Motion- and Aliasing-Compensated Prediction for Hybrid Video Coding

Thomas Wedi and Hans Georg Musmann

***Abstract*—In order to reduce the bit rate of video signals, the standardized coding techniques apply motion-compensated prediction in combination with transform coding of the prediction error. By mathematical analysis, it is shown that aliasing components are deteriorating the prediction efficiency. In order to compensate the aliasing, two-dimensional (2-D) and three-dimensional interpolation filters are developed. As a result, motion- and aliasing-compensated prediction with 1/4-pel displacement vector resolution and a separable 2-D Wiener interpolation filter provide a coding gain of up to 2 dB when compared to 1/2-pel displacement vector resolution as it is used in H.263 or MPEG-2. An additional coding gain of 1 dB can be obtained with 1/8-pel displacement vector resolution when compared to 1/4-pel displacement vector resolution. In consequence of the significantly improved coding efficiency, a Wiener interpolation filter and 1/4-pel displacement vector resolution is applied in H.264/AVC and in MPEG-4 (advanced simple profile).**

***Index Terms*—Aliasing compensation, H264/AVC, hybrid video coding, motion-compensated prediction, Wiener interpolation filter.**

I. 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 in combination with transform coding of the prediction error. Motion-compensated prediction is performed block-wise using displacement vectors with 1/2-pel resolution in MPEG-2 [1] and H.263 [2]. In the past, the use of displacement vectors with higher resolution did not improve the coding efficiency and therefore has not been considered in those standards.

Experimental investigations of the prediction error show that even in the case of translational motion of the image content, no perfect motion compensation is achieved. Prediction errors containing camera noise and other signal components can be observed. Werner [3] supposes that these additional components originate from aliasing and proposes a Wiener interpolation filter for reducing the impact of aliasing.

In order to prove the assumption that aliasing is interfering the prediction, a mathematical analysis of the prediction error is presented in this paper and verified by experimental investigations. Furthermore, concepts for reducing the impact of aliasing on the prediction efficiency are derived from the results of the analysis.

Manuscript received February 9, 2002; revised May 10, 2003.

The authors are with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, 30167 Hannover, Germany (e-mail: wedi@tnt.uni-hannover.de; musec@tnt.uni-hannover.de).

Digital Object Identifier 10.1109/TCSVT.2003.815171

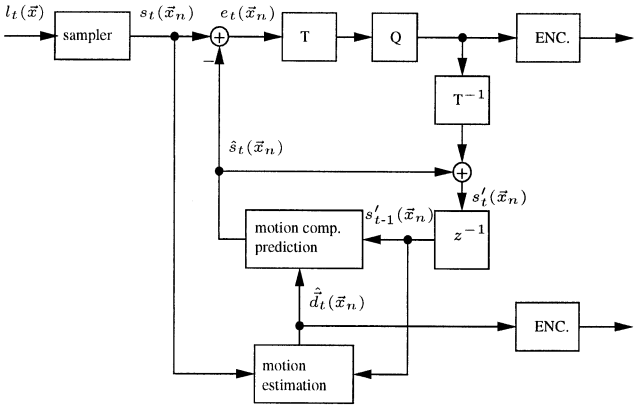

Fig. 1. Block diagram of a hybrid video encoder based on motion-compensated prediction.

In a first step, time invariant two-dimensional (2-D) filters are proposed to attenuate the aliasing components in the prediction signal. In a second step, a concept for motion-adaptive filters are developed in order to compensate the varying aliasing components of the input signal. For demonstrating the improvements of the coding efficiency, the operational rate distortion functions are measured using the peak signal-to-noise ratio (PSNR) criterion.

As a result, 1/4-pel displacement vector resolution was incorporated into the recent video coding standards H.264/AVC [4], [5] and MPEG-4 (advanced simple profile) [6].

Section II introduces the hybrid video coding scheme. Section III presents the mathematical analysis of the prediction error with respect to aliasing. In Section IV, 2-D and three-dimensional (3-D) filters for motion- and aliasing-compensated prediction are described. Experimental results are given in Section V. Section VI provides the conclusions.

II. Hybrid Video Coding Based on Motion-Compensated Prediction

Standardized hybrid video coding techniques like H.263, H.264/AVC, and MPEG-1, -2, and -4 are based on motion-compensated prediction. Fig. 1 shows the generalized block diagram of such a hybrid video encoder. The input image $l_t(\vec{x})$ at time instance t with the space-continuous coordinate $\vec{x} = (x, y)$ is sampled to yield the time- and space-discrete image $s_t(\vec{x}_n)$ with the space-discrete coordinate $\vec{x}_n = (x_n, y_n)$. Image $s_t(\vec{x}_n)$ is predicted by a motion-compensated prediction from an already transmitted reference image $s'_{t-1}(\vec{x}_n)$. The result of

the motion-compensated prediction is image $\hat{s}_t(\vec{x}_n)$. Only the prediction error image

$$e_t(\vec{x}_n) = s_t(\vec{x}_n) - \hat{s}_t(\vec{x}_n) \tag{1}$$

and the motion information $\hat{\vec{d}}_t(\vec{x}_n)$ are encoded and transmitted.

For motion-compensated prediction and coding, the current image is partitioned into blocks. A displacement vector $\hat{\vec{d}}_t(\vec{x}_n)$ is estimated and transmitted for each block that refers to the corresponding position of its image signal in an already transmitted reference image. In former MPEG standards, this reference image is the most recent previous frame $s_{t-1}(\vec{x}_n)$ at time instance $t-1$. In H.263++ and H.264/AVC, it is possible to refer also to preceding images like $s_{t-2}(\vec{x}_n), s_{t-3}(\vec{x}_n), \ldots$. This technique is denoted as motion-compensated prediction with multiple reference frames [7].

The displacement vectors have a fractional-pel resolution. Coding standards like H.263, MPEG-1,2, and MPEG-4 (simple profile) are based on a fractional-pel displacement vector resolution of 1/2 pel. In MPEG-4 (advanced simple profile) [6] and in H.264/AVC [4], 1/4-pel displacement vector resolutions are applied. Displacement vectors with fractional-pel resolution may refer to positions in the reference image which are spatially located between the sampled positions of its image signal. In the following, these positions are called subpel positions. In order to estimate and compensate fractional-pel displacements, the image signal of the reference image has to be generated on subpel positions by interpolation.

Due to nonideal low-pass filters in the image acquisition process aliasing is generated, which deteriorates the interpolation and the motion-compensated prediction. The influence of aliasing on motion-compensated prediction is analyzed in Section III.

III. The Impact of Aliasing on Motion-Compensated Prediction

In this section, the impact of aliasing on the motion-compensated prediction in hybrid video coding (Fig. 1) is analyzed. Therefore, the aliasing components in the input signal, in the prediction signal, and in the resulting prediction error signal have to be considered and are analytically described.

A. Aliasing in the Input Signal

In order to investigate aliasing in the input signal and its influence on motion-compensated prediction, the space-continuous signals $l_{t-1}(\vec{x})$, $l_t(\vec{x})$, and their space-discrete versions $s_{t-1}(\vec{x}_n)$, $s_t(\vec{x}_n)$ are analyzed in the frequency domain. For mathematical simplifications, the calculations are restricted to one spacial coordinate x_n. It is assumed that signal $l_t(x)$ at time instance t is a translationally displaced version of signal $l_{t-1}(x)$ at the previous time instance $t-1$. The displacement is d_x. These space-continuous signals and their spatial Fourier transforms are denoted by

$$\begin{aligned} l_{t-1}(x) &\Rightarrow L_{t-1}(j\omega_x) \\ l_t(x) = l_{t-1}(x-d_x) &\Rightarrow L_t(j\omega_x) = L_{t-1}(j\omega_x)\cdot e^{-jd_x\omega_x}. \end{aligned} \tag{2}$$

According to (2) a translational shift with a displacement d_x in the space-domain leads to a multiplication by $e^{-jd_x\omega_x}$ in the frequency domain [8]. The magnitude $|L_t(j\omega_x)|$ and the phase $\measuredangle L_t(j\omega_x)$ are given by

$$|L_t(j\omega_x)| = |L_{t-1}(j\omega_x)\cdot e^{-j\omega_x d_x}| = |L_{t-1}(j\omega_x)| \tag{3}$$

$$\begin{aligned} \measuredangle L_t(j\omega_x) &= \measuredangle\left(L_{t-1}(j\omega_x)\cdot e^{-j\omega_x d_x}\right) \\ &= \measuredangle L_{t-1}(j\omega_x) - \omega_x \cdot d_x. \end{aligned} \tag{4}$$

Equations (3) and (4) show that the magnitudes of both signals $L_{t-1}(j\omega_x)$ and $L_t(j\omega_x)$ are the same and that their phases differ by minus $\omega_x \cdot d_x$. Thus, the difference of the phase between $L_{t-1}(j\omega_x)$ and $L_t(j\omega_x)$ is a function of the displacement d_x.

The space-continuous input signal $l_{t-1}(x)$ is sampled with a sampling interval X to yield the space-discrete signal $s_{t-1}(x_n)$. The spatial Fourier transform $S_{t-1}(j\omega_x)$ of the space-discrete signal $s_{t-1}(x_n)$ is given by

$$S_{t-1}(j\omega_x) = \frac{1}{X}\sum_{k=-\infty}^{\infty} L_{t-1}\left(j\frac{\omega_x}{X} - j\frac{2\pi k}{X}\right). \tag{5}$$

According to (5), the Fourier transform of the space-discrete signal consists of periodically repeated copies of the Fourier transform of the space-continuous signal [8]. In order to simplify the notations, it is assumed that the spatial sampling interval is $X = 1$. With the normalized frequencies $\Omega = \omega_x/X$ and $\Omega_s = 2\pi/X$, the Fourier transform $S_{t-1}(j\omega_x)$ is given by

$$S_{t-1}(j\Omega) = \sum_{k=-\infty}^{\infty} L_{t-1}\left(j\Omega - jk\Omega_s\right). \tag{6}$$

Correspondingly, the Fourier transform $S_t(j\omega_x)$ of the space-discrete signal $s_t(x_n)$ is given by

$$\begin{aligned} S_t(j\Omega) &= \sum_{k=-\infty}^{\infty} L_t\left(j\Omega - jk\Omega_s\right) \\ &= \sum_{k=-\infty}^{\infty} L_{t-1}\left(j\Omega - jk\Omega_s\right)\cdot e^{-jd_x(\Omega - k\Omega_s)}. \end{aligned} \tag{7}$$

Equation (6) and (7) express the Fourier transforms of the space-discrete signals $S_{t-1}(j\Omega)$ and $S_t(j\Omega)$ in terms of the Fourier transform of the space-continuous signal $L_{t-1}(j\Omega)$. Aliasing can be avoided by ensuring that $L_{t-1}(j\Omega)$ is bandlimited, i.e., $L_{t-1}(j\Omega) = 0$ for $|\Omega| \geq \Omega_s/2$ [8]. Due to nonideal low-pass filters in the image acquisition process, this bandlimitation is not fulfilled

$$L_{t-1}(j\Omega) \neq 0, \quad |\Omega| \geq \frac{\Omega_s}{2}. \tag{8}$$

Thus, aliasing is introduced in the space-discrete signals. However, it is assumed that the nonideal low-pass filters at least lead to the bandlimitation

$$L_{t-1}(j\Omega) = 0, \quad |\Omega| \geq \Omega_s. \tag{9}$$

In the following, we restrict the region of support for the Fourier transforms of the space-discrete signals to the baseband. The baseband is defined as the frequency range $-\Omega_s/2 < \Omega <$

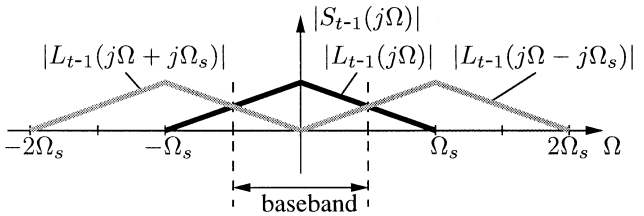

Fig. 2. Fourier transform of the input signal $S_{t-1}(j\Omega)$ in dependency on the frequency Ω. The graph shows the three terms that are necessary to describe the baseband signal.

$\Omega_s/2$ [8]. This restriction greatly simplifies the following mathematics without sacrificing generality.

With the assumption of (9), only three terms of (6) and (7) have to be considered in the baseband (see Fig. 2). The first term to be considered is the term for $k = 0$. This term is called the underlying original signal. The two other terms that have to be considered are the terms for $k = -1$ and $k = 1$. These two terms represent the aliasing components. Thus, the input signal in the baseband $-\Omega_s/2 < \Omega < \Omega_s/2$ is given by

$$\begin{aligned} S_{t-1}(j\Omega) =& L_{t-1}(j\Omega) + L_{t-1}(j\Omega + j\Omega_s) \\ &+ L_{t-1}(j\Omega - j\Omega_s) \end{aligned} \tag{10}$$

$$\begin{aligned} S_t(j\Omega) =& L_{t-1}(j\Omega) \cdot e^{-jd_x\Omega} \\ &+ L_{t-1}(j\Omega + j\Omega_s) \cdot e^{-jd_x(\Omega+\Omega_s)} \\ &+ L_{t-1}(j\Omega - j\Omega_s) \cdot e^{-jd_x(\Omega-\Omega_s)}. \end{aligned} \tag{11}$$

Due to (9), the two aliasing terms do not overlap. Thus, these terms can be summarized to one aliasing term

$$\begin{aligned} A_{t-1}(j\Omega) =& L_{t-1}(j\Omega + j\Omega_s) + L_{t-1}(j\Omega - j\Omega_s) \\ =& L_{t-1}\Big(j\Omega \cdot \Big(1 - \frac{\Omega_s}{|\Omega|}\Big)\Big). \end{aligned} \tag{12}$$

With (12), the space-discrete input signals from (10) and (11) are given by

$$S_{t-1}(j\Omega) = L_{t-1}(j\Omega) + A_{t-1}(j\Omega) \tag{13}$$

$$\begin{aligned} S_t(j\Omega) =& L_{t-1}(j\Omega) \cdot e^{-jd_x\Omega} \\ &+ A_{t-1}(j\Omega) \cdot e^{-jd_x\left(1-\frac{\Omega_s}{|\Omega|}\right)\Omega}. \end{aligned} \tag{14}$$

From (14), it can be recognized that $S_t(j\Omega)$ contains components with different displacements. For the underlying original signal, the displacement is d_x, while for the aliasing components it is

$$d_{xa} = -d_x \cdot \left(\frac{\Omega_s}{|\Omega|} - 1\right), \quad \text{for} -\frac{\Omega_s}{2} < \Omega < \frac{\Omega_s}{2}. \tag{15}$$

Thus, there are two main problems with aliasing in the motion-compensated prediction. At first, (15) shows that the displacement of the aliasing component is a function of Ω. Furthermore, this equation shows that d_{xa} and d_x have different signs. Thus, the aliasing components are moving in the opposite direction than the underlying original signal. In the motion-compensated prediction, only one displacement for the whole signal—including the aliasing components—is considered. Thus, prediction errors due to aliasing remain that have to be coded, even if the displacement of the underlying original signal is estimated perfectly.

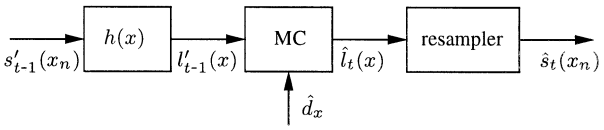

Fig. 3. Generalized block diagram of the motion-compensated predictor.

The prediction signal and the prediction error signal caused by aliasing are calculated in Sections III-B and III-C.

B. Aliasing in the Motion-Compensated Prediction Signal

Fig. 3 shows a block diagram of the motion-compensated prediction. The prediction signal $\hat{s}_t(x_n)$ at time instance t is obtained from the samples of the reconstructed, previous signal $s'_{t-1}(x_n)$ at time instance $t-1$. The motion-compensated predictor consists of three blocks. In the first block, a space-continuous signal $l'_{t-1}(x)$ is generated by applying a reconstruction filter $h(x)$. The Fourier transform of $l'_{t-1}(x)$ is given by

$$L'_{t-1}(j\Omega) = H(j\Omega) \cdot S'_{t-1}(j\Omega). \tag{16}$$

In the second block, the space-continuous signal $l'_{t-1}(x)$ is shifted according to the estimated displacement $\hat{d}_x$ yielding the space-continuous motion-compensated prediction signal $\hat{l}_t(x)$ with the corresponding Fourier transform

$$\hat{L}_t(j\Omega) = L'_{t-1}(j\Omega) \cdot e^{-j\hat{d}_x\Omega}. \tag{17}$$

It is assumed that the estimated displacement is given by the displacement of the underlying original signal

$$\hat{d}_x = d_x. \tag{18}$$

In the third block, $\hat{l}_t(x)$ is resampled with the sampling interval $X = 1$. The result is the space-discrete prediction signal $\hat{s}_t(x_n)$ with the Fourier transform

$$\hat{S}_t(j\Omega) = \sum_{k=-\infty}^{\infty} \hat{L}_t\left(j\Omega - jk\Omega_s\right). \tag{19}$$

Assuming

$$S'_{t-1}(j\Omega) = S_{t-1}(j\Omega) \tag{20}$$

which means that quantization errors are neglected, and with an ideal low-pass filter

$$H(j\Omega) = \begin{cases} 1, & \text{for } |\Omega| < \frac{\Omega_s}{2} \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

the baseband signal of $\hat{S}_t(j\Omega)$ is given by

$$\begin{aligned} \hat{S}_t(j\Omega) =& L_{t-1}(j\Omega + j\Omega_s) \cdot e^{-jd_x\Omega} \\ &+ L_{t-1}(j\Omega) \cdot e^{-jd_x\Omega} \\ &+ L_{t-1}(j\Omega - j\Omega_s) \cdot e^{-jd_x\Omega}. \end{aligned} \tag{22}$$

Using (12), the motion-compensated prediction signal is

$$\hat{S}_t(j\Omega) = L_{t-1}(j\Omega) \cdot e^{-jd_x\Omega} + A_{t-1}(j\Omega) \cdot e^{-jd_x\Omega}. \tag{23}$$

The second term of (23) denotes the aliasing in the prediction signal.

C. The Prediction Error Signal

In order to show the impact of aliasing on the motion-compensated prediction, the prediction error signal is calculated. The Fourier transform of the prediction error signal $e_t(x_n)$ from (1) is

$$E_t(j\Omega) = S_t(j\Omega) - \hat{S}_t(j\Omega) \tag{24}$$

and with (14) and (23)

$$E_t(j\Omega) = A_{t-1}(j\Omega) \cdot \left[e^{-jd_x(1-(\Omega_s/|\Omega|))\Omega} - e^{-jd_x\Omega}\right]. \tag{25}$$

With $\sin(z) = \frac{1}{2j}\left(e^{jz} - e^{-jz}\right)$, (25) can be calculated to

$$\begin{aligned} E_t(j\Omega) =& A_{t-1}(j\Omega)\left[e^{-jd_x\Omega}\left(e^{+jd_x\Omega_s\Omega/|\Omega|} - 1\right)\right] \\ =& A_{t-1}(j\Omega)\left[e^{-jd_x\Omega} \cdot e^{j\frac{1}{2}d_x\Omega_s\Omega/|\Omega|}\right. \\ & \left. \cdot \left(e^{j\frac{1}{2}d_x\Omega_s\Omega/|\Omega|} - e^{-j\frac{1}{2}d_x\Omega_s\Omega/|\Omega|}\right)\right] \\ =& A_{t-1}(j\Omega)\left[e^{-jd_x\left(\Omega-\frac{1}{2}\Omega_s\Omega/|\Omega|\right)} \cdot 2j\right. \\ & \left. \cdot \sin\left(d_x\frac{\Omega_s}{2}\frac{\Omega}{|\Omega|}\right)\right] \\ =& 2A_{t-1}(j\Omega)\sin\left(d_x\frac{\Omega_s}{2}\frac{\Omega}{|\Omega|}\right) \\ & \cdot \left[e^{-jd_x\left(\Omega-\frac{1}{2}\Omega_s\Omega/|\Omega|\right)+\frac{j\pi}{2}}\right]. \end{aligned} \tag{26}$$

Equation (26) denotes the Fourier transform of the prediction error signal caused by aliasing. Due to assumption (18), no displacement estimation errors, and due to assumption (20), no quantization errors are considered. Equation (26) shows that the prediction error signal caused by aliasing is a function of the displacement d_x. Considering $\Omega_s = 2\pi$, the magnitude of the prediction error signal is given by

$$\begin{aligned} |E_t(j\Omega)| =& 2 \cdot |A_{t-1}(j\Omega)| \cdot \left|\sin\left(d_x \cdot \pi \cdot \frac{\Omega}{|\Omega|}\right)\right| \\ =& 2 \cdot |A_{t-1}(j\Omega)| \cdot |\sin(d_x \cdot \pi)|. \end{aligned} \tag{27}$$

The factor 2 in the prediction error signal of (27) results from the addition of the aliasing components in the original signal $S_t(j\Omega)$ of (14) and the incorrect predicted aliasing components in the prediction signal $\hat{S}_t(j\Omega)$ of (23).

In Fig. 4, the normalized magnitude of the prediction error in dependency on the displacement d_x is depicted. There are two main features of the prediction error caused by aliasing:

1) the impact of aliasing on the prediction error vanishes at full-pel displacements ($d_x = 0, \pm 1, \pm 2, \ldots$);
2) the impact of aliasing on the prediction error is maximal at half-pel displacements.

IV. Motion- and Aliasing-Compensated Prediction

In order to reduce the impact of aliasing on the motion-compensated prediction error, only the prediction signal $\hat{s}_t(\vec{x}_n)$ can be modified. Therefore, the aliasing term in (23) has to be matched to that of the input signal in (14). Since the two aliasing components are moving in opposite directions, in a first step, the impact of aliasing on the prediction error signal can be reduced by attenuating the aliasing in the prediction signal. Since fractional-pel displacements generate a prediction error caused by aliasing (Fig. 4), the filter $h(x)$ in Fig. 3, which interpolates the samples $s'_{t-1}(x_n)$ at fractional-pel positions, can be used in order to reduce the impact of aliasing. In the following, two techniques for motion and aliasing compensation are presented. In Section IV-A, a separable 2-D Wiener filter is applied for attenuating the aliasing in the prediction signal. Section IV-B presents a technique for compensating the aliasing in the prediction error signal by a motion adaptive 3-D filter.

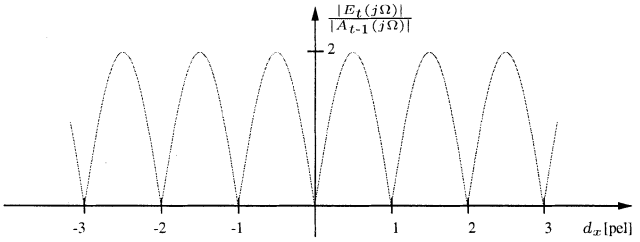

Fig. 4. Normalized Fourier transform of the prediction error magnitude caused by aliasing in dependency on the displacement d_x.

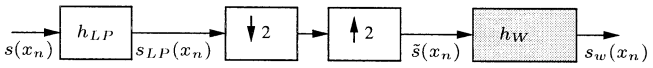

Fig. 5. Block diagram to illustrate the design of the Wiener filter h_W.

A. Motion and Aliasing Compensation Using a Separable 2-D Wiener Filter

In [3], a separable 2-D Wiener filter is proposed to reduce drift due to aliasing for multiresolution hybrid video coding. Here, this filter is applied to attenuate aliasing in the prediction signal for single resolution hybrid video coding with displacement vector resolutions of 1/4 and 1/8 pel.

The design of this Wiener filter will be described using Fig. 5. In this figure, $s(x_n)$ denotes the sampled input signal, which is filtered by a low-pass filter h_{LP}, resulting in $s_{\mathrm{LP}}(x_n)$. Thereafter, the sampling rate of $s_{\mathrm{LP}}(x_n)$ is successively reduced and increased by the factor of 2. The result is signal $\tilde{s}(x_n)$, which is given by

$$\tilde{s}(x_n) = \begin{cases} s_{\mathrm{LP}}(x_n), & \text{for } x_n = 0, \pm 2, \pm 4, \ldots \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Assuming a nonideal low-pass filter h_{LP}, aliasing is introduced in $\tilde{s}(x_n)$. Once h_{LP} has been selected, the corresponding interpolation filter h_W used to reconstruct $s_{\mathrm{LP}}(x_n)$ from $\tilde{s}(x_n)$ can be determined by linear mean squared estimation [9]. The result is a Wiener filter.

In the following, a Wiener filter h_W of finite impulse response described by $2L_W$ coefficients is assumed. In this case, the Wiener filter can also be obtained by solving the Wiener–Hopf equation [9]. With the Autocorrelation function of $s_{\mathrm{LP}}(x_n)$

$$R_{\mathrm{LP}}(n) = E[s_{\mathrm{LP}}(x_n) \cdot s_{\mathrm{LP}}(x_n - n)] \tag{29}$$

TABLE I
FILTER COEFFICIENTS FOR ORIGINAL AND MODIFIED WIENER INTERPOLATION FILTERS AND A BILINEAR FILTER

Wiener filter type	Filter coefficients
original from [3]	(-8,23,-48,161,161,-48,23,-8)/256
mod. for MPEG-4	(-1,3,-6,20,20,-6,3,-1)/32
mod. for H.264/AVC	(1,-5,20,20,-5,1)/32
bilinear	(1,1)/2

the Wiener–Hopf equation of the interpolation problem of Fig. 5 is given by

$$\sum_{l=-L_W}^{(L_W-1)} h_W(l) \cdot R_{\mathrm{LP}}(2(i-l)) = R_{\mathrm{LP}}(2i+1), \quad \text{for } i = -L_W, \ldots, (L_W-1). \quad (30)$$

For a stationary signal $s_{\mathrm{LP}}(x_n)$, it could be shown that, due to the symmetry of the autocorrelation function $R_{\mathrm{LP}}(n)$, the coefficients of the Wiener filter are also symmetric

$$h_W(l) = h_W(-l-1), \text{ for } 0 \le l \le L_W - 1. \quad (31)$$

Thus, the Wiener–Hopf equation from (30) can be expressed by

$$\sum_{l=0}^{(L_W-1)} h_W(l) \cdot [R_{\mathrm{LP}}(2(i-l)) + R_{\mathrm{LP}}(2(i+l+1))] = R_{\mathrm{LP}}(2i+1), \quad \text{for } i = 0, \ldots, (L_W-1). \quad (32)$$

In order to get an unbiased estimate, i.e.,

$$E[s_{\mathrm{LP}}] = E[s_W] \quad (33)$$

the coefficients $h_W(l)$ have to fulfill the condition

$$\sum_{l=-L_W}^{(L_W-1)} h_W(l) = 1. \quad (34)$$

With (31) and (34), the number of coefficients $h_W(l)$ which have to be determined by solving the Wiener–Hopf equation system of (32) is reduced from $2 \cdot L_W$ to $(L_W - 1)$.

In [3], such a Wiener filter was determined. Table I shows the original Wiener interpolation filter from [3], two modified versions, and the bilinear filter. For MPEG-4 (advanced simple profile), the original filter was modified in order to allow an easier hardware implementation [10]. For H.264/AVC, it was further modified from an 8- to a 6-tap filter [11]. For a filter length of $L_W = 1$, the Wiener filter is equivalent to the bilinear filter [3], which is applied in coding standards like H.263, MPEG-1, and MPEG-2. Fig. 6 shows the frequency responses of the modified 6- and 8-tap Wiener interpolation filters and the bilinear filter from Table I. The bilinear filter is far away from an ideal low-pass filter. It has a significant passband attenuation and a significant stopband permeability. The Wiener filters have filter characteristics more similar to that of an ideal low-pass filter. They differ due to the aliasing considered by the Wiener filters.

For the mathematical analysis in Section III, displacement vectors with continuous amplitude resolution are assumed. Real coding schemes like H.264/AVC or MPEG-4 are based on space-discrete displacement vectors with fractional-pel

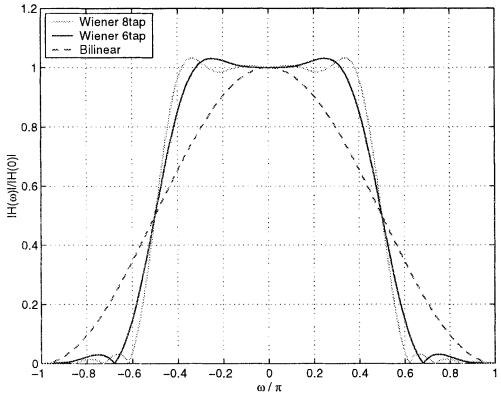

Fig. 6. Normalized frequency response of different interpolation filters.

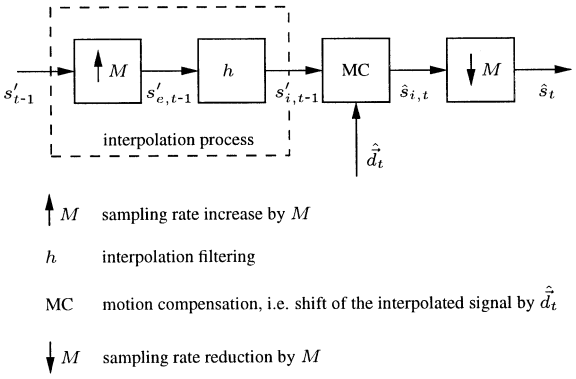

Fig. 7. Block diagram of the motion-compensated predictor using a displacement vector amplitude resolution of $1/M$ pel.

amplitude resolutions of 1/2 or 1/4 pel. In Fig. 7, the block diagram of a motion-compensated predictor with fractional-pel amplitude resolution of $1/M$ pel is shown. For example, with $M = 4$, a motion-compensated predictor with 1/4-pel displacement vector resolution is realized.

The motion-compensated prediction of Fig. 7 consists of two steps. The first step is the interpolation process, where the sampling rate of the already transmitted and reconstructed image $s'_{t-1}(\vec{x}_n)$ is increased by a factor of M and filtered with an interpolation filter h. The result of this first step is the interpolated image $s'_{i,t-1}(\vec{x}_n)$. In the second step, the interpolated signal is shifted according to the estimated displacement vector $\hat{\vec{d}}_t(\vec{x}_n)$ and the sampling rate is reduced by the factor of M. The result is the motion-compensated image $\hat{s}_t(\vec{x}_n)$.

The Wiener filters of Table I are developed for motion-compensated prediction with 1/2-pel displacement vector resolution $(M = 2)$. In order to use these filters for motion-compensated prediction with $M > 2$, the interpolation process is split into two or more steps. Therefore, Fig. 8 shows the block diagram of the interpolation process for an interpolation by an exemplary factor of $M = 4$. In this figure, two representations for this interpolation are given. In Fig. 8(a) the interpolation process is performed in one step with one interpolation filter h. This is the same representation that is also used in Fig. 7. Fig. 8(b) shows a representation where the interpolation is performed in two steps with two interpolation filters h_1 and h_2. In MPEG-4 (advanced simple profile), h_1 is the 8-tap and in H.264/AVC h_1 is the 6-tap Wiener filter of Table I. In both standards, filter h_2 is the simple

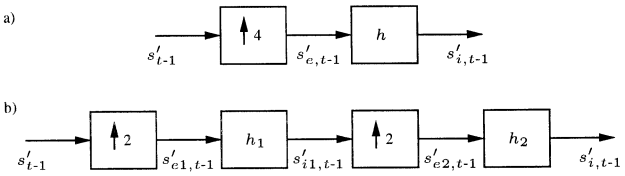

Fig. 8. Two representations for a sampling rate increase by a factor of 4 and interpolation filtering. (a) One-step realization with one interpolation filter h (b) Two-step realization with two interpolation filters h_1 and h_2.

bilinear filter. In order to use a motion-compensated prediction with displacement vector resolutions of 1/8 and 1/16 pel, the interpolation process of Fig. 8(b) is extended by one or two more steps, respectively. In this case, only in the last interpolation step the bilinear filter is used. Note that the representation (b) can be transferred to representation (a). Analyses on increased computational complexity due to motion-compensated prediction with increased motion vector resolutions can be found in [12] and [13].

First results with adaptive 2-D filters indicate that further improvements can be achieved when the 2-D filter is adapted from frame to frame [14]–[16].

B. Motion and Aliasing Compensation Using a Motion-Adaptive 3-D Filter

Here, a concept for a 3-D interpolation filter is described that exploits the fact that the impact of aliasing on the prediction error vanishes at full-pel displacements.

In Fig. 9, a generalized example of a motion-compensated predictor using such a 3-D filter is shown. In order to simplify the explanations, this example is restricted to one spacial coordinate x_n. Thus, Fig. 9 shows one-dimensional image lines instead of 2-D images. Due to the restriction to one dimensional image lines in Fig. 9, the 3-D filter in this figure is restricted to a 2-D filter. Furthermore, a translational motion of 1/2-pel from frame to frame is assumed in this example.

The image $s_t(x_n)$ is predicted from the interpolated reference image $s'_{i,t-1}(x_n)$ using the displacements $d_{x,t}$. For this prediction, samples at subpel positions of $s'_{i,t-1}(x_n)$ have to be interpolated. If these samples are interpolated with a 2-D Wiener filter, a prediction error with reduced but not compensated aliasing is generated. Thus, a 3-D motion adaptive filter is proposed in [17] that uses samples of former images in order to interpolate $s'_{i,t-1}(x_n)$. In the example of Fig. 9, the displacements $d_{x,t-1}$ is considered in order to use reference samples of $s'_{i,t-2}(x_n)$ for the interpolation of the samples at subpel positions of $s'_{i,t-1}(x_n)$. If these interpolated samples are used to predict $s_t(x_n)$, this could be interpreted as a prediction with a full-pel displacement between $s'_{i,t-2}(x_n)$ and $s_t(x_n)$. Since the impact of aliasing vanishes for full-pel displacements according to Fig. 4, the motion-adaptive 3-D interpolation filter compensates also the aliasing. If no reference samples with full-pel displacements are available for interpolation, the 2-D Wiener filter is applied. It is proposed that this 3-D interpolation filter replaces the current H.264/AVC interpolation filter. In [18], a similar method was used for high-resolution video mosaicing. First results with adaptive 3-D filters indicate that further improvements can be achieved when the 3-D filter is adapted from frame to frame [19].

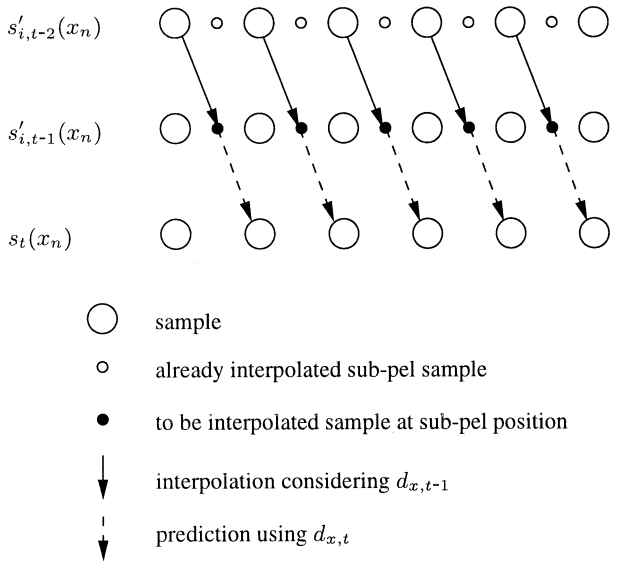

Fig. 9. Example of a motion-compensated predictor applying a 3-D filter. In this example, 1/2-pel displacement vector resolution is assumed. Corresponding lines of three consecutive frames are shown: The frame to be predicted $s_t(x_n)$, the frame to be interpolated $s'_{i,t-1}(x_n)$, and the preceeding interpolated frame $s'_{i,t-2}(x_n)$.

There is a relationship between this motion and aliasing-compensated prediction and the prediction with multiple reference frames [7] or the prediction with generalized B-frames [20]. The main difference is that the 3-D filters generate only one single reference frame with samples at subpel positions and therefore need no side information about the reference frames used for prediction.

V. Experimental Results

For experimental investigations, the coding system H.26L (TML4) and various kinds of test-sequences, each at CIF format and 30 Hz, are applied. Although the results are achieved with TML4, the obtained insights are also valid for the current H.264/AVC coding standard [4].

In order to verify the analytically derived results, the mean squared error (mse) of the prediction as a function of the magnitude of the displacement has been measured in a first step using the test sequence *Flower Garden* with its relatively uniform motion. The result is shown in Fig. 10.

Fig. 10(a) presents the results for different displacement vector resolutions. In order not to affect the aliasing, an almost ideal low-pass filter is used for interpolation. For a comparison with the analytical results in Fig. 4, the additional prediction error originating from camera noise and the restricted displacement vector resolution of the coding system have to be considered. An estimate of the mean squared prediction error generated by camera noise is $\mathrm{mse}_n = 0.3$. The mean squared prediction error at integer-pel displacements is not affected by aliasing and therefore provides estimates of the prediction error generated by the quantization or so-called resolution of the displacement vector. The residual mean squared prediction error $\Delta\mathrm{mse}_{a,s}$ indicates the impact of the aliasing. Due to the increase of the displacement vector resolution from 1/2 to 1/4 pel, a mse reduction of $\Delta\mathrm{mse}_1 = 3.0$ is achieved at $d_x = 3.0$. Due to the increase from 1/4 to 1/8 pel, a reduction

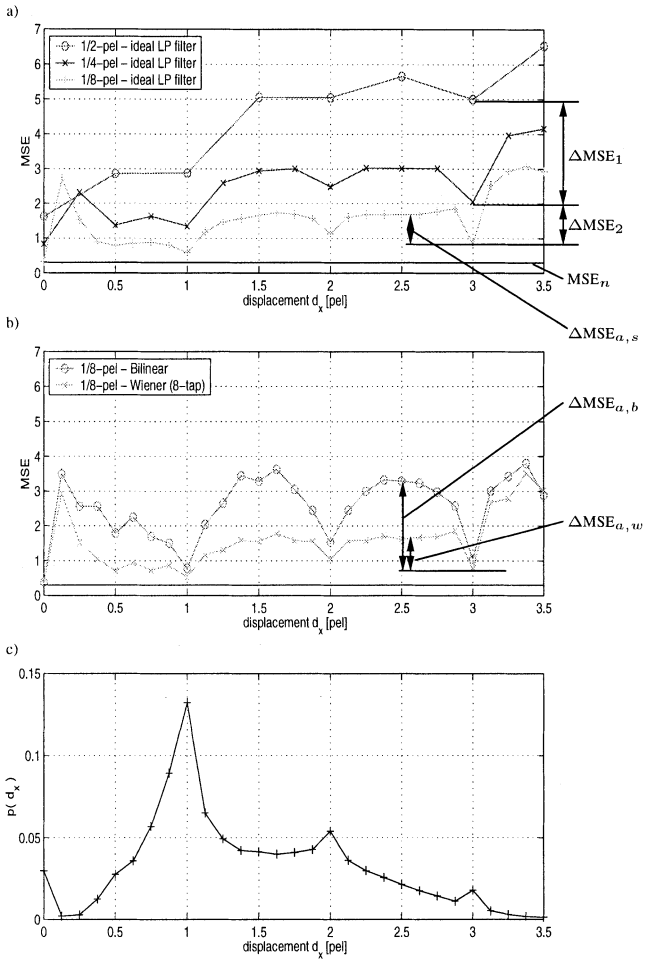

Fig. 10. Measured mse and probability density function $p(d_x)$ in dependency on the displacement d_x for the test sequence *Flower Garden*: (a) mse for different displacement vector resolutions using an almost ideal low-pass (LP) filter; (b) mse for 1/8-pel displacement vectors using different filters; and (c) probability density function $p(d_x)$.

of $\Delta\text{mse}_2 = 1.0$ is achieved. In order to show the aliasing compensation effect of the Wiener filter, Fig. 10(b) presents the mse for a Wiener and a bilinear filter each with 1/8-pel displacement vector resolution. Since the mse at integer-pel displacements is not affected by aliasing, the prediction error difference between the mse of the Wiener and the bilinear filter at these positions is almost zero. But, due to the aliasing compensation effect of the Wiener filter, a mse reduction of $\Delta\text{mse}_{a,b} - \Delta\text{mse}_{a,w} = 1.5$ is obtained at $d_x = 2.5$. In Fig. 10(c), the probability density function $p(d_x)$ is shown. Due to the low probability of $d_x < 0.5$, the mse in this area shows irregularities. In contrast to the analytical results in Fig. 4, the mean squared prediction error curves of Fig. 10 are not normalized.

In order to investigate the influence of the interpolation filters on the coding efficiency, Fig. 11 shows operational rate distortion curves for bilinear and Wiener filter at different displacement vector resolutions of 1/2, 1/4, and 1/8 pel.

The results show that the Wiener filter outperforms the bilinear filter for all displacement vector resolutions. The higher the displacement vector resolution, the more important is the use of a Wiener filter that considers the aliasing. The coding gain increases with the bit rate. The different performance of the Wiener filter and the bilinear filter is due to the longer filter tap size and the fact that the Wiener filter was developed to reduce the aliasing components that are deteriorating the motion-compensated prediction.

In order to compare the coding efficiency for different displacement vector resolutions, Figs. 12 and 13 show two graphs with operational rate distortion curves for the two test sequences *Mobile & Calendar* and *Flower Garden*.

In these figures, the parameters are the displacement vector resolutions 1/2, 1/4, 1/8, and 1/16 pel, each obtained with the Wiener interpolation filter. For comparison, the operational rate distortion curves of 1/2-pel displacement vector resolution with bilinear interpolation are shown that are used in coding standards like H.263, MPEG-1, -2, and -4 (simple profile). The figures show a significant coding gain when the displacement vector resolution is increased up to a resolution of 1/8 pel. The 1/16-pel resolution does not provide a significant additional coding gain. Compared to the 1/2-pel and bilinear interpolation filter as it is used in H.263, MPEG-1, -2, and -4 (simple profile), a coding gain up to 3.0 dB is obtained. Compared to 1/4 pel and the Wiener interpolation filter as it is used in MPEG-4 (advanced simple profile), a gain up to 1.0 dB is obtained by 1/8-pel displacement vector resolution.

The influence of the displacement vector resolution on the coding efficiency can also be seen in Figs. 14 and 15, where the bit-rate reduction in dependence on the displacement vector resolution is shown for different test-sequences at PSNR values of 38 dB (Fig. 14) and of 28 dB (Fig. 15). The PSNR of 28 dB represents low quality and the PSNR of 38 dB represents high quality of a reconstructed sequence. In these figures, test sequences with uniform motion, like *Mobile & Calendar, Flower Garden, Tempete, Waterfall,* and *Bus* are marked with solid lines. Test sequences with nonuniform motion, like *Foreman, Husky, News,* and *Paris* are marked with dashed lines. The graphs show the following.

1) The relative bit-rate reduction for test sequences with uniform motion (solid lines) is higher than for test sequences with nonuniform motion (dashed lines). The reason is that the nonuniform motion inside a block disturbs the motion-compensated prediction.
2) The relative bit-rate reduction is higher for high PSNR values. At a PSNR of 38 dB, a bit-rate reduction between 10%–48% compared to the bit rate required at 1/2-pel displacement vector resolution is achieved. At a PSNR of 28 dB (Fig. 15), the quantization errors disturb the motion-compensated prediction. As a consequence, the side information required for transmitting the displacement vectors can result a negative bit-rate reduction.

VI. Conclusions

By theoretical and experimental investigations, the impact of aliasing on the prediction error of a motion-compensated predictor is analyzed and verified.

In order to reduce the impairment of the motion-compensated prediction by the aliasing, 2-D and 3-D interpolation filters are proposed, which aim to reduce or to compensate the aliasing.

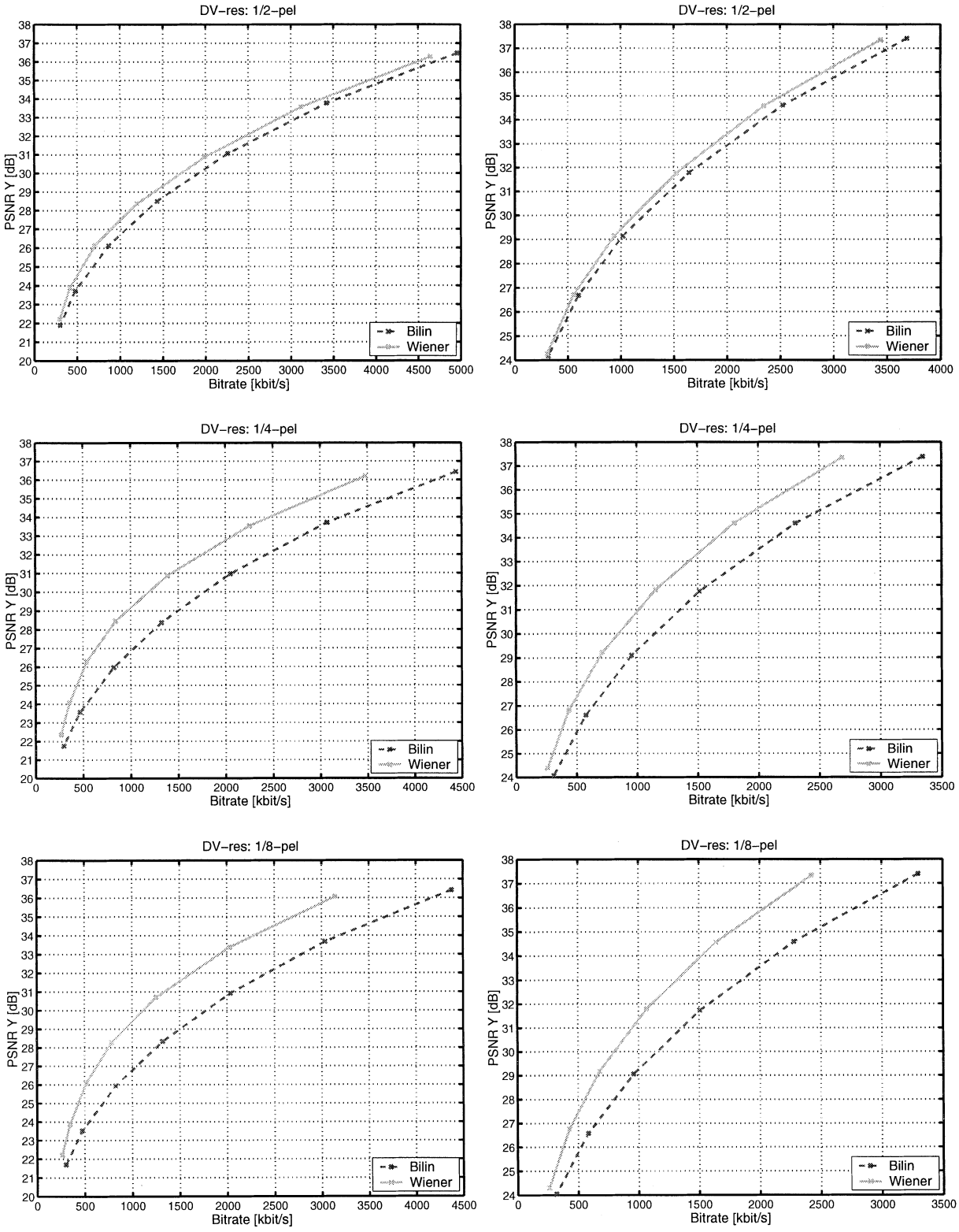

Fig. 11. Operational rate-distortion curves for different interpolation filters and different displacement vector resolutions (DV-res) of 1/2, 1/4, and 1/8 pel for test sequences (a) *Mobile & Calendar* and (b) *Flower Garden*.

As a result, motion- and aliasing-compensated prediction with 1/4- or 1/8-pel displacement vector resolution improve the PSNR of the hybrid coding system H.264/AVC by 2 or 3 dB when compared to 1/2-pel displacement vector resolution and a bilinear interpolation filter as it is applied in H.263, MPEG-2, and MPEG-4 (simple profile). This corresponds to a bit-rate savings of 30% and 40%.

The motion- and aliasing-compensated prediction using Wiener interpolation filters in combination with increased displacement vector resolutions has been proposed to the

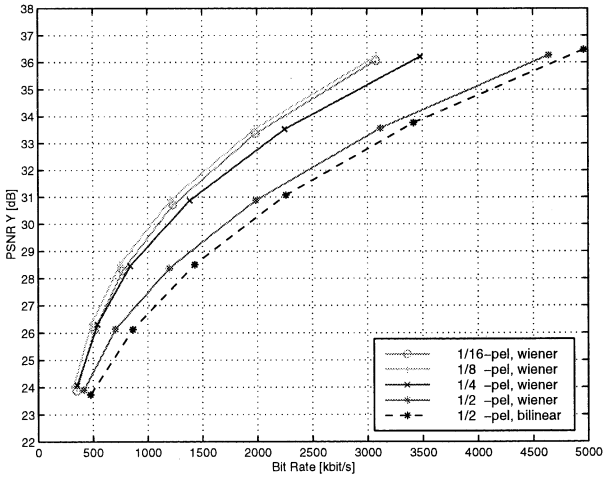

Fig. 12. Operational rate-distortion curves for test sequence *Mobile & Calendar*. Parameter is the displacement vector resolution.

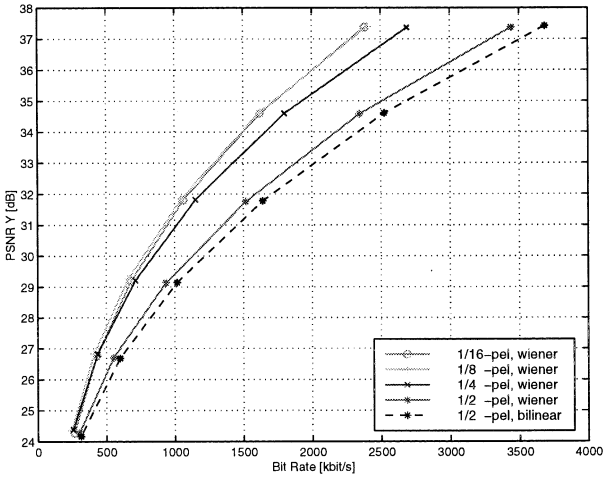

Fig. 13. Operational rate-distortion curves for test sequence *Flower Garden*. Parameter is the displacement vector resolution.

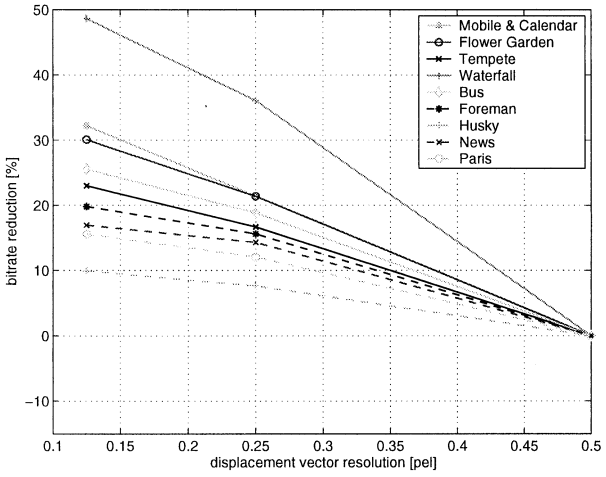

Fig. 14. Bit-rate reduction in dependence on the displacement vector resolution for a PSNR of 38.0 dB.

MPEG-4 [21], [10], and H.264/AVC [22], [23] standardization. As a consequence of the significantly improved coding efficiency, the 6-tap Wiener filter and 1/4-pel displacement vector resolution is applied in H.264/AVC [4] and the 8-tap Wiener filter and 1/4-pel displacement vector resolution is applied in MPEG-4 (advanced simple profile) [6].

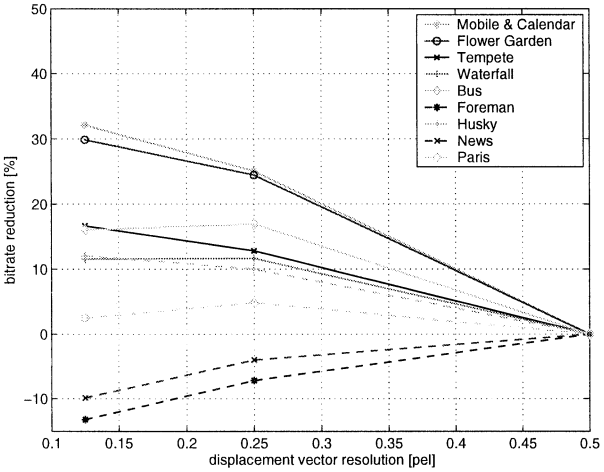

Fig. 15. Bit-rate reduction in dependence on the displacement vector resolution for a PSNR of 28.0 dB.

The motion-adaptive 3-D filter for aliasing compensation is still under investigation. First results obtained with adaptive filters indicate that further improvements can be expected when the filters are adapted from frame to frame.

REFERENCES

[1] MPEG-2: ISO/IEC JTC1/SC29/WG11 and ITU-T, "Revised text for ITU-T recommendation H.262—ISO/IEC 13 818-2: Information technology-generic coding of moving pictures and associated audio information: Video," ISO/IEC and ITU-T, Genf, Switzerland, 1995.

[2] H.263: International Telecommunication Union, "Recommendation ITU-T H.263: Video Coding for Low Bit Rate Communication," International Telecommunication Union, Genf, Switzerland, 1998.

[3] O. Werner, "Drift analysis and, drift reduction for multiresolution hybrid video coding," *Signal Processing: Image Commun.*, vol. 8, no. 5, July 1996.

[4] Joint Video Team (JVT) of ISO/IEC MPEG, ITU-T VCEG (ISO/IEC JTC1/SC29/WGII, and ITU-T SG16 Q.6, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T rec. H.264—ISO/IEC 14 496-10 AVC)," 7th Meeting, Pattaya, Thailand, Mar. 7–14, 2003.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, July 2003.

[6] MPEG-4: ISO/IEC JTCI/SC29/WG11, "ISO/IEC 14 496:2000-2: Information on technology—coding of audio–visual objects—Part 2: Visual," ISO/IEC, Genf, Switzerland, Dec. 2000.

[7] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion- compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 70–84, Feb. 1999.

[8] R. W. Schaefer and A. V. Oppenheim, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[9] A. Papoulis, *Prohahility, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

[10] U. Benzler, "Performance evaluation of a reduced complexity implementation for quarter pel motion compensation," ISO/IEC JTC1/SC29/WG11 MPEG-4: N3146, Jan. 1998.

[11] G. Bjontegaard, "Motion compensation with 1/4-pel accuracy," ITU-T SG16/Q15 (VCEG), Doc. D.361, [Online] Available at: ftp://standard.pictel.com/video-site, Feb. 2000.

[12] S. Saponara, C. Blanch, K. Denolf, and J. Bormans, "Data transfer and storage complexity analysis of the AVC/JVT codec on a tool-by-tool basis," Joint Video Team (JVT), Doc. JVT-Dl38, Klagenfurt, Austria, July 2002.

[13] T. Wedi, "Results on Complexity Performance Investigations: Displacement vector resolution and interpolation filter tap size," ITU-T SG16/Q6 (VCEG), Doc. VCEG-M26, Austin, TX, [Online] Available at: ftp://standard.pictel.com/video-site, Apr. 2001.
[14] ——, "Adaptive interpolation filter for motion compensated prediction," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Rochester, NY, Sept. 2002.
[15] ——, "Adaptive interpolation filter for motion compensated hybrid video coding," in *Proc. Picture Coding Symp. (PCS)*, Seoul, Korea, Jan. 2001.
[16] ——, "Adaptive interpolation filter for H.26L," ITU-T SG16/Q6 (VCEG), Doc. VCEG-N28, Santa Barbara, CA, [Online] Available at: ftp://standard.pictel.com/video-site, Sept. 2001.
[17] ——, "A time-recursive interpolation filter for motion compensated prediction considering aliasing," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Kobe, Japan, Oct. 1999.
[18] A. Smolic and T. Wiegand, "High-resolution video mosaicing," in *Proc. IEEE Int. Conf. Image Processing (JCIP)*, Thessaloniki, Greece, Oct. 2001.
[19] T. Wedi, "Adaptive interpolation filter for motion and aliasing compensated prediction," in *Proc. Visual Communications and Image Processing (VCIP)*, San Jose, CA, Jan. 2002.
[20] M. Flierl and B. Girod, "Generalized B pictures and the draft H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 587–597, July 2003.
[21] U. Benzler, "Results of core experiment P8 (motion and aliasing compensated prediction)," ISO/IEC JTC1/SC29IWG1 I MPEG-4: N2625, Oct. 1997.
[22] T. Wedi, "Motion and aliasing compensated prediction for H.26L," ITU-T SG16/Q15, Doc. Q15I35, Red Bank, NJ, [Online] Available at: ftp://standard.pictel.com/video-site, Oct. 1999.
[23] ——, "1/8-pel Motion Vector Resolution for H.261," ITU-T Q15/SG16, Doc. Q15-K-21, Portland, OR, [Online] Available: ftp://standard.pictel.com/video-sites, Aug. 2000.

Thomas Wedi received the Dipl.-Ing. degree in 1999 from the University of Hannover, Hannover, Germany, where he is currently working toward the Ph.D. degree with research focused on motion- and aliasing-compensated prediction for hybrid video coding.

He has been with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, as Research Scientist and Teaching Assistant. In 2001, he also became a Senior Engineer. His further research interests include video coding and transmission, 3-D image and video processing, and audio-visual communications. He is an active contributor to the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC/ITU-T Joint Video Team (JVT), where the H.264/AVC video coding standard is developed. In both standardization groups he chaired an *ad-hoc* group on interpolation filtering. In cooperation with the Robert Bosch GmbH, he holds several international patents in the area of video compression.

Hans Georg Musmann received the Dipl.-Ing. degree in electrical engineering in 1962 and the Dr.-Ing. degree in 1966, both from the Technische Universität Braunschweig, Braunschweig, Germany.

From 1966 to 1973, he was with the Institut für Nachrichtentechnik, Technische Universität Braunschweig, where he was involved in studies of digital communication systems with emphasis on source encoding. In 1970, he became an Associate Professor. Since 1973, he has been a Full Professor and Head of the Institut für Theoretische Nachrichtentechnik und Informations verarbeitung, University of Hannover, Hannover, Germany. In 1988, he also became a member of the Board of Directors of the Laboratorium für Informationstechnologie of the University of Hannover. His main field of research is video and audio coding for future communication systems. He is author of more than 50 research papers and holder of 15 patents. From 1988 to 1992, he Chaired the ISO MPEG/Audio group, which developed the ISO Audio Coding Standard MP3.