filter coefficient $h_{i,j}$. The number of equations is equal to the number of filter coefficients used for current sub-pel position SP.

$$\begin{aligned} 0 &= \frac{\partial \left(e^{SP}\right)^2}{\partial h_{0,0}^{SP}} \\ &= \sum_{x,y}\left(S_{x,y} - \sum_{i,j} h_{i,j}^{SP} P_{\tilde{x}+i,\tilde{y}+j}\right) P_{\tilde{x},\tilde{y}} \\ &\vdots \end{aligned}$$

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, systems of 6 equations have to be solved. This results in 360 filter coefficients (9 2D filter sets with 36 coefficients per set and 6 1D filter sets with 6 coefficients per set).

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

Since transmitting 360 filter coefficients may result in a high additional bit rate, the coding gain can be drastically reduced, especially for small resolution video sequences. 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.). For a more detailed description, see [2].

3. LOCALLY ADAPTIVE INTERPOLATION FILTER

In order to allow not only an average improvement, but also an adjustment to local video statistics, we developed a locally adaptive filter scheme. The standard Wiener filter might be an approximately ideal filter for macroblocks where the Nyquist Sampling Theorem is not violated. Thus, such macroblocks disturb the filter coefficients for macroblocks influenced by aliasing. For that reason, macroblocks with only low-frequent characteristics are predicted using a standard interpolation filter and consequently, excluded from the adaptive filter estimation step.

To decide, wheather a macroblock is interpolated using an adaptive or the standard filter, three steps are carried out: In the first step, the displacement vectors are estimated for each macroblock, applying the standard Wiener filter for interpolation purpose. For each macroblock, the rate-distortion costs are stored. In the second step, the adaptive filter coefficients are calculated for the entire image, as described in the previous Section. All reference images are interpolated with the new filter. In the last step, the coding process is performed again. For each macroblock coded, using the standard and the adaptive filter, the rate-distortion costs are compared. For all macroblocks where the adaptive filter gives an improvement with respect to rate-distortion costs, a new final adaptive interpolation filter is calculated.

In Figure 2, an image of the sequence *Concrete*, a magnitude response of a corresponding adaptive and the standard filter and a macroblock partitioning (white - adaptive filter, grey - standard filter, black - intra) are depicted. It is obvious, that the encoder chose the standard filter for the macroblocks with low-frequent characteristics, namely the macroblocks on the newscaster. At the same time, for the pebbly wall in the background, with noticeable high-frequency components, the adaptive filter is used for interpolation. The frequency response of the optimal adaptive filter does not have a typical low-pass characteristic. The low-frequency components are distorted due to mirroring effects of aliasing, which has to be reduced. As a side-effect, the filter lets some high-frequent components pass due to symmetric properties of the impulse response.

Applying the locally adaptive filter on a macroblock basis, requires the transmission of additional side information, for each macroblock coded in a INTER-mode with fractional-pel displacement. Since the measured entropy of the side information varies between 0.6 and 0.9 bits per macroblock, it is coded using CABAC.

4. EXPERIMENTAL RESULTS

In our experiments for evaluating the coding efficiency of a locally adaptive interpolation filter, we coded several HDTV and CIF sequences. All simulations were performed using the Main Profile of H.264/AVC (reference software 10.1) using CABAC, IPPP... scheme and five reference frames. The new approach outperforms the approach presented in [2] for all sequences and all bit rates. For CIF sequences, coding gains of up to 0.2 dB, compared to [2] and up to 0.6 dB compared, to the standard H.264/AVC, have been achieved. In Figure 3, three different curves are depicted representing the H.264/AVC standard and enhancements with globally and locally adaptive interpolation filter, applied to the CIF sequences *Concrete*, *Mobile&Calendar*, *Flowergarden* and *Bus*. For HDTV sequences, the number of bits per macroblock is generally lower than for CIF sequences at the same quality, resulting in high costs of side information. Due to the adaptive use of the locally adaptive filter, the coding gain for HDTV sequences is as good as in [2] or slightly better.

5. CONCLUSIONS

A two-dimensional locally adaptive interpolation filter for motion and aliasing compensated prediction is presented. The calculation of the coefficients is carried out analytically by minimising the prediction error energy of the current frame. In order to compensate the aliasing, only the macroblocks are considered for the calculation of the filter coefficients, which are distorted by the aliasing. For the remaining macroblocks, the standard Wiener interpolation filter is applied. As a result, a coding gain of up to 0.6 dB, compared to the standard H.264/AVC or up to 0.2 dB, compared to [2], is obtained. Another advantage is a slightly decreased decoder complexity, compared to [2], since on average 30% of the macroblocks are coded using the standard Wiener filter.

6. REFERENCES

[1] T. Wedi and H. G. Musmann, "Motion- and Aliasing-Compensated Prediction for Hybrid Video Coding", *IEEE transactions on circuits and systems for video technology*, vol. 13, No. 7, July 2003.

[2] Y. Vatis, B. Edler, D. T. Nguyen, J. Ostermann, "Motion- and Aliasing-Compensated Prediction Using a Two-dimensional Non-separable Adaptive Wiener Interpolation Filter", *Proc. ICIP 2005, IEEE International Conference on Image Processing*, Genova, Italy, September 2005.

[3] ITU-T Recommendation H.264 & ISO/IEC 14496-10 (MPEG-4) AVC, "Advanced Video Coding fro Generic Audiovisual Services", version 3: 2005.

[4] R. W. Schaefer, A. V. Oppenheim, *Discrete-Time Signal Processing*, Prentice Hall, 1989.

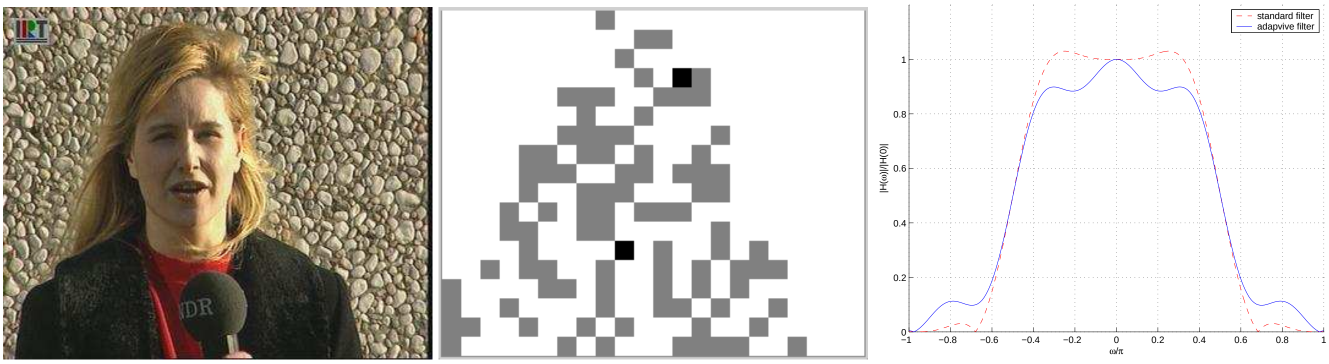

Fig. 2. Original image (left), macroblock partitioning (white - coded using adaptive, grey - coded using standard filter, black - intra) and magnitude responses of adaptive and standard filter for sub-pel positions b and h (right) of a sequence *Concrete*.

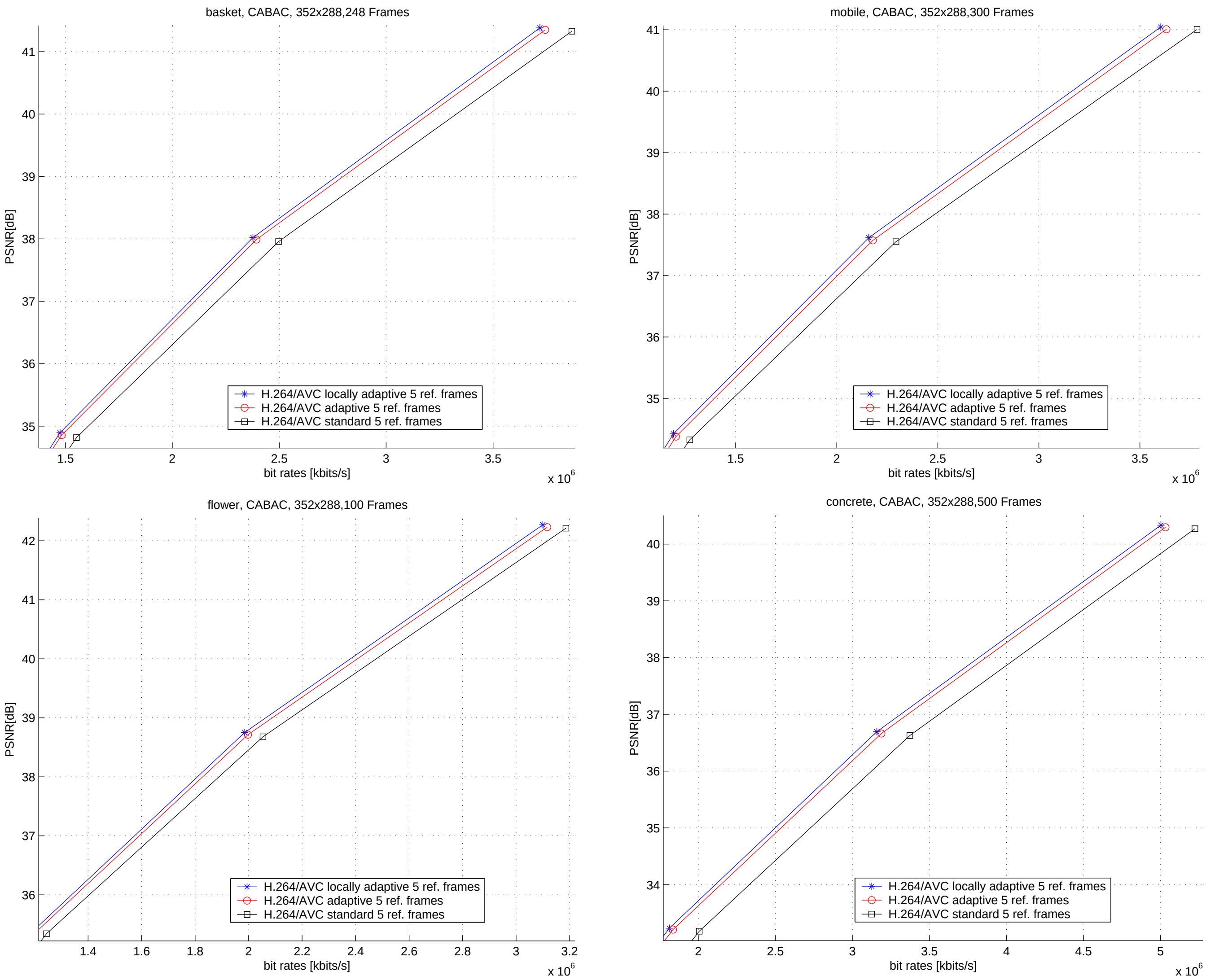

Fig. 3. Bit rates, provided by locally adaptive, adaptive and standard interpolation filter of H.264/AVC for CIF sequences *Basketball* (top left), *Mobile&Calendar* (top right), *Flowergarden* (down left) and *Concrete* (down right).