Mode-Dependent Intra Frame Interpolation for H.264/AVC Compressed Video

Xinwei Gao, Xiaopeng Fan, Debin Zhao
School of Computer Science and Technology, Harbin Institute of Technology
Harbin, China, 150001
{xwgao.cs, fxp, dbzhao}@hit.edu.cn

***Abstract*—In this paper, a mode-dependent intra frame interpolation method is proposed for H.264/AVC compressed video. The intra prediction mode information is taken into account in the interpolation filter design. For each intra prediction mode, an optimal Wiener filter is trained based on the representative video sequences. Therefore the trained filter is adaptive to the intra prediction mode. Furthermore, the quantization parameter is also explored as context information for filter selection. Extensive experiments demonstrate that the proposed method achieves better performance than the traditional methods such as Bicubic, Bilinear, LAZA and NEDI, while keeping low computational complexity.**

I. INTRODUCTION

Image interpolation, which addresses the problem of rescaling a low resolution (LR) image to a high resolution (HR) image, is one of the most elementary research topics in image processing. Image interpolation has a wide range of applications in digital photography, video communication, satellite remote sensing, object recognition, medical analysis, and consumer electronics. Image interpolation is an ill-posed problem due to the fact that there are generally multiple HR images that can be downsampled to the same LR image.

A number of image interpolation methods have been developed. The simplest techniques for image interpolation among these existing methods are based on classical fixed linear filters, such as the Bilinear and Bicubic [1]. These linear filters are efficient for flatten regions, but may not be efficient for edges and texture regions. To improve the efficiency, a spatially adaptive interpolation algorithm called LAZA is proposed in [2], which performs interpolation along local edge directions. LAZA uses simple rules and configurable thresholds to explicitly detect edges and updates the interpolation process accordingly. In [3], a fusion based method is proposed, it first interpolates the missing pixel in the preset multiple directions, gets multiple interpolation results, and then fuses these multiple results by minimum mean square-error estimation (MMSE). Li and Orchard propose a new edge-directed interpolation (NEDI) method [4], in which the linear regression model is used to estimate coefficients to adapt the interpolation at the HR image. Furthermore, Zhang and Wu propose the SAI algorithm [5], which learns and adapts varying scene structures using a 2-D piecewise AR model by a soft-decision manner. In [6], multi-frame is considered for image super resolution and the support vector regression is applied in [7]. Hardie proposes to train Wiener filters based on the motion position of an observation window in image [8]. All the aforementioned methods deal with the uncompressed image/video, but few works focus on the compressed image/video. In [9], the authors introduce a Bayesian super resolution reconstruction technique to model compression and exploit the quantization step information for MPEG-2, H.261, and DV. As [9] describes, "*Super-Resolution algorithms designed for original video don't perform well when directly applied to decompressed image sequences, especially for low compression bit-stream*".

In this work we propose a mode-dependent intra frame interpolation for H.264/AVC compressed video. The intra prediction mode information is taken into account in the filter design. For each intra prediction mode, an optimal Wiener filter is trained based on the representative video sequences. Furthermore, the quantization parameter is also explored as context information. Note that in [10], an adaptive Wiener filter has been proposed for the fractional pixel motion compensated prediction in video coding. Different from [9] and [10], the trained optimal filter in our method is adaptive to the intra prediction mode.

The rest of this paper is organized as follows. Section II presents the framework of the mode-dependent intra frame interpolation method. Experimental results are provided in Section III. Section IV concludes this paper.

II. PROPOSED INTRA FRAME INTERPOLATION

In this section, we first give a brief introduction about H.264/AVC intra prediction. Then we present the framework of mode-dependent intra frame interpolation method. At the last, the Wiener filtering training method is given.

A. H.264/AVC Intra Prediction Mode

In the new video coding standard H.264/AVC, an intra block (I block) is coded using intra prediction without referring to any data outside the current frame. Intra prediction uses pixels from adjacent, previously coded block to predict the values in the current block as Fig. 1. There are nine intra prediction modes, named as *Vertical*, *Horizontal*, *DC*, *Diagonal Down-Left*, *Diagonal Down-Right*, *Vertical-Left*, *Horizontal-Down*, *Vertical-Right*, and *Horizontal-Up* respectively. As illustrated in Fig. 2, the image on the right describes the directional stripe of intra prediction mode. These directional stripes could approximately describe the edge of the image on the left. As indicated in [11], interpolation along edge direction is very effective. This is because, based on geometric constraint of edges, estimation along the edge orientation is optimal in the sense of best inferring unknown pixels. Therefore, a mode-dependent intra frame interpolation method is proposed in the following.

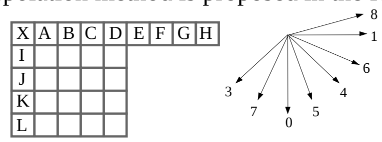


Figure 1. Labeling of prediction samples, 4×4 prediction.

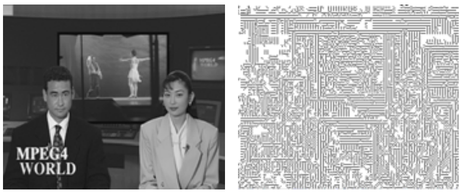


Figure 2. One frame of *News* and its corresponding intra prediction mode's spatial distribution.

B. Mode-Dependent Intra Frame Interpolation

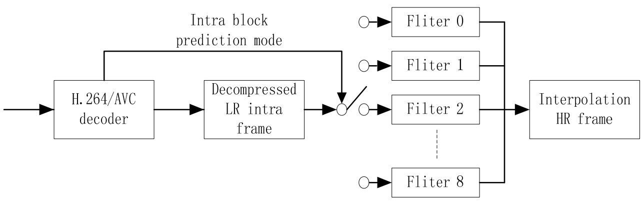


Figure 3. Intra frame interpolation flow chart.

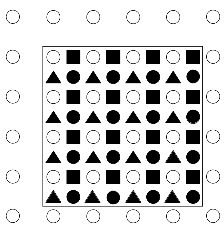


Figure 4. Intra block interpolation of proposed method.

The proposed method is illustrated by taking a single interpolation frame with 2x2 scaling. Fig. 3 shows the flow chart, I frames are decompressed by H.264/AVC decoder, then the filter is chosen by the prediction mode of each intra block for interpolation. Let $\hat{g}$ be a rectangular decompressed LR intra frame. $\hat{f}$ is the rectangular corresponding HR frame to be interpolated. The intra block interpolation is depicted in Fig. 4. The white dots represent the decompressed pixels which we will use to interpolate other pixels. The black triangles represent the pixels in the vertical direction, and they are interpolated by

$$\hat{f}(2x, 2y+1) = \mathrm{G}_{(x,y)} \mathrm{W}_{(k,qp,0)}. \tag{1}$$

The black squares represent the pixels in the horizontal direction. They are interpolated by

$$\hat{f}(2x+1, 2y) = \mathrm{G}_{(x,y)} \mathrm{W}_{(k,qp,1)}. \tag{2}$$

The black dots represent pixels in the diagonal direction. They are interpolated by

$$\hat{f}(2x+1, 2y+1) = \mathrm{G}_{(x,y)} \mathrm{W}_{(k,qp,2)}. \tag{3}$$

In (1)-(3), $\mathrm{G}_{(x,y)} = \left(\hat{g}(x-L+1, y-L+1), \ldots, \hat{g}(x+L, y+L)\right)$ represents the intensity values in the intra compressed frame and $\mathrm{W}_{(k,qp,p)} = \left(w_{(k,qp,p)}(0,0), \ldots, w_{(k,qp,p)}(2L-1, 2L-1)\right)^T$ is the weight vector of the Wiener filtering, p is the subpixel position and p=0,1,2 corresponds to the black triangle pixels, the black square pixels and the black dot pixels respectively. k is the intra block prediction mode. qp is the quantization parameter, which is also considered as context information for Wiener filtering.

C. Wiener Filtering Training

We use the mean square error (MSE) to measure the performance of the intra block interpolation as follows:

$$MSE = E\left(\left\|f - \hat{f}\right\|^2\right) = \sum_{(m,n)\in block}\left[\left(f(m,n) - \hat{f}(m,n)\right)^2\right], \tag{4}$$

where $\|.\|$ denotes the L_2 norm, f is the original HR frame. The optimum weights W should be the one minimizing the MSE in (4). However, such W is unavailable because the actual pixels in f are not available at the decoder. Therefore, we use Wiener filtering training method [12] to offline calculate the W based on some training set. The best coefficient vector $\mathrm{W}_{(k,qp,p)}$ is computed in the training set by MSE Wiener-Hopf equation:

$$\frac{d(MSE)}{d\mathrm{W}_{(k,qp,p)}} = 0. \tag{5}$$

The obtained $\mathrm{W}_{(k,qp,p)}$ is used in (1)-(3) to calculate the $\hat{f}$.

In the training, we take eight CIF sequences: *News*, *Tempete*, *Mobile*, *Football*, *Bus*, *Stefan*, *Foreman* and *Mother*, 250 frames each sequence as the training set. All the frames are coded in I frames, with four different QP = 24, 28, 32, 36. So we finally got the Wiener filters for intra prediction modes (*Vertical*, *Horizontal*, *DC*, *Diagonal Down-Left*, *Diagonal Down-Right*, *Vertical-Left*, *Horizontal-Down*, *Vertical-Right*, and *Horizontal-Up)* under the four quantization parameters. In Fig. 5, we take the weight vectors ($L = 3$) for the intra prediction modes: 3 (Diagonal Down-Left) and 4 (Diagonal Down-Right) as an example.

$$\begin{pmatrix} -0.01 & -0.01 & 0.02 & 0.02 & -0.02 & 0.01 \\ -0.02 & 0.02 & -0.05 & -0.10 & 0.04 & -0.03 \\ 0.03 & -0.07 & 0.28 & 0.47 & -0.14 & 0.06 \\ 0.05 & -0.13 & 0.43 & 0.30 & -0.08 & 0.03 \\ -0.01 & 0.03 & -0.06 & -0.06 & 0.03 & -0.02 \\ 0.01 & -0.01 & -0.01 & 0.02 & -0.01 & 0.01 \end{pmatrix} \begin{pmatrix} 0.01 & -0.02 & 0.01 & 0.01 & -0.01 & -0.01 \\ -0.03 & 0.03 & -0.07 & -0.08 & 0.03 & -0.02 \\ 0.04 & -0.13 & 0.43 & 0.33 & -0.11 & 0.05 \\ 0.05 & -0.11 & 0.33 & 0.44 & -0.12 & 0.03 \\ -0.03 & 0.026 & -0.06 & -0.07 & 0.04 & -0.02 \\ -0.01 & 0.01 & 0.01 & 0.01 & -0.01 & 0.01 \end{pmatrix}$$

Figure 5. The weight vectors for intra modes: 3(Left) and 4(Right).

It can be seen from Fig. 4, the weights of the positions along the intra prediction mode's direction are much greater than other weights in these vectors.

III. Experimental Results

To evaluate the proposed method, extensive experiments were carried out in this section. For thoroughness and fairness of our comparison study, we exploit some widely used CIF sequences: *News*, *Tempete*, *Mobile*, *Football*, *Bus*, *Stefan* (6 sequences in the training set) and *Basket*, *Akiyo*, *Container*, *Funfair*, *Novel* of the size 352x288, 4CIF sequences: *Crew* and *Harbour* of the size 704x576, and 720P sequences: *City* and *Cyclists* of the size 1280x720 (9 sequences in the testing set), 30 frames each sequence. First the MPEG-B down-sampling is used in our experiments (each image is filtered and then down-sampled by the direct-subsampling method. The filter coefficient is set to be [2, 0, -4, -3, 5, 19, 26, 19, 5, -3, -4, 0, 2]/64 [13]). These video sequences are compressed by H.264/AVC in the form of all I frames. The proposed interpolation method is performed at the decoder. In our experiments, the W presents the optimal Wiener filter vector with size of $2L \times 2L$ ($L = 3$), which is adaptive to the intra prediction mode.

The performance is measured by PSNR and SSIM [14] between original video and interpolated video acquired both in the training set and the testing set. Our method is compared with some representative work in the literature: (1) bicubic interpolation [1], (2) bilinear interpolation, (3) locally-adaptive zooming algorithm (LAZA) [2], and (4) new edge-directed interpolation (NEDI) [4].

Since the original HR images are known in the simulation, we can compare the interpolation results with the true sequences and measure the objective and subjective quality of them. Tables I-II tabulate the objective quality comparison with respect to PSNR of the five different methods when applied to the six test sequences in training set. It can be observed that for all instances the proposed algorithm consistently works better than other methods. From Tables I and II, the proposed method can improve the objective quality of generated HR frames. The average gains in Tables I and II are 0.40dB and 0.26dB compared to Bicubic respectively. Compared to Bilinear, the average gains are more than 0.6dB. Our method also outperforms the edge detection based local methods: LAZA and NEDI. The gains are 1.05dB and 0.7dB in Tables I and II compared to LAZA. Compared to NEDI, the average gains are more than 1dB.

PSNR can measure the intensity difference between two videos, but it may fail to describe the visual perception quality of the video.

TABLE I. Comparison of PSNR on QP=24

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
News	28.46	27.78	27.74	27.63	29.07
Tempete	26.05	25.72	25.65	25.33	26.23
Mobile	21.98	21.63	21.60	21.22	22.33
Football	28.59	27.93	27.84	27.23	29.16
Bus	25.23	24.83	24.77	24.27	25.56
Stefan	26.02	25.50	25.40	24.35	26.37
Average	**26.05**	**25.50**	**25.40**	**24.35**	**26.45**

TABLE II. Comparison of PSNR on QP=32

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
News	27.73	27.18	27.14	27.06	28.19
Tempete	25.29	25.04	24.98	24.75	25.37
Mobile	21.60	21.29	21.26	20.95	21.84
Football	27.29	26.83	26.77	26.34	27.62
Bus	24.58	24.25	24.22	23.82	24.79
Stefan	25.34	24.92	24.84	23.95	25.57
Average	**25.30**	**24.91**	**24.86**	**24.47**	**25.56**

The SSIM index is one of the most commonly used measures for image visual quality assessment. We further use SSIM to measure the average visual quality of all the frames of these interpolation methods. The higher SSIM value means the better visual quality. From Tables III-IV, it could be seen that proposed algorithm again achieves the highest average SSIM scores among the competing methods. It means our method can achieve better performance on the image visual quality.

TABLE III. Comparison of SSIM on QP=24

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
News	0.9069	0.8995	0.8990	0.8977	0.9087
Tempete	0.8264	0.8118	0.8089	0.8003	0.8332
Mobile	0.7421	0.7257	0.7245	0.7107	0.7534
Football	0.8563	0.8418	0.8381	0.8290	0.8617
Bus	0.8054	0.7908	0.7871	0.7693	0.8160
Stefan	0.8661	0.8524	0.8500	0.8350	0.8747
Average	**0.8338**	**0.8203**	**0.8179**	**0.8070**	**0.8412**

TABLE IV. Comparison of SSIM on QP=32

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
News	0.8759	0.8696	0.8693	0.8684	0.8767
Tempete	0.7734	0.7610	0.7587	0.7526	0.7771
Mobile	0.7041	0.6890	0.6881	0.6766	0.7134
Football	0.7685	0.7573	0.7553	0.7504	0.7720
Bus	0.7405	0.7286	0.7236	0.7131	0.7480
Stefan	0.8377	0.8246	0.8226	0.8089	0.8446
Average	**0.7833**	**0.7716**	**0.7696**	**0.7616**	**0.7886**

Table V tabulates the objective quality comparison with respect to PSNR of the five different methods when applied to these nine test sequences in the testing set. Table VI shows the image visual quality assessment comparison with respect to SSIM in the testing set. Compared with the other four methods, the proposed method can also improve both the objective quality and the visual quality of generated HR frames only with little loss when compared with the performance on the training set.

From these experimental results, we found an interesting phenomenon that the NEDI and the LAZA methods do not show better performances than the Bicubic and the Bilinear method in the compressed frames.

TABLE V. TESTING SET PSNR ON QP=24

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
Akiyo	24.17	23.96	23.89	23.36	24.33
Basket	33.04	32.52	32.52	32.90	33.52
Container	26.69	26.38	26.34	24.81	26.96
Funfair	25.03	24.79	24.76	24.25	25.25
Novel	28.75	28.66	28.67	28.34	28.82
Crew	34.86	34.45	34.41	34.41	35.20
Harbour	30.54	29.47	29.39	39.10	31.60
City	31.12	30.83	30.79	30.44	31.28
Cyclists	37.01	36.38	36.33	36.43	37.13
Average	**30.13**	**29.71**	**29.67**	**29.33**	**30.45**

TABLE VI. TESTING SET SSIM ON QP=24

Video	Bicubic	Bilinear	LAZA	NEDI	Proposed
Akiyo	0.9354	0.9320	0.9320	0.9352	0.9346
Basket	0.7694	0.7585	0.7539	0.7367	0.7771
Container	0.8405	0.8359	0.8345	0.8247	0.8407
Funfair	0.8123	0.8011	0.7992	0.7900	0.8190
Novel	0.8366	0.8332	0.8331	0.8328	0.8367
Crew	0.9026	0.8995	0.8986	0.8979	0.9027
Harbour	0.8991	0.8808	0.8780	0.8718	0.9144
City	0.8650	0.8585	0.8558	0.8369	0.8665
Cyclists	0.9222	0.9207	0.9202	0.9205	0.9218
Average	**0.8647**	**0.8578**	**0.8561**	**0.8496**	**0.8682**

Fig. 6 shows the subjective quality comparison. The proposed method produces better visually pleasant results among these competing methods.

IV. CONCLUSION

A mode-dependent intra frame interpolation method is proposed for H.264/AVC compressed video in this paper. In the proposed method, each pixel to be interpolated is approximated as the weighted combination of its spatial neighborhood and all pixels to be interpolated in one intra block share the same weights. Unlike other traditional interpolation methods, the weights are intra prediction mode-dependent and trained by Wiener filtering on the representative video sequences in terms of different intra prediction modes. In addition, the quantization parameter is further utilized as the context information for the proposed adaptive filter. Extensive experiments demonstrate that the proposed method achieves better performance than the traditional methods while keeping low computational complexity.

ACKNOWLEDGMENT

This work was partly supported by the National Science Foundation of China, 60736043 and the Major State Basic Research Development Program of China, 973 Program 2009CB320905.

REFERENCES

[1] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 29, no. 6, pp. 1153-1160, Dec. 1981.

[2] S. Battiato, G. Gallo, and F. Stanco. "A locally-adaptive zooming algorithm for digital images," *Image and Vision Computing*, vol. 20, no. 11, pp. 805–812, Sept. 2002.

[3] L. Zhang and X. Wu, "An edge-guided image interpolation via directional filtering and data fusion," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp.2226-2238, Aug. 2006.

[4] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521-1527, Oct. 2001.

[5] L. Zhang, X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," *IEEE Transactions on Image Processing*, vol.17, no. 6, pp. 887-896, June. 2008.

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327-1344, Oct. 2004.

[7] K. S. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1596-1610, June. 2007.

[8] R. Hardie, "A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter," *IEEE Transactions on Image Processing*, vol. 16, no.12, pp. 2953-2964, Dec. 2007.

[9] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Transactions on Image Processing*, vol. 13, no. 1, pp. 33-43, Jan. 2004.

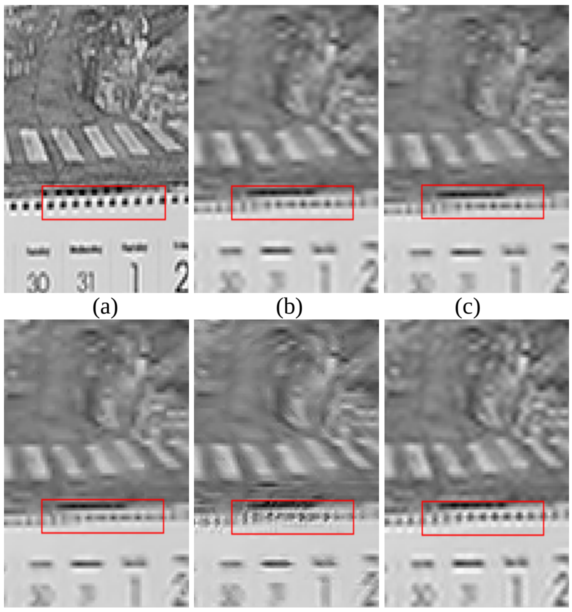
[10] S. Wittmann, T. Wedi, "Separable adaptive interpolation filter for video coding," *IEEE International Conference on Image Processing*, vol. 8, pp. 2500-2503, Oct, 2008.

[11] Xianming Liu, Debin Zhao, Ruiqin Xiong, Siwei Ma, Wen Gao, and Huifang Sun, "Image Interpolation via Regularized Local Linear Regression," accepted by *IEEE Transactions on Image Processing*, unpublished.

[12] F. Jin, P. Fieguth, L. Winger, and E. Jernigan, "Adaptive Wiener filtering of noisy images and image sequences," *IEEE International Conference on Image Processing*, vol. 3, pp. III-349-52, Sept. 2003.

[13] "Spatial scalability filters," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, July. 2005.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.



(a) (b) (c)

(d) (e) (f)

Figure 6. Comparison of different methods for sequence: *Mobile.* (a) original frame; (b) bicubic; (c) bilinear; (d) LAZA [2]; (e) NEDI [4]; (f) proposed method.