Block Adaptive Super Resolution Video Coding

Siwei Ma[1], Li Zhang[1], Xinfeng Zhang[2], and Wen Gao[1]

[1] Institute of Digital Media, Peking University, Beijing, 100871, China
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China
`{swma,xfzhang,wgao}@jdl.ac.cn`

Abstract. Super resolution technique was first proposed for enhancing the image resolution, and then it was expanded to video sequence for obtaining a higher resolution video from low resolution input. Recently, super-resolution based video coding has emerged as an important research topic as the image resolution increases rapidly and the downsampling coding is very efficient for bit rate reduction. With the super-resolution algorithm, we can encode the input video with low resolution at lower bitrate and reconstruct a high resolution video efficiently at the decoder side. In this paper, a block adaptive super resolution based coding framework is proposed for video coding. In the proposed scheme, block adaptive downsampling and upsampling with super-resolution is selected based on the rate-distortion cost decision, where the distortion caused by super-resolution algorithm in the reconstruction process is also included. Experimental results show that the proposed scheme is very promising for high resolution coding.

Keywords: Super-resolution, interpolation, rate-distortion cost.

1 Introduction

Super resolution (SR) was first proposed for enhancing the image resolution, through which a high resolution image can be reconstructed from one or more low resolution images. It has been widely researched because of a variety of applications, such as how to get better visual effect for displaying a SD (Standard Definition) video on a HD (High Definition) TV system. Many super resolution algorithms have been proposed in the recent years, and according to the number of images used for super-resolution reconstruction, it can be classified into two classes, single frame super resolution and multi-frame super resolution. For single frame super resolution, interpolation is the most intuitive approach for SR image reconstruction. In [1]-[4], adaptive interpolation techniques are used to spatially adapt the interpolation coefficients to match the local structures around the edges well. Wang and Ward [5] proposed to use the gradient to get the edge direction and the best correlation pixels with the interpolated pixel. Zhang and Wu [6] proposed to adaptively fuse two interpolation results in two mutually orthogonal directions using the statistics of a local window. Hong et al. [7] proposed to identify only five edge directions to maintain the geometry and sharpness of image and perform an adaptive combined zero order, bilinear interpolation and an adaptive Gaussian filter. In [8], Li and Orchard proposed to use the covariance of the low-resolution image to estimate the

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 1048–1057, 2009.

high-resolution image covariance, which represents the edge direction information to some extent, and proposed a Wiener-filtering like interpolation scheme. The main advantage of interpolation SR is that the computational complexity is low, making real-time applications possible, but it will cause blurring or staircase artifacts in edge area [9]. For multi-frame super-resolution, besides interpolation motion estimation takes an important role in the quality of super-resolution image. Barreto et al. [10] gave an in-depth performance evaluation between the block-matching and the optical flow based motion estimation algorithm. After motion estimation, joint temporal-spatial interpolation is used to get high resolution image. Ur and Gross [11] used the generalized multichannel sampling theorem of Papoulis and Brown to perform a non-uniform interpolation. Alam et al. [12] utilized a weighted nearest neighbor interpolation method. Nguyen and Milanfar [13] proposed a wavelet interpolation for interlaced two dimensional data.

Later, super-resolution was expanded to the video coding. Callico et al. [14] modified an existing hybrid video encoder platform to add SR based on a non-uniform interpolation method. Yet it is more like a post processing for super resolution than a real super-resolution encoder, because the reconstructed super-resolution image is not involved into the encoding framework. D. Barreto [15] proposed a region based super-resolution video coding framework. The authors found that flat areas, particularly when there is no motion, can be recovered using simple interpolation techniques. However, areas with no motion and texture are most affected by the downsampling and compression process. So the authors proposed to segment the image into flat, motion or texture region, and different super resolution scheme is used for downsampling and upsampling. In [15], the segmentation information will be transferred to the decoder side. It can be seen that the proposed region based super resolution coding scheme has better adaptivity, yet it is still not adaptive enough because it encodes the whole sequence with downsampling, and super-resolution doesn't always work well for all blocks in the region. In [16] and [17], block based adaptive downsampling/upsampling coding has shown better performance for low bitrate coding. The performance of downsampling coding was also analyzed well in [16], which is more suitable for low bitrate coding as the distortion of discarded information is very difficult to compensating at high bitrate. More efficient downsampling/upsampling method will do help to improve the performance of high bitrate coding, but the previous researches do not pay much attention to this point.

In this paper, a block adaptive super-resolution coding is proposed. In the proposed scheme, block adaptive downsampling coding is selected based on the rate-distortion cost decision, where the distortion caused by super-resolution algorithm in the reconstruction process is also included. In the proposed scheme, the blocks are classified into flat, motion and non-motion blocks by using the spatial or motion features of the reconstructed downsampled block. As the decoder can also get the reconstructed downsampled block, the segment information need not be transmitted to the decoder as [15] and the bits saving is achieved. Experimental results show that the proposed block adaptive super resolution coding scheme can achieve better performance than the traditional coding without downsampling, which may be a very promising scheme for future high resolution coding.

The rest of this paper is organized as follows, in Section 2, the block adaptive low complexity super resolution coding scheme is detailed; experimental results are provided in Section 3, and Section 4 concludes the paper.

2 Block Adaptive Super Resolution Video Coding

Figure 1 shows the proposed block-adaptive super resolution coding framework. The input macroblock may be downsampled in horizontal, vertical, or both directions respectively, and the downsampled macroblock will be coded as a new macroblock mode as [17]. But different from [17] where only the rate and distortion of the downsampled block was estimated, in the proposed scheme the downsampling mode is selected with rate-distortion optimization (RDO) mode decision method. In the proposed RDO mode decision process, the distortion of the downsampled block and the distortion of reconstructed block which was discarded in the downsampling process both are computed. As referred in [15], the performance of super-resolution does relate with the region characteristics. In the proposed scheme the blocks are classified into three kinds of blocks: flat, motion and non-motion blocks. For flat blocks, bilinear interpolation is used to reconstruct the discarded blocks. For motion blocks, an adapted SR method is used. For non-motion blocks, the edge preserved algorithm in [8] is used. As the block classification is based on the reconstructed downsampled blocks, the decoder can do the same processing correspondingly and the classification information need not be transmitted to the decoder as side information.

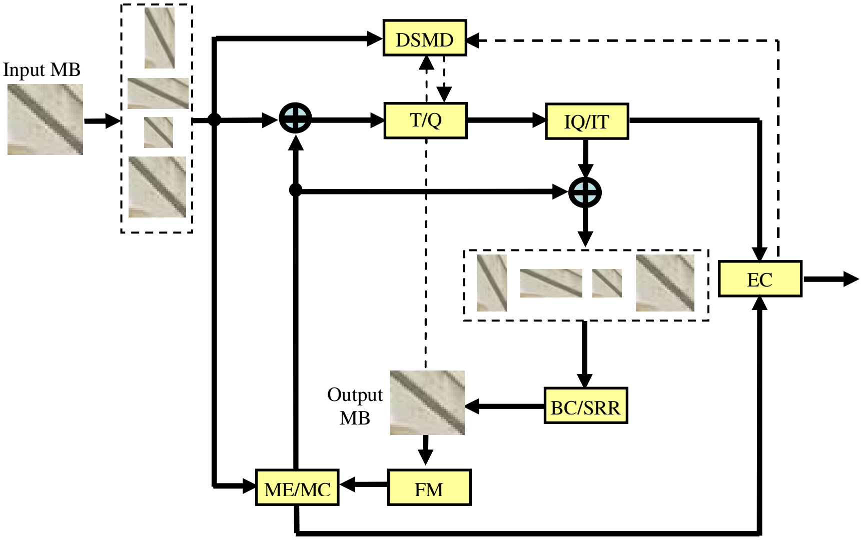


Fig. 1. Proposed block adaptive super-resolution coding framework. DSMD: Downsampling Mode Decision; T/Q: Transform/Quantization; IQ/IT: Inverse Quantization/Inverse Transform; EC : Entropy Coding; BC: block classification; SRR: Super-resolution Reconstruction; FM: Frame Memory; ME/MC: Motion Estimation/Motion Compensation.

The proposed scheme can be implemented in the following steps:

Step 1. Downsampling and coding of the downsampled block.

As shown in Table 1, three kinds of downsampling mode are defined: only horizontal or vertical downsampling (8x16, 16x8) mode, and both horizontal and vertical downsampling (8x8) mode [16][17]. Here, simple decimation filter is used for downsampling, for example, for horizontal downsampling, every other sample in the horizontal direction is kept for downsampling coding.

Table 1. Macroblock based downsampling Mode

Downsampling Mode	Description
0	No downsampling
1	One 16x8 block after vertical downsampling
2	One 8x16 block after horizontal downsampling
3	8x8 block after both horizontal and vertical downsampling

The block adaptive super resolution coding method is implemented into AVS (Audio Video coding Standard) reference software (http://www.avs.org.cn/fruits/software/rm52j_r1.zip). AVS is a video coding standard established by China AVS working group, which has the similar framework as the state-of-the-art video coding standards, such as H.264. In this paper, the block adaptive super resolution coding is only used for P frame coding, as it is easy for implementation and performance verification. The downsampling coding mode is coded jointly with the macroblock prediction mode as shown in Table 2. For P_L0_16x8_down coding mode, only one 16x8 macroblock partition is coded, including one 16x8 motion vector and coefficients of two 8x8 DCT blocks. After prediction, transform and entropy coding, the RD cost of downsampled block is calculated as:

$$JD_i = D_i + \lambda \times R_i \tag{1}$$

where i is the downsampling mode as listed in Table 1 from 1 to 3. When i is 0, the RD cost of original macroblock mode is calculated. D_i is MSE between the original blocks and the reconstructed ones in the downsampled macroblock. R_i is the bits used to encode the downsampled block, including header bits and coefficient bits.

Table 2. Macroblock modes for block adaptive super resolution coding

Macroblock Mode	Description
0	P_L0_16x16
1	P_L0_L0_16x8
2	P_L0_L0_8x16
3	P_8x8
4	P_8x8ref0
5	P_L0_16x8_down
6	P_L0_8x16_down
7	P_L0_8x8_down

Step 2. Block classification.

The block classification is based on the features of the reconstructed downsampled block. The block is classified into three categories: flat, motion and non-motion blocks. A joint spatial and temporal motion features [15] [18] block classification method is used in this paper. For each reconstructed downsampled 8x8 block, 8x8 transform is used to calculate the sum of AC coefficients.

$$S_j = \sum_{i=2}^{63} | AC_{j,i} |, \tag{2}$$

where j is the 8x8 block number in the downsampled macroblock. i is the coefficient index in the zig-zag scan order. If S_j is less than a threshold T_f, the block is identified as flat block, otherwise the block is classified into non-flat block. In the paper, the threshold value of T_f is 100. For non-flat blocks, we need further differentiate them as motion or non-motion blocks. As for non-motion blocks, the spatial information is more important for super-resolution reconstruction, while for motion blocks the temporal motion information should be considered too. To differentiate the motion and non-motion blocks, the motion difference between the reconstructed block and the prediction block is calculated as:

$$SAD_j = \sum_{k,l=0}^{7} | \mathrm{Y}_{j,k,l} - \mathrm{X}_{j,k,l} |, \tag{3}$$

where $Y_{j,k,l}$ is the reconstructed block signal, and $X_{j,k,l}$ is the predicted 8x8 block sample derived with the transmitted motion vector. After the calculation of (3), the block is classified into motion block which is well predicted, if SAD_j is in a region $[T_{ml,} \; T_{mu}]$, otherwise it is viewed as non-motion block, which may be a stationary block or non well-predicted block. In the paper, the threshold value of T_{ml} and T_{mu} is dependent on quantization step. The flowchart of the block classification algorithm is also shown in Fig. 2.

Step 3. Super-resolution reconstruction and distortion calculation of the reduced block.

As said above, for different kinds of blocks different super-resolution is used. For flat blocks, bilinear interpolation is used for upsampling. For non-motion blocks, the edge preserved algorithm proposed in [1] is used, because of the low complexity for real time video coding. For motion blocks, adapted super-resolution reconstruction algorithm is used for upsampling. As shown in Fig. 3, the samples of downsampled block are projected to the corresponding high resolution grid first. For the reduced samples, it is constructed as:

$$SR(x, y) = R(x, y) + g(x', y') - R(x'', y'') = R(x,y) + e(x',y') \tag{4}$$

(x, y) is the reduced sample position in the high resolution image. $SR(x, y)$ is the reconstructed reduced sample. $R(x, y)$ is the reference sample for the reduced sample derived with the motion vector of downsampled block. $g(x', y')$ is the downsampled block sample paired with (x, y) sample in the high resolution image and $R(x'', y'')$ is the corresponding reference sample.

Step 4. Rate distortion optimization mode decision for downsampling coding.

After the super-resolution reconstruction of the reduced block, the whole rate-distortion cost of the downsampling coded macroblock can be calculated,

$$JD_i = D_i + \lambda \times R_i + D_u ,$$

where D_u is the distortion between the reduced block and the reconstructed one. The best mode minimizing the RD cost is selected.

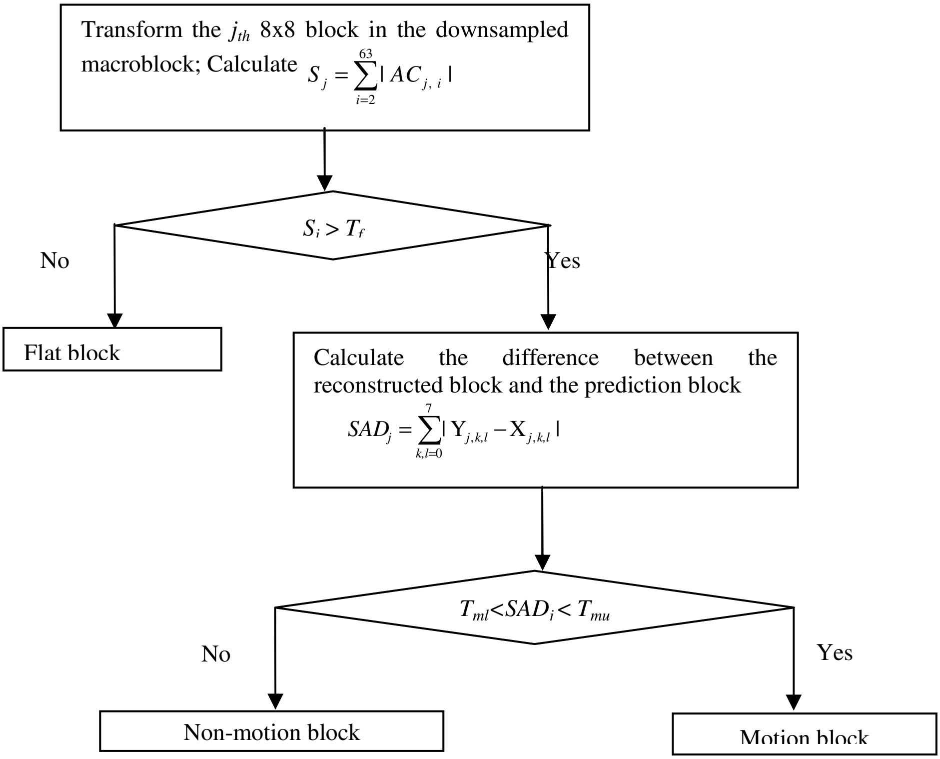


Fig. 2. Proposed block classification scheme

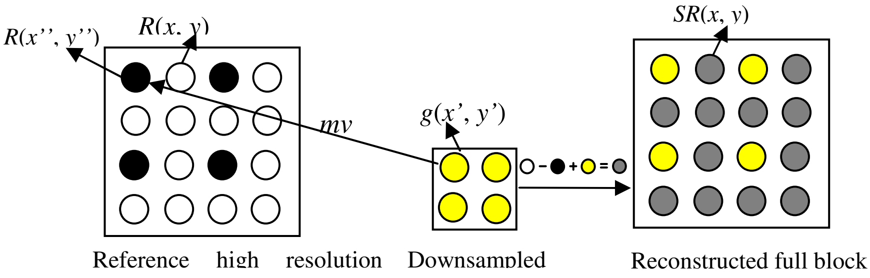
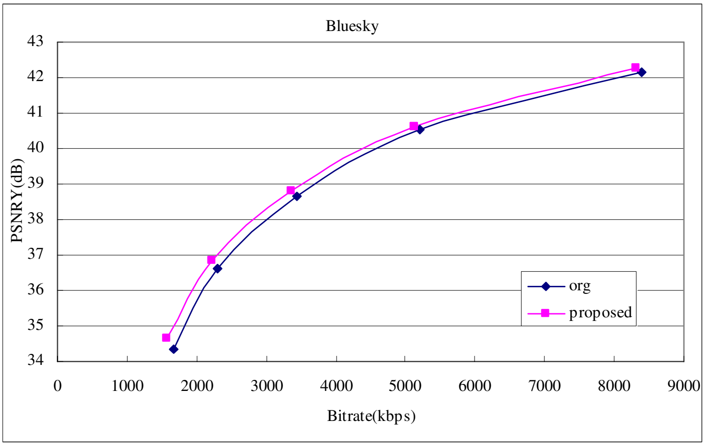


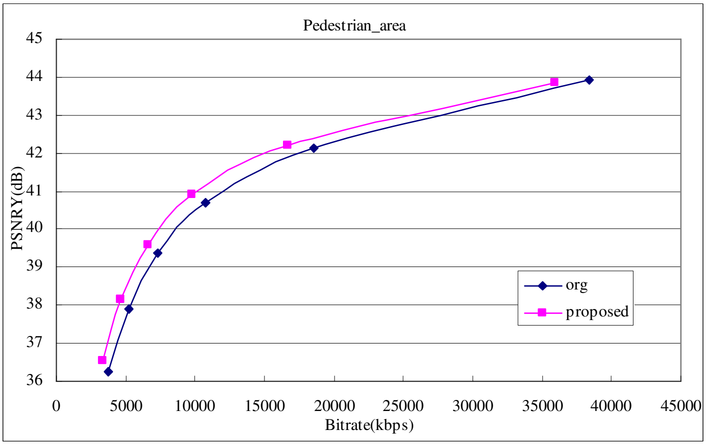
Fig. 3. Adapted super resolution scheme for motion blocks

3 Experimental Results

The proposed block adaptive super resolution coding method is implemented into AVS reference software (http://www.avs.org.cn/fruits/software/rm52j_r1.zip) for P frame coding. It was tested on 720p and 1080p test sequences. 100 frames are coded for each test sequence. The first frame is coded as I frame and remained frames are coded as P frame and 2 reference frames are used. Intra mode is disabled in P frame as only inter mode super-resolution coding is supported so far.



(a)



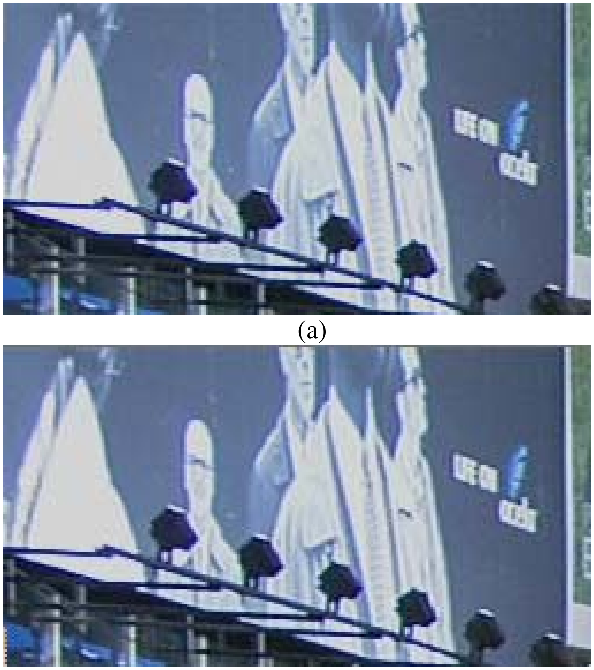
(b)

Fig. 4. R-D curves on 1080p test sequences (a) Bluesky (b)Pedestrian_area

Fig. 4 shows the RD curves for test sequences. From the curves, it can be seen that the proposed scheme can achieve coding gain at low bitrate as shown in [17], moreover it also shows obvious coding gain at high bitrate, where the PSNR is even higher than 40dB. Table 3 shows test results on more test sequences.

Table 3. Test results on more test sequences

Sequence	Format	Bitrate saving (%)	BDPSNR(dB)
Bluesky	1080p	4.78	0.20
Pedestrian_area	1080p	13.62	0.37
Night	720p	5.49	0.19
Harbour	720p	1.29	0.06
Crew	720p	21.91	0.79
Average		9.42	0.32



(a)

(b)

0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	2	0	0	3	0	0	0	0	0	2	1	3	0
0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(c)

Fig. 5. Visual quality comparison between the original sub-image and super-resolution reconstructed sub-image for Night sequence. QP=28. (a) original 256x128 sub image; (b) super-resolution reconstructed sub image; (c) down-sampling mode map for the sub-image.

Fig. 5 shows the subjective results between the original sub-image and the reconstructed sub-image with super-resolution. The sub-image size is 256x128, and Fig. 5 (c) shows the downsampling mode map for the sub-image. The super-resolution reconstructed macroblocks also show good visual quality as the original ones.

For the complexity of the proposed scheme, as the super-resolution method for flat, motion, and non-motion blocks is simple interpolation operation, so the complexity of the proposed scheme is not high, it is suitable for real time video coding.

4 Conclusions

This paper proposed a block adaptive low complexity super resolution video coding scheme. In the proposed scheme, block adaptive downsampling and upsampling is selected based on the rate-distortion cost decision, where the distortion caused by super-resolution reconstruction is also included. The block upsampling mode is adaptive with the block spatial and motion features, which is decided according to the reconstructed downsampled block and need not be transmitted to the decoder side as overhead information. Experimental results show that the proposed scheme can achieve significant performance improvement at low bitrate coding, up to 0.79dB. It may be a very promising scheme for future high resolution coding. In the future, we will extend it to B frame coding, and more efficient super-resolution reconstruction method for motion blocks will be studied.

Acknowledgements

This work was supported in part by National Science Foundation of China (60803068), and National Basic Research Program of China (973 Program, 2009CB320904).

References

1. Battiato, S., Gallom, G., Stanco, F.: A locally-adaptive zooming algorithm for digital images. Image and Vision Computing 20(11), 805–812 (2002)
2. Jensen, K., Anastassiou, D.: Subpixel edge localization and the interpolation of still images. IEEE Trans. on Image Processing 4, 285–295 (1995)
3. Carrato, S., Ramponi, G., Marsi, S.: A simple edge-sensitive image interpolation filter. In: Proc. IEEE Int. Conf. Image Processing, vol. 3, pp. 711–714 (1996)
4. Morse, B.S., Schwartzwald, D.: Isophote-based interpolation. In: Proc. IEEE Int. Conf. Image Processing, vol. 3, pp. 227–231 (1998)
5. Wang, Q., Ward, R.: A new edge-directed image expansion scheme. In: Proc. IEEE Int. Conf. Image Processing, vol. 3, pp. 899–902 (2001)
6. Zhang, L., Wu, X.: An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans. on Image Processing 15, 2226–2238 (2006)
7. Li, X., Orchard, M.T.: New edge-directed interpolation. IEEE Trans. Image Process. 10, 1521 (2001)
8. Asuni, N., Giachetti, A.: Accuracy improvements and artifacts removal in edge based image interpolation. In: Proc. 3rd Int. Conf. Computer Vision Theory and Applications, VISAPP 2008 (2008)

9. van Ouwerkerk, J.D.: Image super-resolution survey. Image and Vision Computing 24(10), 1039–1052 (2006)
10. Barreto, D., Alvarez, L.D., Abad, J.: Motion Estimation Techniques in Super-Resolution Image Reconstruction. A Performance Evaluation. In: Tsvetkov, M., Golev, V., Murtagh, F., Molina, R. (eds.) VIRTUAL OBSERVATORY: Plate Content Digitization, Archive Mining & Image Sequence Processing. Heron Press, Sofia (2005)
11. Ur, H., Gross, D.: Improved Resolution from Sub-pixel Shifted Pictures. CVGIP: Graphical Models and Image Processing 54, 181–186 (1992)
12. Alam, M.S., Bognar, J.G., Hardie, R.C., Yasuda, B.J.: Infrared Image Registration and High-Resolution Reconstruction using Multiple Translationally Shifted Aliased Video Frames. In: IEEE Trans. Instrum. Meas., vol. 49, pp. 915–923 (2000)
13. Nguyen, N., Milanfar, P.: An Efficient Wavelet-Based Algorithm for Image Superresolution. In: Proc. Int. Conf. Image Processing, vol. 2, pp. 351–354 (2000)
14. Callico, G.M., Nunez, A., Llopis, R.P., Sethuraman, R., de Beeck, M.O.: A Low-Cost Implementation of Super-Resolution Based on a Video Encoder. In: IEEE 28th Annual Conf. of the Industrial Elec. Society, November 2002, vol. 2, pp. 1439–1444 (2002)
15. Barreto, D., Alvarez, L.D., Molina, R., Katsaggelos, A.K., Callicó, G.M.: Region-Based Super-Resolution for Compression. Multidimensional Systems and Signal Processing 18(2-3), 59–81 (2007)
16. Lin, W., Dong, L.: Adaptive Downsampling to Improve Image Compression at Low Bit Rates. IEEE Transactions on Image Processing 15(9), 2513–2521 (2006)
17. Nguyen, V., Tan, Y., Lin, W.: Adaptive Downsampling/Upsampling for Better Video, Compression at Low Bit Rate. In: Proc. IEEE Int. Symposium on Circuits and Systems, May 2008, pp. 1624–1627 (2008)
18. Lin, L.: Video Bi-Rate Control with Spline Approximated Rate-Distortion Characteristics. Dissertation (May 1997)