Robust Multi-Frame Super-Resolution Based on Spatially Weighted Half-Quadratic Estimation and Adaptive BTV Regularization

Xiaohong Liu, *Student Member, IEEE*, Lei Chen, *Student Member, IEEE*, Wenyi Wang, and Jiying Zhao, *Member, IEEE*

***Abstract*— Multi-frame image super-resolution focuses on reconstructing a high-resolution image from a set of low-resolution images with high similarity. Combining image prior knowledge with fidelity model, the Bayesian-based methods have been considered as an effective technique in super-resolution. The minimization function derived from maximum *a posteriori* probability (MAP) is composed of a fidelity term and a regularization term. In this paper, based on the MAP estimation, we propose a novel initialization method for super-resolution imaging. For the fidelity term in our proposed method, the half-quadratic estimation is used to choose error norm adaptively instead of using fixed L_1 and L_2 norms. Besides, a spatial weight matrix is used as a confidence map to scale the estimation result. For the regularization term, we propose a novel regularization method based on adaptive bilateral total variation (ABTV). Both the fidelity term and the ABTV regularization guarantee the robustness of our framework. The fidelity term is mainly responsible for dealing with misregistration, blur, and other kinds of large errors, while the ABTV regularization aims at edge preservation and noise removal. The proposed scheme is tested on both synthetic data and real data. The experimental results illustrate the superiority of our proposed method in terms of edge preservation and noise removal over the state-of-the-art algorithms.**

***Index Terms*— Multi-frame super-resolution, median operator based initialization, spatial weight, half-quadratic estimation, adaptive bilateral total variation (ABTV).**

I. Introduction

SUPER-RESOLUTION (SR) has been a promising technique to increase the image resolution without modifying the sensor of a camera. Different from single image super-resolution, multi-frame super-resolution aims to reconstruct a high-resolution (HR) image from a set of low-resolution (LR) images taken from the same scene.

In a multi-frame SR process, how to accurately extract the image texture existing in different LR frames is a vital challenge to reconstruct an HR image with good quality [1]. In this case, image registration and blur identification should be taken into consideration [2], [3]. Image registration is used to estimate the displacements among LR images, which directly influences the quality of final output. For blur identification, point-spread function is incorporated to model the blur kernel. Moreover, during the reconstruction step, if the LR images have non-redundant information, the ill-posed nature of SR problem can be over-determined by adding more LR images to the objective function. The pixels in LR images can be aligned on an HR grid according to the sub-pixel shifting with respect to the reference LR image.

However, in practical applications, the irregular pixel movement and unknown blurring can directly influence the super-resolved result. Moreover, the LR images are not always non-redundant, which limits the performance of the simple image reconstruction model. Besides, the blur kernel of a camera is usually unknown, which makes the multi-frame image SR more challenging.

In this work, in order to solve the above mentioned problems, we propose a new robust multi-frame SR method based on spatially weighted half-quadratic estimation and adaptive BTV regularization. In our proposed algorithm, there are three major contributions which effectively improve the quality of the final estimated HR image:

1) A novel initialization method based on median operator is introduced. Different from the commonly used bilinear and bicubic interpolation, our proposed initialization method uses image registration techniques to align LR images referenced by the objective LR image. After alignment, the median operator is used to generate a composed LR image insensitive to outliers, and then the initial HR image is created by upsampling the composed LR image.
2) A novel fidelity term based on spatially weighted half-quadratic estimation is proposed. The spatial weight matrix is determined by the frame-wise and pixel-wise information obtained from the observation errors. The half-quadratic estimation is used to choose error

Manuscript received October 25, 2016; revised October 2, 2017 and April 15, 2018; accepted May 31, 2018. Date of publication June 15, 2018; date of current version July 9, 2018. This work was supported by the Natural Sciences and Engineering Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Amit K. Roy Chowdhury. *(Corresponding author: Jiying Zhao.)*

X. Liu is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: liux173@mcmaster.ca).

L. Chen and J. Zhao are with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: lchen148@uottawa.ca; jzhao@uottawa.ca).

W. Wang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wangwenyi@uestc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2848113

norm adaptively instead of using fixed L_1 or L_2 norm. In our proposed method, the half-quadratic estimation is adjusted by the spatial weight to achieve better performance.

3) A novel adaptive BTV regularization method is proposed. Due to the drawbacks of traditional regularization methods, an adaptive matrix is used to adjust the relevant value in the BTV regularization. Therefore, the performance of the BTV regularization can be controlled by this matrix based on gradient operator.

The rest of this paper is organized as follows. Section II reviews some related works on multi-frame SR. Section III introduces the observation model and basic framework of multi-frame SR. Section IV details the proposed algorithm based on spatially weighted half-quadratic estimation and adaptive BTV regularization. Section V presents the experimental results and Section VI concludes this paper.

II. Related Works

Super-resolution techniques have been extensively studied in the past three decades. According to the number of input images, they can be classified into three principal categories: single image SR, multi-frame image SR and video SR. In this paper, we mainly focus on the multi-frame image SR.

In general, the single image SR is about recovering high-frequency components from the observed LR image. Yang *et al.* [4] proposed a typical model based on sparse representation, which trained a joint over-complete dictionary pair to reconstruct LR images. Through edge detection and feature selection, Chan *et al.* [5] proposed an extended neighbor embedding based super-resolution method to address the inappropriate choices of sizes and training patches. Timofte *et al.* [6] proposed an Anchored Neighborhood Regression (ANR) model to reconstruct the LR images more efficiently by utilizing neighbor-based dictionaries. Their subsequent A+ model [7] combined ANR and Simple Function (SF) together to further improve the quality of output HR images. Dong *et al.* [8] first introduced a deep learning method for single image SR based on Convolutional Neural Networks (CNN). The experimental results demonstrate the effectiveness of their proposed method. For video SR, Kapperler *et al.* [9] proposed a VSRnet framework which regarded input frames as independent images and fed the images to a SRCNN-inspired network with three specified combination methods. Caballero *et al.* [10] proposed a Video Efficient Sub-Pixel Convolutional Neural Network (VESPCN) that applied motion compensation to input frames and utilized subpixel-shuffle to upsample LR images. Based on VESPCN, Tao *et al.* [11] proposed a novel sub-pixel motion compensation (SPMC) algorithm to integrate motion compensation and upsampling into one operation. Yang *et al.* [12] proposed a spatial-temporal recurrent residual network to model inter-frame correlation for video SR. Liu *et al.* [13] proposed a temporal adaptive network and a spatial alignment network for video SR. Both the temporal adaptation and the spatial alignment modules were used to increase the robustness to complex motion. Their experimental results demonstrated the superiority of the proposed model in terms of spatial consistency and temporal coherence.

The multi-frame SR was first addressed in paper [14] using a frequency domain algorithm which is easy to implement and computationally cheap. However processing multi-frame SR in frequency domain will introduce serious visual artifacts. Since then, many approaches have been proposed to solve the problem. Due to the drawbacks of frequency domain based methods, algorithms which enhance images in the spatial domain have become increasingly popular [15], [16]. Since super-resolution is an ill-posed problem, regularization techniques are widely used to constrain the minimization function and are also utilized as prior knowledge for the fidelity model. The Bayesian-based spatial domain methods can effectively solve this ill-posed problem and are widely used in image super-resolution. Spatial domain based multi-frame image SR usually reconstructs the HR image from the related LR images by exploiting the sub-pixel displacements [17]. In practical applications, except for the affine movement, the sub-pixel displacements can also be partial movement, non-rigid movement and occlusion which are difficult to estimate.

In general, the framework of multi-frame image SR in spatial domain mainly contains the fidelity term and the regularization term. The fidelity term is used to maintain the fidelity between the HR frame and LR frames. And the regularization term aims at regularizing the minimization function. Because the noise in the observation model usually fits the Gaussian distribution, choosing L_2 norm for the fidelity term can obtain good results. But in practical applications, the observation model suffers from various noises and errors introduced by inaccurate estimation of registration and blurring kernels. Farsiu *et al.* first used L_1 norm in the fidelity term and achieved better results than using L_2 norm [17]. Although the L_1 norm is robust to outliers, it may introduce more observation errors than L_2 norm while the estimation of images is accurate. The drawbacks of fixed norms motivated researchers to combine the advantages of L_1 and L_2 norms. Some M-estimators such as Huber function [18] were proposed to replace the fixed norms. Yue *et al.* [19] proposed a locally adaptive L_1, L_2 norm to handle images with mixed noises and outliers. But setting a threshold to choose L_1 or L_2 norm makes the minimization function non-derivable. Zeng and Yang [20] proposed a new adaptive norm based on half-quadratic estimation. It combined the advantage of L_1 and L_2 norms and the minimization function can be derivable at every point. Therefore, the simple optimization methods such as steepest decent still guaranteed the convergence of the minimization function. Köhler *et al.* [21] proposed an Iteratively Re-weighted (IRW) multi-frame SR method based on MAP estimation. Two adaptive matrices were generated to weigh the L_2 norm in the fidelity term and the L_1 norm in the regularization term respectively. Diverging from the Bayesian-based methods, Huang *et al.* [22], [23] proposed a novel multi-frame SR method based on bidirectional recurrent convolutional neural network. In the model, the traditional recurrent full connections were replaced with weight-sharing convolutional connections, and conditional convolutional connections were added for temporal dependency modelling.

Liao *et al.* [24] proposed a deep draft-ensemble learning to address the multi-frame/video SR. In this model, multiple SR drafts were generated to improve the motion estimation, and a deep convolutional neural network was used to reconstruct the super-resolved result from these SR drafts. Their method is currently achieving the best multi-frame SR results among deep learning based algorithms.

Within the regularization techniques, one of the commonly used methods is Tikhonov regularization based on L_2 norm [25]. However, L_2 norm is particularly sensitive to outliers such that it can introduce some artifacts into images. Nowadays, sparse prior is very popular in single image SR. But for multi-frame SR, the redundant information among the LR frames in the spatial domain is more reliable than in the sparse domain. Total variation (TV) family such as bilateral total variation (BTV) [17] is another popular regularization technique. Unlike Tikhonov regularization, the BTV uses L_1 norm to handle the outliers. Farsiu *et al.* showed that the BTV regularization is robust to outliers and can retain more detailed information than Tikhonov regularization.

III. Preliminaries

A. Observation Model of Multi-Frame Super-Resolution

The observation model formulates the relationship between an HR frame and a sequence of LR frames with high similarity. Therefore, an accurate observation model is vital for the multi-frame SR algorithm. According to the study on camera sensor, some assumptions have been made to describe the observation model, which directly affect the performance of the final result.

In general, the LR frames can be regarded as the corresponding HR frame going through the geometric motion operator, blurring operator and down-sampling operator successively. Therefore, considering all the degradative operators, the observation model can be formulated as follows

$$\mathbf{Y}_k = \mathbf{D}\mathbf{B}_k\mathbf{M}_k\mathbf{X} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{X}$ is the HR frame and expressed in lexicographic order as $\mathbf{X} = [x_1, x_2, \ldots, x_N]^T$, where N is the total number of pixels in HR frame which equals to $rm \times rn$ and r is the down-sampling factor. Therefore, the size of $\mathbf{X}$ is $rmrn \times 1$. Similar to the definition of $\mathbf{X}$, $\mathbf{Y}_k = [y_{k,1}, y_{k,2}, \ldots, y_{k,L}]^T$, which represents the kth LR frame with the size of $mn \times 1$, where $k = 1, 2, \ldots, K$. K is the number of LR frames and $L = m \times n$. $\mathbf{M}_k$ represents the geometric motion matrix between HR frame and kth LR frame with the size of $rmrn \times rmrn$. $\mathbf{B}_k$ is the blurring matrix for the kth LR frame with the size of $rmrn \times rmrn$ and $\mathbf{D}$ is the down-sampling matrix with the size of $mn \times rmrn$. In general, image noise should be taken into consideration. $\mathbf{n}_k$ represents the noise added into the kth LR frame with the size of $mn \times 1$.

In order to simplify Equ. (1), $\mathbf{D}\mathbf{B}_k\mathbf{M}_k$ can be regarded as a system matrix $\mathbf{W}_k$ as proposed in paper [26]. As mentioned above, each LR pixel can be obtained via a weighted sum of the relevant HR pixels and the mapping weights are saved in $\mathbf{W}_k$ in row-wise order. By combining the different transformations as a united system matrix $\mathbf{W}_k$, Equ. (1) can be rewritten as follows

$$\mathbf{Y}_k = \mathbf{W}_k\mathbf{X} + \mathbf{n}_k, \tag{2}$$

B. Basic Framework of Multi-Frame Super-Resolution

The basic framework of multi-frame SR contains a fidelity term and a regularization term. For the fidelity term, the M-estimator is introduced to minimize the residual between the estimated HR frame and the given LR frames. The regularization term aims at constraining the minimization function so that the reconstructed image can reach a robust state. The traditional framework of multi-frame SR can be formulated as follows

$$\hat{\mathbf{X}} = arg\min_{\mathbf{X}} \left\{ \sum_{k=1}^{K} \|\mathbf{Y}_k - \mathbf{W}_k\mathbf{X}\|_p^p + \lambda\Upsilon(\mathbf{X}) \right\}, \tag{3}$$

where $\Upsilon(\mathbf{X})$ is the regularization term with respect to $\mathbf{X}$. K is the total number of LR images, λ is the trade-off parameter between the two terms and p represents the choice of Lp norm.

For the regularization term $\Upsilon(\mathbf{X})$, Tikhonov and TV family are usually used as image prior knowledge. In this paper, bilateral total variation (BTV) is adopted since it is computationally cheap and easy to implement. The formula of BTV regularization is expressed as follows

$$\Upsilon_{BTV}(\mathbf{X}) = \sum_{n=-P}^{P}\sum_{m=-P}^{P} a^{|m|+|n|} \left\| \mathbf{X} - \mathbf{S}_x^n\mathbf{S}_y^m\mathbf{X} \right\|_1, \tag{4}$$

where $\mathbf{S}_x^n$ shifts $\mathbf{X}$ by n pixels in vertical direction and $\mathbf{S}_y^m$ shifts $\mathbf{X}$ by m pixels in horizontal direction. a is a scaled weight with the range of $0 < a < 1$ and P is a parameter used to control the decaying effect on the summation of the BTV regularization.

The BTV regularization term preserves the image texture by penalizing the first-order gradient magnitudes. Although it can suppress noise, the BTV can remove a lot of texture information as well. In order to solve this problem, we combine the BTV regularization with the gradient operator to preserve image texture and suppress noise simultaneously.

IV. Proposed Multi-Frame SR Algorithm

In this section, we introduce our proposed algorithm in detail. For the image initialization, we use a novel method based on the median operator to generate an outlier-insensitive HR image as our initial setting. For the fidelity term, the half-quadratic estimation is used to choose error norms adaptively. Besides, a spatial weight matrix based on frame-wise and pixel-wise observation errors is established to scale the result of half-quadratic estimation. For the regularization term, we propose an adaptive BTV regularization method to suppress image noises and preserve image texture simultaneously. Compared with the traditional BTV regularization method, our proposed regularization method assigns each pixel an adaptive weight. If a pixel is in edge areas, the corresponding weight will be small to preserve image edges. Otherwise, the weight will be large to suppress image noise. Therefore,

our regularization method can adaptively suppress noise and preserve image edges simultaneously by imposing different weightings on different pixels.

A. Estimation of Initial High-Resolution Image

Multi-frame SR algorithm typically utilizes optimization methods to minimize the objective function so that the final HR image can be reconstructed when the function reaches the minimal point. In this case, if the initial HR image is estimated accurately, the minimization function can reach a stable state rapidly to reconstruct the HR image with good quality. In other words, the image initialization directly influences the quality of final results. In traditional multi-frame SR methods, the initial HR estimation $\mathbf{X}_0$ is obtained by using bicubic or bilinear interpolation on the reference LR image. If the quality of the reference LR image is bad, the quality of the initial HR will be unsatisfactory. In our proposed algorithm, a novel initialization method based on image warping and median operator is used to solve this problem. Each LR frame is first warped according to the shape of the reference LR frame. Then a composed LR image is reconstructed by using median operator. Finally, the initial HR image is obtained by interpolating the composed LR image. Unlike the traditional initialization which only considers the reference frame, our proposed initial method utilizes all LR frames to generate the initial HR image.

Image warping transforms an image from one plane to another plane based on some mathematical functions [27]. In our initialization, every non-reference LR image is warped as the shape of the reference LR image. Therefore, the pixels in the same location of registrated LR images have the same details. If we choose the first LR frame $\mathbf{Y}_1$ as the reference image, the warping procedure can be expressed as

$$\mathbf{Y}'_k = P_w(\mathbf{Y}_k, \mathbf{Y}_1), \quad k = 2, 3, 4, \cdots, K, \tag{5}$$

where $P_w(\cdot)$ is the projection function, which maps the non-reference LR images to the reference image. K is the total number of LR frames in sequence.

After the process of image warping, the LR images are aligned and the composed LR image is generated to keep the texture information. If the mean operator is chosen to generate the composed LR image, the expression can be formulated as

$$\mathbf{Y}_c = f_{mean}(\mathbf{Y}_1, \mathbf{Y}'_2, \mathbf{Y}'_3, \cdots, \mathbf{Y}'_K), \tag{6}$$

where f_{mean} denotes the pixel-wise mean operator and $\mathbf{Y}_c$ represents the composed LR image. The value of the $(i, j)th$ pixel in composed image is calculated by

$$\mathbf{Y}_c^{(i,j)} = \frac{\mathbf{Y}_1^{(i,j)} + \mathbf{Y}_2^{\prime(i,j)} + \mathbf{Y}_3^{\prime(i,j)} + \cdots + \mathbf{Y}_K^{\prime(i,j)}}{K}. \tag{7}$$

Different from the mean operator, the median operator searches the median value of pixels in the same location of registrated LR images, which can be expressed as

$$\mathbf{Y}_c = f_{med}(\mathbf{Y}_1, \mathbf{Y}'_2, \mathbf{Y}'_3, \cdots, \mathbf{Y}'_K), \tag{8}$$

where f_{med} represents the median operator. The $(i, j)th$ pixel in the composed LR image is the median value of all $(i, j)th$ pixels in relevant LR images. Fig. 1 shows the composed LR image by using mean and median operators respectively. The LR sequence is generated by shifting, rotating and blurring the original *Cameraman* image. The mixed Gaussian and Salt&Pepper noises have also been added in each LR frame.

(a) (b)

Fig. 1. Example of using mean and median operator to generate composed LR image respectively. (a) Mean operator. (b) Median operator.

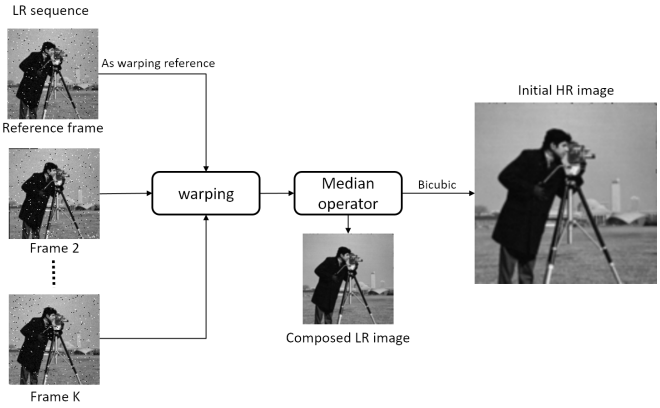


Fig. 2. Framework of generating initial HR image $\mathbf{X}_0$.

Comparing these two operators, the mean operator calculates the mean value of relevant LR pixels. However, due to the different levels of distortion, the LR images are different from each other. If we use the mean operator to reconstruct the composed LR image, the distorted pixels will have the same contribution to the fine pixels, which may lead to undesirable result. Distinct from the mean operator, the median operator extracts the median value of the corresponding pixels. Even though there are some distorted pixels, they barely affect the median value. Therefore, the median operator is more robust to outliers in the sequence of corresponding pixels. Accordingly, we choose the median operator to generate our composed LR image.

In addition, the initial HR image is generated by interpolating the composed LR image. Since our initial algorithm can eliminate most of outliers by searching the median value of all relevant pixels, our initial estimation has better quality than other methods. The whole framework of generating the initial HR image is shown in Fig. 2.

B. Proposed Spatially Weighted Fidelity Term Based on Half-Quadratic Estimation

Our proposed multi-frame SR algorithm is based on maximum a posteriori (MAP) estimation. According to the

Bayesian theorem, the expression of MAP estimation can be formulated as

$$\hat{\mathbf{X}}_{MAP} = arg\max_{\mathbf{X}} \; P(\mathbf{X}|\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_K) = arg\max_{\mathbf{X}} \; \frac{P(\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_K|\mathbf{X})P(\mathbf{X})}{P(\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_K)}, \quad (9)$$

where $\mathbf{X}$ represents the estimated HR image, $\mathbf{Y}_k$ is the kth LR image and K is the total number of LR images. Since $P(\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_K)$ has no influence on the maximization function $\hat{\mathbf{X}}_{MAP}$, the above MAP estimation can be rewritten as

$$\hat{\mathbf{X}}_{MAP} = arg\max_{\mathbf{X}} P(\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_K|\mathbf{X})P(\mathbf{X}). \quad (10)$$

In general, the observation of each LR image $\mathbf{Y}_k$ is independent. Therefore, Equ. (10) can be simplified to

$$\hat{\mathbf{X}}_{MAP} = arg\max_{\mathbf{X}} \prod_{k=1}^{K} P(\mathbf{Y}_k|\mathbf{X}) \cdot P(\mathbf{X}), \quad (11)$$

where $P(\mathbf{X})$ describes the prior probability for an HR image and $P(\mathbf{Y}_k|\mathbf{X})$ represents the conditional probability of the LR image $\mathbf{Y}_k$ given an HR image $\mathbf{X}$. The distribution of $P(\mathbf{Y}_k|\mathbf{X})$ is derived from the observation model. Therefore, if the observation error is assumed to be the independent and identically distributed (i.i.d) Gaussian noise with zero mean, the distribution of $P(\mathbf{Y_k}|\mathbf{X})$ can be expressed as

$$P(\mathbf{Y}_k|\mathbf{X}) \propto \exp\left\{-\frac{(\mathbf{Y}_k - \mathbf{W}_k\mathbf{X})^T(\mathbf{Y}_k - \mathbf{W}_k\mathbf{X})}{2\sigma_g^2}\right\}, \quad (12)$$

where σ_g is the standard derivation of the Gaussian noise. In practical applications, the noises existing in LR frames are usually mixed. Besides, inaccurate motion estimation and invalid pixels should be taken into consideration. In order to solve these problems, the robust SR methods based on Laplacian distribution and M-estimators were proposed in the papers [17], [20]. However, the distribution of the observation error in these methods was assumed to be space invariant, which limits their performance in real-world applications [21]. In our proposed method, the distribution of observation error is space-variant since each image pixel is scaled by a spatial weight respectively. For each LR frame $\mathbf{Y}_k$, the observation error $\mathbf{r}_k$ can be formulated as

$$\mathbf{r}_k = \mathbf{Y}_k - \mathbf{W}_k\mathbf{X}, \quad (13)$$

where $\mathbf{r}_k = [r_{k,1}, r_{k,2}, \cdots, r_{k,L}]^T$ represents the residual vector.

The half-quadratic (HQ) function was first proposed in paper [28] as a potential function, which combines the advantages of L_1 and L_2 norms. With the parameter α, the half-quadratic function can reduce the effect of different kinds of errors such as large registration errors and small Gaussian errors. The half-quadratic estimation is defined as

$$f(x, \alpha) = \alpha\sqrt{\alpha^2 + x^2}, \quad (14)$$

where α is a positive constant and x represents the observation error. The half-quadratic function is strictly convex and twice continuously differentiable so that any convex optimization algorithms can easily obtain the optimum value. The first derivative of $f(x, \alpha)$ is approximately linearly proportional to small errors and then gradually approaches to a constant. The first derivative function with respect to x is shown below

$$f'(x, \alpha) = \frac{\alpha x}{\sqrt{\alpha^2 + x^2}}. \quad (15)$$

If $\alpha = 1$, the half-quadratic function and its derivative is shown in Fig. 3.

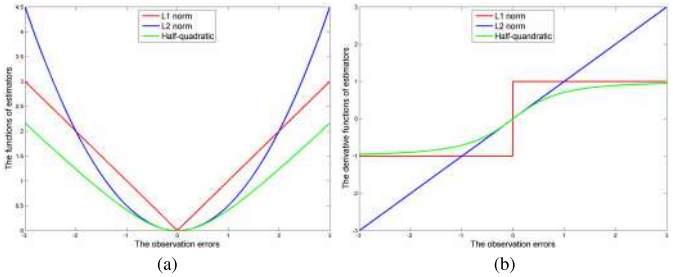


Fig. 3. Error norms. (a) Norm functions of L_1, L_2 and $f(x, \alpha)$, (b) Their corresponding derivative norm functions.

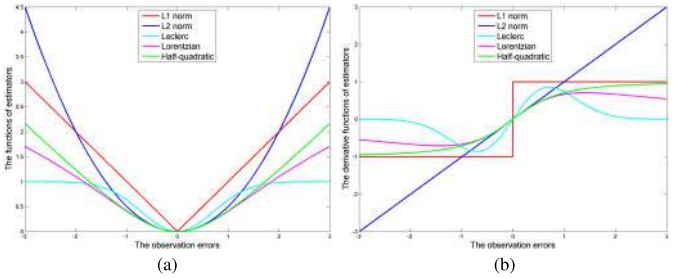


Fig. 4. Error norms. (a) Norm functions of L_1, L_2, Leclerc, Lorentzian and half-quadratic estimation, (b) Their corresponding derivative norm functions.

The derivative of half-quadratic function performs like L_2 norm when the observation errors are small, and then gradually performs like L_1 norm when the observation errors are large to suppress the outliers such as image noise and mis-registrations. Fig. 4 shows the norm functions and their derivative functions of L_1 norm, L_2 norm, Leclerc, Lorentzian and half-quadratic estimation respectively. The thresholds of Leclerc and Lorentzian are both set to 1. Compared with other M-estimators such as Leclerc and Lorentzian, the half-quadratic estimation has the best performance.

As shown in Fig. 4, the traditional fixed norm function such as L_1 and L_2 norms can not adjust their output due to different observation errors. For the commonly used M-estimators, the Leclerc and Lorentzian can fit L_1 and L_2 norms adaptively with different inputs. However, they both have extremum as the observation error increases, which makes them non-monotonic. Unlike the Leclerc and Lorentzian estimators, the half-quadratic estimation is monotonically increasing. Therefore, it is robust to the observation errors. By using the half-quadratic function as an adaptive norm, the spatially weighted fidelity term is

$$\hat{\mathbf{X}} = arg\min_{\mathbf{X}} \left\{\sum_{k=1}^{K}\sum_{l=1}^{L} \beta_{k,l} \cdot \alpha_k\sqrt{\alpha_k^2 + r_{k,l}^2}\right\}, \quad (16)$$

where α_k is the half-quadratic parameter for kth frame and $r_{k,l}$ represents the lth observation error of kth frame. $\boldsymbol{\beta}_k = [\beta_{k,1}, \beta_{k,2}, \cdots, \beta_{k,L}]^T$ is the spatial weight vector for the kth observation error. In our algorithm, it is adaptively determined by multiplying two weighting functions which control the global weight and local weight respectively. The expression of $\boldsymbol{\beta}_k$ can be formulated as

$$\boldsymbol{\beta_k} = \boldsymbol{\beta}_k^{global} \cdot \boldsymbol{\beta}_k^{local}, \tag{17}$$

where $\boldsymbol{\beta}_k^{global}$ and $\boldsymbol{\beta}_k^{local}$ are the global and local weighting functions respectively. For the global weighting function, since the quality of each LR frame is different, $\boldsymbol{\beta}_k^{global}$ gives every LR frame a global weight according to the average observation error. The definition of the average observation error $\overline{r_k}$ is expressed as follows

$$\overline{r_k} = \sum_{l=1}^{L} |r_{k,l}|/L, \tag{18}$$

where $|\cdot|$ represents the absolute operator and L is the total number of pixels in the LR frame. When the average observation error is large, the relevant weight should be small to decay the effect of this frame. When the error is small, the observation model can accurately estimate the HR image from the LR frame. Therefore, the corresponding weight should be large. Based on the above analysis, the global weighting function is expressed as

$$\boldsymbol{\beta}_k^{global} = \frac{1/\overline{r_k}}{\max(1/\overline{\boldsymbol{r}})}, \tag{19}$$

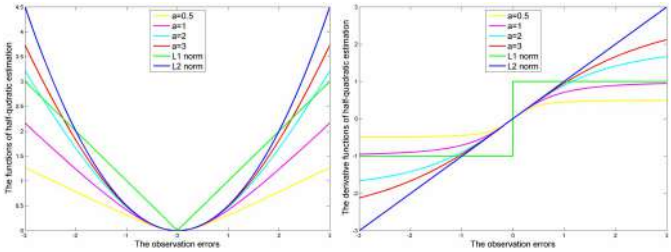
where $\overline{\boldsymbol{r}} = [\overline{r_1}, \overline{r_2}, \cdots, \overline{r_K}]$. For the local weighting function, it is mainly used to eliminate outliers from inliers in our algorithm. If we assume that the inlier pixels still follow the Gaussian distribution, the weighting function $\boldsymbol{\beta}_k^{local}$ can be defined as [21]

$$\boldsymbol{\beta}_k^{local} = \begin{cases} 1 & \text{if } |r_{k,l}| \le c\sigma_g, \\ \dfrac{c\sigma_g}{|r_{k,l}|} & \text{otherwise}, \end{cases} \tag{20}$$

where c is a positive constant to distinguish outliers from inliers. In our algorithm, c is set to 2. The extent of $2\sigma_g$ includes nearly 95% of the whole pixels. If most of the inliers are assumed to be within this range, the rest of them should be outliers. In this case, lower weights are assigned to decay the effect of these outliers. Besides, the selection of Gaussian deviation σ_g is not fixed. It is automatically determined in each iteration t. Köhler *et al.* [21] proposed a method to estimate σ_g by using the median absolute deviation (MAD) [29] and the MAD is derived from the weighted median operator. In our algorithm, this method is used to estimate σ_g adaptively, which can be expressed as

$$\begin{aligned} \sigma_g^t &= \sigma_0 \cdot MAD(\mathbf{r}^{t-1}|\boldsymbol{\beta}^{t-1}) \\ &= \sigma_0 \cdot MED\Big(\big|\mathbf{r}^{t-1} - MED(\mathbf{r}^{t-1}|\boldsymbol{\beta}^{t-1})\big|\Big|\boldsymbol{\beta}^{t-1}\Big), \end{aligned} \tag{21}$$

where $\boldsymbol{\beta}$, $\mathbf{r}$ are the confidence matrix and residual matrix assembled from $\boldsymbol{\beta}_k$ and $\mathbf{r}_k$ respectively with the column order. Besides, σ_0 is a constant scale factor, which depends on the distribution of the residual $\mathbf{r}$. In our method, σ_0 is set to 1.4826.



(a) (b)

Fig. 5. Error norms. (a) Half-quadratic estimation $f(x, \alpha)$ with different α values, (b) Their corresponding derivative functions.

Additionally, the parameter α_k can be adaptively chosen. In general, accurate registration of sub-pixel displacement and estimation of the Point Spread Function (PSF) are difficult to achieve in real applications. For each LR frame, the accuracy level of the PSF estimation and the registration may be different. The frame with large residual error should have less contribution to the final recovered HR image. In contrast, if the frame has small observation error, it should have more contribution to the final result. Thus, the parameter α_k should be adaptively determined according to the observation error of each LR frame.

Fig. 5 shows the performance of the half-quadratic estimation with respect to some different α values. When the parameter α tends to 0, the half-quadratic function performs like L_1 norm. With increasing parameter α, the adaptive error norm performs gradually close to L_2 norm. In order to define the accuracy level of each LR frame, the averaged observation error $\overline{r_k}$ shown in Equ. (18) is used. In general, $\overline{r_k}$ has a small value when the estimation of HR image is accurate. In this case, the observation error fits the Gaussian distribution. The parameter a_k should be large to perform like L_2 norm. In contrast, for those LR frames with outliers and mis-registrations, $\overline{r_k}$ is large. The parameter a_k should be small to perform like L_1 norm which can suppress these kinds of errors. Consequently, the parameter α_k should be positive and inversely proportional to $\overline{r_k}$, which is defined as

$$\alpha_k = \frac{\max(\overline{r_k})}{\overline{r_k}}. \tag{22}$$

C. Proposed Adaptive BTV Regularization Term

The traditional regularization terms such as Tikhonov and TV family can not distinguish edges. Therefore, although the noise is eliminated, the texture is suppressed as well, which limits the performance of these traditional regularization methods.

In our proposed regularization term, we introduce an adaptive weight matrix $\mathbf{W_G}$ with the same size as the HR image. Therefore, every element in BTV regularization is controlled by a relevant weight. The adaptive BTV regularization term can be expressed as

$$\Upsilon(\mathbf{X}) = \mathbf{W_G} \sum_{n=-P}^{P} \sum_{m=-P}^{P} a^{|m|+|n|} \left\| \mathbf{X} - \mathbf{S}_x^n \mathbf{S}_y^m \mathbf{X} \right\|_1, \tag{23}$$

(a) (b) (c)

Fig. 6. Visualization of the adaptive weight $\mathbf{W_G}$ for BTV regularization in different iterations. (a) Ground truth, (b) In the first iteration, (c) In the last iteration.

where P is a control parameter and a is a scaled weight with the range of $0 < a < 1$, which controls the decaying effect to the summation. In our experiment, a and P are set to 0.7 and 2 respectively. The weight matrix $\mathbf{W_G}$ should have small values in edge areas to preserve the detailed information. In order to extract image edges, the gradient operator is used in our algorithm, which can be formulated as

$$\mathbf{G} = \sqrt{\mathbf{I}_x^2 + \mathbf{I}_y^2}, \tag{24}$$

where $\mathbf{I}_x$ and $\mathbf{I}_y$ are the first-order gradients of the estimated HR image in the vertical and horizontal directions respectively. $\mathbf{G}$ is the gradient operator which extracts edges from the HR image. From the above analysis, $\mathbf{W_G}$ should be inversely proportional to $\mathbf{G}$. Thus, we define the weight matrix $\mathbf{W_G}$ as

$$\mathbf{W_G} = \frac{1}{w + \sqrt{\frac{\mathbf{G}}{G_{max}}}}, \tag{25}$$

where w is a positive constant as a tuning parameter to adjust the extent of $\mathbf{W_G}$. The default value of w is 0.5 in our algorithm. G_{max} represents the maximum value of $\mathbf{G}$ and $\frac{\mathbf{G}}{G_{max}}$ scales $\mathbf{G}$ to the range of 0 to 1. The square root operator is used to extend the difference among the values.

The weight matrix $\mathbf{W_G}$ is adaptively updated in every iteration according to the recent estimation of HR image. Fig. 6 shows the visualization of the adaptive weight $\mathbf{W_G}$ in the first and last iterations. The adaptive weight has a small value in image texture areas and a large value in other areas to preserve the edges and eliminate noises adaptively. Due to the blurring effect, the extraction of image texture is inaccurate at first. But after some iterations, the blurring effect is gradually eliminated so that the edge extraction becomes progressively more accurate.

Compared to the traditional regularization, the proposed adaptive BTV regularization term preserves image edges and suppresses noise simultaneously according to a weight matrix based on a gradient operator.

D. Framework of Our Proposed Algorithm

In our proposed framework, the spatially weighted fidelity term based on half-quadratic estimation and the proposed adaptive BTV (ABTV) regularization term are combined to estimate the HR image from a sequence of LR frames. Fig. 7 shows the whole framework of our proposed multi-frame SR algorithm.

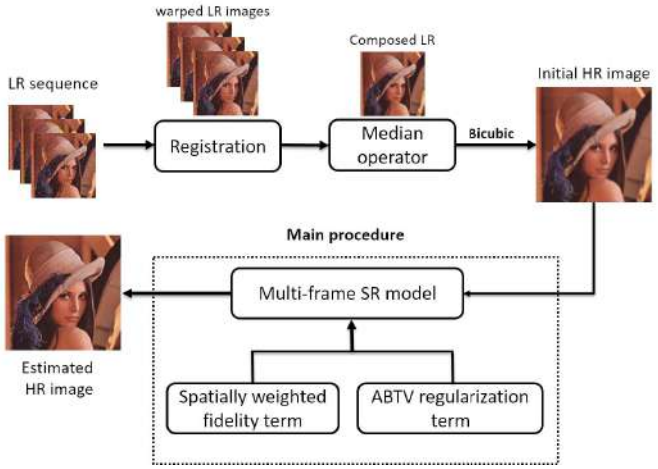


Fig. 7. Framework of proposed multi-frame super-resolution algorithm.

The initial step of our proposed algorithm mainly generates a composed LR image as a reference to initialize the HR image $\mathbf{X}$. The composed LR image has less outliers and more detailed information than any original LR image. The main procedure contains the spatially weighted fidelity term and the proposed ABTV regularization term. The final estimated HR image is reconstructed when the objective function has the minimum value. Therefore, the recovered HR image is formulated as

$$\hat{\mathbf{X}} = arg\min_{\mathbf{X}} \left\{ \sum_{k=1}^{K}\sum_{l=1}^{L} \beta_{k,l} \cdot \alpha_k \sqrt{\alpha_k^2 + r_{k,l}^2} + \lambda \cdot \mathbf{W_G} \sum_{n=-P}^{P}\sum_{m=-P}^{P} a^{|m|+|n|} \left\| \mathbf{X} - \mathbf{S}_x^n \mathbf{S}_y^m \mathbf{X} \right\|_1 \right\}, \tag{26}$$

where λ is the trade-off parameter to control the balance between the fidelity term and regularization term. Since the first-order gradient of objective minimization function $f(x)$ is needed in the optimization step, the expression of the first-order derivative function with respect to $\mathbf{X}$ is calculated as

$$f'(\mathbf{X}) = \sum_{k=1}^{K} \boldsymbol{\beta}_k \cdot \alpha_k \frac{\mathbf{W}_k^T(\mathbf{W}_k\mathbf{X} - \mathbf{Y}_k)}{\sqrt{\alpha_k^2 + (\mathbf{W}_k\mathbf{X} - \mathbf{Y}_k)^2}} + \lambda \mathbf{W_G} \times \sum_{n=-P}^{P}\sum_{m=-P}^{P} a^{|m|+|n|}(\mathbf{I} - \mathbf{S}_y^{-m}\mathbf{S}_x^{-n}) sign(\mathbf{X} - \mathbf{S}_x^n \mathbf{S}_y^m \mathbf{X}), \tag{27}$$

where $f'(\mathbf{X})$ denotes the derivative of $f(\mathbf{X})$. $\mathbf{I}$ is an identity matrix. $\mathbf{S}_x^{-n}$ and $\mathbf{S}_y^{-m}$ are the transposes of matrices $\mathbf{S}_x^n$ and $\mathbf{S}_y^m$. They shift image $\mathbf{X}$ in the opposite directions as $\mathbf{S}_x^n$ and $\mathbf{S}_y^m$ do respectively.

The quality of estimated HR image is gradually improved with the increase of iterations. For the multi-frame SR problem, there are many optimization methods to solve the minimization problem such as Steepest Decent (SD) [30] and Conjugate Gradient (CG) [31]. In our proposed algorithm, we use Scaled Conjugate Gradient (SCG) to solve the function expressed in Equ. (26). Compared with other optimization

Algorithm 1 Proposed Multi-Frame Image Super-Resolution

Initialization:

Set initial HR image $\mathbf{X}_0$ from median operator based initialization method;

Set $\boldsymbol{\alpha}_0 = 1$, $\boldsymbol{\beta}_0 = 1$, outer-loop iteration $t_1 = 1$;

Set the SCG terminal criterion parameter $\eta = 10^{-4}$.

while the terminal criterion expressed in (28) is not fulfilled and the outer-loop iteration $t_1 \leq T_{max}$ **do**:

- Compute σ_g^t from $\mathbf{X}^{t-1}$ and $\boldsymbol{\beta}^{t-1}$ according to (21)
- Compute $\overline{r_k}^{t-1}$ from r_k^{t-1} according to (18)
- Compute $(\boldsymbol{\beta}_k^{global})^t$ from $\overline{r_k}^{t-1}$ according to (19)
- Compute $(\boldsymbol{\beta}_k^{local})^t$ from σ_g^t and r_k^{t-1} according to (20)
- Compute $\boldsymbol{\alpha}^t$ from $\overline{r_k}^{t-1}$ according to (22)
- Compute $\mathbf{W_G}$ from $\mathbf{G}$ according to (25)
- Set $t_2 = 1$ and $\mathbf{X}^t = \mathbf{X}^{t-1}$
- **while** the terminal criterion expressed in (28) is not fulfilled and the inner-loop iteration $t_2 \leq T_S$ **do**:
  1) update $\mathbf{X}^t$ according to (26) using SCG iteration
  2) Set $t_2 = t_2 + 1$
- **end while**
- Set $t_1 = t_1 + 1$

end while

Output the estimated HR image $\mathbf{X}$

methods, the SCG can adaptively adjust the step size according to the approximative speed in the gradient direction [32]. Moreover, the convergence of SCG is faster than SD and CG due to the scaled step size. The SCG optimization is terminated if the maximum iteration is reached or the maximum absolute difference between $\mathbf{X}^t$ and $\mathbf{X}^{t-1}$ is lower than the terminal parameter η, which can be expressed as

$$\max_{i=1,\cdots,N} |\mathbf{X}_i^t - \mathbf{X}_i^{t-1}| < \eta, \tag{28}$$

where η is set to 10^{-4}. The proposed multi-frame SR algorithm can be summarized in Algorithm 1.

V. Experimental Results and Analysis

In this section, we conduct experiments to test the performance of our proposed multi-frame SR algorithm. we first compare the quality of our initial HR image with other traditional methods. Then the performance of spatially weighted half-quadratic estimation is shown in detail. Furthermore, the proposed method and other 7 multi-frame SR methods are used to estimate the HR images from two kinds of LR frames. One kind of the LR frames are synthetically generated from known HR images and the other kind is directly photographed from the low-resolution cameras [33].

A. Experimental Setup

In practice, the performance of SR methods cannot be numerically evaluated since the ground truth HR images are always not available. In our experiment, the proposed method is first tested on synthetic data since the ground truth images for synthetic LR frames are available. We generate 16 LR frames from one HR image and the displacement of every frame is simulated as a rigid motion. Therefore, the HR image is displaced by uniform distributed random translations and rotations. The range of random translations is from -2 to 2 pixels and the rotation angles are randomly changing from -1° to 1°. Then the displaced HR frames are blurred by a 4×4 Gaussian kernel with $\sigma = 0.4$ and subsampled with factor r. Gaussian noise and Salt&Pepper noise are added in the simulated LR sequence simultaneously as mixed noises to increase the difficulty of accurate estimation. The variance of additive Gaussian noise and Salt&Pepper noise are both set to 0.02. Besides, the PSNR and SSIM are used to measure the quality of our estimated HR images.

TABLE I

Average PSNR/SSIM Results for Synthetic Data (Generated From Set 5 and Set 14) With Respect to Different λ Parameters Under Ratio = 2

Data \ λ	0.0003	0.0004	0.0005	0.0006	0.0007
Set 5	34.0950	34.6153	**34.8288**	34.6778	34.5862
	0.9698	0.9740	**0.9755**	0.9754	0.9750
Set 14	32.0116	32.0723	**32.1782**	31.7661	31.5103
	0.9493	0.9508	**0.9515**	0.9474	0.9450

In our experiments, the first frame of LR sequence is chosen as our reference frame and the initial HR image is obtained by our novel initialization method. The intensity range of our test images is set to [0, 1]. For color images, we use RGB model as our color model and apply our algorithm to all channels. The regularization parameter λ is a vital parameter used to balance the fidelity term and the regularization term. The value of λ is determined empirically based on numerous experiments. In our experiments, we first generated many synthetic data by using Set 5 [34] and Set 14 [35] datasets. Then, the most appropriate λ value can be found by identifying the value that produces the best performance. Table I demonstrates that when the λ is set to 0.0005, the proposed method has the best performance in terms of PSNR/SSIM values. Therefore, λ is set to 0.0005 in our experiments. SCG is used to minimize the objective function. The termination criterion is set to $\eta = 10^{-4}$. Since the minimization function usually converges within 25 iterations, the maximum iteration number is set to $T_S = 25$. In general, the sampling factor is commonly set to 2. Therefore, we present the simulated images with $r = 2$ in our expriments. For the images with higher zoom factors such as 3 and 4, we present them in real data experiments. For motion estimation, the Enhanced Correlation Coefficient (ECC) [36], [37] is used to estimate the subpixel movement between two LR frames.

B. Quality of the Estimated Initial HR Image

In our proposed algorithm, we use a novel method to estimate the initial HR image. Different from the traditional methods that only initialize the HR image from one LR frame, our proposed method estimates the initial HR image based on more comprehensive information by considering the warped

TABLE II

PSNR AND SSIM RESULTS FOR SET 5 AND SET 14 WITH DIFFERENT INITIALIZATIONS UNDER GAUSSIAN NOISE AND SALT&PEPPER NOISE (RATIO = 2)

Noise	Bilinear		Bicubic		Our initialization	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set 5						
Gaussian	26.6135	0.8682	26.0221	0.8415	**28.7267**	**0.9320**
Salt&Pepper	23.7867	0.8419	22.4617	0.8229	**29.0668**	**0.9434**
Set 14						
Gaussian	24.8262	0.7881	24.5465	0.7705	**26.1077**	**0.8465**
Salt&Pepper	22.9991	0.7663	22.0305	0.7583	**26.2036**	**0.8564**

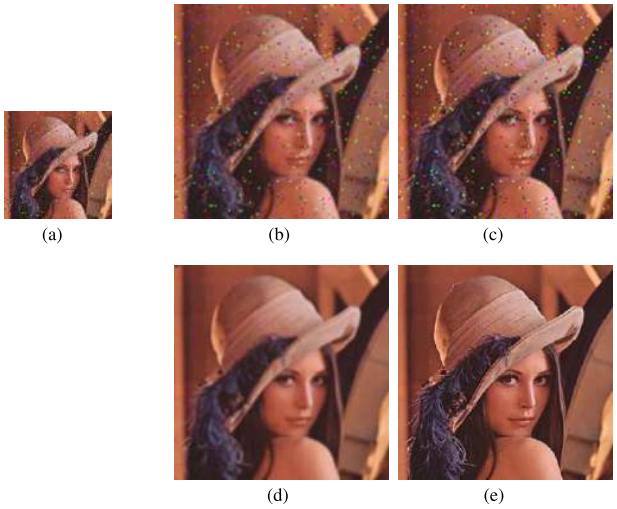
(a) (b) (c) (d) (e)

Fig. 8. The comparison of different initial estimations under Gaussian noise. (a) First frame of LR sequence. (b) Bilinear interpolation. (c) Bicubic interpolation. (d) Our initialization. (e) Original Lena image.

LR sequence. The LR frames are first warped according to the shape of the reference LR image. Then a median operator is used to eliminate the noise and reconstruct a composed LR image. Finally, the initial HR image is obtained by bicubic interpolation.

We use the synthetic data generated by Set 5 and Set 14 datasets to illustrate the superiority of the proposed initialization method. Bilinear and bicubic interpolations are used for comparisons. The way to generate the synthetic data is consistent with Section V-A except that the Gaussian noise with a variance of 0.05 and Salt&Pepper noise with a variance of 0.02 are added respectively. Table II gives the values of PSNR and SSIM on two datasets for different initializations with noise added. Table II demonstrates that our proposed initialization has higher PSNR and SSIM values than the conventional interpolation methods, which shows the robustness from leveraging all LR frames to compensate noise effects.

Fig. 8 and Fig. 9 show the visual comparison of our novel initialization, bilinear interpolation and bicubic interpolation under Gaussian noise and Salt&Pepper noise respectively. From the two figures, the conventional interpolation methods have limited performance to suppress the noise effects. Conversely, our novel initialization is quite robust to noises and does not introduce any unnatural artifact from the median



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

Fig. 9. The comparison of different initial estimations under Salt&Pepper noise. (a) First frame of LR sequence. (b) Bilinear interpolation. (c) Bicubic interpolation. (d) Our initialization. (e) Original Lena image.

based noise compensation. Generated by our proposed initialization method, the initial HR image has superior quality that helps the subsequent reconstruction steps.

C. Performance Analysis

In this section, we discuss and justify the concrete improvement of our innovation terms including the novel initialization method (INIT), the spatial weighted half-quadratic estimation (SWHQ) and the adaptive BTV regularization (ABTV). All the three innovation terms are stepwise added to the basic multi-frame SR framework, which uses the bicubic interpolation as its initialization, half-quadratic estimation as its fidelity term and BTV as its regularization. We generated five image sequences by using Set 5 dataset and named them with $Baby$, $Bird$, $Butterfly$, Kid, $Woman$ to test the performance of each innovation term.

Table III gives the performance comparison in terms of PSNR and SSIM. In the table, all the methods use the same parameter setting as determined in Section V-A For the $Bicubic + HQ + BTV$ and $INIT + HQ + BTV$ methods, the concrete improvement of our initialization method is justified since the only difference between them is their initializations. The $INIT + SWHQ + BTV$ method improves the image quality by 0.51 dB over the $INIT + HQ + BTV$ method in PSNR. With the comparison of $INIT + HQ + ABTV$ and $INIT + HQ + BTV$ methods, the average PSNR improvement is 0.44 dB by replacing the BTV regularization with the ABTV regularization. The total average improvement of the $INIT + SWHQ + ABTV$ method is 0.71 dB over the $INIT + HQ + BTV$ method. Furthermore, we use the bicubic interpolation as the initialization method to analyse the performance of the SWHQ and ABTV terms without good initialization. In this case, the $Bicubic + SWHQ + ABTV$ method has an average improvement of 3.10 dB in PSNR over the $Bicubic + HQ + BTV$ method, which confirms the good

TABLE III
PSNR AND SSIM RESULTS TO ILLUSTRATE THE GOOD PERFORMANCE OF OUR PROPOSED INNOVATION TERMS

Images	Bicubic+HQ+BTV	Bicubic+SWHQ+ABTV	INIT+HQ+BTV	INIT+SWHQ+BTV	INIT+HQ+ABTV	INIT+SWHQ+ABTV
Baby	30.5946	35.7190	35.9614	36.5230	36.4624	**36.8254**
	0.9281	0.9727	0.9748	0.9781	0.9783	**0.9792**
Bird	32.9675	35.6467	35.8080	36.7167	36.9949	**37.2389**
	0.9675	0.9851	0.9859	0.9904	0.9909	**0.9913**
Butterfly	28.9079	32.8839	32.4935	32.9695	32.5938	**32.9770**
	0.9706	0.9890	0.9878	0.9895	0.9886	**0.9895**
Kid	30.6584	31.9433	32.0856	32.1556	32.1685	**32.2456**
	0.8940	0.9273	0.9291	0.9311	0.9317	**0.9324**
Woman	31.7692	34.1884	34.2496	34.7908	34.5833	**34.8825**
	0.9647	0.9810	0.9804	0.9851	0.9850	**0.9856**
Average	30.9795	34.0763	34.1196	34.6311	34.5606	**34.8339**
	0.9450	0.9710	0.9716	0.9748	0.9749	**0.9756**

Fig. 10. Ten images for synthetic data. (a) Motorbike. (b) Butterfly. (c) Lena. (d) Hat. (e) Parrot. (f) Raccoon. (g) Cameraman. (h) House. (i) Pepper. (j) Barbara.

performance of the SWHQ fidelity term and the ABTV regularization.

D. Experiments on Synthetic Data

In our experiments, 10 commonly used HR images are chosen to generate our synthetic LR sequences, which are shown in Fig. 10. For all of our synthetic data, the mixed noises are added. Table IV demonstrates the quantitative comparisons of 8 different algorithms using PSNR and SSIM metrics. These 8 algorithms are bicubic interpolation, L_2 + Tikhonov [25], L_2 + BTV [17], L_1 + BTV [17], the deep draft-ensemble learning based SR (DeepSR) [24], the Bilateral Edge Preserving (BEP) algorithm [20], the Iteratively Re-weighted (IRW) algorithm [21] and our proposed algorithm. For DeepSR, the proposed initialization is used to eliminate noise effect in LR frames since deep learning based SR methods usually do not consider the effect of mixed noises on synthetic data. From Table IV, our proposed method has the highest PSNR and SSIM values in most cases, which validates the superiority of our algorithm.

Except for numerical comparison, Fig. 11, Fig. 12 and Fig. 13 present the visual comparison of our algorithm and other 7 algorithms. From these figures, the L_2 based methods such as $L_2+Tikhonov$ and L_2+BTV have poor performance to suppress the Salt&Pepper noise and registration errors. On the contrary, L_1+BTV is more robust to them. Therefore, compared with $L_2 + Tikhonov$, the results of $L_1 + BTV$ have better quality in most cases. Benefit from using the adaptive error norm in its fidelity and regularization terms, the BEP algorithm has ability to suppress both Salt&Pepper and Gaussian noise. Our experiments show that the results of BEP method are better than the traditional fixed-norm methods, but some noises still exist due to its simple initial estimation. In general, CNN and RNN based methods are quite robust to scaling and translation, but not robust to noise and rotation. One reason is that it is difficult to contain all noise degrees and rotation angles in their training dataset due to its finite size. In our experiments on synthetic data, mixed noises and rotations are used to make the SR problem more challenging, which makes the CNN based DeepSR method

TABLE IV
PSNR/SSIM RESULTS OF MULTI-FRAME SUPER-RESOLVED IMAGES FROM 8 DIFFERENT ALGORITHMS UNDER RATIO = 2 AND MIXED NOISES

Images	Bicubic	L2+Tikhonov	L2+BTV	L1+BTV	DeepSR	BEP	IRW	Proposed
Motorbike	20.6443	25.1011	25.1625	28.1603	27.6493	28.7804	29.8364	**30.1432**
	0.7545	0.8714	0.8834	0.9247	0.9225	0.9495	0.9602	**0.9636**
Butterfly	21.2965	25.9730	27.6616	28.9103	28.9352	30.4342	32.1017	**32.7968**
	0.8938	0.9432	0.9653	0.9674	0.9772	0.9820	0.9883	**0.9893**
Lena	22.5737	27.3270	29.4337	29.6870	31.3271	30.7712	33.0384	**33.4546**
	0.8448	0.9094	0.9409	0.9426	0.9631	0.9596	0.9757	**0.9768**
Hat	22.5973	27.1948	29.5654	30.0563	31.3314	31.6839	33.1078	**33.3577**
	0.8702	0.9107	0.9632	0.9436	0.9731	0.9750	0.9820	**0.9834**
Parrot	22.2565	27.2338	29.7220	29.9128	33.1376	32.0377	34.4372	**34.4912**
	0.8318	0.8585	0.9363	0.9053	0.9710	0.9570	**0.9766**	0.9758
Raccoon	22.5195	27.2443	27.7986	31.1457	28.2073	30.3571	32.0525	**32.5648**
	0.7960	0.8771	0.8895	0.9467	0.9156	0.9468	0.9628	**0.9665**
Cameraman	21.7009	25.6756	25.2985	27.3820	29.0710	28.7688	29.9739	**30.9628**
	0.6222	0.6152	0.7566	0.8427	0.8851	0.8722	0.8888	**0.9106**
House	23.6665	28.2798	31.6099	30.9898	33.6750	32.8833	35.0856	**35.7729**
	0.6435	0.6215	0.8213	0.7246	0.8915	0.8638	0.9077	**0.9090**
Pepper	22.2396	25.7537	29.1594	29.3307	31.1285	31.0188	31.9347	**32.1581**
	0.6708	0.6586	0.8301	0.7825	0.8897	0.8940	**0.9082**	0.9071
Barbara	21.8615	25.2559	25.2462	28.5042	27.4903	27.1066	29.3687	**32.4591**
	0.6371	0.6920	0.7411	0.8573	0.8437	0.8419	0.9088	**0.9277**
Average	22.1356	26.5039	28.0658	29.4079	30.1953	30.3842	32.0937	**32.8161**
	0.7565	0.7958	0.8728	0.8837	0.9233	0.9242	0.9459	**0.9510**

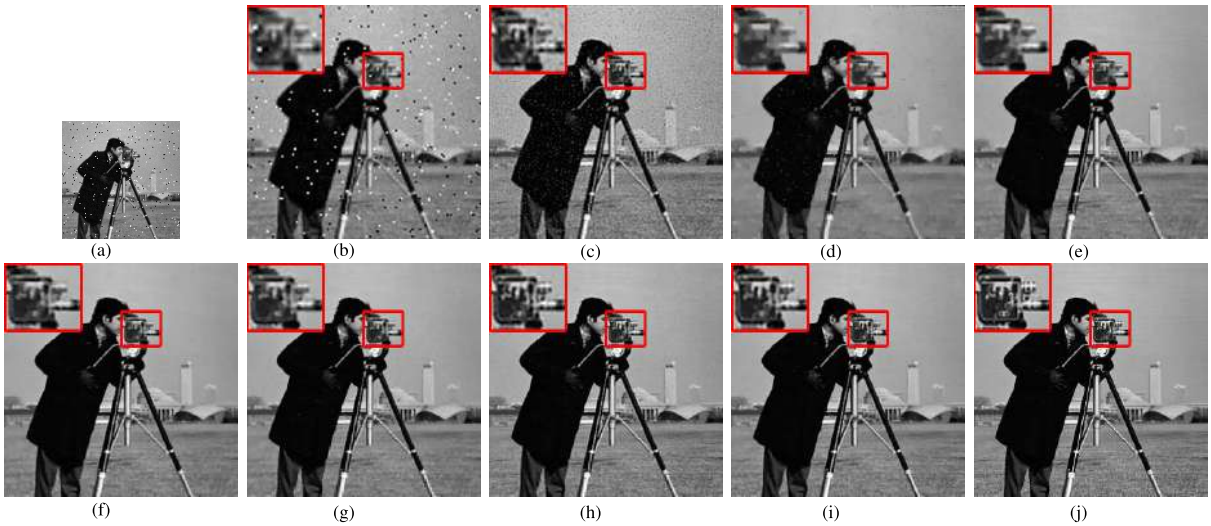


Fig. 11. Visual comparison of multi-frame super-resolved results from different methods for *Cameraman* image with mixed noises. (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed. (j) Ground truth.

not perform well. As an iteratively re-weighted SR method, the IRW algorithm has the best performance of suppressing mixed noises and preserving image edges compared with the existing SR algorithms.

Therefore, we will only describe the visual comparison between our proposed algorithm and the IRW algorithm. For all the places in the test images, our algorithm performs better or equally well. For example, in *Cameraman* image, some noises remain in background area for the IRW algorithm. However, our algorithm eliminates most of the noises and smooths the background. For *Barbara* image, the IRW algorithm has many artifacts on the scarf. In contrast, our algorithm recovers the detailed information correctly. For *Motorbike* image, above the brand 'CR', our algorithm generates the three vertical lines more clearly than the IRW algorithm.

E. Computational Complexity Analysis

In this section, we perform the detailed analysis of the computational complexity of our proposed method and the

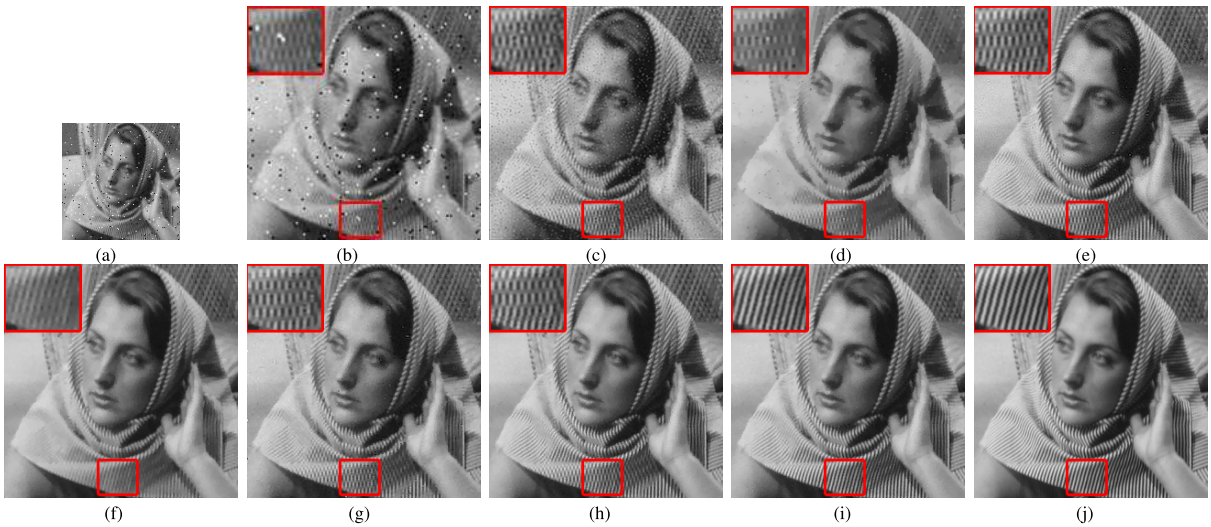


Fig. 12. Visual comparison of multi-frame super-resolved results from different methods for *Barbara* image with mixed noises. (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed. (j) Ground truth.

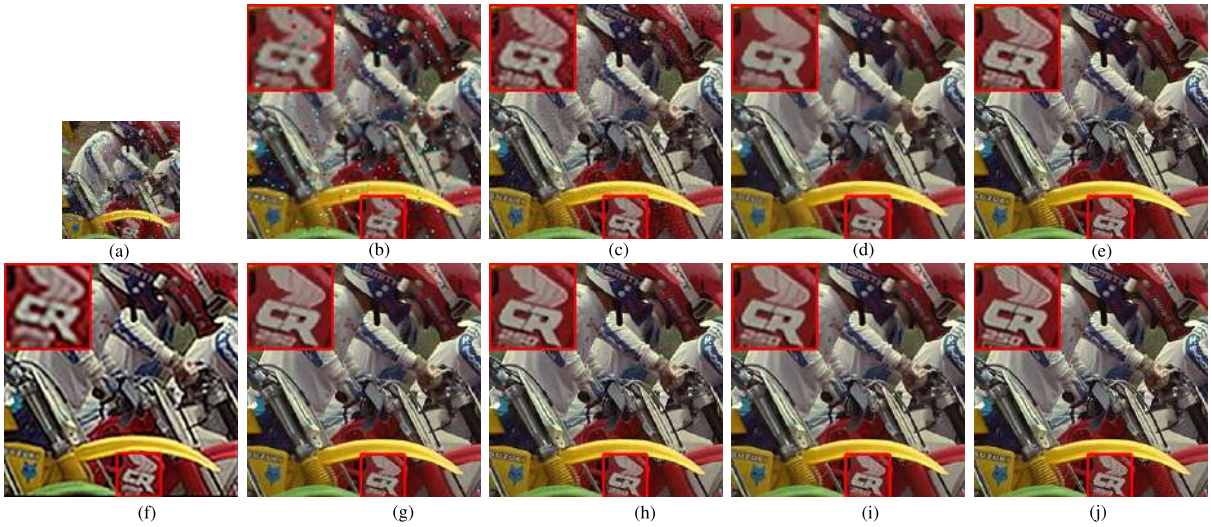


Fig. 13. Visual comparison of multi-frame super-resolved results from different methods for *Motorbike* image with mixed noises. (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed. (j) Ground truth.

TABLE V
COMPUTATIONAL COMPLEXITY ANALYSIS OF OUR PROPOSED METHOD AND THE COMPARED METHODS

Method	Bicubic	L2+Tikhonov	L2+BTV	L1+BTV	BEP	IRW	DeepSR	Proposed
Time (s)	0.0061	34.6039	18.8070	24.8449	111.9170	96.8499	544.7716	46.8632

compared methods in terms of run-time. The run-time evaluation was implemented on a laptop computer with the Intel i7-4710HQ CPU and NIVIDIA GeForce GTX 860M GPU established on the Matlab environment. The test LR sequence has 16 frames, each of which is a color image with size of 128×128×3 and the SR ratio is set to 2. Table V gives the run-time complexity of our proposed method and the compared methods. Compared with the state-of-the-art SR methods such as the DeepSR, BEP and IRW, the proposed method has lower computational complexity since our good initialization method can accelerate the convergence of the objective minimization function. DeepSR has the highest computational complexity since multiple SR drafts have to be generated to obtain the final result. The $L_2 + Tikhonov$, $L_2 + BTV$ and $L_1 + BTV$ methods have lower computational complexity than our proposed method. However, the quality of our super-resolved images is much better than theirs in terms of both visual evaluation and PSNR/SSIM results.

F. Robustness Analysis

In this section, we analyse and justify the robustness of our proposed method when the input LR sequences are corrupted

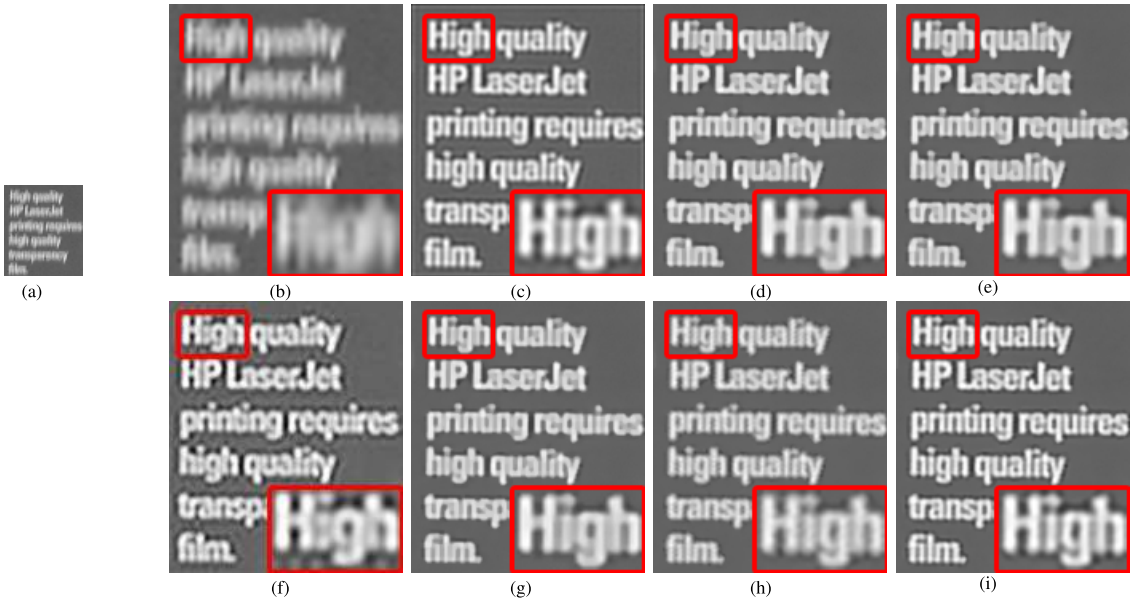


Fig. 14. Visual comparison of multi-frame super-resolved results from different algorithms for *text* frames (r = 3). (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov. (d) L2 + BTV (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed.

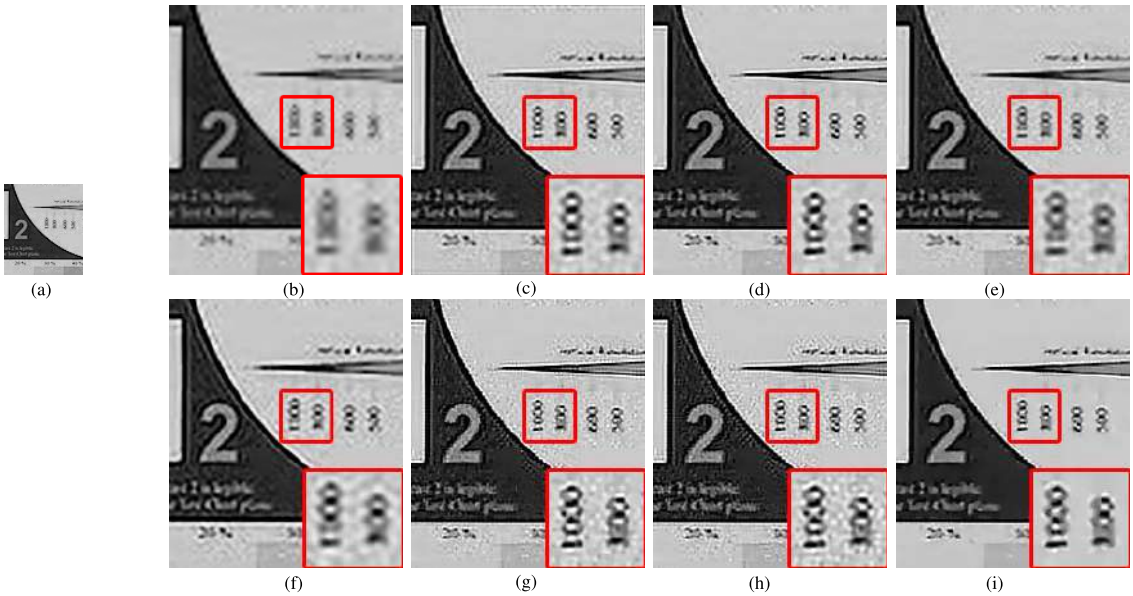


Fig. 15. Visual comparison of multi-frame super-resolved results from different algorithms for *adyoron* frames (r = 3). (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov. (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed.

with different degrees of blurring and noising respectively. The other parameters have the same values as the previous experiments. Two robust SR methods, the BEP and IRW methods, are compared with our proposed method. The test synthetic data is generated by $Bird$ image from Set 5. Table VI demonstrates the SR results with different blurring parameter σ from 0.2 to 0.6. Although the quality of estimated SR image gradually decreases when the σ increases, the other two robust methods show larger reduction. Our proposed method has the best performance of robustness under blur corruption in terms of PSNR and SSIM results compared with the others. In Table VII, we use mixed noises including Gaussian noise and Salt&Pepper noise to test the performance of the three robust methods under noise corruption. The variance of mixed noises is set from 0.01 to 0.05 respectively. Compared with the other two methods, our proposed method has the best performance of suppressing the mixed noises, which demonstrates the robustness of our method in terms of noise corruption.

G. Experiments on Real Data

In this section, we use real data to test our proposed algorithm in practical applications. The real data obtained from Multi-Dimensional Signal Processing Research Group (MDSP) [33] is the most widely used dataset to test

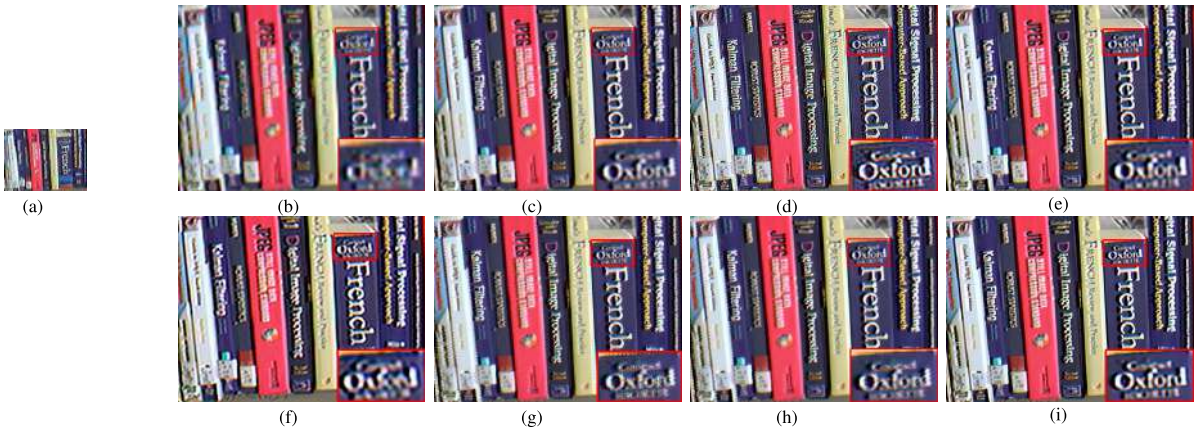


Fig. 16. Visual comparison of multi-frame super-resolved results from different algorithms for *book* frames (r = 3). (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov. (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed.

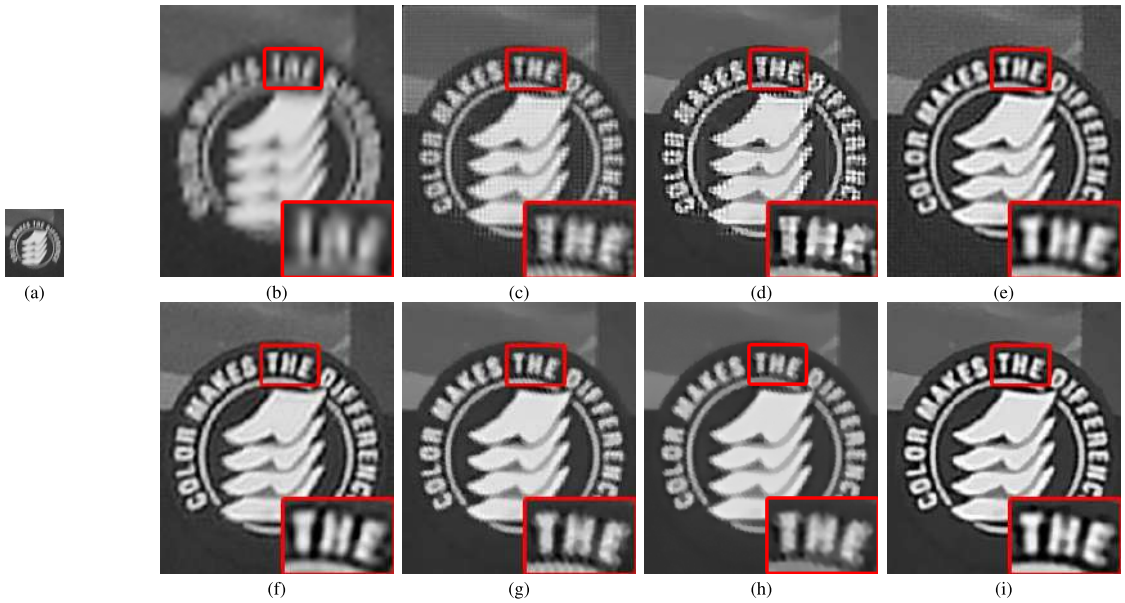


Fig. 17. Visual comparison of multi-frame super-resolved results from different algorithms for *disk* frames (r = 4). (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov. (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed.

TABLE VI
PSNR/SSIM RESULTS WITH DIFFERENT BLURRING DEGREES

Method \ σ	0.2	0.3	0.4	0.5	0.6
BEP	34.4817	35.3082	33.7265	33.8097	32.1872
	0.9868	0.9878	0.9834	0.9846	0.9798
IRW	37.5326	36.4166	36.1033	36.0561	34.9902
	0.9918	0.9899	0.9899	0.9900	0.9879
Proposed	**38.2315**	**37.2075**	**37.1791**	**36.2093**	**35.0945**
	0.9932	**0.9919**	**0.9912**	**0.9909**	**0.9893**

TABLE VII
PSNR/SSIM RESULTS WITH DIFFERENT NOISE DEGREES

Method \ Variance	0.01	0.02	0.03	0.04	0.05
BEP	35.5163	33.1350	32.3433	29.5547	29.0386
	0.9861	0.9781	0.9688	0.9492	0.9459
IRW	38.9311	36.1072	34.3216	32.5285	30.9014
	0.9945	0.9900	0.9837	0.9759	0.9641
Proposed	**39.3273**	**37.1177**	**34.9488**	**33.8386**	**32.8755**
	0.9952	**0.9912**	**0.9865**	**0.9815**	**0.9752**

the performance of multi-frame SR methods. Since this data set is shot by real camera, there are no ground truth images. Therefore, the image assessment matrices such as PSNR and SSIM can not be used to evaluate the quality of real images. In this paper, we only use visual comparison to assess image quality for real data. Moreover, for real data, the PSF kernel is unknown. To simplify this blind deblurring problem, we assume the unknown PSF kernel is a 4×4 Gaussian kernel

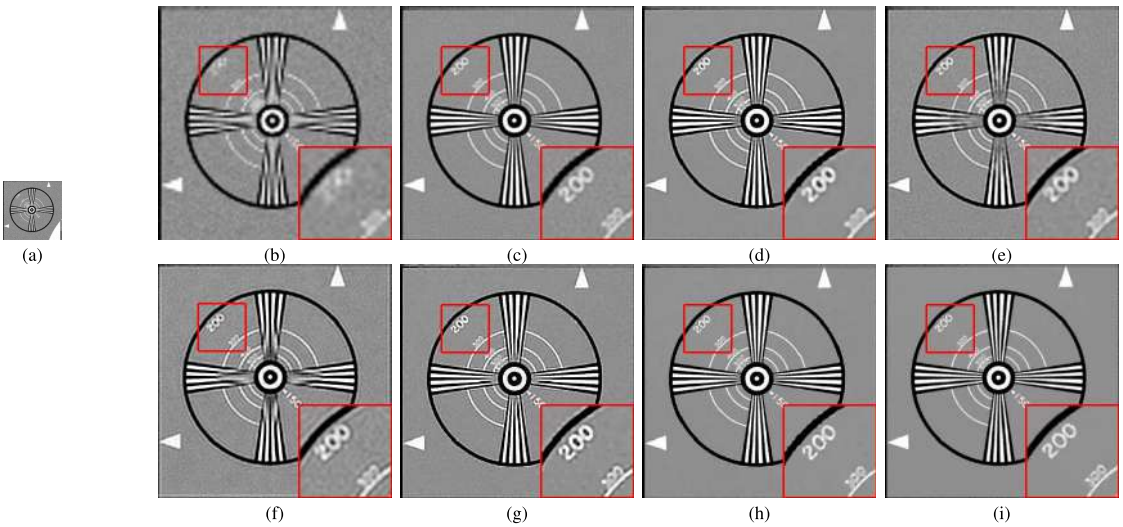


Fig. 18. Visual comparison of multi-frame super-resolved results from different algorithms for *EIA* frames (r = 4). (a) Reference LR image. (b) Bicubic interpolation. (c) L2 + Tikhonov. (d) L2 + BTV. (e) L1 + BTV. (f) DeepSR. (g) BEP. (h) IRW. (i) Proposed.

with $\sigma = 0.4$. This PSF kernel is determined empirically based on numerous experiments via visual comparison.

For the real data, large scale factors such as 3 and 4 are used to reconstruct our HR images. Fig. 14, Fig. 15 and Fig. 16 show the visual comparison of super-resolved results of *text*, *adyoron* and *book* respectively from different multi-frame SR methods with the scale factor of 3. In Fig. 17 and Fig. 18, the results of *disk* and *EIA* are presented with the scale factor of 4. Compared with other 7 methods, our estimated HR images can effectively suppress the errors caused by noise, registration and bad estimation of unknown PSF kernels. Besides, the detailed information in real images is preserved well, which shows that our super-resolved images have better quality in visual comparison.

VI. Conclusions

In this paper, we proposed a robust multi-frame image SR algorithm based on spatially weighted half-quadratic estimation and adaptive BTV regularization. A novel initial method based on median operator is used to generate an outlier-insensitive HR image as the initial value. For the fidelity term, the half-quadratic estimation is introduced to choose norm adaptively instead of using fixed L_1 and L_2 norms. Besides, a spatial weight matrix is used as a confidence map to scale the result of half-quadratic estimation. For the regularization term, an adaptive regularization method based on bilateral total variation (BTV) is proposed to suppress image noise and preserve image edges simultaneously. Both the simulated data and real data are tested to evaluate the performance of the proposed method. The experimental results demonstrate that our method outperforms the state-of-the-art algorithms with better visual quality and higher values in quality metrics. As for future work, we are trying to extend our algorithm to blind SR.

References

[1] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Process.*, vol. 90, no. 2, pp. 405–414, 2010.

[2] C. Liu and D. Sun, "On Bayesian adaptive video super resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 346–360, Feb. 2014.

[3] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia, and E. Wu, "Handling motion blur in multi-frame super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5224–5232.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognit. Lett.*, vol. 30, no. 5, pp. 494–502, 2009.

[6] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1920–1927.

[7] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis. (ACCV)*. Cham, Switzerland: Springer, 2014, pp. 111–126.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[9] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, "Video super-resolution with convolutional neural networks," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 109–122, Jun. 2016.

[10] J. Caballero *et al.*, "Real-time video super-resolution with spatio-temporal networks and motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2848–2857.

[11] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. (2017). "Detail-revealing deep video super-resolution." [Online]. Available: https://arxiv.org/abs/1704.02738

[12] W. Yang, J. Feng, G. Xie, J. Liu, Z. Guo, and S. Yan, "Video super-resolution based on spatial-temporal recurrent residual networks," *Comput. Vis. Image Understand.*, vol. 168, pp. 79–92, Mar. 2017.

[13] D. Liu *et al.*, "Learning temporal dynamics for video super-resolution: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3432–3445, Jul. 2018.

[14] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, no. 2, pp. 317–339, 1984.

[15] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

[16] M.-C. Chiang and T. E. Boult, "Efficient super-resolution via image warping," *Image Vis. Comput.*, vol. 18, no. 10, pp. 761–771, 2000.

[17] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[18] V. Patanavijit and S. Jitapunkul, "A robust iterative multiframe super-resolution reconstruction using a Huber Bayesian approach with Huber-Tikhonov regularization," in *Proc. IEEE Conf. Int. Symp. Intell. Signal Process. Commun.*, Dec. 2006, pp. 13–16.

[19] L. Yue, H. Shen, Q. Yuan, and L. Zhang, "A locally adaptive L_1-L_2 norm for multi-frame super-resolution of images with mixed noise and outliers," *Signal Process.*, vol. 105, pp. 156–174, Dec. 2014.

[20] X. Zeng and L. Yang, "A robust multiframe super-resolution algorithm based on half-quadratic estimation with modified BTV regularization," *Digit. Signal Process.*, vol. 23, no. 1, pp. 98–109, 2013.

[21] T. Köhler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger, "Robust multiframe super-resolution employing iteratively re-weighted minimization," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 42–58, Mar. 2016.

[22] Y. Huang, W. Wang, and L. Wang, "Bidirectional recurrent convolutional networks for multi-frame super-resolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 235–243.

[23] Y. Huang, W. Wang, and L. Wang, "Video super-resolution via bidirectional recurrent convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1015–1028, Apr. 2018.

[24] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia, "Video super-resolution via deep draft-ensemble learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 531–539.

[25] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[26] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Overcoming registration uncertainty in image super-resolution: maximize or marginalize?" *EURASIP J. Adv. Signal Process.*, vol. 2007, p. 023565, Dec. 2007.

[27] G. Wolberg, *Digital Image Warping*, vol. 10662. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990.

[28] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[29] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, no. 4, p. 1071, 1988.

[30] R. Battiti, "First-and second-order methods for learning: Between steepest descent and Newton's method," *Neural Comput.*, vol. 4, no. 2, pp. 141–166, 1992.

[31] T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization," *SIAM J. Numer. Anal.*, vol. 20, no. 3, pp. 626–637, 1983.

[32] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Nov. 1993.

[33] S. Farsiu, *MDSP Super-Resolution and Demosaicing Datasets*. Accessed: Oct. 1, 2017. [Online]. Available: https://users.soe.ucsc.edu/ milanfar/software/sr-datasets.html

[34] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Guildford, U.K., 2012.

[35] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surfaces*. Berlin, Germany: Springer, 2010, pp. 711–730.

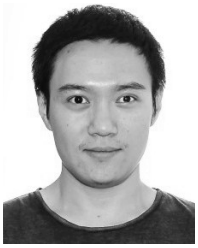
[36] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.

[37] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1858–1865, Oct. 2008.

Xiaohong Liu (S'18) received the B.E. degree in telecommunication engineering from Southwest Jiaotong University, China, and the M.A.Sc. degree in electrical and computer engineering from the University of Ottawa, Canada, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, McMaster University. His research interests include image and video processing, computer vision, and machine learning.

Lei Chen (S'15) received the B.E. and M.E. degrees in electrical engineering from Shandong University, China, in 2010 and 2013, respectively.

He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, University of Ottawa. His research interests are on image and video processing.



Wenyi Wang received the B.S. degree from Wuhan University, China, in 2009, and the M.S. and Ph.D. degrees from the University of Ottawa, Canada, in 2011 and 2016, respectively. Since 2017, he has been a Faculty Member with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include computer vision, pattern recognition, and video processing.

Jiying Zhao (M'00) received the Ph.D. degree in electrical engineering from North China Electric Power University, and the Ph.D. degree in computer engineering from Keio University. He is currently a Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada. His research interests include image and video processing and multimedia communications. He is a member of the Professional Engineers Ontario.