ADAPTIVE INTERPOLATION FILTER FOR MOTION COMPENSATED PREDICTION

Thomas Wedi

Institute of Communication Theory and Signal Processing
University of Hannover, Appelstr. 9a, 30167 Hannover, Germany
E-Mail: wedi@tnt.uni-hannover.de

ABSTRACT

Standardized hybrid video coding systems are based on motion compensated prediction with fractional-pel displacement vector resolution. In the recent JVT video coding scheme (MPEG-4 part 10, H.264) displacement vector resolutions of 1/4- or 1/8-pel are applied. In order to estimate and compensate these fractional-pel displacements, interpolation filters are used. So far, these interpolation filters are invariant. The same filter coefficients are applied for all sequences and for all images of a sequence. Therefore it is not possible to consider non-stationary statistical properties of video signals in the interpolation process. In this paper an adaptive interpolation scheme is presented. This interpolation scheme uses filter coefficients that are adapted once per image to the non-stationary statistical properties of the video signal. The filter-coefficients are coded and transmitted. Due to the adaptive interpolation filter a coding gain up to 0.8 dB PSNR is obtained in the JVT coding scheme.

1. INTRODUCTION

Standardized hybrid video coding systems like the recent JVT coding scheme [1] are based on motion compensated prediction. Fig. 1 shows the generalized block diagram of such an hybrid video encoder. The current image s_t at time instance t is predicted by a motion compensated prediction (MCP) from an already transmitted image $s'_{t\text{-}1}$. The result of the motion compensated prediction is image $\hat{s}_t$. Only the prediction error e_t and the motion information $\vec{d}_t$ are coded and transmitted.

For the motion compensated prediction, the current image is partitioned into blocks. A displacement vector $\vec{d}_t$ is assigned to each block that refers to the corresponding position of its image signal in an already transmitted reference image $s'_{t\text{-}1}$. The displacement vectors have a fractional-pel resolution. In the JVT coding scheme displacement vectors with 1/4- or 1/8-pel resolution are applied

Displacement vectors with fractional-pel resolution may refer to positions in the reference image $s'_{t\text{-}1}$ that are located between the sampled positions. In the following these positions are called subpel positions. In order to estimate and compensate fractional-pel displacements, the image signal on subpel positions has to be generated by interpolation. Up to now this interpolation is done by an invariant interpolation filter h. The same filter is used for all sequences and for all images of a sequence. Therefore it is not possible to consider non-stationary statistical properties of video signals in the interpolation process.

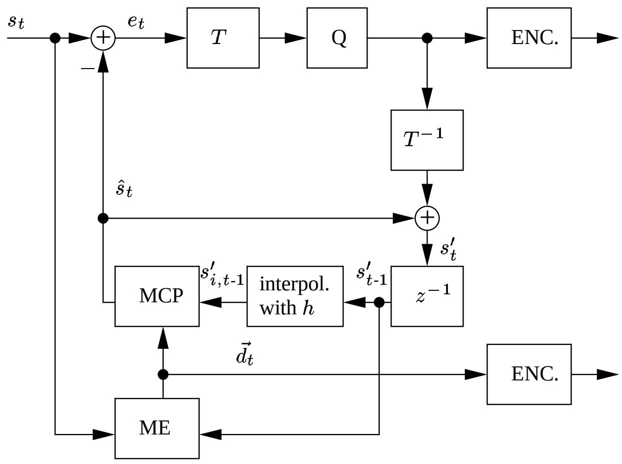

Fig. 1. Generalized block diagram of a hybrid video encoder based on motion compensated prediction.

In this paper an adaptive interpolation filter for motion compensated prediction is presented in order to reduce the prediction error and improve the coding efficiency. The adaptive interpolation filter h_t is based on filter-coefficients that are adapted once per image.

In Section 2 of this paper, a theoretical basis of the adaptive interpolation scheme is introduced. Section 3 describes details of the implementation for the JVT coding scheme. Experimental results are given in Section 4. The paper closes with a summary.

2. THEORETICAL BASIS OF THE ADAPTIVE INTERPOLATION

In this section a theoretical basis of the adaptive interpolation scheme is given using Fig. 2. This figure illustrates the motion compensated prediction using an adaptive interpolation filter $h_t(l)$. In order to simplify the explanations, this

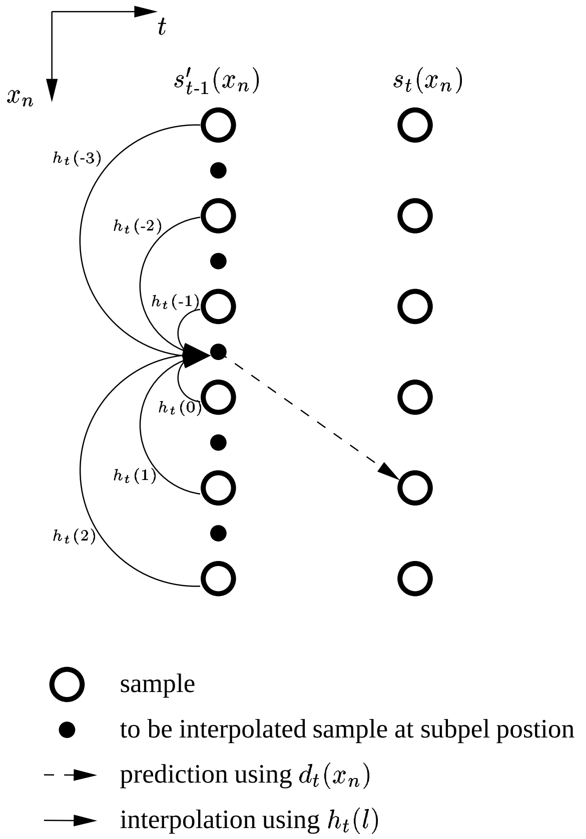

Fig. 2. Illustration of motion compensated prediction for a fractional-pel displacement using an adaptive interpolation filter $h_t(l)$. Corresponding lines of two consecutive images are shown: The reconstructed reference image $s'_{t\text{-}1}(x_n)$ and the image to be predicted $s_t(x_n)$.

example is restricted to one spacial coordinate x_n. Thus, Fig. 2 shows one dimensional image lines instead of two dimensional images. The image $s_t(x_n)$ is predicted from the reference image $s'_{t\text{-}1}(x_n)$ using the displacement $d_t(x_n)$. It is assumed that the fractional-pel displacement vector resolution is 1/2-pel. In the example of Fig. 2, the displacement vector refers to a subpel position of image $s'_{t\text{-}1}(x_n)$ that has to be generated by interpolation. For the interpolation purpose an interpolation filter $h_t(l)$ is used. This filter generates the sample on subpel position that is used to predict the corresponding sample of image $s_t(x_n)$.

In the following, an adaptive filter $h_t(l)$ of finite impulse response described by $2L$ coefficients is assumed:

$$h_t(l) = \begin{cases} \neq 0, & -L \leq l \leq L-1 \\ = 0, & \text{otherwise} \end{cases} \qquad (1)$$

In case of a fractional pel displacement (e.g. $d_t(x_n) = \pm 0.5$ pel, ± 1.5 pel, ...), the prediction signal $\hat{s}_t(x_n)$ is given by

$$\hat{s}_t(x_n) = \sum_{l=-L}^{L-1} h_t(l) \cdot s'_{t\text{-}1}(x_n - \tilde{d}_t(x_n) + l) \qquad (2)$$

where

$$\tilde{d}_t(x_n) = \lfloor d_t(x_n) \rfloor \qquad (3)$$

is the displacement that is rounded to the nearest integer towards minus infinity. This rounding is necessary, because $d_t(x_n)$ refers to a subpel position in image $s'_{t\text{-}1}(x_n)$ (see Fig. 2). Since the filter $h_t(l)$ has to weight the samples of $s'_{t\text{-}1}(x_n)$ in order to calculate the prediction image $\hat{s}_t(x_n)$, the rounded displacement from (3) has to be used in equation (2).

In case of a fullpel displacement (e.g. $d_t(x_n) = \pm 1$ pel, ± 2 pel, ...) no interpolation has to be performed. Thus, the prediction signal for fullpel displacements is given by

$$\hat{s}_t(x_n) = s'_{t\text{-}1}(x_n - d_t(x_n)). \qquad (4)$$

The coefficients of the adaptive interpolation filter are estimated by minimizing the prediction error

$$e_t(x_n) = s_t(x_n) - \hat{s}_t(x_n) \qquad (5)$$

that has to be coded (Fig. 1). Thus, the adaptive interpolation filter $h_t(l)$ can be interpreted as an adaptive prediction filter and the filter coefficients can be determined by solving the Wiener-Hopf equation [2]. With the Autocorrelation function of the reference signal $s'_{t\text{-}1}(x_n)$

$$R_{t\text{-}1,t\text{-}1}(n) = E[s'_{t\text{-}1}(x_n) \cdot s'_{t\text{-}1}(x_n - n)] \qquad (6)$$

and the crosscorrelation function of the signal to be predicted $s_t(x_n)$ and the displaced reconstructed reference signal $s'_{t\text{-}1}(x_n - \tilde{d}(x_n))$

$$R_{t\text{-}1,t}(n) = E[s'_{t\text{-}1}(x_n - \tilde{d}(x_n)) \cdot s_t(x_n - n)] \qquad (7)$$

the Wiener Hopf Equation is given by

$$\sum_{l=-L}^{(L-1)} h_t(l) \cdot R_{t\text{-}1,t\text{-}1}(i-l) = R_{t\text{-}1,t}(i), \quad \text{for } i = -L, \ldots, (L-1). \qquad (8)$$

For stationary signals $s_t(x_n)$ and $s'_{t\text{-}1}(x_n)$ it could be shown that, due to the symmetry of the autocorrelation and crosscorrelation function, the coefficients of the filter are also symmetric

$$h_t(l) = h_t(-l-1), \quad \text{for} \quad 0 \le l \le L-1. \qquad (9)$$

Thus, the Wiener-Hopf equation from (8) can be expressed by

$$\sum_{l=0}^{(L-1)} h_t(l)\left[R_{t\text{-}1,t\text{-}1}(i\text{-}l) + R_{t\text{-}1,t\text{-}1}(i\text{+}l\text{+}1)\right] = R_{t\text{-}1,t}(i),$$

$$\text{for } i = 0, \ldots, (L-1). \quad (10)$$

With (9), the number of coefficients of $h_t(l)$ that have to be determined by solving the Wiener Hopf equation is reduced from $2L$ to L.

The assumption

$$\sum_{l=-L}^{(L-1)} h_t(l) = 1 \qquad (11)$$

is not made. Thus, video signals with lightness changes or fading sequences where $E[s_{t\text{-}1}] \neq E[s_t]$ may also be predicted with the Adaptive Filter .

Due to equations (6) and (7) the correlation functions of the Wiener-Hopf equation (10) are calculated from the image to be coded s_t, the reference image $s'_{t\text{-}1}$, and the estimated displacement vectors $\vec{d_t}$. Thus, influences that deteriorate the accuracy of the prediction like

- *Aliasing* in s_t and $s'_{t\text{-}1}$,
- *Quantization errors* in $s'_{t\text{-}1}$, and
- *Displacement estimation errors* in $\vec{d_t}$

are considered in the estimation of the filter coefficients. Due to these considerations, the prediction error is reduced and the coding efficiency is improved by using the adaptive interpolation filter.

3. ADAPTIVE INTERPOLATION FILTER FOR THE JVT CODING SCHEME

Fig. 3 shows the block diagram of the interpolation process that is applied in the JVT coding scheme for 1/4-pel displacement vector resolution. A similar interpolation process is applied in MPEG-4 (AS-profile). The image is interpolated in two steps. In the first interpolation step the resolution of the reconstructed image $s'_{t\text{-}1}$ is increased by a factor of 2 and filtered by an interpolation filter h_1. In the JVT coding scheme, filter h_1 is an invariant 6-tap Wiener filter with the following coefficients: [1,-5,20,20,-5,1]/32 [1]. Thus, the corresponding filter h_1 of Fig. 3 is h_1=[1,0,-5,0,20,32,20,0,-5,0,1]/32. In the second interpolation step

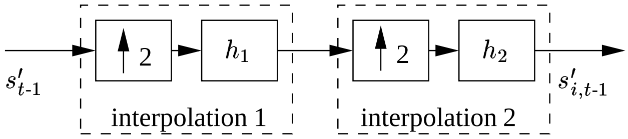

Fig. 3. Interpolation process for 1/4-pel displacement vector resolution that is used in the JVT coding scheme. It is based on two steps with two interpolation filters h_1 and h_2.

the resulting image is again sampled up by a factor of 2 and filtered by filter h_2. In the JVT coding scheme h_2 is a simple bilinear interpolation filter.

In the adaptive interpolation scheme a symmetric 6-tap filter h_1=$h_{1,t}$ is used in the first interpolation step that is adapted once per image. In the second interpolation step, an invariant bilinear interpolation filter is applied. Thus, only 3 coefficients for h_1 have to be estimated and transmitted.

The coefficient estimation and the motion compensated prediction are performed in the following steps:

1. Displacement vectors $\vec{d_t}$ are estimated for the image to be coded.

2. Filter coefficients are estimated by minimizing the energy of the prediction error e_t=s_t-$\hat{s}_t$ when performing the motion compensated prediction with displacement vectors $\vec{d_t}$ from step 1.

3. The current image is predicted by motion compensated prediction. For this purpose the filter coefficients of step 2 and the displacement vectors of step 1 are applied.

The estimated coefficients of the adaptive interpolation filter are coded and transmitted. For this purpose, a differential coding scheme is applied. The coefficients are quantized with 8 bit and the differences to the filter coefficients of the preceding image are transmitted.

4. EXPERIMENTAL RESULTS

For experimental investigations the JM-2 (Joint Model 2) of the JVT coding system is used. The applied test-sequences are *Foreman* and the VQEG test-sequence *Waterfall* each at CIF format and 30 Hz.

In Fig. 4 operational rate distortion (RD) curves for both sequences are given. In each graph the RD-curve of the reference JVT coding scheme with the invariant 6-tap filter and the JVT coding scheme with the 6-tap Adaptive Interpolation Filter (JVT-AIF) are shown. Due to the adaptive interpolation a coding gain of 0.5 dB for the test sequence *Foreman* and 0.8 dB for *Waterfall* are obtained.

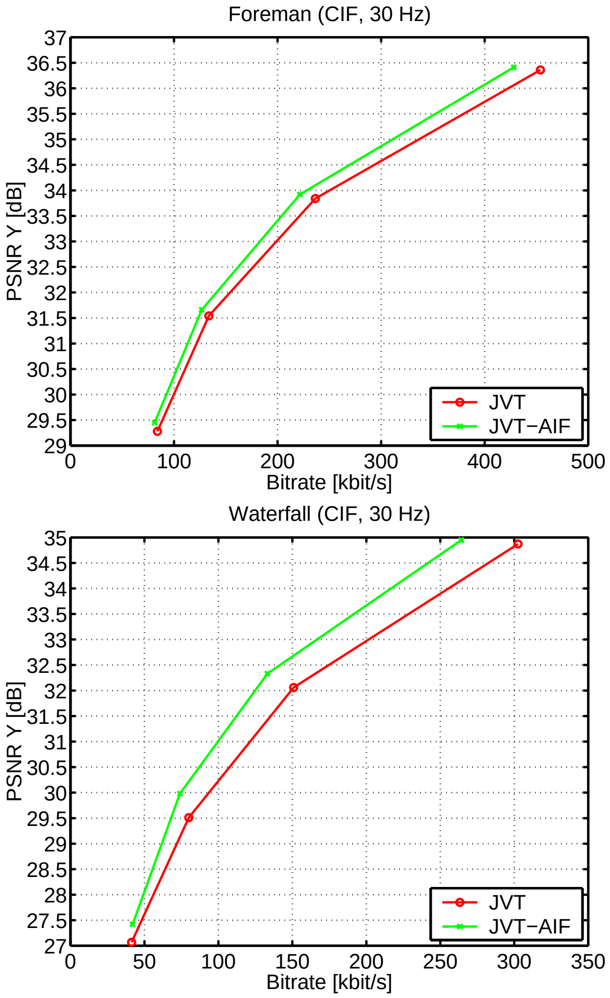

Fig. 4. Operational rate distortion curves for test-sequences *Foreman* and *Waterfall*.

In section 2 it was shown that, due to the adaptive interpolation filter, the prediction error e_t is reduced by the implicit consideration of displacement estimation errors and aliasing in the coefficient estimation process. In order to verify this, Fig. 5 shows the prediction gain due to the adaptive interpolation filter g in dependence on the image number for the test-sequence *Foreman*. The gain g is the SAD reduction of the prediction error e_t for each image in percent of the SAD reduction accumulated for the whole sequence. Thus, the figure shows the contributed gain for each image to the overall gain.

In Fig. 5 three different parts of the sequence with high coding gains are pointed out. The video signal and the estimated displacements in these parts have particular statistical properties, which are analyzed in the following. The first part with high coding gain (images 50-70) is a part where the person in the sequence is in high motion. This indicates high displacement estimation errors. The second part (images 170-220) is characterized by a fast camera pan, which leads to motion blur and indicates high displacement estimation errors. The third part is a part with little global motion and therefore with little displacement estimation errors. But, due to high frequency components aliasing disturbs the prediction in this part.

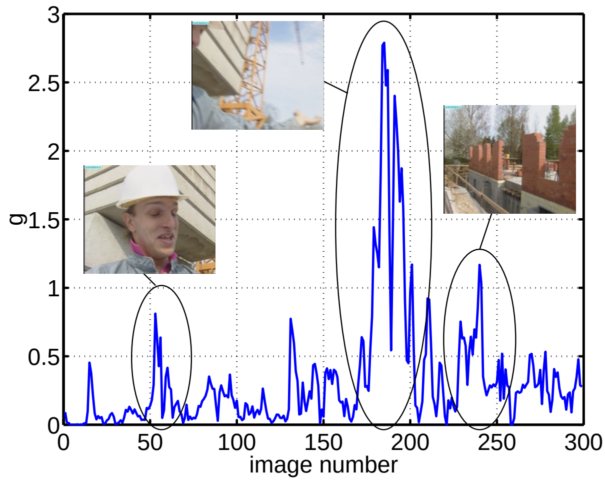

Fig. 5. Prediction gain due to the adaptive interpolation filter g in dependence on the image number for the test-sequence foreman. In this graph three different parts of the sequence each represented by a typical image are pointed out.

5. SUMMARY

An Adaptive Interpolation Filter for motion compensated prediction is presented. In a theoretical analysis the adaptive interpolation filter is interpreted as an adaptive prediction filter. Thus, the coefficients of this filter are estimated by solving a Wiener-Hopf equation. Due to the prediction character of the filter aliasing, quantization errors and displacement estimation errors are implicitly considered in the coefficient estimation process. Due to the adaptive interpolation filter a gain of 0.8 dB PSNR is obtained in the JVT coding scheme.

6. REFERENCES

[1] JVT: ISO/IEC and ITU-T, "Committee Draft of Joint Video Specification (ITU-T Rec. H.264 / ISO/IEC 14496-10 AVC)", Doc. JVT-C167, Fairfax, Virginia, USA, May 2002

[2] A. Papoulis, "Probability, Random Variables, and Stochastic Processes", McGraw-Hill, New York, 1991