# A Fast Kernel Regression Framework for Video Super-Resolution

**Wen-sen Yu[1,2], Ming-hui Wang[1], Hua-wen Chang[3], and Shu-qing Chen[1,4]**
[1] College of Computer Science, Sichuan University
Chengdu, 610064 – P.R.China
[e-mail:wensenyu@gmail.com, wangminghui@scu.edu.cn, changhuawen@gmail.com, talk2csq@hotmail.com ]
[2] College of Mathematics and Computer Science, WuYi University
Wuyishan, 354300 – P.R.China
[3] College of Computer and Communication Engineering, Zhengzhou University of Light Industry
Zhengzhou, 450002 – P.R.China
[4] Department of Electronics and Information Engineering, Putian University
Putian, 351100 – P.R.China
*Corresponding author: Wen-sen Yu

*Received April 16, 2013; revised August 16, 2013; revised December 11, 2013; accepted January 4, 2014; published January 29, 2014*

## Abstract

A series of kernel regression (KR) algorithms, such as the classic kernel regression (CKR), the 2- and 3-D steering kernel regression (SKR), have been proposed for image and video super-resolution. In existing KR frameworks, a single algorithm is usually adopted and applied for a whole image/video, regardless of region characteristics. However, their performances and computational efficiencies can differ in regions of different characteristics. To take full advantage of the KR algorithms and avoid their disadvantage, this paper proposes a kernel regression framework for video super-resolution. In this framework, each video frame is first analyzed and divided into three types of regions: flat, non-flat-stationary, and non-flat-moving regions. Then different KR algorithm is selected according to the region type. The CKR and 2-D SKR algorithms are applied to flat and non-flat-stationary regions, respectively. For non-flat-moving regions, this paper proposes a similarity-assisted steering kernel regression (SASKR) algorithm, which can give better performance and higher computational efficiency than the 3-D SKR algorithm. Experimental results demonstrate that the computational efficiency of the proposed framework is greatly improved without apparent degradation in performance.

***Keywords:*** video super-resolution, kernel regression framework, Similarity-assisted Steering Kernel Regression

This research was supported by National Natural Science Foundation of China (Grant No.61071162) , Scientific Research Fund of Fujian Provincial Department of Education (Grant No. JA10271).

**http://dx.doi.org/10.3837/tiis.2014.01.014**

## 1. Introduction

Super-resolution (SR) is a process of reconstructing a high-resolution (HR) image from multiple low-resolution (LR) inputs. Its basic idea is to enhance the resolution of reference image by making full use of the information contained in both reference and auxiliary images. A variety of algorithms have been presented to solve the SR problem. The frequency domain method was firstly introduced by Tsai and Huang [1], and extended by their successors [2][3]. However, the performance of the frequency domain methods is usually limited by the global translational motion and spatially invariant degradation. Thus, a variety of the spatial domain SR methods have been proposed. The iteration back-projection (IBP) algorithm [4] yields a HR image by iteratively back-projecting the error between the simulated LR images and the observed ones. The maximum a posteriori (MAP) method utilizes the spatial domain observation model and the prior knowledge of the target HR image to estimate the target HR image under a Bayesian theorem framework [5][6][7][8][9]. The projection on convex sets (POCS) method tends to incorporate the prior knowledge of the target HR image into the convex constraint sets and to restrict the SR solution to be a member of the convex sets [10][11]. An extensive review of the SR methods can be seen in [12] [13].

Although the SR technique has been extensively studied in the past three decades, the super-resolution on general video sequences still remains an open problem. In existing video super-resolution (VSR) algorithms, either the motion models are oversimplified, or the computational efficiency is unsatisfactory. Several VSR algorithms presented in [14][15][16] limit their motion model to the case of translational motion. As a result, these algorithms cannot achieve good performance on general video sequences with arbitrary motion model. The fundamental difficulty for the super-resolution on general video sequences is to provide accurate subpixel motion estimations. Recent progresses have focused on two types of VSR methods. One type is simultaneous VSR method. Keller et al.[17] presented a VSR algorithm, which simultaneously estimates a HR sequence and its motion field via the calculus of variations. Liu et al. [18] proposed an adaptive VSR algorithm, which simiultaneously estimates HR frame, motion field, blur kernel and noise level in a Bayesian framework. However, their performance is affected by the accuracy of optical flow estimation. Another type is non-motion-estimation-based VSR method. Danielyan et al. [19] created a VSR algorithm by extending the block-matching 3-D filter, in which the explicit motion estimation is avoided by classifying the image patches using block matching. Protter et al. [20] generalized the non-local means (NLM) algorithm (a denoising algorithm) to enhance the resolution of general video sequences without explicit motion estimation. Takeda et al. [21] extended the 2-D steering kernel regression (SKR) approach[22] to 3-D for video super-resolution. With similar ideas, K.Zhang et al.[23] extended the 2-D normalized convolution approach to 3-D case for video super-resolution. H.Zhang et al.[24][25] presented a nonlocal kernel regression (NL-KR) framework and applied it to SR reconstruction. The NL-KR framework exploits both the nonlocal self-similarity and local structure regularity for a more reliable and robust estimation. Their works provide new thinking and methods to achieve super-resolution on general video sequences. However, these approaches share a common defect: low computational efficiency.

Takeda et al. [21][22] proposed a series of kernel regression (KR) algorithms for image and video super-resolution, such as the classic kernel regression (CKR), the 2- and 3-D steering kernel regression. The CKR algorithm is computationally efficient but its

performance on the edge regions is poor. The 2- and 3-D SKR algorithms improve the performance on the edge regions by exploiting 2- and 3-D local radiometric structure information. However, the use of more information leads to higher computational cost. In order to improve computational efficiency, this paper proposes a fast kernel regression framework for video super-resolution, which takes full advantage of three KR algorithms and avoids their disadvantage. In this framework, each video frame is first analyzed and divided into three types of regions: flat, non-flat-stationary, and non-flat-moving regions. Then different KR algorithm is selected according to the region type. The CKR and 2-D SKR algorithms are applied to flat and non-flat-stationary regions, respectively. For non-flat-moving regions, this paper proposes a similarity-assisted steering kernel regression (SASKR) algorithm, which is an extension of the NL-KR algorithm. The SASKR algorithm exploits the supplementary information from local spatial and temporal orientations separately. It consists of two parts: the local SKR and non-local SKR terms. The local SKR term exploits the supplementary information contained in the local spatial orientations while the non-local SKR term makes use of the supplementary information contained in the local temporal orientation. The SASKR algorithm can provide better performance and higher computational efficiency than the 3-D SKR algorithm.

The remainder of this paper is organized as follows. Section 2 briefly reviews several KR algorithms and presents a similarity-assisted steering kernel regression algorithm. A fast kernel regression framework for video super-resolution is described in Section 3 and experimental results are illustrated in Section 4. Finally, conclusions are summarized in Section 5.

## 2. Similarity-assisted Steering Kernel Regression Algorithm

In this section, a brief technical review of several KR algorithms is firstly presented. Then an extension of non-local kernel regression algorithm, called similarity-assisted steering kernel regression, is proposed.

### 2.1 Classic Kernel Regression Algorithm for Image Super-resolution

The classic kernel regression (CKR) algorithm usually performs in a local manner, i.e., a pixel value of interest is estimated from the samples within a small neighborhood of that pixel. For two dimensional cases, the regression model is

$$y_i = z(\mathbf{x}_i) + n_i \qquad i = 1, \cdots, p \tag{1}$$

where $y_i$ is a noise sample at position $\mathbf{x}_i = [x_{i,1}, x_{i,2}]^T$ ( $x_{i,1}$ and $x_{i,2}$ are spatial coordinates), $z(\cdot)$ is a regression function, $n_i$ is a zero mean additive Gaussian noise, and $p$ is the total number of samples within the neighborhood. The generalization of kernel regression estimate $\hat{z}(\mathbf{X})$ is given by solving the following weighted least squares problem [21].

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \left[ (\mathbf{Y} - \mathbf{\Phi A})^T \mathbf{K} (\mathbf{Y} - \mathbf{\Phi A}) \right] \tag{2}$$

where

$$\mathbf{Y} = [y_1, y_2, \cdots, y_p]^T$$
$$\mathbf{A} = [\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1^T, \cdots, \boldsymbol{\alpha}_N^T]^T$$
$$\mathbf{K} = diag[k(\mathbf{x}_1 - \mathbf{x}), k(\mathbf{x}_2 - \mathbf{x}), \cdots, k(\mathbf{x}_p - \mathbf{x})]^T$$
$$\boldsymbol{\Phi} = \begin{bmatrix} 1, & (\mathbf{x}_1 - \mathbf{x})^T, & vect\{(\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})^T\}^T, & \cdots \\ 1, & (\mathbf{x}_2 - \mathbf{x})^T, & vect\{(\mathbf{x}_2 - \mathbf{x})(\mathbf{x}_2 - \mathbf{x})^T\}^T, & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1, & (\mathbf{x}_p - \mathbf{x})^T, & vect\{(\mathbf{x}_p - \mathbf{x})(\mathbf{x}_p - \mathbf{x})^T\}^T, & \cdots \end{bmatrix}$$

$\boldsymbol{\alpha}_n$ ( $n = 0 \cdots N$ ) is the regression coefficient, and $\boldsymbol{\alpha}_0$ is the desired pixel value estimation $\hat{z}(\mathbf{X})$ . $\boldsymbol{\Phi}$ is the regression base, $vect(\cdot)$ is an operator that extracts the lower-triangular part of a symmetric matrix and lexicographically orders it into a column vector, and $k(\mathbf{x}_i - \mathbf{x})$ is a kernel function which represents a weight for each sample $y_i$. The Gaussian kernel function is defined as

$$k(\mathbf{x}_i - \mathbf{x}) = \frac{1}{2\pi h^2} \exp\left\{ -\frac{(\mathbf{x}_i - \mathbf{x})^T (\mathbf{x}_i - \mathbf{x})}{2h^2} \right\} \tag{3}$$

where $h$ is the global smoothing parameter.

**Fig. 1** illustrates how the CKR algorithm is applied to image super-resolution. The input image is firstly upsampled into the HR grid. Then each missing pixel (denoted as white circle) value is estimated from the samples (denoted as black circles) within a small neighborhood of that pixel. The CKR algorithm is simple and computationally efficient. However, its performance on the edge regions is poor.

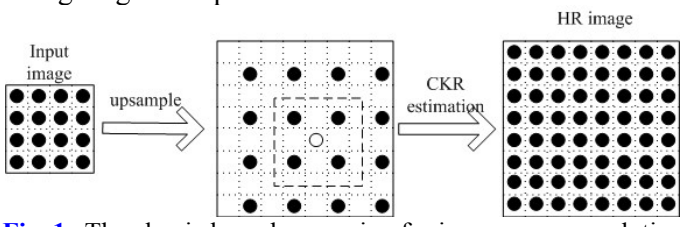


**Fig. 1.** The classic kernel regression for image super-resolution

## 2.2 2-D Steering Kernel Regression Algorithm

The 2-D steering kernel regression (SKR) algorithm is proposed to improve the performance on the edge regions [21]. The 2-D SKR algorithm defines its kernel function as

$$k(\mathbf{x}_i - \mathbf{x}) = \frac{\sqrt{\det(\mathbf{C}_i)}}{2\pi h^2} \exp\{ -\frac{(\mathbf{x}_i - \mathbf{x})^T \mathbf{C}_i (\mathbf{x}_i - \mathbf{x})}{2h^2} \} \tag{4}$$

where $\mathbf{C}_i$ is estimated as the covariance matrix of gradients of 2-D neighboring pixels.

$$\mathbf{C}_i = \mathbf{J}_i^T \mathbf{J}_i$$

with

$$\mathbf{J}_i = \begin{bmatrix} z_1'(\widetilde{\mathbf{x}}_1) & z_2'(\widetilde{\mathbf{x}}_1) \\ \vdots & \vdots \\ z_1'(\widetilde{\mathbf{x}}_m) & z_2'(\widetilde{\mathbf{x}}_m) \end{bmatrix}$$

where $z_1'(\cdot)$ and $z_2'(\cdot)$ are the first-order derivative along the directions of two coordinate axes, respectively. $\widetilde{\mathbf{x}}_j$ ( $j = 1 \cdots m$ ) is a sample position that fall into the analysis window centered on $\mathbf{x}_i$, and $m$ is the total number of samples within the analysis window. The 2-D SKR algorithm captures the local radiometric structures and feeds the structure information to the kernel function. Thus, this algorithm gives better performance on the edge regions but lower computational efficiency than the CKR algorithm.

In order to process video super-resolution, Takeda et al. [21] generalized the 2-D SKR to the 3-D SKR algorithm. The 3-D SKR algorithm exploits the supplementary information contained in local spatial and temporal orientations to achieve good video super-resolution results. However, the 3-D SKR algorithm has an inherent limit: the size of the spatiotemporal neighborhood must be small. Thus, many auxiliary frames, which contain supplementary information but are far away from the reference frame, can not be exploited.

### 2.3 Similarity-assisted Steering Kernel Regression Algorithm

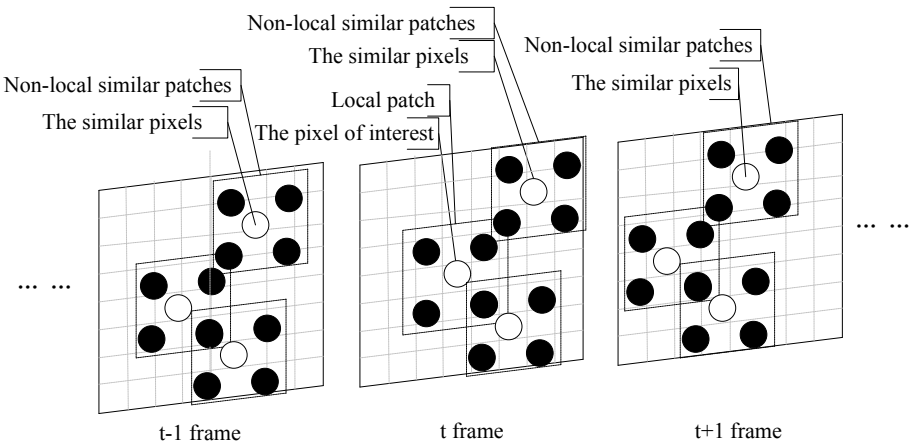


**Fig. 2.** The NL-KR algorithm for image and video restoration

In proposed framework, a similarity-assisted steering kernel regression (SASKR) algorithm is introduced as a replacement to the 3-D SKR algorithm. The SASKR algorithm is similar with the NL-KR algorithm. As shown in **Fig. 2**, the NL-KR algorithm makes use of the local patch and the non-local similar patches to estimate a pixel value of interest via kernel regression method. The NL-KR algorithm can give a more reliable and robust estimation. However, it is computationally heavy. the SASKR algorithm has two advantages over NL-KR algorithm. Firstly, the SASKR algorithm improves the computational efficiency. As shown in **Fig. 3**, the SASKR algorithm exploits only similar pixels along motion trajectory instead of all non-local similar pixels. Thus, the SASKR algorithm can achieve higher computational efficiency. Secondly, the SASKR algorithm adopts the SKR technique in the local and non-local terms, which can give better performance.

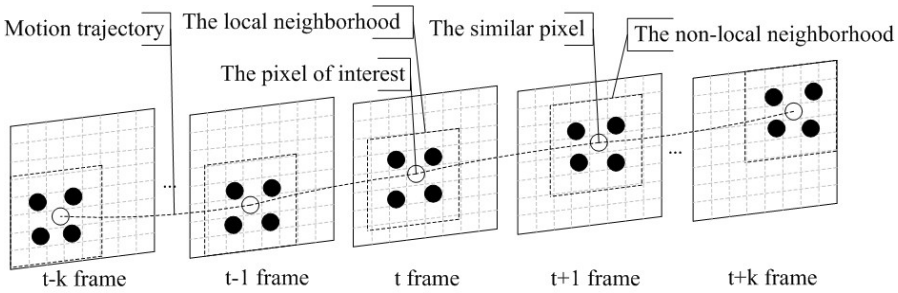


**Fig. 3.** The similarity-assisted steering kernel regression algorithm for video super-resolution

Next, a detailed description of the SASKR algorithm is given. As shown in **Fig. 3**, the pixel of interest is estimated from the sample values within the local and non-local neighborhoods via 2-D SKR technique. A local neighborhood is a region centered on the pixel of interest and a non-local neighborhood is a region centered on each similar pixel along motion trajectory. Mathematically, the SASKR algorithm can be formulated into an optimization problem:

$$\begin{aligned}\hat{\mathbf{A}} &= \arg\min_{\mathbf{A}} \overbrace{(\mathbf{Y}_0 - \mathbf{\Phi A})\mathbf{K}_0(\mathbf{Y}_0 - \mathbf{\Phi A})}^{local} + \overbrace{\sum_{t=-k,t\neq 0}^{k}(\mathbf{Y}_t - \mathbf{\Phi A})^T \mathbf{K}_t(\mathbf{Y}_t - \mathbf{\Phi A})w_t}^{non-local} \\ &= \arg\min_{\mathbf{A}} \sum_{t=-k}^{k}(\mathbf{Y}_t - \mathbf{\Phi A})^T \mathbf{K}_t(\mathbf{Y}_t - \mathbf{\Phi A})w_t \end{aligned} \tag{5}$$

where $\mathbf{Y}_0 = [y_{0,1}, y_{0,2}, \cdots, y_{0,p}]^T$ and $\mathbf{Y}_t = [y_{t,1}, y_{t,2}, \cdots, y_{t,p}]^T$ ( $t \neq 0$ ) are column vectors composed of the sample values within the local and non-local neighborhoods, respectively. $w_t$ is a weight for the similar pixel at frame $t$, which is calculated by measuring the similarity between $\mathbf{Y}_0$ and $\mathbf{Y}_t$.

$$w_t = \exp\left(-\frac{\|\mathbf{Y}_0 - \mathbf{Y}_t\|_2^2}{2\sigma^2}\right) \tag{6}$$

The Eq.(5) includes two parts: the local SKR and non-local SKR terms. The proposed algorithm exploits the supplementary information contained in the spatial and temporal orientations by the two terms. The first element of the regression coefficients $\hat{A}$ is taken as an estimate value of the pixel of interest.

$$\hat{z}(x) = \alpha_0 = e_1^T \hat{A} = e_1^T \left[\mathbf{\Phi}^T\left(\sum_{t=-k}^{k} w_t K_t\right)\mathbf{\Phi}\right]^{-1} \mathbf{\Phi}^T \sum_{t=-k}^{k} w_t K_t Y_t \tag{7}$$

By the above analysis, it can be seen that the SASKR algorithm has the capability to exploit all the auxiliary frames containing supplementary information. Thus, it breaks the inherent limit of the 3-D SKR algorithm. The experimental results in Section 4.2 verify the effectiveness of the SASKR algorithm. As we can see, the SASKR algorithm gives higher PSNR values and offers better visual effect than the 3-D SKR.

## 3. The Fast Kernel Regression Framework for Video Super-Resolution

It can be seen that the performance and computational efficiency of KR algorithms will be different when applied to regions of different characteristics within a single video frame. The CKR algorithm has high computational efficiency but its performance on the edge regions is poor; The SKR algorithm gives better performance on the edge regions but lower computational efficiency. The SASKR algorithm offers the best performance but the lowest computational efficiency. On the basis of the consideration of performance and computational efficiency, this paper proposes a fast kernel regression framework, in which the KR algorithms with different computational complexity are automatically selected for estimating the pixels in different regions.

The chart flow of the proposed framework is shown in **Fig. 4**. The estimation process of each pixel is divided into two stages.

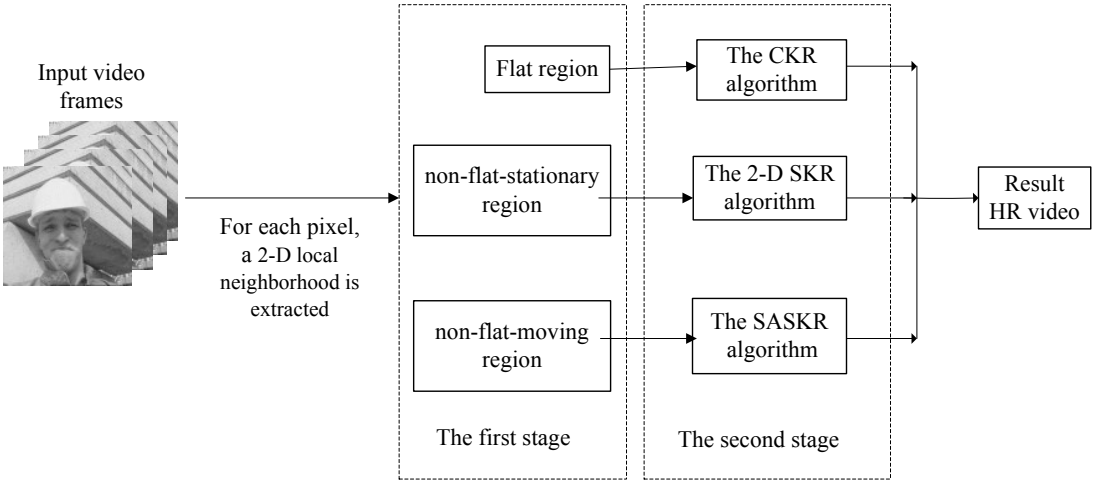


**Fig. 4.** A flow chart of the proposed framework

In the first stage, a 2-D local neighborhood of the pixel of interest is extracted and the region type of the local neighborhood is analyzed. The regions in video frames are classified into three categories: flat, non-flat-stationary, and non-flat-moving regions. The analysis process is divided into two steps.

The first step determines whether the region type of the local neighborhood is flat region by analyzing the 2-D local radiometric structure. As described previously, the local radiometric structure can be captured by the covariance matrix of the spatial gradient vectors within the local neighborhood. The eigenvalues ($\lambda_1$ and $\lambda_2$) of the covariance matrix are a measurement of the gradient strength in two perpendicular directions. Since the constant region can be characterized by $\lambda_1 = \lambda_2 = 0$, the smoothness of a region $\varsigma$, which is defined in [26], can be adopted for distinguishing between the flat and non-flat region.

$$\varsigma = |\lambda_1| + |\lambda_2| \tag{8}$$

The region is a flat region when $\varsigma$ is less than a certain threshold $\varepsilon_\varsigma$.

If the region is not a flat region, the second step further determines whether the local region has movement between video frames. This can be done by many methods. In this paper, a

simple method is adopted, which is based on the intensity difference between the local neighborhood $\mathbf{Y}_0$ and the corresponding region $\mathbf{Y}_1$ in the same position of next or previous frame. The intensity difference is defined as follow:

$$PD = \left| \mathbf{Y}_0 - \mathbf{Y}_1 \right| \tag{9}$$

Ideally, the region without movement can be characterized by $PD = 0$. However, the intensity difference is easily affected by noise. So a certain threshold $\varepsilon_{PD}$ is predefined according to the variance of the noise. The region is a non-flat-stationary region when $PD$ is less than $\varepsilon_{PD}$. Otherwise the region is a non-flat-moving region.

In the second stage, a suitable KR algorithm is selected to estimate the pixel of interest according to the region type of its local neighborhood. The CKR algorithm is used to estimate the value of the pixels in the flat regions aiming to improve computational efficiency. Its implementation is described in subsection 2.1. The 2-D SKR algorithm is used to estimate the value of the pixels in the non-flat-stationary regions aiming to improve the performance on the edge regions. The implementation of this algorithm is divided into three steps. First, the gradients ( $[z_1'(\cdot), z_2'(\cdot)]^T$ ), which are at all the sample positions $\{\mathbf{x}_i\}_{i=1}^{p}$ within the local neighborhood, are estimated. This is so-called "pilot estimate" in [21]. Second, the covariance matrix $\mathbf{C}_i$ of each sample $y_i$ is estimated from the initial "pilot". Finally, the 2-D SKR algorithm is applied to estimate the pixel of interest from the local neighborhood (which is embedded in a HR grid). The SASKR algorithm is used to estimate the value of the pixels in the non-flat-moving regions aiming to make full use of the supplementary information contained in local spatial and temporal orientations. The implementing steps of this algorithm are similar to the 2-D SKR algorithm. But the SASKR algorithm exploits the samples within both the local and non-local neighborhood.

## 4. Experimental Results

Two sets of experiments are carried out in this section. First, in Section 4.1, the computational efficiency of the proposed framework is validated by presenting the computational times of processing several real-world video sequences. Second, in Section 4.2, the performance of the proposed framework is examined by presenting the obtained results of super resolving several video sequences.

In all experiments, the degraded videos are obtained by the following manner: the original videos are blurred using a $3\times3$ uniform point spread function (PSF), spatially downsampled by a factor of 3:1 in the horizontal and vertical directions, and then contaminated by an additive white Gaussian noise with standard deviation 2. The all experiments are performed using MATLAB on Intel Core i7-3770 CPU 3.4GHz Microsoft windows 7 platform.

### 4.1 Computational Efficiency

In order to validate the computational efficiency of the proposed framework, three experiments are implemented. In first experiment, six real-world videos, namely, "Foreman"(288×351×30), "gsaleman" (288×351×30), "Miss America" (270×189×30), "Suzie" (240×351×30), "gbus" (288×351×30), "coastguard"(144×174×30), are used as original videos. The proposed framework is compared with the 3-D SKR and SASKR algorithms. The implementation code of the 3-D SKR algorithm is downloaded from the

author's website. The parameter setting of the 3-D SKR algorithm is the same as the literature [21]. As for the parameter settings of the SASKR algorithm and the proposed framework, the value of the global smoothing parameter $h$ is set to 1.5; the size of the local neighborhood is fixed to $7\times7$; the support of the similarity searching is fixed to be a $15\times15\times11$ local cubicle centered on the pixel of interest; the thresholds $\varepsilon_{\varsigma}=10$ and $\varepsilon_{PD}=20$.

The computational times of three comparison algorithms for whole video (30 frames) are summarized in **Table 1**. As we can see, the speed of the proposed framework is much faster than other two algorithms. The main reason is that the KR algorithms with different computational complexity are applied to estimate the pixels in the regions of different categories. The results indicate that the proposed framework greatly improves the computational efficiency. Note that the improvement of computational efficiency of different videos with same size and resolution is also different. For instance, the computational time of "Foreman" sequence is 434.84 seconds, whereas that of "gsalesman" sequence is 82.93 second. The main reason is that the number of the regions of different categories contained in each video sequence is different. In addition, the speed of the SASKR algorithm also is faster than the 3-D SKR. The increase of computational speed of the SAPSKR algorithm is due to the decrease of the computational cost of the kernel function. The kernel function of the SASKR algorithm is 2-D instead of 3-D.

**Table 1.** The computaional times (second) of three comparison algorithms for six videos

| Video name | 3-D SKR | SASKR | The proposed framework |
|---|---|---|---|
| Foreman | 1.46E+04 | 1.17E+04 | 434.84 |
| gsalesman | 3.68E+04 | 3.03E+04 | 82.93 |
| Miss America | 1.53E+04 | 1.38E+04 | 11.97 |
| Suzie | 1.35E+04 | 1.18E+04 | 20.27 |
| gbus | 3.98E+04 | 3.30E+04 | 346.07 |
| coastguard | 1.40E+03 | 1.13E+03 | 5.92 |

In second experiment, The proposed framework is compared with some other state-of-the-art methods[18][20]. Since the codes for these methods are not currently available publicly, only limited comparisons are possible. Under same experimental conditions, the method of [20] requires approximately 20 seconds per frame when super resolving the "Suzie" sequences with high-resolution frame size of $210\times250$ pixels, whereas the proposed framework needs only 0.5146 seconds. The C++ implementation of the method in [18] takes about two hours when super resolving a $720\times480$ frame using 30 adjacent frames at an up-sampling factor of 4, whereas the MATLAB implementation of the proposed framework only takes 3.7308 seconds.

The third experiment is to study how the computational efficiency of the proposed framework is sensitive to the scene in the videos. In this experiment, the whole "gbus" sequence($288\times351\times150$) is used as original videos.The whole "gbus" sequence includes some segments in which the scene changes frequently. A cropped sequence is obtained by taking the top 30 frames of the whole sequence. The scene in the cropped sequence changes slowly. The average computational time on the whole sequence is 31.664 second per frame, whereas the average computational time on the cropped sequence is 11.5359 second per frame. therefore, the computational effciency of the proposed framework is parameter sensitive to the scene in the videos. The main reason is that the method of determining the

region types in the proposed framework is simple. The propblem can be solved by adopting some more complex methods of determining the region types.

### 4.2 Performance

In this section, the performance of the proposed framework is examined. Performance comparisons are implemented with related state-of-the-art algorithms. For a fair comparison, the TV-based deblurring algorithm [27] is used for image deblurring. The first experiment is to compare the proposed framework with the 3-D SKR and SASKR algorithms on six video sequences. Both the objective and subjective quality assessments are adopted to evaluate different algorithms. For the objective quality assessment, the Peak-Signal-to-Noise Ratio (PSNR) and the Structural Similarity (SSIM) index [28] are adopted as the evaluation metric. The graphs in **Fig.5-10** illustrate the frame-by-frame PSNR values of reconstructed videos by three comparison algorithms. The average PSNR values are summarized in **Table 2** and the average SSIM values are summarized in **Table 3**. As can be seen, the performance of three algorithms is very close.

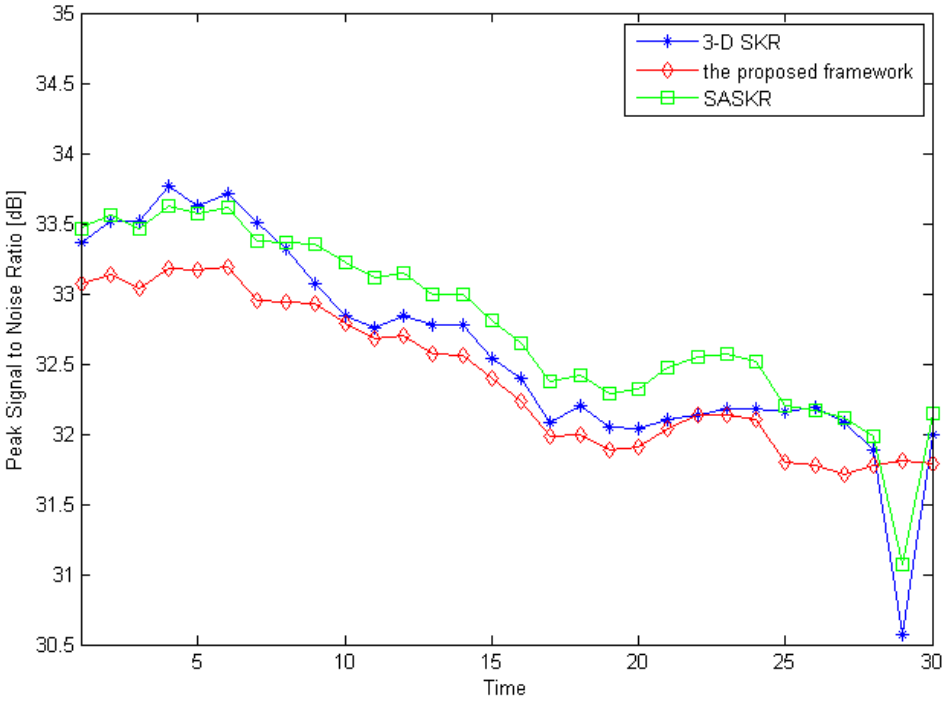


**Fig. 5.** PSNR values of each reconstructed HR frame by three algorithms for the Foreman sequences

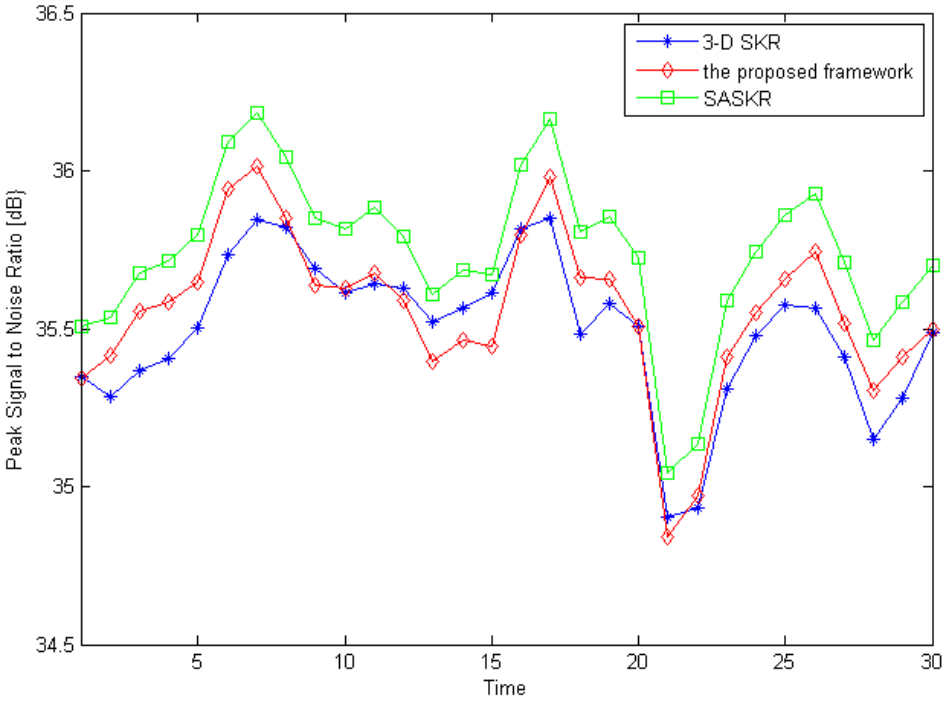


**Fig. 6.** PSNR values of each reconstructed HR frame by three algorithms for the Miss America sequences

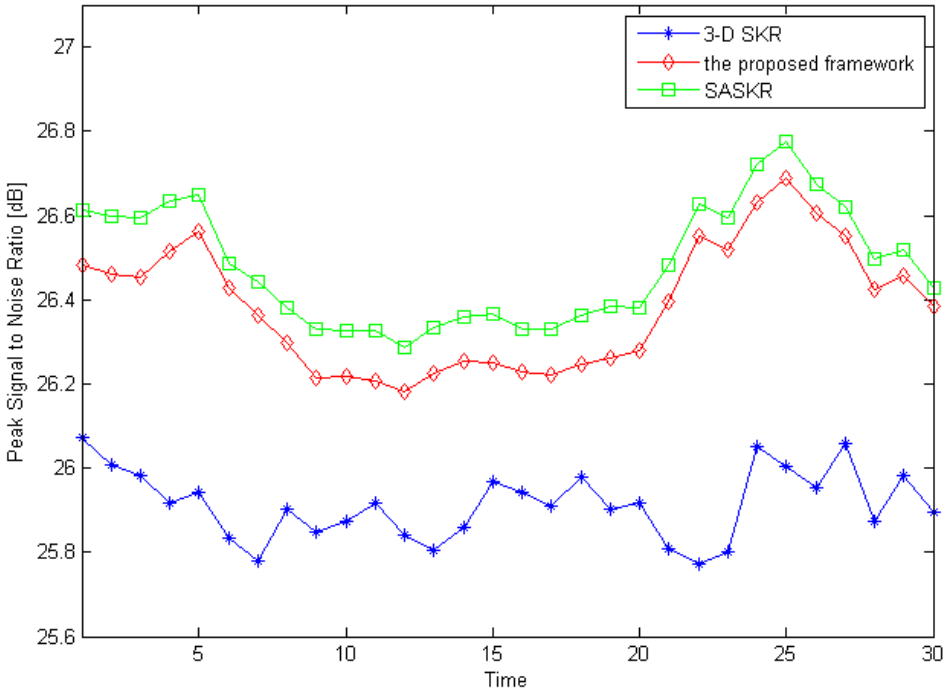


**Fig. 7.** PSNR values of each reconstructed HR frame by three algorithms for the gsalesman sequences

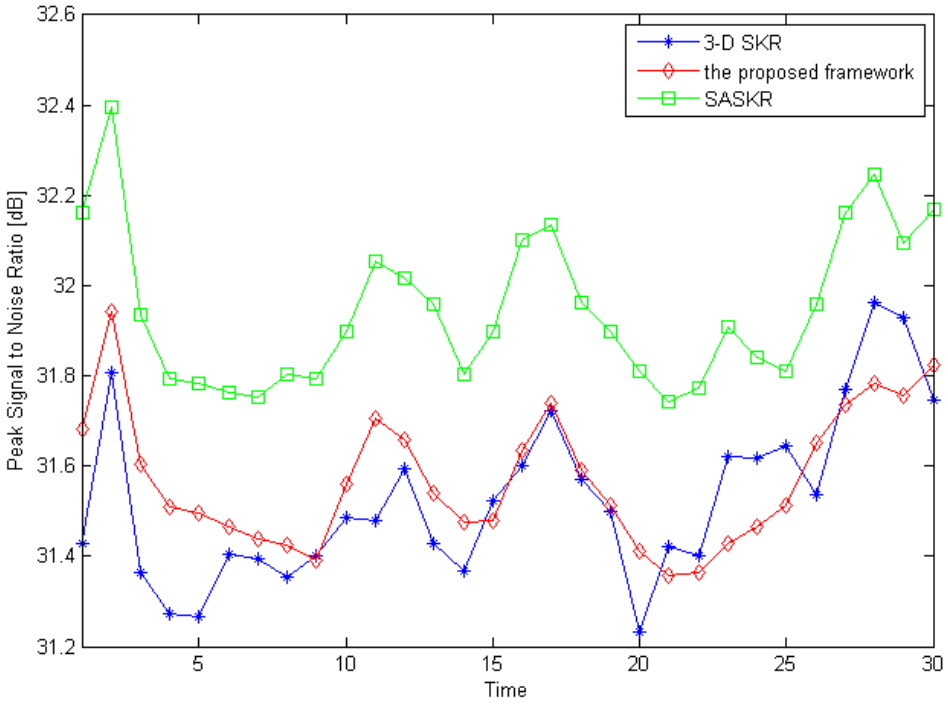


**Fig. 8.** PSNR values of each reconstructed HR frame by three algorithms for the Suzie sequences

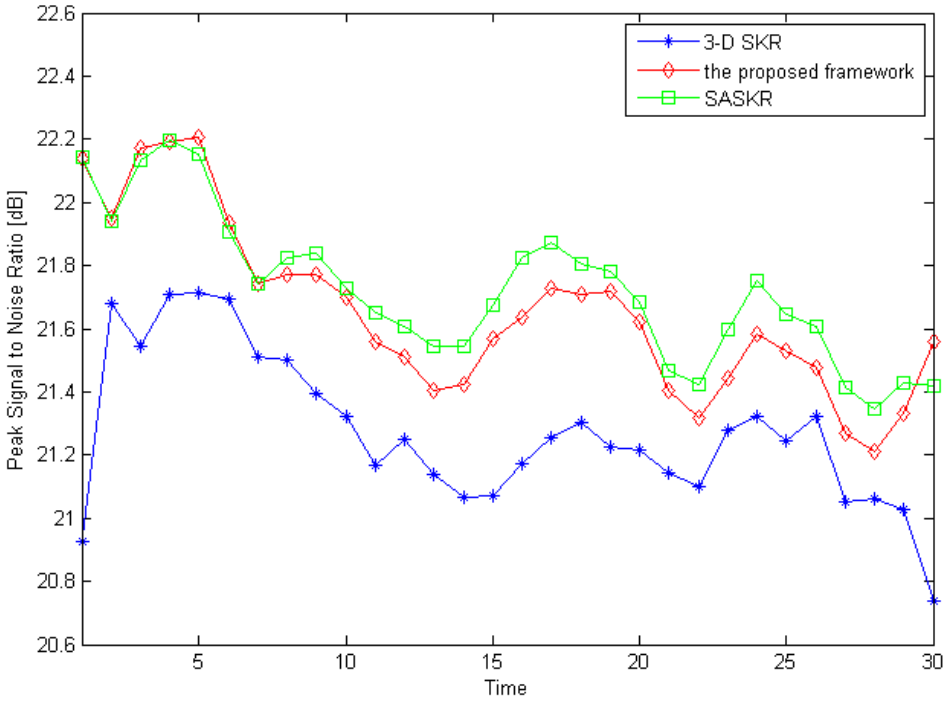


**Fig. 9.** PSNR values of each reconstructed HR frame by three algorithms for the gbus sequences

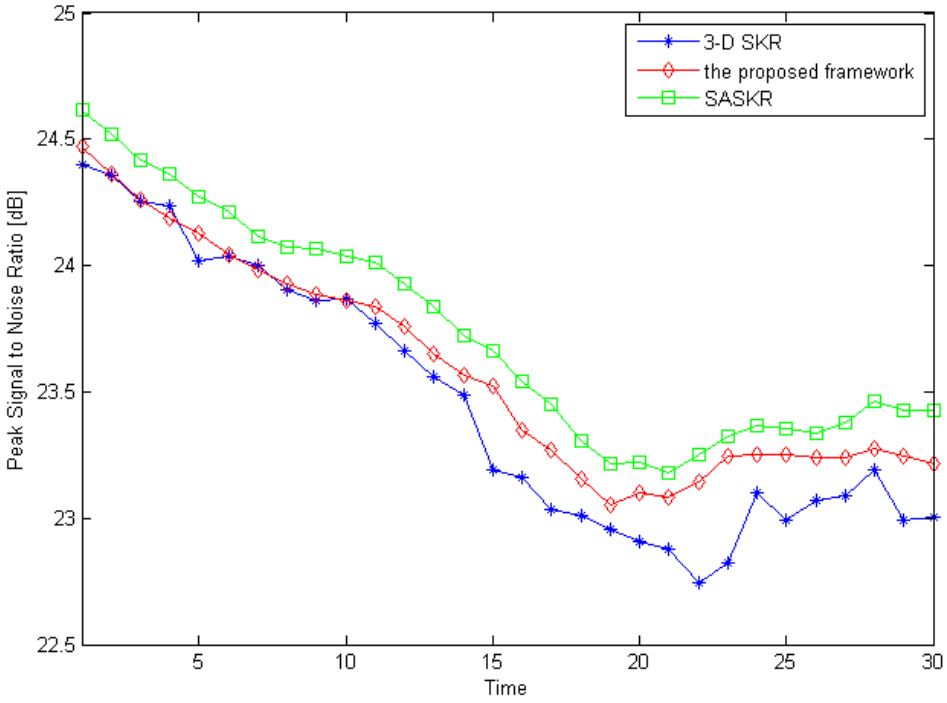


**Fig. 10.** PSNR values of each reconstructed HR frame by three algorithms for the coastguard sequences

**Table 2.** The average PSNR values for six reconstructed HR videos

| Video name | 3-D SKR | SASKR | The proposed framework |
|---|---|---|---|
| Foreman | 32.6058 | 32.7842 | 32.4128 |
| Miss America | 35.4970 | 35.7398 | 35.5566 |
| gsalesman | 25.9132 | 26.4817 | 26.3850 |
| Suzie | 31.5282 | 31.9476 | 31.5716 |
| gbus | 21.2714 | 21.7230 | 21.6525 |
| coastguard | 23.4519 | 23.7368 | 23.5856 |

**Table 3.** The average SSIM values for six reconstructed HR videos

| Video name | 3-D SKR | SASKR | The proposed framework |
|---|---|---|---|
| Foreman | 0.8869 | 0.8891 | 0.8839 |
| Miss America | 0.9141 | 0.9153 | 0.9148 |
| gsalesman | 0.7191 | 0.7565 | 0.7522 |
| Suzie | 0.8342 | 0.8481 | 0.8371 |
| gbus | 0.6123 | 0.6361 | 0.6793 |
| coastguard | 0.5639 | 0.5916 | 0.6035 |

As for the subjective quality assessment, the assessment is obtained from human visual system. The SR results on Foreman and MissAmerica sequences are shown in **Fig. 11** for visual comparison. The difference between the reconstructed frames by three algorithms is also not apparent.

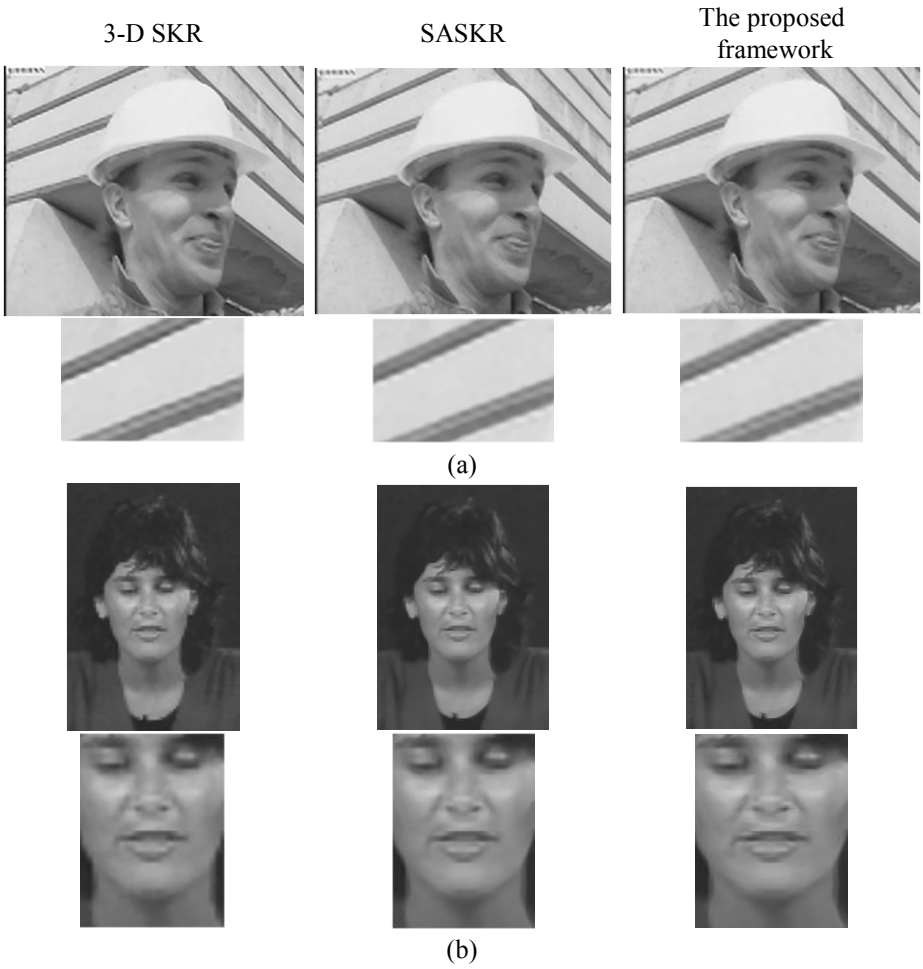


**Fig. 11.** Samples of the reconstructed frames by three algorithms. (a) foreman, (b) MissAmerica , the first column is the results of the 3-D SKR algorithm, the second column is the results of the SASKR algorithm, and the third column is the results of the proposed framework

The second experiment is to compare the proposed framework with some other state-of-the-art methods such as BM3D [19], GNLM [20], NL-KR [25]. Similarly, since the codes for these methods are not currently available, only limited comparisons are implemented. Here, the PSNR and SSIM index are used for objective evaluation. The average PSNR values are summarized in **Table 4** and the average SSIM values are summarized in **Table 5**. Note that the results from three comparison algorithms are cited directly from [25]. Since different Foreman sequences are adopted, the results of the proposed framework on Foreman sequences are different from **Table 2** and **Table 3**. For a fair comparison, Foreman sequences being used in [20] are adopted in this experiment, which are size of 288×312×30. As we can see from **Table 4** and **Table 5**, the performance of the proposed framework is slightly decreased as compared with the BM3D and NL-KR algorithms.

**Table 4.** The average PSNR values for two reconstructed HR videos

| Video name | BM3D | GNLM | NL-KR | The proposed framework |
|---|---|---|---|---|
| Foreman | 33.50 | 32.82 | 34.01 | 33.1040 |
| Miss America | 36.30 | 35.35 | 36.44 | 35.5566 |

**Table 5.** The average SSIM values for two reconstructed HR videos

| Video name | BM3D | GNLM | NL-KR | The proposed framework |
|---|---|---|---|---|
| Foreman | - | 0.9025 | 0.9120 | 0.9062 |
| Miss America | - | 0.9136 | 0.9164 | 0.9148 |

In conclusion, although the performance of the proposed framework is slightly degraded as compared with the BM3D and NL-KR algorithms, its computational efficiency is greatly improved. Therefore, it is believe that the proposed framework strikes a good balance on the computational efficiency and performance.

## 5. Conclusion

A fast kernel regression framework is proposed in this paper. In this framework, the video regions are classified into three categories: flat, non-flat-stationary, and non-flat-moving regions. The KR algorithms with different computational complexity can be adaptively selected for estimating each high-resolution pixel in different regions. Thus, the proposed framework makes the best use of the advantages of different KR algorithms and greatly improves the computational efficiency. In addition, a similarity-assisted steering kernel regression algorithm is proposed for estimating the pixels in the non-flat-moving regions. The SASKR can give better performance and higher computational efficiency than the 3-D SKR algorithm.

## References

[1] R. Y. Tsai and T. S. Huang, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Processing,* vol. 1, pp. 317-339, 1984.

[2] S. P. Kim, N. K. Bose and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 38, pp. 1013-1027,Jun,1990. Article (CrossRef Link)

[3] E.Kaltenbacher and R.C.Hardie, "High Resolution Infrared Image Reconstruction Using Multiple, Low Resolution, Aliased Frames," in *Proc. of IEEE Nat. Aerospace Electronics Conf*, vol.2, pp.702-709, May, 1996. Article (CrossRef Link)

[4] M.Irani and S.Peleg, "Improving resolution by image registration," CVGIP*: Graphical Models and Image Processing*, vol.53, no.3, pp.231-239, May, 1991. Article (CrossRef Link)

[5] S.Farsiu, N.D, Robinson, M.Elad and P.Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Processing*, vol. 13, no.10, pp. 1327-1344. October, 2004. Article (CrossRef Link)

[6] H.Shen, L.Zhang, B.Huang and P.Li, "A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution," *IEEE Trans. Image Processing*. vol.16, no.2, pp.479-490, February, 2007. Article (CrossRef Link)

[7] G.Chantas, N.Galatsanos and N.Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Trans. Image Processing*, vol.16, no.7, pp.1821-1830, July, 2007. Article (CrossRef Link)

[8] X.Li, Y.Hu, X.Gao and D.Tao, "A multi-frame image super-resolution method," *Signal Processing*, vol.90, no.2, pp.405-414, February, 2010. Article (CrossRef Link)
[9] L.Zhang, H.Zhang, H.Shen and P.Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing*, vol.90, no.3, pp.848-859, March, 2010. Article (CrossRef Link)
[10] A.M.Tekalp, M.K.Ozkan and M.I.Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," *in Proc. IEEE Int. Conf. Acousics, Speech, Signal Processing*, vol.3, pp.169-172, March, 1992. Article (CrossRef Link)
[11] C.Fan, J.Zhu, J.Gong and C.Kuang, "POCS super-resolution sequence image reconstruction based on improvement approach of keren registration method," *Intelligent Systems Design and Applications*, vol.2, pp.333-337, October, 2006. Article (CrossRef Link)
[12] S.Borman and R.L.Strevenson, "Super-Resolution from Image Sequences – A Review," *In Proc.1998 Midwest Symp.Circuits and Systems*, pp.374-378, 1999. Article (CrossRef Link)
[13] S.C.Park, M.K.Park and M.G.Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Trans. Signal Processing Magazine*, vol.20, no.3, pp.21-36, 2003. Article (CrossRef Link)
[14] S.Farsiu, M.Elad and P.Milanfar, "Video-to-Video Dynamic Super-Resolution for Grayscale and Color Sequences," *EURASIP Journal on Applied Signal Processing*, vol.2006, pp.1-15.Article (CrossRef Link)
[15] R.Hardie, "A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter," *IEEE Trans. Image Processing*, vol.16, no.12, pp.2953-2964, December, 2007. Article (CrossRef Link)
[16] M.M.Islam, V.K.Asari, M.N. Islam and M.A.Karim, "Super-Resolution Enhancement Technique for Low Resolution Video," *IEEE Trans.Consumer Electron.*, vol.56, no.2, pp.919-924, May, 2010. Article (CrossRef Link)
[17] S.H.Keller, F.Lauze, and M.Nielsen, "Video Super-Resolution Using Simultaneous motion and Intensity Calculations," *IEEE Trans. image processing*, vol.20, no.7, pp.1870-1884, July, 2011. Article (CrossRef Link)
[18] C.Liu and D.Sun, "A Bayesian Approach to Adaptive Video Super Resolution," in *Proc. of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pp.209-216, June, 2011. Article (CrossRef Link)
[19] A.Danielyan, A.Foi, V.Katkovnik and K.Egiazarian, "Image and video super-resolution via spatially adaptive block-matching filtering," *Presented at the Int.Workshop on Local and Non-Local Approximation in Image Processing, Lausanne, Switzerland*, August, 2008. Article (CrossRef Link)
[20] M.Protter, M.Elad, H.Takeda and P.Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction," *IEEE Trans. image processing*, vol.18, no.1, pp.36-51, January, 2009. Article (CrossRef Link)
[21] H.Takeda, P.Milanfar, M.Protter and M.Elad, "Super-Resolution without Explicit Subpixel Motion Estimation," *IEEE Trans.image processing*, vol.18, no.9, pp.1958-1975, September, 2009. Article (CrossRef Link)
[22] H.Takeda, S.Farsiu and P.Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Trans.image processing*, vol.16, no.2, pp.349-366, February, 2007. Article (CrossRef Link)
[23] K.Zhang, G.Mu, Y.Yan, X.Gao, and D.tao, "Video super-resolution with 3D adaptive normalized convolution," *Neurocomputing*, vol.94, no.1, pp.140-151, October 2012. Article (CrossRef Link)
[24] H.Zhang, J.Yang, Y.Zhang and T.Huang, "Non-Local Kernel Regression for Image and Video Restoration.," *Lecture Notes in Computer Science*, 2010, vol.6313, *Computer Vision-ECCV 2010*, pp.566-579. Article (CrossRef Link)
[25] H.Zhang, J.Yang, Y.Zhang and T.Huang, "Image and Video Restorations via Nonlocal Kernel Regression," *IEEE Transactions on cybernetics*, vol.43, no.3, pp. 1035 - 1046, June, 2013. Article (CrossRef Link)
[26] H.Su, L.Tang, Y.Wu, D.Tretter and J.Zhou, "Spatially Adaptive Block-based Super-resolution," *IEEE Transactions on Image Processing*, vol.21, no.3, pp.1031-1045, March, 2012. Article (CrossRef Link)

[27] P.Getreuer , 2012.[Online].Available:http://www.getreuer.info/home/tvreg.
[28] 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 Process*, vol.13, no.4, pp.600-612, Apr.2004. Article (CrossRef Link)

**Wen-sen Yu** is currently a Ph.D. student in College of Computer Science at Sichuan University, Chengdu, China. He is working as an associate professor of College of Mathematics and Computer Science in WuYi University. He received his M. S. degree in software engineering from Jiangxi Normal University, Nanchang, China, in 2004. His current research interests include image and video super-resolution, image fusion, sparse representation.

**Ming-hui Wang** was born in Xi'an, China, in 1971. He is working as a professor of College of Computer Science in Sichuan University. He has been working as a post-doctoral research fellow of information and communication engineering in Tsinghua University. He received the Ph.D and M. S. degrees from Northwestern Polytechnical University and Xidian University, respectively. His current research interests are information fusion and computer vision.

**Hua-wen Chang** is currently a Ph.D. student in College of Computer Science at Sichuan University, China. He is working as a lecturer of College of Computer and Communication Engineering in Zhengzhou University of Light Industry. He received his M. S. degree in computer science from Guilin University of Technology, China, in 2007. His current research interests include image and video quality assessment, image fusion, sparse representation, independent component analysis and wavelet analysis.

**Shu-qing Chen** received her M.S. degree from South-central University for nationalities, China, in 2005. Currently, she is working for her Ph.D degree in Sichuan University and also a lecturer with the Department of Electronics and Information Engineering in Putian University, China. Her research interests include image & video feature extraction, matching and alignment, and reconstructions.