Hybrid Video Coding Based on High-Resolution Displacement Vectors

Thomas Wedi

Institut fuer Theoretische Nachrichtentechnik und Informationsverarbeitung
Universitaet Hannover, Appelstr. 9a, 30167 Hannover, Germany
E-Mail: wedi@tnt.uni-hannover.de

ABSTRACT

Common motion compensated hybrid video coding standards such as H.263, MPEG-1, MPEG-2, MPEG-4 are based on a fractional-pel displacement vector resolution of 1/2-pel. Recent approaches like MPEG-4 ACE and H.26L (TML5) use a 1/4-pel displacement vector resolution. In order to estimate and compensate fractional-pel displacements, the image signal has to be interpolated. Therefore different interpolation filters are used in the standards. In this paper an enhanced motion compensated hybrid video codec is presented, which is based on high-resolution displacement vectors. For this purpose, displacement vector resolutions of 1/8- and 1/16-pel are used in order to improve the motion compensated prediction and the coding efficiency. The coding results for different resolutions are presented and the dependence on different interpolation filters is analysed. It turned out that the higher the displacement vector resolution is, the higher the influence of the filter on the coding gain is. A gain up to 3.0 dB PSNR is obtained compared to a hybrid video codec, which is based on 1/2-pel resolution and bilinear interpolation like H.263, MPEG-1,2,4. Compared to 1/4-pel displacement vector resolution and a wiener interpolation filter as it is used in MPEG-4 ACE and H.26L (TML5), a gain up to 1.0 dB PSNR is obtained.

Keywords: displacement vector resolution, interpolation, motion compensated prediction, H.26L, MPEG, hybrid video coding

1. INTRODUCTION

Standardised hybrid video codecs like H.261, H.263, MPEG-2, MPEG-4 etc. are based on motion compensated prediction.[1,2,4,5] Figure 1 shows the block diagram of such a hybrid video encoder, where the current image to be coded $s(t)$ is predicted by a motion compensated prediction from an already transmitted image $s'(t\text{-}1)$.

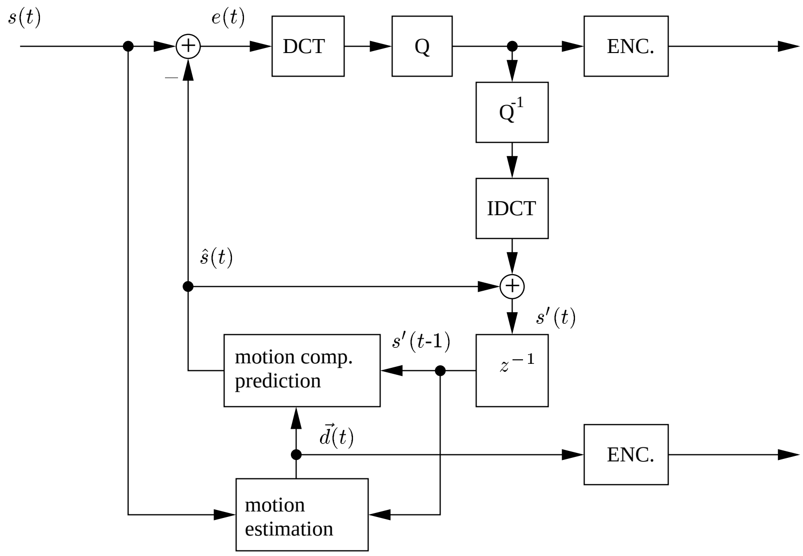


Figure 1. Block diagram of a hybrid video encoder based on motion compensated prediction.

The result of the motion compensated prediction is image $\hat{s}(t)$. Only the prediction error $e(t)$ and the motion information $\vec{d}(t)$ is coded and transmitted. For the motion compensated prediction purpose, the current image is partitioned into blocks. A displacement vector $\vec{d}(t)$ is assigned to each block, which 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, and therefore may refer to a position in an reference image, which is located between the sampled positions of its image signal. In order to estimate and compensate fractional-pel displacements, the image signal on subpel positions has to be generated by interpolation.

The most frequently used fractional-pel resolution of the displacement vectors $\vec{d}(t)$ is 1/2-pel (H.263, MPEG-1, MPEG-2, MPEG-4). Recent approaches like MPEG-4 ACE and H.26L (TML5) are based on 1/4-pel displacement vector resolution. Compared to 1/2-pel resolution the use of 1/4-pel resolution improves the motion compensated prediction and the coding efficiency.[9,10]

In this paper a further increase of the displacement vector resolution is proposed in order to further improve the motion compensated prediction and the coding efficiency. An enhanced motion compensated prediction scheme for hybrid video coding is presented, which is based on high-resolution displacement vectors of 1/8- and 1/16-pel. Therefore the motion compensated prediction scheme has to be adapted to the higher displacement vector resolution. Using higher displacement vector resolutions the interpolation scheme has to be adapted too. For this purpose, different interpolation filters are analysed. The enhanced motion compensated hybrid video codec with high-resolution displacement vectors is tested with different test sequences and the coding result is compared to the results of video coding standards.

In Section 2 the motion compensated prediction with fractional-pel displacement vector resolution is introduced in detail. Section 3 is focused on the interpolation process for different displacement vector resolutions and with different interpolation filters. In Section 4 the enhanced motion compensated hybrid video codec with high-resolution displacement vectors is presented. Experimental results are presented in Section 5 and in Section 6 a summary is given.

2. MOTION COMPENSATED PREDICTION WITH FRACTIONAL-PEL DISPLACEMENT VECTOR RESOLUTIONS

In this Section the motion compensated prediction module of Figure 1 is described in detail. This module uses the already transmitted signal $s'(t\text{-}1)$ and the estimated displacement vectors $\vec{d}(t)$ in order to create the prediction image $\hat{s}(t)$. Figure 2 shows the block diagram of the motion compensated prediction module for a displacement vector resolution of $1/M$. The

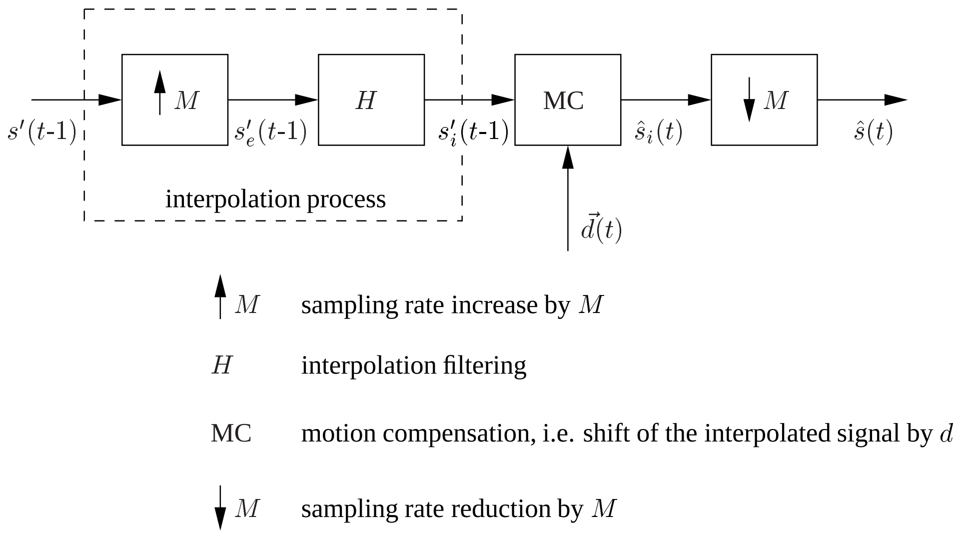


Figure 2. Block diagram of the motion compensated prediction based on a displacement vector resolution of $1/M$.

motion compensated prediction is performed in two steps. The first step is the interpolation process, where the sampling rate of the already transmitted image $s'(t\text{-}1)$ is increased by a factor of M and filtered with an interpolation filter H. The result of this first step is the interpolated image $s'_i(t\text{-}1)$. In the second step the interpolated signal is shifted according to the estimated displacement vector $\vec{d}(t)$ and the sampling rate is reduced by the factor of M. The result is the motion compensated image $\hat{s}(t)$. The next Section describes the interpolation process in detail.

3. INTERPOLATION PROCESS

Figure 3 shows the block diagram of the interpolation process for an interpolation by an exemplary factor of 4. In this Figure two representations for this interpolation are given. In part a) of Figure 3 the interpolation process is performed in one step

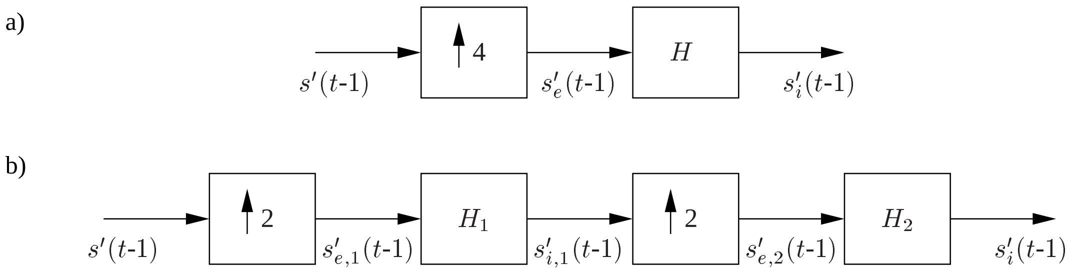


Figure 3. Two representations for a sampling rate increase by a factor of 4 and interpolation filtering. a) Based on one step with one interpolation filter H b) Based on two steps with two interpolation filters H_1 and H_2

with one interpolation filter H. This is the same representation that is also used in Figure 2. Part b) of Figure 3 shows a representation, where the interpolation is performed in two steps with two interpolation filters H_1 and H_2. Since in the video coding standards MPEG-4 ACE and H.26L the representation of part b) is used, it is also used in this contribution. Note that the representation b) can be transfered to representation a). In order to interpolate displacement vector resolutions of 1/8- and 1/16-pel, the interpolation process of Figure 3 b) is extended by one or two steps, respectively.

In the following table 1 an overview over different video coding standards, their used displacement vector resolution (DV-res) and their interpolation filters H_1 and H_2 are given. The most popular interpolation method is the bilinear interpolation,

Standard	DV-res	M	H_1	H_2
H.261	1/1-pel	1	–	–
H.263	1/2-pel	2	Bilinear	–
MPEG-1,2,4	1/2-pel	2	Bilinear	–
MPEG-4 ACE	1/4-pel	4	Wiener (8 tap)	Bilinear
H.26L (TML5)	1/4-pel	4	Wiener (6 tap)	Bilinear

Table 1. Video coding standards with their motion vector resolution (DV-res), their upsampling factor M and their interpolation filters H_1 and H_2.

which is used in H.263, MPEG-1, MPEG-2 and MPEG-4. Recent approaches like MPEG-4 ACE and H.26L are based on Wiener interpolation filters.[3,4] These Wiener Interpolation Filters were designed to interpolate the image signal while reducing aliasing components, which are deteriorating the motion compensated prediction.[7,8] The different interpolation filters and the dependence between the filter coefficients and the impulse response are depicted in table 2. In order to interpolate the two dimensional image, the one dimensional filters are applied first horizontal and after this vertically.

Name	taps	Filter coefficients	Impulse response of interpolation filter
Bilinear	2	(128,128)/256	(128,256,128)/512
Hamming-Sinc	8	(-2,8,-34,156,156,-34,8,-2)/256	(-2,0,8,0,-34,0,156,256,156,0,-34,0,8,0,-2)/512
Wiener	8	(-8,24,-48,160,160,-48,24,-8)/256	(-8,0,23,0,-48,0,161,256,161,0,-48,0,23,0,-8)/512

Table 2. One dimensional interpolation filters with their number of filter taps, their filter coefficients and their impulse response.

The Bilinear and the Wiener Filter were already mentioned above. The 8-tap Wiener Filter is the same filter, which is used in MPEG-4 ACE. Additionally the 8-tap Hamming-Sinc filter is used, which results from a Hamming windowed Sinc function.[6] The Sinc function is the impulse response of an ideal low-pass filter, which should be used for an interpolation of an image without aliasing.[6] In Figure 4 the frequency responses of these three interpolation filters is depicted. The Figure shows,

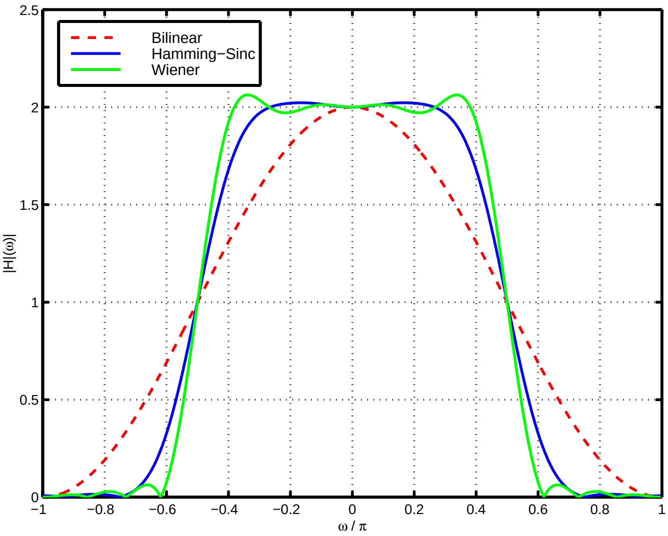


Figure 4. Magnitude responses of different interpolation filters.

that the bilinear filter is faraway from an ideal low-pass filter. It has a significant passband attenuation and also a significant stopband permeability. The Hamming windowed Sinc function has a much more sharper filter characteristic. Thus the filter remains close to the magnitude two over more of the passband and becomes close to zero more rapidly in the stopband. While the magnitude response of the Hamming-Sinc is monotonic in both the passband and the stopband, the Wiener filter has an equiripple behavior. But therefore the Wiener Filter has a sharper filter characteristic than the Hamming-Sinc.

4. THE HYBRID VIDEO CODEC BASED ON HIGH-RESOLUTION DISPLACEMENT VECTORS

The displacement vector resolutions of 1/8- and 1/16-pel and the different interpolation filters are implemented in the H.26L (TML1) codec. In the motion compensated prediction scheme the motion vector candidate has to be estimated for each block. For the 1/8-pel displacement vector resolution this is done by the following 5 steps:

1. Make a full search to find the best 1/1-pel vector.
2. Check the 8 1/2-pel positions around the best 1/1-pel vector in order to find the best 1/2-pel vector.
3. Check the 8 1/4-pel positions around the best 1/2-pel vector in order to find the best 1/4-pel vector.
4. Check the 8 1/8-pel positions around the best 1/4-pel vector in order to find the best 1/8-pel vector.
5. Select the motion vector, which produces the lowest rate-distortion cost.

For the 1/16-pel displacement vector resolution one more step is used. In order to code the displacement vectors with enhanced resolution, the coding scheme for the displacement vectors is adapted.

In order to reduce the possible filter combinations for the experimental results, the same interpolation filter is used in each interpolation step. E.g. for a 1/4-pel interpolation scheme, as it is depicted in Figure 3, this means that $H_1 = H_2$.

5. EXPERIMENTAL RESULTS

The test sequences used for the experimental results are *Mobile & Calendar* and *Flower Garden* each at CIF format and 30 Hz. The results for different interpolation filters and the results for different displacement vector resolutions are presented and discussed separately. The coding gain has been measured by the PSNR of the reconstructed image signal.

In order to investigate the influence of the interpolation filters on the coding efficiency Figure 5 shows the operational rate distortion curves for the three filters of table 2 and for the different displacement vector resolutions of 1/2-, 1/4- and 1/8-pel. It turned out, that the Wiener filter leads to the best results for all displacement vector resolutions. The higher the displacement vector resolution is the higher the difference between the Wiener filter and the bilinear and Hamming-Sinc is. The gain in PSNR between the results obtained with the Wiener Filter and the bilinear Filter is up to 0.8 dB for 1/2-pel displacement vector resolution, up to 1.5 dB for 1/4-pel displacement vector resolution and up to 2.3 dB for 1/8-pel displacement vector resolution. The highest gain is achieved for higher bit rates and the gain is lower for the lower bit rates. Furthermore a significant difference between the Wiener Filter and the Hamming-Sinc filter is obtained. This is due to the fact that the Wiener filter was developed to reduce the aliasing components that are deteriorating the motion compensated prediction. Since the Wiener Filter leads to the best performance for all resolutions and sequences it is used for the further comparisons of the displacement vector resolutions.

In order to investigate the results for different displacement vector resolutions Figure 6 shows two graphs with operational rate distortion curves for the test sequences used. In addition to the curves for the displacement vector resolutions 1/2, 1/4, 1/8, 1/16 with the Wiener interpolation filter a curve tor the 1/2-pel resolution with the bilinear Interpolation filter (H.263, MPEG-2, MPEG-4) is shown in dashed. The curves show, that there is 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 lead to further gains. Compared to 1/2-pel and bilinear interpolation filter as it is used in H.263, MPEG-1,2,4 a gain up to 3.0 dB is obtained. Compared to 1/4-pel and the 8-tap Wiener interpolation filter as it is used in MPEG-4 ACE a gain up to 1.0 dB is obtained.

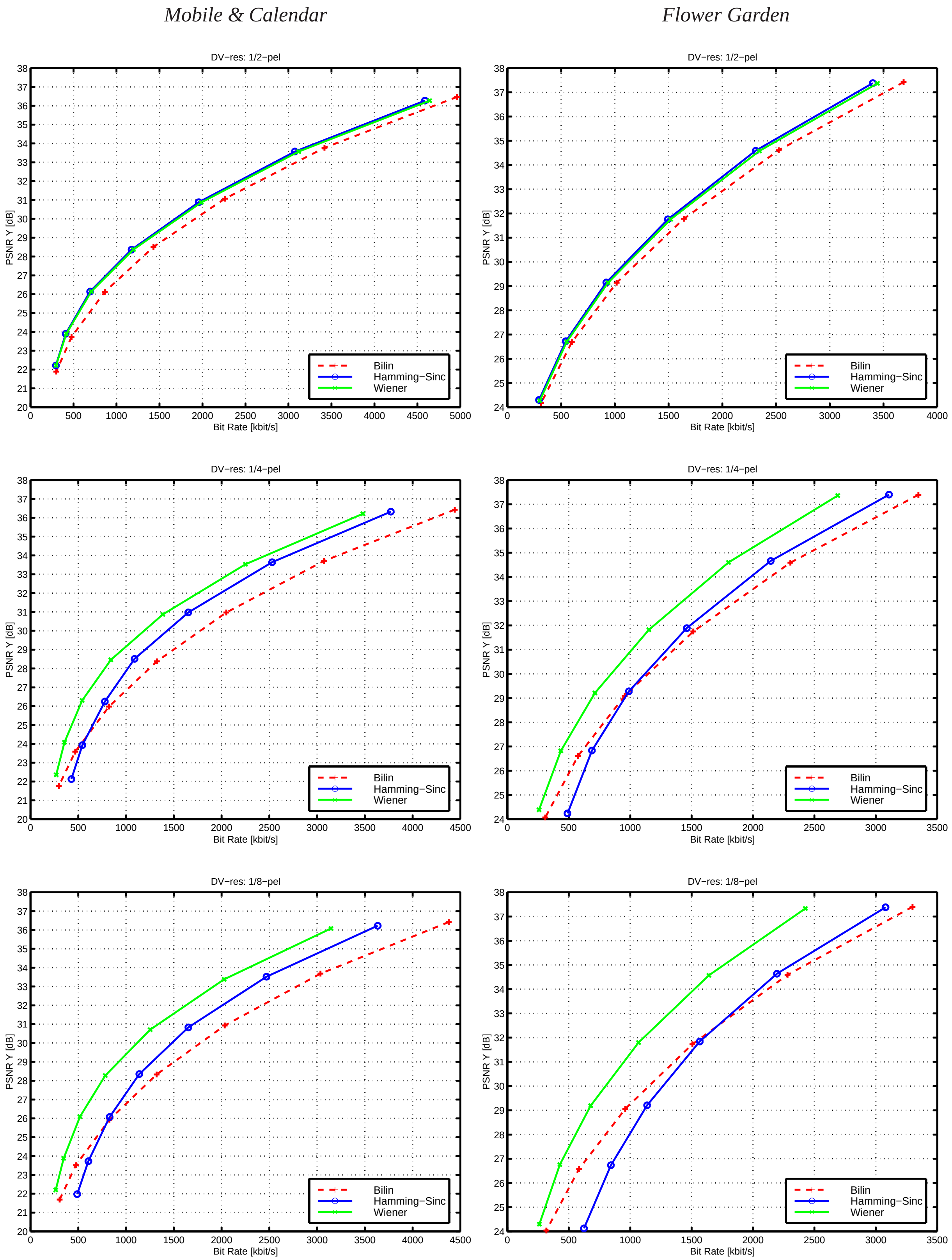
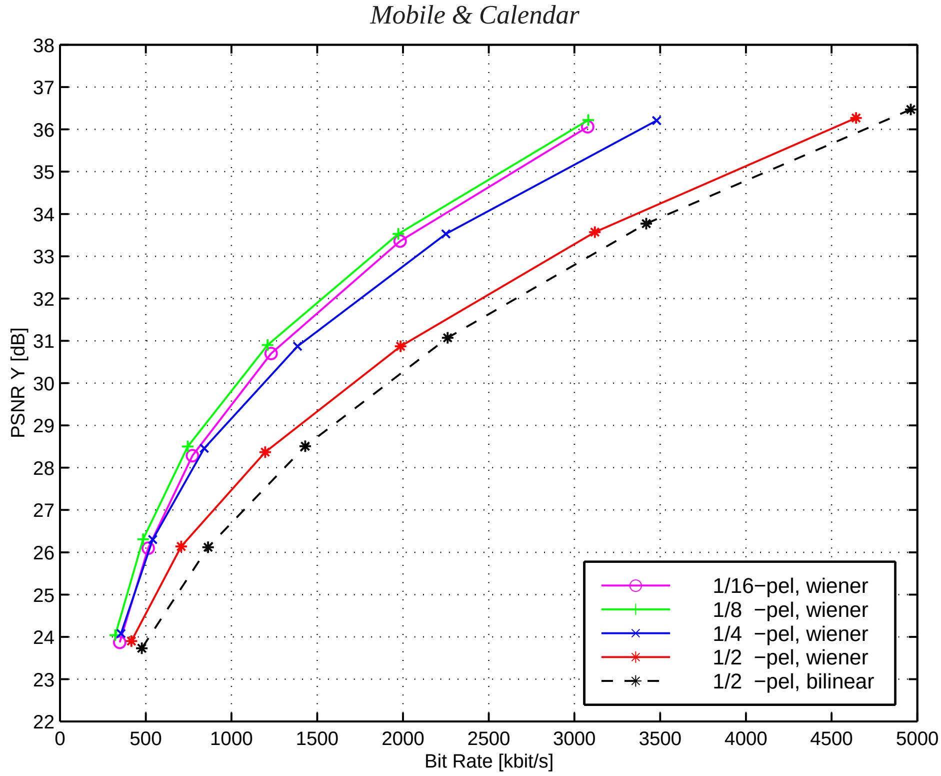


Figure 5. Operational rate distortion curves for the different interpolation filters and different displacement vector resolutions (DV-res) of 1/2-, 1/4- and 1/8-pel for test sequences *Mobile & Calendar* (left) and *Flower Garden* (right).



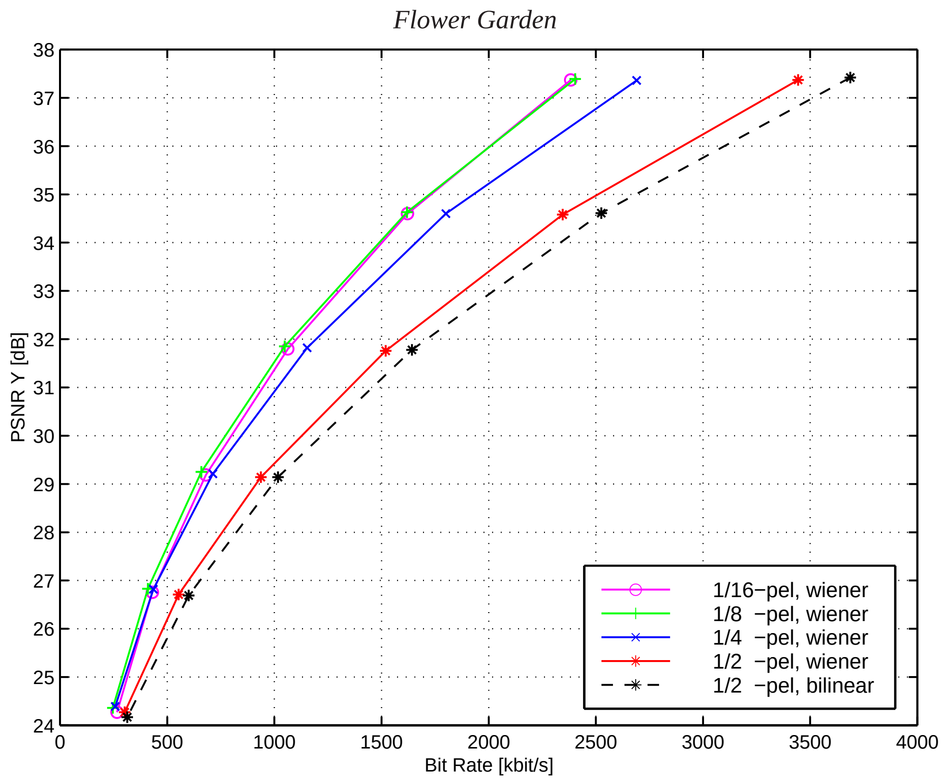


Figure 6. Operational rate distortion curves for test sequences *Mobile & Calendar* and *Flower Garden*.

6. SUMMARY

A motion compensated hybrid video codec based on high-resolution displacement vectors is presented. Therefore displacement vector resolutions of 1/8- and 1/16-pel are used. Furthermore three different interpolation filters are applied and analysed. The very simple bilinear filter, an 8-tap Hamming windowed Sinc filter and an 8-tap Wiener interpolation filter. The codec was tested with the sequences *Mobile & Calendar* and *Flower Garden* each at CIF format and 30 Hz.

An analyse for the different interpolation filters shows that the Wiener filter leads to best coding efficiency. The higher the displacement vector resolution is, the higher the influence of the filter on the coding gain is. For a displacement vector resolution of 1/2-pel the wiener filter leads to a gain up to 0.8 dB compared to the bilinear filter and almost no gain compared to the hamming Sinc. For a displacement vector resolution of 1/8-pel the wiener filter leads to a gain up to 2.3 dB compared to the bilinear filter and 1.5 dB compared to the hamming Sinc filter.

The results for the different displacement vector resolutions showed a significant coding gain while using high-resolution displacement vectors of 1/8 pel. The 1/16-pel displacement vector resolution does not lead an additional gains compared to 1/8-pel resolution. Compared to 1/2-pel and bilinear interpolation filter as it is used in H.263, MPEG-1,2,4 a 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 ACE a gain up to 1.0 dB is obtained.

REFERENCES

1. ITU Telecom. Standardization Sector of ITU, ”Video Codec for audiovisual services at $p \times 64$ kbit/s ”, ITU-T Recommendation H.261; Version 1, Nov. 1990; Version 2, Mar. 1993
2. ITU Telecom. Standardization Sector of ITU, ”Video coding for low bitrate communication”, ITU-T Recommendation H.263; Version 1, Nov. 1995; Version 2, Jan. 1996
3. ITU-T SG16/Q15 (G. Bjontegaard), ”H.26L Test Model Long Term Number 5 (TML-5)”, ITU-T SG16/Q15 doc. Q15-K-59, (downloadable via ftp://standard.pictel.com/video-site), Aug. 2000
4. MPEG-4: ISO/IEC, ”Final Committee Draft for ISO/IEC 14496-2”, Doc. ISO/IEC/JTC1/SC29/WG11 N2202, March 1998.
5. MPEG-4: ISO/IEC, ”Final Proposed Draft Amendment 1 for ISO/IEC 14496-2”, Doc. ISO/IEC/JTC1/SC29/WG11 N2802, Vancouver, July 1999.
6. A.V. Oppenheim, R.W. Schaefer, ”Discrete-Time Signal Processing”, Prentice Hall, ISBN 0-13-216292-X, 1989.
7. O. Werner, ”Drift analysis and drift reduction for multiresolution hybrid video coding”, Signal Processing: Image Communication, Vol. 8, No. 5, July 1996, p.387-409, ISSN: 0923-5965
8. U. Benzler, O. Werner, ”Improving multiresolution motion compensating hybrid coding by drift reduction”, Picture Coding Symposium 1996, March 1996, Melbourne.
9. MPEG-4: ISO/IEC (U. Benzler), ”Results of core experiment P8 (Motion and Aliasing Compensated Predicition”, Doc. ISO/IEC/JTC1/SC29/WG11 N2625, October 1997.
10. ITU-T SG16/Q15 (T. Wedi), ”Motion and Aliasing Comensated prediction for H.26L”, ITU-T SG16/Q15 doc. Q15-i-35, (downloadable via ftp://standard.pictel.com/video-site), Oct. 1999