Blocking Artefact Free Video Coding Based on a Bilinear Forward Image Warping Model

Guido Heising, Kai Uwe Barthel, Wiebke Johannsen, and Christoph Steinbach

Institut für Nachrichtentechnik und Theoretische Elektrotechnik
Technische Universität Berlin, Einsteinufer 25, D-10587 Berlin, Germany
E-mail: heising@ee.tu-berlin.de

Abstract

In this paper a new very low bit rate video coding scheme is presented. It is based on a forward motion estimation technique using an image warping model, thus delivering the basis for temporal tracking of regions or objects. In order to deal with new or disappearing image contents a new algorithm for the adaptation of a tracked control grid is proposed.

To improve the coding efficiency a high quality pixel interpolation in the warping based predictor has to be used, which unfortunately also emphasizes the high frequency coding artefacts. To reduce these artefacts the prediction error is coded by a biorthogonal wavelet transform. The proposed coder yields an improvement of the PSNR of ca. 0.7 dB compared to the advanced H.263 coder. In addition the visual quality is better as any blocking artefacts in the coded images are avoided.

1. Introduction

Image warping based coders using backward motion compensation in most cases reach a higher coding gain compared to block matching based schemes [1-3]. This is due to the employed piecewise affine or bilinear motion model [4]. A further advantage of the image warping model is the possibility to use it in a forward motion estimation scheme, since no block-overlaps or holes will be introduced in the predicted image. Therefore a grid, which is distorted according to the estimated motion (fig. 1a), is overlaid on the image thus enabling tracking of regions over time. The advantages are two-fold, firstly this provides the means for an object oriented coding scheme, in which different objects are tracked. Hereby some of the functionalities required by MPEG-4 are easy to implement, i.e. object zooming or transfiguration [5,6]. Secondly the temporal correlation of motion vectors may be exploited to improve the motion estimation step and the motion coding.

Due to a bad motion estimation or new or disappearing image content, the forward motion estimation may lead to an over distorted grid, which yields to a reduced prediction quality. A common solution is to adapt the grid to the edges of the image in each frame thus loosing the ability to track the motion [7,8]. In [8,9] the deletion and insertion of node points is introduced in order to handle new and disappearing image content.

In this paper a forward motion estimation scheme is presented which is based on the bilinear motion model using a grid of quadrangles. A new scheme for the adaptation of the quadrangle based grid is presented. Furthermore it is shown that combining an warping based predictor with a wavelet coder leads to images free of blocking artefacts.

2. Forward motion estimation using a bilinear image warping model

For motion estimation the previous picture, denoted by index a, is subdivided into quadrangles, thus creating a control grid. For every control point (vertex of a quadrangle) a motion vector is estimated by octagonal matching, i.e. finding the minimal mean square error between the predicted and the current four surrounding quadrangles of a control point (fig. 1a). In doing so the motion vectors of the eight neighbor control points are fixed and only the motion vector of the center control point is changed. The motion vectors are iteratively refined using a 4-step search algorithm in each iteration. The prediction is made by warping [4] the quadrangles of the previous picture to the predicted deformed quadrangles.

2.1 Motion interpolation

For warping based prediction the motion vectors inside the quadrangles are bilinearly interpolated from their 4 vertex motion vectors (fig. 1b, eq. 1-4). According to the forward motion estimation scheme the motion vectors are

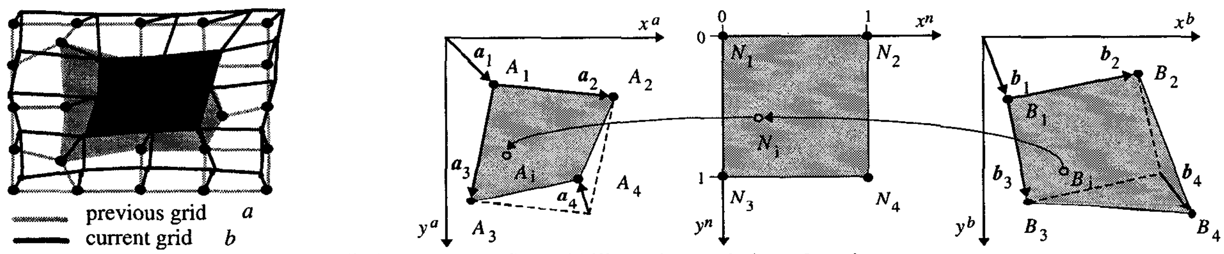


Fig. 1: a) Principle of image warping prediction b) Spatial bilinear interpolation of motion vectors

0-8186-8183-7/97 $10.00

directed from control point positions A of the previous frame a to new positions B in the current frame b. Using this new grid for motion estimation of the next frame enables tracking of control points and region motion over time.

For a given set of motion vectors, $MV_{1..4}=B_{1..4}-A_{1..4}$, the bilinear warp is done in two steps (fig. 1b). At first the coordinates at integer pixel positions of the current frame b are mapped onto the normalized plane n by an inverse bilinear transform (eq. 1).

$$x^n_{i1/2} = \frac{x_i^b - b_{11} - b_{31} y^n_{i1/2}}{b_{21} + b_{41} y^n_{i1/2}}, \quad y^n_{i1/2} = -\frac{c_1}{2c_2} \pm \sqrt{\frac{c_1^2}{4c_2^2} - \frac{c_0}{c_2}} \tag{1}$$

with

$$\begin{aligned} c_0 &= b_{21}\left(b_{12} - y_i^b\right) \\ c_1 &= b_{41}\left(b_{12} - y_i^b\right) + b_{42}\left(x_i^b - b_{11}\right) + b_{12}b_{41} - b_{31}b_{22} \\ c_2 &= b_{41}b_{32} - b_{31}b_{42} \end{aligned} \tag{2}$$

$$\begin{aligned} &\boldsymbol{b}_1 = B_1, \quad \boldsymbol{b}_2 = B_2 - B_1, \quad \boldsymbol{b}_3 = B_3 - B_1, \\ &\boldsymbol{b}_4 = B_4 - B_2 + B_1 - B_3 \quad \text{with } \boldsymbol{b}_j = \begin{bmatrix} b_{j1} \\ b_{j2} \end{bmatrix} \; j \in \{1,\dots,4\} \end{aligned} \tag{3}$$

This inverse transform is not unique, but only one set of the solutions maps the quadrangle of b inside the normalized square of n. In the second step the coordinates of the normalized plane n are mapped onto the quadrangle of frame a by a bilinear transform (eq. 4).

$$\begin{aligned} x_i^a &= a_{11} + a_{21}x_i^n + a_{31}y_i^n + a_{41}x_i^n y_i^n \\ y_i^a &= a_{12} + a_{22}x_i^n + a_{32}y_i^n + a_{42}x_i^n y_i^n \end{aligned} \tag{4}$$

The eight bilinear motion parameters a_{jk} are estimated by using eq. 3 with points $A_{1..4}$ instead of $B_{1..4}$ of fig. 1b. A gray value can be predicted for every pixel on the sample grid of frame b by using eq. 1 and 4. As the corresponding point in frame a lies on a non integer position, the gray value can be found by a bilinear interpolation of the four surrounding pixels on the sample grid.

2.2 Pixel interpolation

To reduce the strong lowpass effect of the first order filtering of the bilinear interpolation the last reconstructed frame a is upsampled prior to the motion estimation step. The advantage of this procedure is that there is no need for a repeated computation of a high quality interpolation during the motion estimation step. Best results were achieved by computing the unknown pixels at the non integer positions by a one dimensional 10-tap hamming weighted sinc interpolation filter with an upsampling factor of z=4 for both directions. By applying the bilinear pixel interpolation on this upsampled grid the predicted image gets much sharper. In addition less high frequency transform coefficients of the prediction error have to be transmitted, thereby a lower bit rate is achieved.

In fig. 2 the coding efficiency of the in [3] presented backward warping prediction based scheme using a DCT

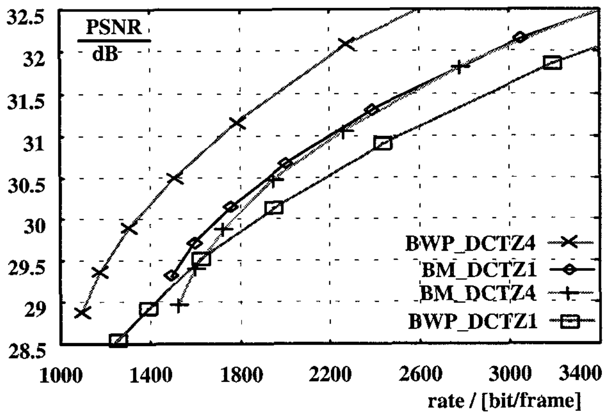
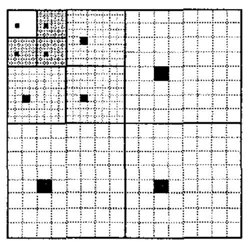


Fig. 2: Comparison of the proposed coder using either backward warping prediction (BWP) or blockmatching (BM), DCT coding and an upsampling factor of z=1 or 4 for the interpolation (for *foreman*)

error coding with (z=4) and without (z=1) upsampling, named *BWP_DCTZ4* and *BWP_DCTZ1* respectively, is compared to the here presented coder using a full-search blockmatching prediction, named *BMA_DCT1* and *BMA_DCT4*. By employing the above mentioned upsampling scheme in a warping based coder the PSNR is increased by 1.0-1.5 dB. As this scheme emphasizes the high frequency image components and with them the included coding artefacts the performance of the blockmatching based coder using upsampling is even worse. Although the PSNR in the warping based coder is increased the blocking and ringing artefacts are intensified too [10]. To reduce these visually annoying effects the usage of a lapped transform error coding is proposed.

3. Wavelet based error coding

Due to the underlying continuous motion model of the motion compensation step, the predicted image is free of blocking artefacts. In order to avoid any blocking artefacts it is proposed to use a 9/7-tap biorthogonal block overlapping wavelet coder [11] for the prediction error coding. After a 3-stage wavelet transform of the error image the quantizer and huffman code tables of the H.263 coder [12] are used, which are optimized for DCT coding. Therefore the coefficients are ordered according to their spatial dependencies in blocks of 8x8 (fig. 3a). In order to exploit the spatial and spectral correlations the scan shown in fig. 3b is applied for run length coding. Compared to the DCT based coder the bit rate is slightly increased but the visual quality is extremely improved (fig. 6a, 6c), as in the



1	2	5	6	17	18	21	22
3	4	7	8	19	20	23	24
9	11	13	14	25	26	29	30
10	12	15	16	27	28	31	32
33	35	37	39	49	50	53	54
34	36	38	40	51	52	55	56
41	43	45	47	57	58	61	62
42	44	46	48	59	60	63	64

Fig. 3: Coefficients: a) wavelet domain b) run length scan

wavelet coder no blocking artefacts are accumulated by the feedback loop of the predictor.

4. Automatic grid adaptation

In contrast to the block matching motion estimation scheme the image warping based scheme enables us to make a forward prediction. Hereby neither overlaps nor holes are introduced in the predicted frame. But when the motion is tracked over several frames, the grid gets more and more distorted (fig. 4a) and the prediction becomes worse. So it is proposed to perform a grid adaptation prior to the motion estimation.

At first the geometry of the current grid is evaluated and afterwards a local optimization is performed. This can also be done at the decoder side without transmission of additional side information.

For the detection of an over-distorted grid area the following three criteria are employed:

- *size of a quadrangle*
  If the size exceeds (falls short of) a threshold the quadrangle will be over-distorted.
- *minimal freedom of motion of the control point*
  If the minimal freedom of motion of the control point falls short of a threshold the quadrangle will be over-distorted. The gray quad in fig. 5a shows the freedom of motion of control point *P*, the black line the minimal freedom of motion d_{min}.
- *geometric form of the octagon made out of the surrounding quadrangles of a control point*
  If an inner angle of a direct neighbor control point exceeds (falls short of) a threshold, the octagon will be over-distorted (fig. 5b).

In the following four possible solutions for the grid adaptation by editing control points are introduced:

- *shift of an affected control point*
- *deletion of an affected control point*
- *insertion of an effected control point*
- *initializing the whole grid with squares*

At first all control points and respective quadrangles are detected that do not meet the quality criteria mentioned above. Then the affected points are shifted one after another, always testing the quality criteria. In those areas that still do not satisfy the quality criteria single points are deleted or inserted. In doing so the required local

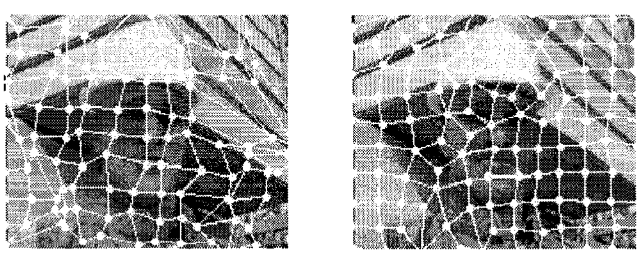


Fig. 4: a) Distorted grid without adaptation b) Adapted grid

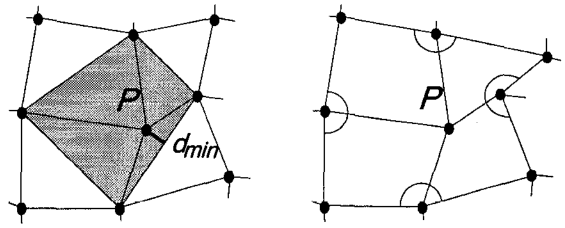


Fig. 5: a) Freedom of motion b) Over-distorted octagon

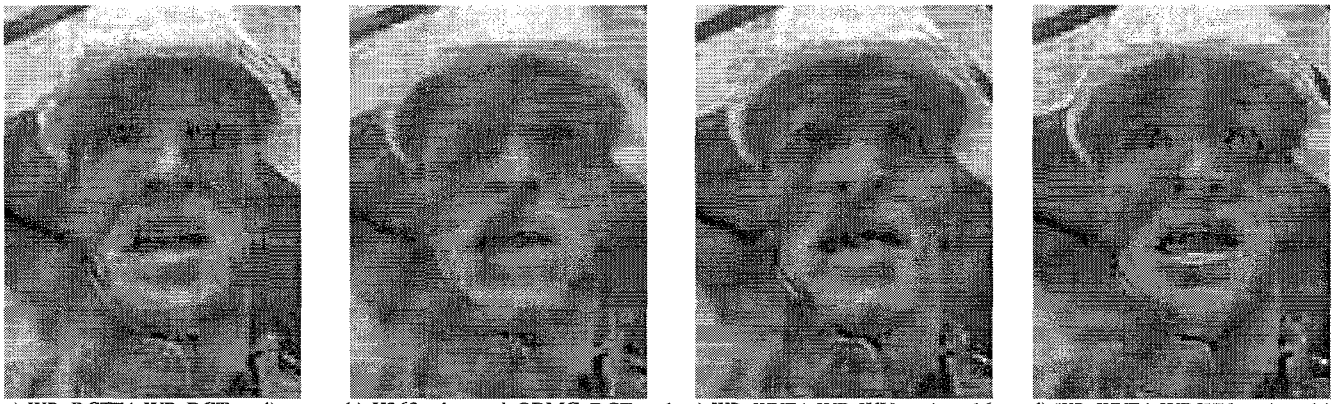
triangulation step is only allowed to produce quadrangles so that bilinear warping can still be performed. This can only be achieved if the scheme does not stick to the 4-neighborhood of each control point. This implies that a control point may have less or more than four surrounding quadrangles (fig. 4b). After a change the quality criteria are evaluated for the affected region. Iteratively the insertion/deletion and criteria evaluation is done until all points match the criteria or a maximum number of iterations is reached. In the latter case the grid is ill conditioned and a regular grid is used.

4. Simulation results

In the following simulation results concerning the coding efficiency and visual quality of the proposed coder without grid adaptation compared to other coders are shown. Up to now the coding scheme of the new coder is not able to deal with an adapted grid, as there is a varying number of motion vectors to code. So these bit rates will be presented at the conference.

Simulations were carried out with the monochrome 30 Hz QCIF sequence *foreman* (frames: 1-150, size: 176x144, block size: 16x16, vector range: ±15.5 pixel). Different fixed quantizer step sizes for error coding were used. A ITU H.263 coder of Telenor R&D, the *tmn 1.5* coder, was used as a reference system. For a fair comparison the bits for the first frame of all coders and the side information of the H.263 coders, which were not used, i.e. MCBPC, header and crominance, are not included in the statistics. All coders used the H.263 huffman tables. The averaged values of the PSNR over the bit rate are presented in fig. 7.

Fig. 7 shows that the performances of the proposed forward motion estimation scheme (*FWP_WVZ4*) using the wavelet prediction error coding and the scheme using the DCT (*FWP_DCTZ4*) are similar, but the visual quality is extremely improved. This is shown in fig. 6a and 6c where coding results of the two schemes are presented at comparable bit rates and PSNRs. Further one can see that the presented wavelet coder yields an improvement of about 1.3 dB compared to the H.263 coder and of ca. 0.7 dB compared to the advanced H.263 coder (*H263 adv_pred*). Additionally the reconstructed frames of the new coder are much sharper and free of any blocking artefacts. Fig. 6b and 6c show coding results at comparable PSNRs and 6b and 6d at comparable bit rates. Comparing the backward, (*BWP_WVZ4*) [3], and forward warping based schemes the latter performs slightly worse (fig. 7).



a) *WP_DCTZ4*: WP, DCT, z=4)
∅: PSNR=29.90, bit/frame=1309
frame 66: PSNR=29.46

b) *H263 adv_pred*: OBMC, DCT, z=1
∅: PSNR=29.85, bit/frame=1602
frame 66: PSNR=29.54

c) *WP_WVZ4*: WP, WV, z=4, q=16
∅: PSNR=30.01, bit/frame=1368
frame 66: PSNR=29.68

d) *WP_WVZ4*: WP,WV, z=4, q=14
∅: PSNR=30.59, bit/frame=1595
frame 66: PSNR=30.31

Fig. 6: Coding results of the 66. frame of *foreman* (contrast increased by 2); (∅: averaged values for the first 150 frames; BM: blockmatching; WP: warping prediction; OBMC: overlapped block motion compensation; DCT: discrete cosine transform; WV: wavelet transform; z: upsampling factor: q: half quantizer step size)

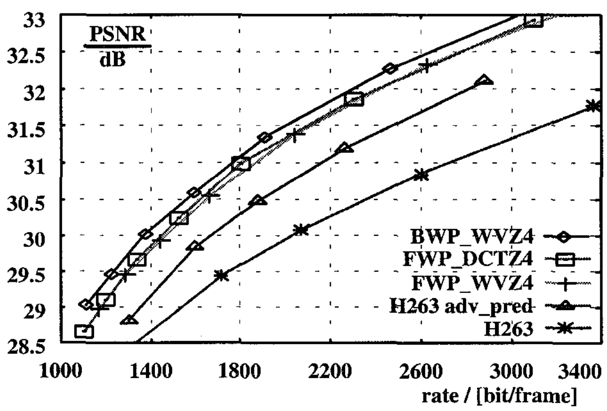


Fig. 7: Comparison of the proposed coder (FWP_WVZ4) with the ITU H.263 coders for the first 150 frames of *foreman*

5. Conclusions

A new blocking artefact free video coding scheme based on a forward image warping motion model has been presented. It is shown that a high quality pixel interpolation has to be used in a warping based predictor to reach a higher PSNR. But this kind of interpolation unfortunately also emphasizes the blocking and ringing artefacts. It is shown that it is essential to use a lapped transform error coding in a warping based coder in order to avoid blocking artefacts. A new coder combining warping based prediction with wavelet error coding is proposed. This coder outperforms the warping based prediction coder using a DCT as far as the visual quality is concerned and the ITU H.263 coder in both, the visual quality and the PSNR. A gain of the PSNR of about 0.7 dB is achievable compared to the advanced version of the H.263 coder.

Furthermore a new scheme for the adaptation of the used warping grid is proposed, which delivers the basis for region tracking over time. Thus it is well suited for an object oriented coding scheme. The current work is focused on changing the coding algorithms to meet the requirements of the grid adaptation based scheme.

References

[1] G. J. Sullivan, and R. L. Baker, "Motion compensation for video compression using control grid interpolation," *Proc. Inter. Conf. on Acoust. Speech Signal Processing (ICASSP'91)*, Toronto, Canada, May 1991, pp. 2713-2716.

[2] C.-L. Huang, and C.-Y. Hsu, "A new motion compensation method for image sequence coding using hierarchical grid interpolation," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 4, No. 1, February 1994, pp. 42-52.

[3] G. Heising, and G. Ruhl, "Video coding using spatial extrapolation based motion field segmentation," *IEEE Inter. Conf. on Image Process.*, Vol. II, Sep. 1996, pp. 482-485.

[4] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, CA, 1990.

[5] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 7, No. 1, February 1997, pp. 19-31.

[6] C. Toklu, A. T. Erdem, M. I. Sezan, and A. M. Tekalp, "2-D Mesh Tracking for Synthetic Transfiguration," *IEEE Inter. Conf. on Image Process.*, Vol. III, 1995, pp. 536-539.

[7] O. Lee and Y. Wang, "A deformable block matching algorithm for motion compensated prediction," *Picture Coding Symposium (PCS 94)*, Sep. 1994, pp. 223-226.

[8] Salembier, P.; Ayuso, X.: Very low bit rate video coding using active triangular mesh. Proc. Inter. Conf. on Acoust. Speech Signal Processing (1996), S. 2060-2063.

[9] Y. Altunbasak, A. M. Tekalp, and G. Bozdagi, "Two-dimensional object-based coding using a content-based mesh and affine motion parameterization," *IEEE Inter. Conf. on Image Processing,* Vol. 2, 1995, S. 394-397.

[10] G. Heising, "Blocking artefact free video coding by combining warping based prediction with wavelet error coding," *Picture Coding Symposium (PCS 97)*, Berlin, Sep. 1997.

[11] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, Vol. 1, April 1992, pp. 205-220.

[12] ITU-T Recommendation H.263, "Video coding for low bit rate communication (TMN5)," July 1995.