PSO-based Fusion Method for Video Super-Resolution

Ming-Hui Cheng, Nai-Wei Lin, and Kao-Shing Hwang

***Abstract*—In this study, video super-resolution using particle swarm optimization (PSO) is proposed to super-resolve low-resolution (LR) frames. The proposed super-resolution method consists of three main modules, i.e., supersampling, spatio-temporal classification, and frame fusion using PSO. In the proposed method, the LR frames are super-resolved to high-resolution frames through the fusion of four full-resolution frames. One of four full-resolution frames is obtained using direct spatial interpolation, and the other three are obtained using motion compensation with given reference frames. The essence of the proposed method is the spatio-temporal classification mechanism that exploits the temporal variation between frames and the spatial energy inside the frame. Using the classification results, PSO is used to determine the optimal weights for frame fusion. Simulation results show that the proposed fusion method successfully improves the perceptual quality and the average peak signal-to-noise ratio (PSNR) in super-resolved frames.**

I. Introduction

IMAGE or video super-resolution provides very useful features in many image and video applications, such as video surveillance, medical imaging, and satellite imaging [1]. Traditional image super-resolution methods construct a high-resolution (HR) image from a single image or a set of low-resolution (LR) images additionally captured [2]. Such applications are often referred to as scaling, interpolation, or single-image super-resolution [3]. The construction of an HR frame using several successive frames in a video sequence is called video super-resolution [4].

Brandi et al. [5-6] proposed a method to magnify the blocks in an LR frame using adjacent HR key frames as reference. The super-resolved block is obtained by integrating the high-frequency parts of the HR blocks and low-frequency parts of the original LR blocks. Simonyan et al. [7] proposed a fusion-based approach to super-resolve the LR frames using the edge properties and temporal correlations. In the fusing stage, the authors used the AdaBoost classifier to generate the weights. Although their approach provides a good concept of frame fusion from two available HR frames, it does have drawbacks.

To overcome the three main drawbacks of the previous approach, first, we adopt the spatial interpolation method in [8], which is noise-free and efficient, to obtain finer edge properties to improve the classification and fusion results. We set up a frame-group structure to avoid error propagation and use both forward and backward motion estimations to achieve more accurate compensations. Based on the above-mentioned better spatio-temporal information, a classifier is designed to output a class label. We then issue a training mechanism using particle swarm optimization (PSO) to produce optimal weightings for the fusing operation.

The present paper is organized as follows. In Section 2, the architecture of the proposed PSO-based fusion method for video super-resolution is described. The details of the proposed spatio-temporal classification and frame fusion method using PSO are described in Sections 3 and 4, respectively. Simulation results are presented in Section 5, followed by the concluding remarks.

II. Architecture of Proposed PSO-based Fusion Method

In the present study, the frames in a video sequence are grouped together. As shown in Fig. 1, each group consists of HR and LR frames [5-7]. For each pair of HR frames H^u and H^{u+1}, there are n LR frames l_b^u, $b = 1,2,...,n$ in between. In practice, HR frames have full-resolution, and LR frames are decimated with a factor of two. In this study, the LR frames are super-resolved with the assistance of adjacent HR frames. The proposed super-resolution method has three main steps, i.e., super-sampling, spatio-temporal classification, and frame fusion using PSO. The block diagram is shown in Fig. 2.

The methods in the Supersampling Block are well known; In the method, each LR frame l_b^u between two successive HR frames H^u and H^{u+1} is processed independently. The detailed flow of the Supersampling Block is shown in Fig. 3. The HR frames H^u and H^{u+1} are first sub-sampled with a factor of two to generate two corresponding LR frames h^u and h^{u+1}. We then adopt the spatial interpolation method in [8] to supersample them back to the original solution and denote them as $\hat{H}^u$ and $\hat{H}^{u+1}$. We also supersample the LR frame l_b^u to $\hat{L}_b^u$, which has the same solution as that of $\hat{H}^u$ and $\hat{H}^{u+1}$.

The Motion compensation Block shown in Fig. 3 uses the motion compensation method [10], in which $\hat{L}_b^u$ is treated as the anchor frame. In the motion estimation part, $\hat{H}^u$ and

M. H. Cheng is with the Computer Science and Information Engineering Department, University of National Chung Cheng, Chiayi 621, Taiwan. (phone: +886-920-840877; e-mail: cmhvince@gmail.com)

N. W. Lin is with the Computer Science and Information Engineering Department, University of National Chung Cheng, Chiayi 621, Taiwan. (e-mail: naiwei@cs.ccu.edu.tw)

K. S. Hwang is with the Electrical Engineering Department, University of National Chung Cheng, Chiayi 621, Taiwan. (e-mail: hwang@ccu.edu.tw)

$\hat{H}^{u+1}$ are the reference frames, whereas in the compensation part, the original frames H^u and H^{u+1} are used. There are three compensated frames, namely, forward ${}_{\text{Fw}}L'_b$, backward ${}_{\text{Bw}}L'_b$, and bidirectional compensated ${}_{\text{Bi}}L'_b$ frames. For example, in the forward compensation, frame $\hat{L}^u_b$ is compensated to obtain frame ${}_{\text{Fw}}L'_b$ with $\hat{H}^u$ as the motion estimation reference and H^u as the compensation reference. Here, all the frames have a full-resolution size.

As shown in Fig. 2, the first set of output of the Supersampling block, the supersampled frames $\hat{H}^u$, $\hat{H}^{u+1}$, and $\hat{L}^u_b$, and the LR frame l^u_b are fed into the Spatio-temporal Classification Block to generate a class label C_q. The other set of output, the compensated frames ${}_{\text{Fw}}L'_b$, ${}_{\text{Bw}}L'_b$, and ${}_{\text{Bi}}L'_b$, and the supersampled frame $\hat{L}^u_b$ are fed into the Frame Fusion using PSO Block. Together with the class label, the super-resolved frame $\tilde{L}^u_b$ is generated

III. Spatio-temporal Classification

The spatio-temporal classifier characterizes each pixel in the supersampled frame $\hat{L}^u_b$ into one of the eight classes according to temporal variation and spatial energy. Here, the temporal variation detection is used to determine the object variation in time. Spatial energy detection is employed to analyze the explicit shape area.

A. Temporal Variation Detection

The main propose of the temporal-variation detection is to calculate the variation in time, which will be used as an indicator for subsequent classification. The forward variation E^{Fw}_T is computed from $\hat{L}^u_b$ with respect to $\hat{H}^u$. As shown in Fig. 4, for each pixel (x, y) in $\hat{L}^u_b$, we collect one $3\times3\times2$ volume and denote it as f_m and $m=1\ldots18$. With the pre-specified $3\times3\times2$ temporal filter mask, α_m and $m=1\ldots18$, as shown in Fig. 5, we define the variation intensity at (x, y) as

$$E^{\text{Fw}}_T(x,y)=\left|\sum_{m=1}^{18}\alpha_m f_m\right|. \tag{1}$$

The backward variation $E^{\text{Bw}}_T(x,y)$ is computed from $\hat{L}^u_b$ with respect to $\hat{H}^{u+1}$ in a similar manner. After the variation intensities of all pixels in $\hat{L}^u_b$ are calculated, $E^{\text{Fw}}_T(x,y)$ and $E^{\text{Bw}}_T(x,y)$ are regarded as temporal variation frames, both of which have a full-resolution size.

In the temporal variation frames E^{Fw}_T and E^{Bw}_T, the variation intensities will be set to zero if some conditions are met. Let B be a non-overlapping 8×8 block of the supersampled frame $\hat{L}^u_b$ and B^E be the corresponding block in E_T. In the motion compensation Block, using the motion estimation of $\hat{L}^u_b$ with $\hat{H}^u$ as the reference, we obtain a motion vector and the corresponding SAD denoted by ζ. The strategy of eliminating unnecessary variation intensities is defined by

$$B^E(x,y)=\begin{cases}0, & \text{if } \begin{array}{l}(\sigma^2<700)\wedge(\zeta<350),\\ (700\le\sigma^2\le1400)\wedge(\zeta<700),\\ (\sigma^2>1400)\wedge(\zeta<1000),\end{array}\\ B^E(x,y), & \text{otherwise.}\end{cases} \tag{2}$$

Here, σ^2 indicates the variance of block B and the symbol $\wedge$ denotes the logical AND operation. The corresponding thresholds are empirically determined.

B. Spatial Energy Detection

As the explicit shpaes is an important perceptual feature from the spatial perspective, edge detection results can be regarded as spatial energy. The Laplacian of Gaussian (LoG) mask [11] is applied to the LR frame l^u_b to establish an LR edge-detected frame. The spatial interpolation method [8] is then utilized to interpolate the LR edge-detected frame to construct a full-resolution edge-detected frame, which can be treated as a spatial energy frame E_S. To obtain better results for the spatio-temporal classification utilizing E_S, a simple thresholding is used to ignore edge points with weak edge signals.

C. Classification

When the temporal variation frames E^{Fw}_T and E^{Bw}_T, and the spatial energy frame E_S are obtained, the type of pixels in supersampled frame $\hat{L}^u_b$ can then be determined. We characterize the type according to the intensities (i.e., large or small) by pre-specifying two thresholds θ_T and θ_S for temporal variation and spatial energy, respectively. Each pixel is classified into one of the eight classes C_q, where $q=1,2,...,8$. C_1 and C_5 are the non-motion classes. C_2 and C_3 are specific motion without spatial energy classes. C_6 and C_7 are specific motion with spatial energy classes. C_4 and C_8 are arbitrarily motion classes. The eight classes are determined by the following conditions:

$$C_1:\ (E_S<\theta_S)\wedge(E^{\text{Fw}}_T<\theta_T)\wedge(E^{\text{Bw}}_T<\theta_T). \tag{3}$$
$$C_2:\ (E_S<\theta_S)\wedge(E^{\text{Fw}}_T<\theta_T)\wedge(E^{\text{Bw}}_T>\theta_T). \tag{4}$$
$$C_3:\ (E_S<\theta_S)\wedge(E^{\text{Fw}}_T>\theta_T)\wedge(E^{\text{Bw}}_T<\theta_T). \tag{5}$$
$$C_4:\ (E_S<\theta_S)\wedge(E^{\text{Fw}}_T>\theta_T)\wedge(E^{\text{Bw}}_T>\theta_T). \tag{6}$$
$$C_5:\ (E_S>\theta_S)\wedge(E^{\text{Fw}}_T<\theta_T)\wedge(E^{\text{Bw}}_T<\theta_T). \tag{7}$$
$$C_6:\ (E_S>\theta_S)\wedge(E^{\text{Fw}}_T<\theta_T)\wedge(E^{\text{Bw}}_T>\theta_T). \tag{8}$$
$$C_7:\ (E_S>\theta_S)\wedge(E^{\text{Fw}}_T>\theta_T)\wedge(E^{\text{Bw}}_T<\theta_T). \tag{9}$$

$$C_8: (E_S > \theta_S) \wedge (E_T^{\mathrm{Fw}} > \theta_T) \wedge (E_T^{\mathrm{Bw}} > \theta_T). \quad (10)$$

In Eqs. (3)-(10), the symbol $\wedge$ denotes the logical AND operation.

IV. Frame Fusion using PSO

The supersampling frame $\hat{L}_b^u$ and the three compensated frames ${}_{\mathrm{Fw}}L_b'$, ${}_{\mathrm{Bw}}L_b'$, and ${}_{\mathrm{Bi}}L_b'$ are constructed from the Supersampling block. We then perform a fusing operation on these four frames according to the classification results.

A. Frame Fusion using Fusion Mask

There are two types of fusion operations. One is the $3\times3\times3$ fusion type, which is between the frame $\hat{L}_b^u$ and the group ${}_{\mathrm{Fw}}L_b'$ and ${}_{\mathrm{Bw}}L_b'$, as shown in Fig. 6(a). The other fusion operation is the $3\times3\times2$ fusion type, which is between the frame $\hat{L}_b^u$ and ${}_{\mathrm{Bi}}L_b'$, as shown in Fig. 6(b). There are eight class labels C_q, $q=1,2,...,8$. The four non-motion and arbitrarily motion classes C_1, C_5, C_4, and C_8 are used for the $3\times3\times2$ fusion type. The motion classes C_2, C_3, C_6, and C_7 are used for the $3\times3\times3$ fusion type.

Based on the classification results, each pixel in $\hat{L}_b^u$ collects its own fusion volume A. The fusion operation is the convolution following [12]. Suppose that $\{f_A(x,y,z)\}$ is the fused voxel value from a given fusion volume A, where (x,y) and z are the spatial coordinates and collected frame index, respectively. The fusion operation corresponding to the fusion mask M_q is defined as follows:

$$f_A(x,y,z) = \iiint_{(x',y',z')\in M_q(x,y,z)} w^M(x',y',z') f(x',y',z') dx'dy'dz', \quad (11)$$

where $M_q(x,y,z)$ and $w^M(x',y',z')$ indicate the size and weight of the q-th class's fusion mask M, respectively. The discrete function of the fusion operation with fusion volume A is given by

$$f(A) = \sum_{d=1}^{k} w_d^{M_q} f_d^A, \quad (12)$$

where w_d is the d-th weight in the corresponding q-th class's fusion mask M. The term k is equal to 27 for the $3\times3\times3$ fusion volume and 18 for the $3\times3\times2$ fusion volume.

B. Fusion Mask Determination using PSO

The fusion mask used to perform the fusion operation is constructed through a training process to obtain optimal solutions. There are eight fusion masks to be determined: four $3\times3\times2$ fusion masks for class labels C_1, C_4, C_5, and C_8 and four $3\times3\times3$ fusion masks for C_2, C_3, C_6, and C_7. In the optimization process, we choose two CIF video sequences, Akiyo and Mother-Daughter, as the training sets. We assume that the frame group shown in Fig. 1 is of size 3. That is, every HR frame H of size 352×288 is followed by two LR frames l of size 176×144, which are sub-sampled from the original video sequence.

Each pixel in $\hat{L}_b^u$ collects the fusion volume according to its class label. For example, in class 1, a $3\times3\times2$ fusion volume A is collected, as shown in Fig. 6(b). Assume that there are N fusion volumes in the training set belonging to class 1. Each $3\times3\times2$ fusion volume is regarded as an 18-dimensional (18-D) vector. In PSO, each particle represents a vector of 18 fusion weights. Here, we define the cost F_i for particle i in PSO of population size Z by

$$F_i = \sum_{j=1}^{N} (|P_j - (\sum_{d=1}^{18} f_j^d \cdot s_i^d)|), \quad (13)$$

where s_i^d is the position of particle i at dimension d, $i=1,2,3,\ldots,Z$. The term P_j is the corresponding pixel to be fused with the j-th fusion volume from the original frames.

PSO is a heuristic optimization algorithm that includes random terms; thus, different sub-optimal solutions may be obtained from separate experiments. In the present study, ten experiments are conducted. The ten sets of converged solutions (GBEST) $p_{g,j}^d$, $j=1..10$ are shown in Table 1. The final weights w_d at d-th dimension of the fusion mask is averaged by

$$w_d = \frac{\left\lfloor \frac{1}{5}\left(\sum_{j=1}^{10} p_{g,j}^d\right)\right\rfloor}{\sum_{m=1}^{k}\left(\left\lfloor \frac{1}{5}\left(\sum_{j=1}^{10} p_{g,j}^m\right)\right\rfloor\right)}, \quad (14)$$

where operator $\lfloor a \rfloor$ denotes the nearest integer smaller or equal to a. Similarly, the 18-D fusion weights of classes 4, 5, and 8 and the 27-D fusion mask weights of classes 2, 3, 6, and 7 can be separately determined using PSO. The eight fusion masks corresponding to the class labels are given in Eqs. (15)-(22).

$$M_1 = \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.25&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.75&0\\\hline 0&0&0\\\hline\end{array} \quad (15)$$

$$M_2 = \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.6&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.3&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.1&0\\\hline 0&0&0\\\hline\end{array} \quad (16)$$

$$M_3 = \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.1&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.3&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.6&0\\\hline 0&0&0\\\hline\end{array} \quad (17)$$

$$M_4 = \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.8&0\\\hline 0&0&0\\\hline\end{array}\ \begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0.2&0\\\hline 0&0&0\\\hline\end{array} \quad (18)$$

$$M_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.14 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.86 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (19)$$

$$M_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.56 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.11 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (20)$$

$$M_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.13 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.62 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (21)$$

$$M_8 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.83 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.17 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (22)$$

V. Simulation Results

In this study, the experiments are conducted using Borland C++ Builder on an Intel Core 2 Duo CPU 2.0 GHz-Microsoft Windows platform. Several video sequences are used to evaluate the performance of the proposed fusion method. For comparison purposes, the proposed method and three other methods, i.e., nearest neighbor interpolation (NN), bicubic interpolation (Bicubic), and super-resolution using key frame (SR-key), are implemented [6].

In the proposed method, the decimation factor of the LR frame is set to 2. The frame group consists of an HR frame of size 352×288 and a number of LR frames of size 176×144. To evaluate the objective performance, the frame group size is set to 3, and the corresponding HR frame ratio is 0.33. The HR and LR frames are encoded using H.264 intraframe coding with QP=28. In the motion compensations, the search range is set to 15, and the block size is set to 8×8. In the Spatio-temporal classification Block, the two threshold values θ_T and θ_S are empirically set to 60 and 1, respectively. In the off-line PSO method, the population size Z, inertia weight ω, and learning constants c_1 and c_2 in PSO are empirically set to 100, 0.1, 0.2, and 0.3, respectively.

The quality comparison of the four methods with an HR frame ratio of 0.33 for the Akiyo clips is shown in Fig. 7. The horizontal axis is the frame number, and the vertical axis shows the PSNR quality. As observed, the performance of the proposed method is “much” better than that of the three comparison methods.

For visual comparison, the super-resolved results of the 74th frame of the Akiyo video sequence are shown in Fig. 8, in which the HR frame ratio is 0.33. Fig. 8(a) shows the ground truth, and Fig. 8(b) shows the LR frame. The results of NN, Bicubic, SR-key, and the proposed method are shown in Fig. 8(c)-(f). For overall comparison, the proposed method outperforms the other methods. To emphasize the visual effect, some details, such as the shoulder and face parts, are shown in Figs. 9 and 10, respectively. As observed, the visual effect of the proposed method is “much” better than that of the three comparison methods.

VI. Concluding Remarks

In this study, a PSO-based fusion method for frame resolution enhancement in video super-resolution is proposed. To improve the single frame enlargement performance, the spatial interpolation method in [8] is used to obtain finer edge properties to improve the classification results. The frame group structure is used to avoid error propagation. We use both forward and backward motion estimations to achieve more accurate compensation. The spatio-temporal classification approach is proposed to exploit the temporal variation and spatial energy. Based on the above-mentioned better spatio-temporal information, a classifier is designed to output a class label. We then issue a training mechanism using PSO to produce optimal weightings for the fusing operation. Based on the simulation results, the proposed fusion method indeed improves the average PSNR and the perceptual quality in the super-resolved frame.

References

[1] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing*, vol. 90, pp. 848-859, Mar. 2010.

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21-36, May 2003.

[3] J. D. V. Ouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, vol. 24, pp. 1039-1052, 2006.

[4] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Trans. on Image Processing*, vol. 18, pp. 36-51, Jan. 2009.

[5] F. Brandi, R. D. Queiroz, and D. Mukherjee, "Super-resolution of video using key frames," in *IEEE International Symposium on Circuits and Systems*, 2008, pp. 1608-1611.

[6] F. Brandi, R. D. Queiroz, and D. Mukherjee, "Super-resolution of video using key frames and motion estimation," in *IEEE International Conference on Image Processing*, 2008, pp. 321-324.

[7] K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, "Fast video super-resolution via classification," in *IEEE International Conference on Image Processing*, 2008, pp. 349-352.

[8] H. Y. Chen and J. J. Leou, "Saliency-directed image interpolation using particle swarm optimization," *Signal Processing*, vol. 90, pp. 1676-1692, May 2010.

[9] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, 1995, pp. 1942-1948.

[10] G. J. Sullivan and T. Wieg, "Video compression-from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, pp. 18-31, Jan. 2005.

[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd Edition. New Jersey: Prentice Hall, 2002.

[12] X. Li, "Video processing via implicit and mixture motion models," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, pp. 953-963, Aug. 2007.

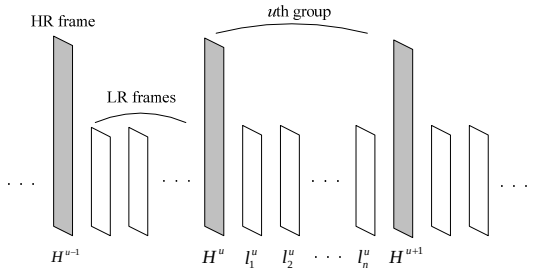


Fig. 1. Structure of video sequence in the proposed method.

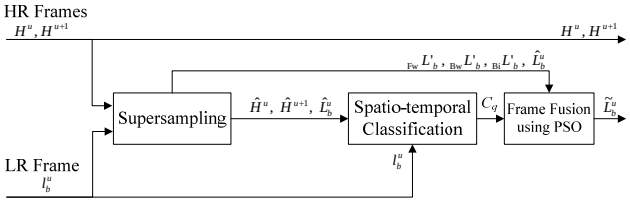


Fig. 2. Block diagram of the proposed PSO-based fusion method.

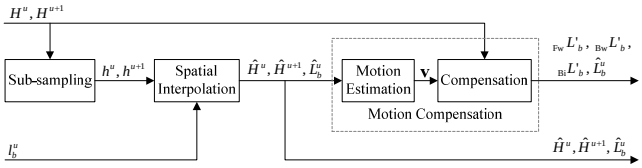


Fig. 3. Block diagram of supersampling.

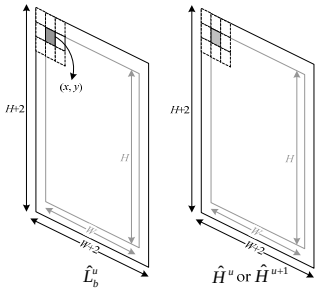


Fig. 4. 3×3×2 volume collected from a supersampled frame $\hat{L}_b^u$ and an adjacent supersampled frame .

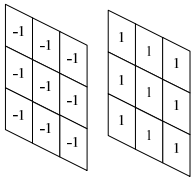


Fig. 5. Proposed 3×3×2 temporal filter mask.

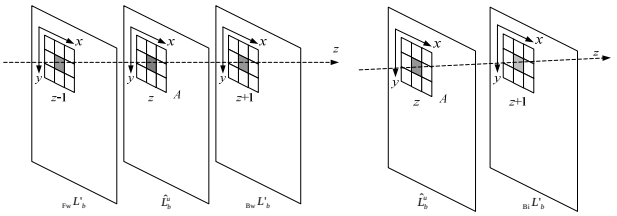


Fig. 6. Collected fusion volume A: (a) 3×3×3 fusion volume and (b) 3×3×2 fusion volume.

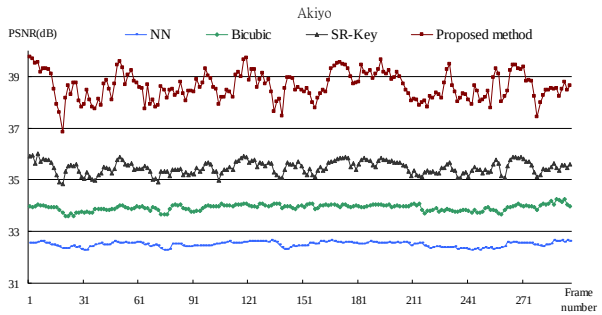
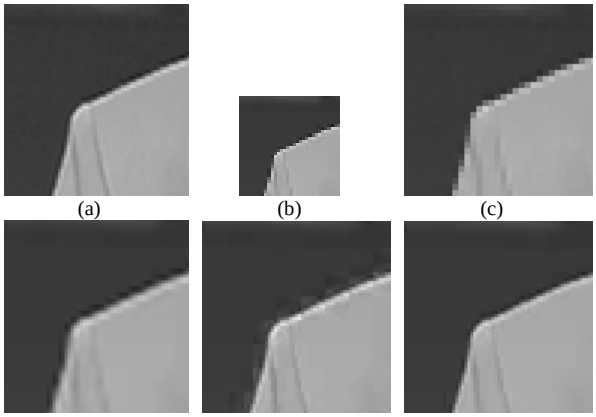


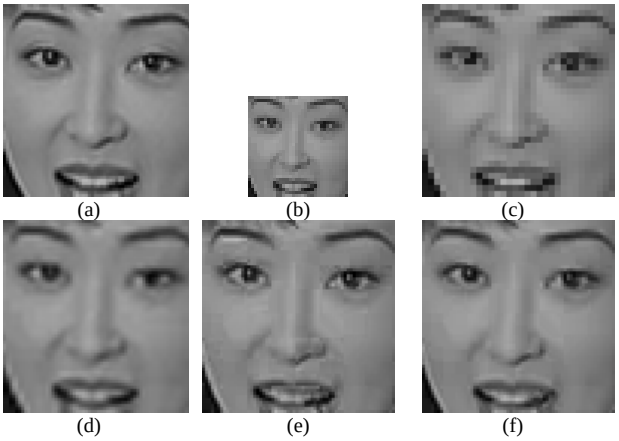
Fig. 7. Comparison of the PSNR of the four methods on the Akiyo video sequence with HR frame ratio of 0.33.

Fig. 8. The 74th LR frame of the Akiyo video sequence with HR frame ratio=0.33: (a) ground truth, (b) LR frame, (c) NN, (d) Bicubic, (e) SR-key, and (f) the proposed method.



(a) (b) (c) (d) (e) (f)

Fig. 9. Shoulder part of the 74th frame of the Akiyo video sequence: (a) ground truth, (b) LR frame, (c) NN, (d) Bicubic, (e) SR-key, and (f) the proposed method.



(a) (b) (c) (d) (e) (f)

Fig. 10. Face part of the 74th frame of the Akiyo video sequence: (a) ground truth, (b) LR frame, (c) NN, (d) Bicubic, (e) SR-key, and (f) the proposed method.

Table 1. PSO final results of the 18 weights of class 1 fusing type in 10 experiments.

experiment	p_g^1	p_g^2	p_g^3	p_g^4	p_g^5	p_g^6	p_g^7	p_g^8	p_g^9	p_g^{10}	p_g^{11}	p_g^{12}	p_g^{13}	p_g^{14}	p_g^{15}	p_g^{16}	p_g^{17}	p_g^{18}
1	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0
2	0	0	0	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	5	0	0	1	0
5	0	0	0	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0
7	0	1	0	0	0	0	0	0	0	0	0	0	0	6	0	0	1	0
8	0	0	0	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0
10	0	0	0	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0