Journal of Modern Optics

ISSN: 0950-0340 (Print) 1362-3044 (Online) Journal homepage: http://www.tandfonline.com/loi/tmop20

Investigation of multi-frame image super-resolution based on adaptive self-learning methods

Shao-Shuo Mu, Ye Zhang & Ping Jia

To cite this article: Shao-Shuo Mu, Ye Zhang & Ping Jia (2017) Investigation of multi-frame image super-resolution based on adaptive self-learning methods, Journal of Modern Optics, 64:2, 111-121, DOI: 10.1080/09500340.2016.1210249

To link to this article: https://doi.org/10.1080/09500340.2016.1210249

Published online: 02 Aug 2016.

JOURNAL OF MODERN OPTICS, 2017
VOL. 64, NO. 2, 111–121
http://dx.doi.org/10.1080/09500340.2016.1210249

Taylor & Francis
Taylor & Francis Group

Investigation of multi-frame image super-resolution based on adaptive self-learning methods

Shao-Shuo Mu[a,b], Ye Zhang[a] and Ping Jia[a]

[a]Key Laboratory of Airborne Optical Imaging and Measurement, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China; [b]School of Daheng, Graduate University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

The goal of this paper is to introduce and demonstrate a new high-performance super-resolution (SR) method for multi-frame images. By combining learning-based and reconstruction-based SR methods, this paper proposes a multi-frame image super-resolution method based on adaptive self-learning. Using the adaptive self-learning method and recovery of high-frequency edge information, an initial high-resolution (HR) image containing effective texture information is obtained. The edge smoothness prior is then used to satisfy the global reconstruction constraint and enhance the quality of the HR image. Our results indicate that this method achieves better performance than several other methods for both simulated data and real-scene images.

ARTICLE HISTORY
Received 3 April 2016
Accepted 30 June 2016

KEYWORDS
Super-resolution; adaptive; self-learning

1. Introduction

Super-resolution has been a hot spot in the field of opto-electronic imaging, especially in the fields of aerospace, remote sensing and target recognition, among others. Because the sampling frequency of the detector is often lower than the Nyquist frequency, it results in aliasing and under-sampling of the image. Moreover, degradation factors such as noise and the point spread function of the optical system also lead to poor image quality. The most direct method of improving the image resolution is reducing the detector pixel size, but this cannot be made infinitely small. When the pixel density increases past a certain point, the SNR will decrease.

In order to enhance image resolution and realize 'super-resolution' imaging in a traditional detector, super-resolution (SR) technology based on a combination of hardware and software has become a point of interest and has been widely used in regular, airborne and board cameras (*1–3*). Xu et al. (*4*) proposed an SR algorithm for multiple terahertz imaging, which can obtain higher resolution than that of the original image. Tomas et al. (*5*) provided a new high-performance reconstruction method for super-resolution structured illumination microscopy based on a posteriori probability estimation, which achieves good suppression of out of focus light, improves spatial resolution, and allows reconstruction of both 2D and 3D images of cells even in the case of weak signals. At present, many SR techniques have been proposed, such as interpolation-based, reconstruction-based and learning-based SR techniques. Interpolation-based SR techniques use a block smoothing principle; this method is simple and easy, but cannot effectively recover considerable amount of high-frequency information. The CEVA Company first combined these techniques with the CEVA-MM3101 low-power imaging and vision platform successfully, which utilized low-resolution (LR) sensors to construct high-resolution (HR) images, and has been applied to camera-enabled mobile devices (http://www.ceva-dsp.com/CEVA-MM3101). Reconstruction-based SR techniques adopt edge smoothness priors and the global reconstruction constraint as a regularization term, as seen for instance in Refs. (*6–14*). Without an explicit need for super-resolution with accuracy in multiple images, Takeda et al. (*7*) presented a novel multidimensional kernel regression methodology, which is based on the notion of consistency between the estimated pixels. This method did not need an explicit computation of complex motions, and improved overall performance. Because the traditional total variation (TV) favours a piecewise constant solution, its results will be poor in the flat regions and will produce some pseudo-edges under high noise intensity. Yuan proposed a spatially weighted TV model and a regional spatially adaptive

CONTACT Shao-Shuo Mu 860600894@qq.com

TV model in (*8*) and (*9*), respectively, which reduce the artefacts and pseudo-edges produced by a TV model and maintain the partial smoothness of the HR image. Jeong et al. (*10*) proposed a super-resolution method robust to noise that involves two phases: the learning phase and the reconstruction phase. In order to reduce the large computation load and realize the local spatially adaptive process of the prior model and regularization parameter, Zhang et al. (*11*) proposed a block-based, local, spatially adaptive reconstruction algorithm. Although this technique can maintain a clear edge, it cannot effectively restore the texture detail information. SR techniques based on learning utilize the similarity redundancy structure of natural images to restore the missing high-frequency detail information of LR images. The learning-based algorithm can effectively restore the texture information, but cannot effectively suppress artefacts and distortion, which easily leads to the construction of unnatural HR images. Therefore, an SR method based on a combination of reconstruction and learning techniques is of interest. Compared with previous algorithms, the learning-based algorithm relies on a larger number of databases. The gradient profile prior is proposed in (*15*), which utilizes the model of gradient profiles of natural images, and produces state-of-the-art results. Yang et al. (*16*) proposed a sparse coding method for dictionary learning. In (*17*), the sparse coding super-resolution (SCSR) method is improved to strengthen the sparse similarity between LR and HR image blocks. This method is simpler and more effective than that in (*16*). In (*18*), a beta process joint dictionary learning approach (BD-JDL) was adopted to make the HR and LR dictionaries more accurate and consistent, gaining better texture results compared with (*17*). All of the above reconstruction methods require a large number of external training images. When external images cannot be accessed, it is necessary to start from the point of the similarity redundancy structure of input images. In (*20–23*), different self-learning methods are proposed. On the basis of (*19*), Zhang et al. (*20*) proposed a self-similarity redundancy algorithm, which selects original LR images and their corresponding degraded images as training images, and uses a modified LLE algorithm to obtain one HR image.

In this paper, we mainly focus on multi-frame super-resolution processing. Focusing on the superiorities of reconstruction-based and learning-based algorithms, we propose an adaptive self-learning SR method, using several LR images to reconstruct one better HR image with rich texture, without the aid of external training images. Experimental results demonstrate that the proposed method performs better than other mainstream algorithms.

This paper is organized as follows: Section 2 gives a discussion of classical SR methods based on LLE-learning. Section 3 presents the framework of our proposed algorithm in detail, followed by experimental results in Section 4. Finally, conclusions are made in Section 5.

2. Problem formulation

In this section, we describe one classical SR method based on LLE-learning (*19*).

2.1. LLE-learning theory

In order to obtain imaging resolution better than that of a traditional detector, various super-resolution algorithms have been proposed. In particular, we consider those based on learning. Chang et al. (*19*) first applied LLE technology to super-resolution reconstruction. Assuming that the low-resolution image and the high-resolution image constitute the feature space on the two manifolds, many external HR natural images are needed to create dictionary blocks. This method requires external HR images as an HR training set. The corresponding degraded frameworks of HR images are adopted as an LR training set. Super-resolution technology based on LLE-learning works mainly as follows:

(1) The HR training images and the corresponding LR training images are divided into the same number of image blocks, the sizes of which are $Sq \times Sq$ and $q \times q$, respectively;
(2) The input image is divided into the same number image blocks ($q \times q$). Every block then searches the K nearest-neighbour image blocks in the LR training set, and solves for each weight ω_i.
(3) Using a linear combination of K HR blocks corresponding to the K nearest neighbour LR image blocks $\sum_K \omega_i h_i$ (where the h_i represents the HR blocks corresponding to the nearest neighbour LR image blocks), the reconstructed HR image block is obtained.

3. The proposed algorithm

Acquiring an HR image from a photoelectric imaging system has become a topic of great interest in the field of computer vision. Using a combination of learning-based and reconstruction-based SR methods, this paper proposes an adaptive self-learning SR method, which is able to reconstruct a richer texture HR image from continuous multi-frame images, without the aid of external training images. The process of the proposed method is as follows: first, dictionary blocks are created from the input images. After a target HR block searches for the K nearest HR training blocks, the target pixel of the HR block is acquired

using the adaptive self-learning reconstruction method in the K corresponding LR blocks. Next, the edge recovery method is used to obtain the missing high-frequency edge information. Finally, by combining with an SR method based on reconstruction, a new edge prior constraint is proposed to meet the global reconstruction constraint and get the final HR image.

3.1. Creating dictionary blocks

In this section, we propose a method to build dictionary blocks, which is different from LLE-learning theory. Assuming that the target HR image resolution is S times that of the input LR image resolution, input images (as LR training images) are first divided into $q \times q$ blocks, with an overlap of $q - 1$ pixels; the corresponding S-times up-sampled images (as HR training images) are divided into $p \times p$ blocks, with an overlap of $p - S$ pixels. The relationship between p and q is given in Equation (1).

$$p = (q-1)S + 1 \tag{1}$$

Equation (1) ensures that LR training blocks correspond to HR training blocks; below is an explanation of why this expression is used. As shown in Definition (1), we can prove that when two one-dimensional matrices f and g (with sizes L and $S \times L$, respectively) are divided into blocks f_1 and g_1(with sizes q and p), there is a one-to-one correspondence. In this section, Equation (1) is used to divide LR training images and HR training images (two-dimensional matrices) into blocks (of sizes $q \times q$ and $p \times p$). Corresponding blocks can thus be obtained (*this paper uses* ×2 *magnification, so S is set to* 2).

Definition 1 : We define 'f' and 'g' as one-dimensional matrices, with sizes L and $S \times L$ ($S \geq 2$, and S is an integer). They are divided into q and p blocks, with overlap regions of $q - 1$ and $p - S$, respectively. Then the number of 'f' blocks f_1 is equal to the number of 'g' blocks g_1.

Proof : The number of 'f' blocks f_1 is described by Equation (2)

$$f_1 = L - (q-1) \tag{2}$$

The number of 'g' blocks g_1 can be expressed using an arithmetic progression, as shown in Equation (3) (where S is the arithmetic spacing)

$$S \times g_1 - (S-1) = S \times L - (p-1) \tag{3}$$

Then, Equation (4) can be obtained from Equation (3), when $\lfloor \bullet \rfloor$ is rounded down to the nearest integer.

$$\begin{aligned} |g_1| &= \left| \frac{S \times L - (p-1) + S - 1}{S} \right| \\ &= \left| L - (q-1) + \left(1 - \frac{1}{S}\right) \right| \\ &= L - (q-1) \end{aligned} \tag{4}$$

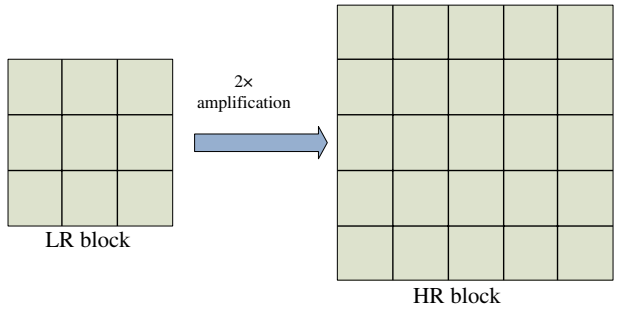

Figure 1. Dividing blocks ($S = 2, p = 5, q = 3$).

Therefore f_1 is equal to g_1. This is still the case if 'f' and 'g' are two-dimensional image matrices. The segmentation of the proposed method is as shown in Figure 1.

3.2. Initial HR estimation by adaptive self-learning super-resolution and high-frequency edge recovery

We will obtain the final HR image in this section. Suppose there are n input images $L_1, L_2, \ldots, L_n$, with a corresponding HR training set $H_1, H_2, \ldots, H_n$. According to Section 3.1, the sets are divided into image blocks $l_1, l_2, \ldots, l_m$ and $h_1, h_2, \ldots, h_m$, respectively. We use one input image L_1 as the target LR image to be reconstructed. First, the nearest neighbour image blocks must be found (*16*), given by Equation (5). Every block h_i of the corresponding HR image H_1 searches for the K nearest blocks $h_{i1}, h_{i2}, \ldots, h_{iK}$ (h_{ik} indicates one nearest neighbour block of h_i in $h_1, h_2, \ldots, h_m$). Their corresponding LR blocks are $l_{i1}, l_{i2}, \ldots, l_{iK}$.

We search for the K nearest neighbour blocks of h_i, as in Equation (5)

$$\begin{gathered} h_i = \sum_{j=1}^{K} \omega_j h_{ij} + \varepsilon_i, \ (\min \varepsilon_i) \\ \left(\text{We presume } K = 9 \text{ in this paper}\right) \end{gathered} \tag{5}$$

This can be turned into a least-squares problem, as in the following:

$$\varepsilon_i = \operatorname{argmin} \left\| h_i - \sum_{j=1}^{K} \omega_j h_{ij} \right\|^2 \text{ s.t. } \sum_{j=1}^{K} \omega_j = 1 \tag{6}$$

Using Equations (5) and (6), we obtain K nearest neighbour blocks $h_{i1}, h_{i2}, \ldots, h_{iK}$ and corresponding LR blocks $l_{i1}, l_{i2}, \ldots, l_K$.

The LR training blocks $l_1, l_2, \ldots, l_m$ all come from the original input images, which contain truer and more effective pixel information than the up-sampled HR training blocks $h_{i1}, h_{i2}, \ldots, h_{iK}$. Therefore, we adopt the K LR training blocks $l_{i1}, l_{i2}, \ldots, l_{iK}$ obtained above to reconstruct the HR image, instead of using $h_{i1}, h_{i2}, \ldots, h_{iK}$ (*20*, *22*). This

paper uses the Nonlocal-Means (NLM) algorithm (6) to fuse the LR training blocks$l_{i1}, l_{i2}, \ldots, l_{iK}$ into the target HR pixel, as shown in Equation (7). The NLM algorithm uses fuzzy motion estimation, which does not require explicit image registration; therefore, this method need not deal with the difficult local motion estimation problem. In addition, the motion model of a multi-frame image cannot be considered.

Consider

$$\eta = \sum_{t=1}^{K} \sum_{(k,l)\in\Omega} \omega(k,l,t) \left\| D_{k,l} R_{k,l} X - y_t \right\|^2 \tag{7}$$

where y_t represents the LR training image blocks $l_{i1}, l_{i2}, \ldots, l_{iK}$ obtained above. $R_{k,l}X$ is one block h_i of H_1 ($R_{k,l}$ is the block extraction operator). $D_{k,l}$ represents the down-sampling matrices and $\omega(k,l,t)$ represents the weight coefficient. The derivation of Equation (7) is shown in Equation (8).

$$\frac{d\eta}{dX} = 2 \sum_{t=1}^{K} \sum_{(k,l)\in\Omega} \omega(k,l,t) (D_{k,l} R_{k,l})^T (D_{k,l} R_{k,l} X - y_t) \tag{8}$$

Setting Equation (8) equal to 0 gives us Equation (9)

$$\begin{aligned} X &= \left[\sum_{t=1}^{K} \sum_{(k,l)\in\Omega} \omega(k,l,t) (D_{k,l} R_{k,l})^T (D_{k,l} R_{k,l}) \right]^{-1} \\ &\quad (D_{k,l} R_{k,l})^T \left(\sum_{(k,l)\in\Omega} \sum_{t=1}^{K} \omega(k,l,t) y_t \right) \\ &= \left[\sum_{t=1}^{K} \sum_{(k,l)\in\Omega} \omega(k,l,t) (R_{k,l})^T D_{k,l}^T D_{k,l} R_{k,l} \right]^{-1} \\ &\quad (D_{k,l} R_{k,l})^T \left(\sum_{(k,l)\in\Omega} \sum_{t=1}^{K} \omega(k,l,t) R_{k,l}^T D_{k,l}^T y_t \right) \end{aligned} \tag{9}$$

For simplification of further calculations, the assumption is made that $R_{k,l}\,X$ extracts the centre pixel of the image block; then $(R_{k,l})^T D_{k,l}^T D_{k,l} R_{k,l} \approx 1$. For $\sum_{j=1}^{K} \omega_j = 1$, Equation (9) turns into Equation (10) below:

$$\begin{aligned} X(k,l) &= \frac{\sum_{t=1}^{K} \omega(t) \overline{y(t)}}{\sum_{t=1}^{K} \omega(t)}, \ \overline{y(t)} = R_{k,l}^T D_{k,l}^T y_t, \ \omega(t) \\ &= \exp\left\{ \frac{-\left\| h_i - y_t \right\|^2}{2\sigma^2} \right\} \quad (\sigma \text{ set to} 1) \end{aligned} \tag{10}$$

where $\overline{y(t)}$ represents the up-sampling of y_t followed by extraction of the centre block pixel. This section utilizes the principle of the weighted average of similar neighbourhood pixels (25), and finally solves for $\overline{y(t)}$. This paper

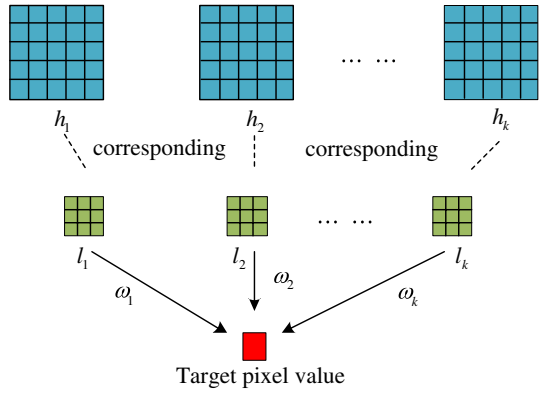

Figure 2. Process of adaptive self-learning super-resolution.

adopts the average filter idea to solve for $\overline{y(t)}$, as shown in Equation (11). Consider

$$\overline{y(t)} = \sum_{N(t)} K(i,t) y(i) \tag{11}$$

where $y(i)$ represents the centre pixel of y_t. $N(t)$ represents the index set consisting of all pixels $y(i)$ ofy_t ($y(i) \in y_t$). $K(i,t)$ represents the similarity weight between $y(i)$ and index set $N(t)$.

The weight coefficient $K(i,\,t)$ is given by

$$K(i,t) = \frac{1}{\eta} \exp\left(\frac{-\delta_{i,j}}{\pi^2} \right) \tag{12}$$

where $\delta_{i,j} = \left\| y(i) - y(j) \right\|^2$indicates the Euclidean distance between pixels 'i' and 'j'. In addition, $\eta = \sum_j \exp\left(\frac{-\delta_{i,j}}{\pi^2}\right)$, where '$\pi$' controls the decay of the exponential function (24) and is set to 1. The process is as shown in Figure 2. This section will obtain an initial HR estimation $H_{initial}$.

The biggest difference between a real HR scene and the LR image is that the high-frequency edge information of the former is richer than that of the latter. Effective reconstruction of the high-frequency edge information can make the reconstructed image closer to the real HR image; therefore, the focus and chief difficulty of the SR method is to recover as much high-frequency edge information as possible. Beginning with the initial HR estimation obtained from the process described in Figure 2, we propose a further high-frequency edge information recovery method.

We employ the initial HR estimation $H_{initial}$ as the testing image to be reconstructed. Meanwhile, the original input images $L_1, L_2, \ldots, L_n$ are employed as LR training images, and corresponding high-frequency images $H_1, H_2, \ldots, H_n$ ($H_i = L_1 - L_i{\downarrow_p}{\uparrow_p}$) as HR training images. The reconstruction process can be expressed by a flow chart (Figure 3), and allows us to obtain a high-frequency edge

reconstruction H_{hf}. Then we gain a final HR estimation $H_{further}$ given by:

$$H_{further} = H_{initial} + H_{hf} \quad (13)$$

3.3. GPU optimization

The Graphics Processing Unit (GPU) is a general-purpose computing processor (GPGPU) with high performance and high memory bandwidth. It is especially suitable for parallel processing of data, in which each element of the data is mapped to a corresponding parallel processing thread in the GPU. Due to the large amounts of data, it is necessary to accelerate the computation speed of large-scale parallel data processing. In the proposed method, reconstruction of an initial HR estimation uses block parallel processing, so that each image block is mapped to each thread in the AMDGPU embedded system platform using OpenCL, and thus parallel optimization is realized.

One kernel function of the initial HR estimate is as follows:

```
kernel void calculate_initialHR( SrcimgLR, sampler, width, height)
 *Retrieve the thread index(ix, iy)
   If ix<width and iy<height
     *Initialization V, W
     For each block centre H(i, j) corresponding to all LR training blocks
       Calculate weight
       Calculate
       Accumulate V+ = weight ×
       W+ = weight
     End for
   Result = V/W
 End
 Write the result to image
```

3.4. Edge prior constraint based on reconstruction

In Section 3.2, we proposed a further high-frequency edge information recovery method. Adding high-frequency reconstructed information to the initial SR estimation, as in Equation (13), may cause an over-smoothing effect and result in an unnatural image. To maintain a clear and natural edge, the SR technique based on reconstruction utilizes edge smoothness priors and the global reconstruction constraint as a regularization term. Therefore, in order to further enhance the quality of the initial HR image, and to obtain a clear and natural HR image, this article proposes a new edge prior constraint $\|\tilde{X} - \tilde{X}_0\|_2^2$, as shown in Equation (14). This constraint not only keeps the high-frequency edge information obtained from the learning-based SR method, but also meets the reconstruction constraints of the whole image (*22*). Consider

$$\hat{X} = \arg\min \|y - DHX\|_2^2 + \mu\|\tilde{X} - \tilde{X}_0\|_2^2 \quad (14)$$

where 'y' is the input image, X_0 is the initial SR image $H_{further}$ obtained from the learning-based SR method, $\tilde{X}_0$ is the high-frequency edge information H_{hf} of X_0, D is the down-sampling matrix and H stands for the optical blurring matrix. The transformation of Equation (14) is:

$$\frac{\partial \hat{X}}{\partial X} = (DH)^T(y - DHX) + \mu(\mathrm{div}(\tilde{X}) - \mathrm{div}(\tilde{X}_0)) \quad (15)$$

which results into the following:

$$X_{n+1} = X_n + \tau[(DH)^T(y - DHX) + \mu(\mathrm{div}(\tilde{X}) - \mathrm{div}(\tilde{X}_0))] \quad (16)$$

where $\mathrm{div} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. The μ is used as a control switch to balance between the two constraints. Larger μ values should be selected to protect edge information; conversely, smaller μ values will produce better image constraints.

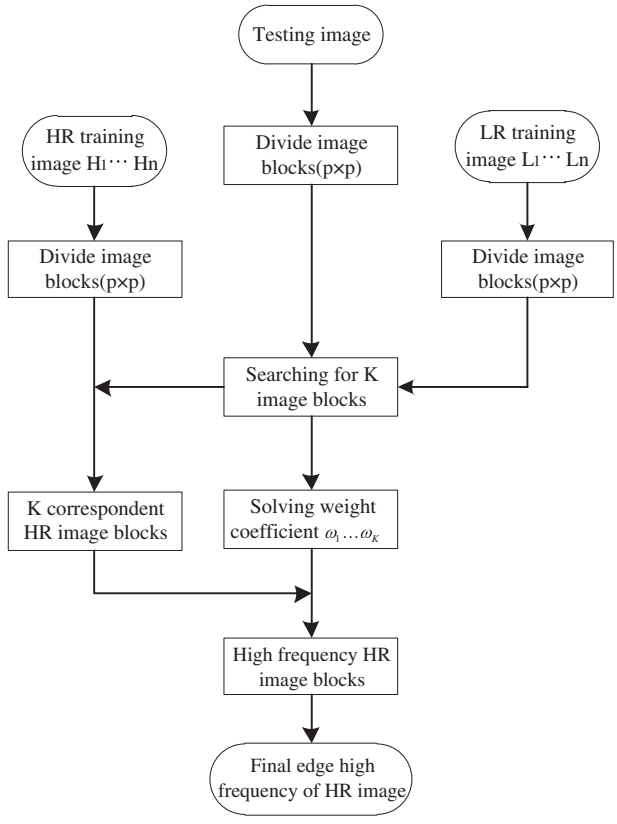

Figure 3. Flowchart for getting high-frequency edge information.

4. Experimental results and analysis

4.1. Lena simulated experiment

To verify the advantages of the proposed algorithm, simulation results for a high resolution image (Figure 4(a)) with a size of 294 × 228 pixels are discussed in this section. First, the original HR image was shifted with sub-pixel displacements to produce four images. The sequence was then convoluted with a 3 × 3 uniform point spread

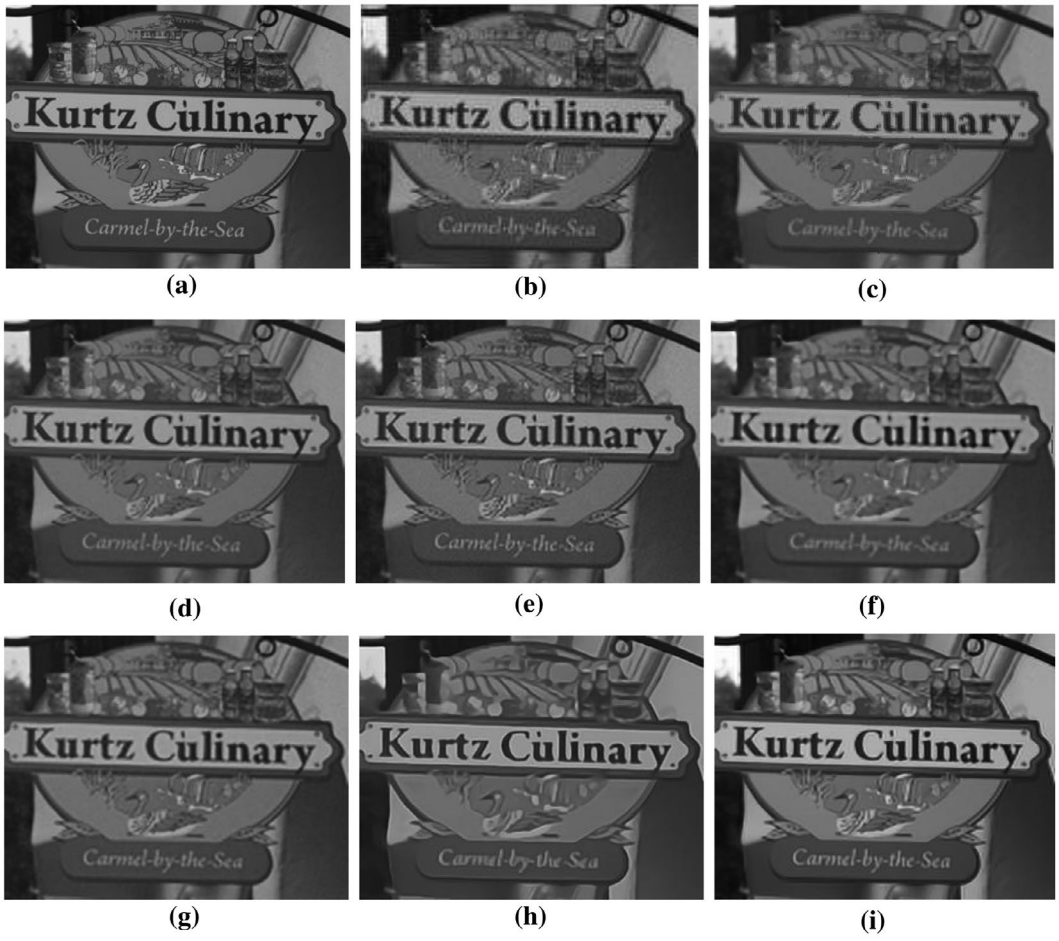

Figure 4. Simulated reconstructed results. (a) HR image (b) LR image (c) NLM method, PSNR = 22.8464, SSIM = 0.7492, Operation time = 53.6721 s. (d) SCSR method, PSNR = 26.3778, SSIM = 0.8649, Operation time = 103.3653 s. (e) BP-JDL method, PSNR = 26.8799, SSIM = 0.8779, Operation time = 197.0814 s. (f) AGMS method, PSNR = 25.3527, SSIM = 0.8219, Operation time = 2.3411 s. (g) Multiple SR method, PSNR = 25.5773, SSIM = 0.8444, Operation time = more than one hour. (h) The proposed method without regularization (Section 3.4). (i) The proposed method, PSNR = 27.3424, SSIM = 0.8988, Operation time = 23.9132 s.

function (PSF), and down-sampled by a factor of 2 in both the vertical and horizontal directions. Finally, Gaussian noise with a variance of 2 was added to the sequence. We apply the NLM algorithm (*6*), SCSR method (*17*), BP-JDL method (*18*), AGMS (adaptive gradient magnitude self-interpolation) method (*14*), multiple SR method (*23*) and the method proposed in this work to reconstruct the simulated LR image. In all experiments using the proposed method (Sections 4.1 and 4.2), the reconstruction factor S was set to 2. Input images are divided into 3×3 blocks, with a 2 pixel overlap between adjacent blocks. The corresponding S-times up-sampled images are divided into 5×5 blocks, with an overlap of 3 pixels. We set $\tau = 0.1$ and $\mu = 0.1$ in Equation (16). Reconstruction results are shown in Figure 4. In order to verify the advantages of the proposed method with regularization (Section 3.4), the final stage without regularization is shown in Figure 4(h). From a visual comparison of the results, Figure 4(i) appears more natural compared with Figure 4(h), and Figure 4(h) appears to be over-smoothed. Because the high-frequency reconstructed information is added to the initial SR estimation (Equation (13)), we may see an unnatural image and over-smoothing effects with the improved edge information recovery. This is also the reason for the edge-prior constraint based on reconstruction. Compared with other SR methods, the result of the proposed method has better edge information, and image details, such as 'Carmel-by-the-Sea', are more exquisite. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are employed to evaluate the objective quality of the reconstructed HR image. The PSNR and SSIM of the result of the proposed method are larger than those of the other results (Figure 4). The proposed method result is closer to the standard HR image and the quality of the reconstruction image is better. At the same time, we used a GPU to run this proposed algorithm and found that the operation speed is also significantly better than that of most

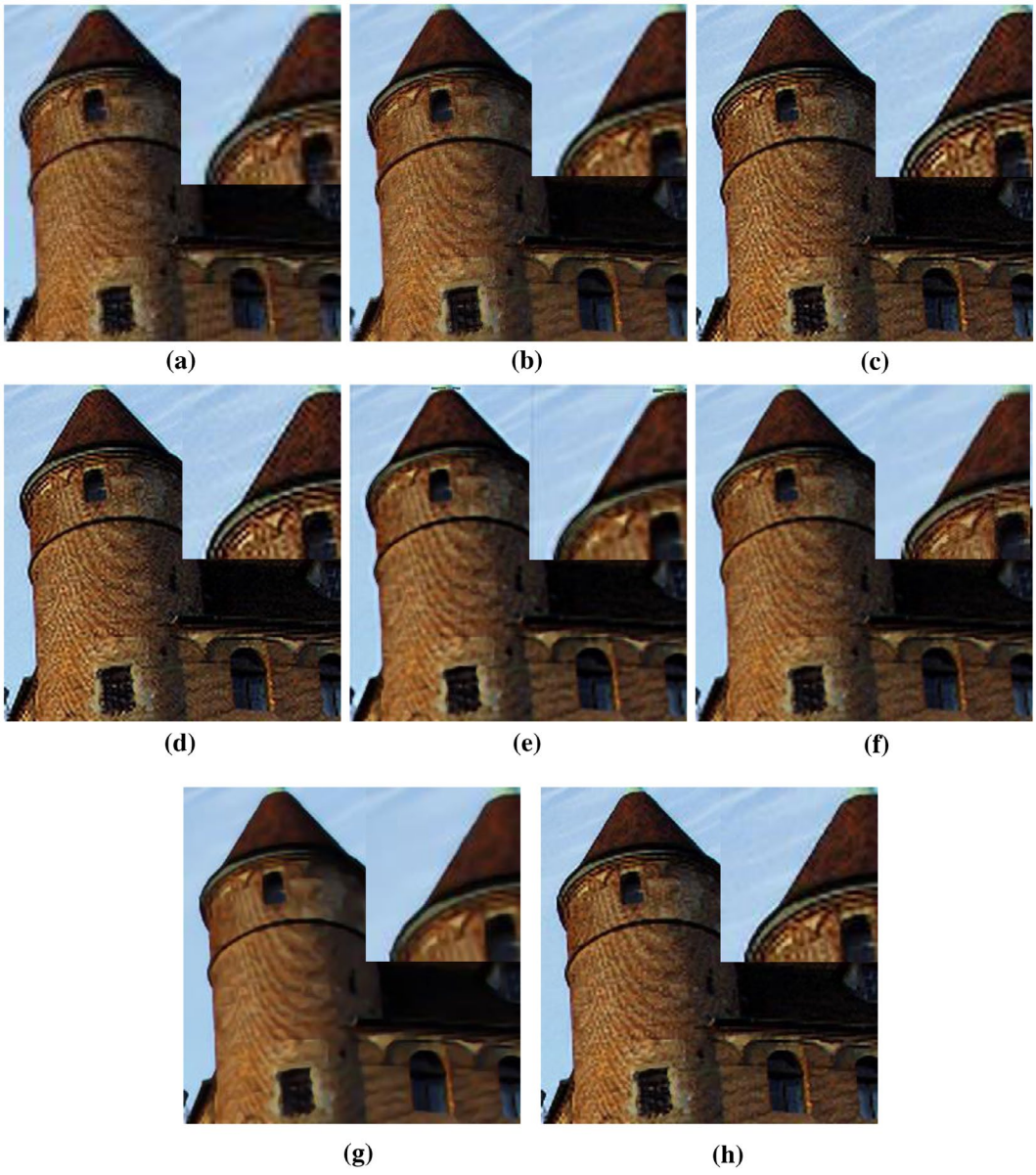

Figure 5. Castle reconstructed results. (a) One of four LR images. (b) NLM method, SNR = 0.9246, IE = 6.3151, AG = 6.1895, Operation time = 68.9213 s. (c) SCSR method, SNR = 0.9193, IE = 7.2669, AG = 9.2214, Operation time = 106.4351 s. (d) BP-JDL method, SNR = 0.9019, IE = 7.2217, AG = 7.0612, Operation time = 198.6566 s. (e) AGMS method, SNR = 0.8646, IE = 7.2558, AG = 5.6988, Operation time = 1.8672 s. (f) Multiple SR method, SNR = 0.9169, IE = 7.2244, AG = 8.8511, Operation time = more than one hour. (g) The proposed method without regularization. (h)The proposed method, SNR = 0.9234, IE = 7.2754, AG = 9.3059, Operation time = 30.4513 s.

other algorithms. The AGMS method adopts a simple edge-prior knowledge, so its operation time is superior to that of our method; however, our reconstructed result (Figure 4(i)) is better than that of the AGMS method (Figure 4(f)).

4.2. Real data experiment

In this section, we present three real data experiments to illustrate the performance of the proposed algorithm, and compare the proposed method with five typical methods: the NLM method, the SCSR method (*17*), the BP-JDL method (*18*), the AGMS (adaptive gradient magnitude self-interpolation) method (*14*) and the multiple SR method (*23*). The first experiment uses 4 frames of degraded castle images (Castle). The second experiment uses three continuous frames of video images of a foreman (Foreman). The third experiment uses 4 frames of real images shot by the camera system.

We are able to prove the superiority of the proposed algorithm from the comparative visual and objective quality of the reconstructed HR image. Measures of the signal-to-noise ratio (SNR), average gradient (AG) and information entropy (IE) are all employed to evaluate

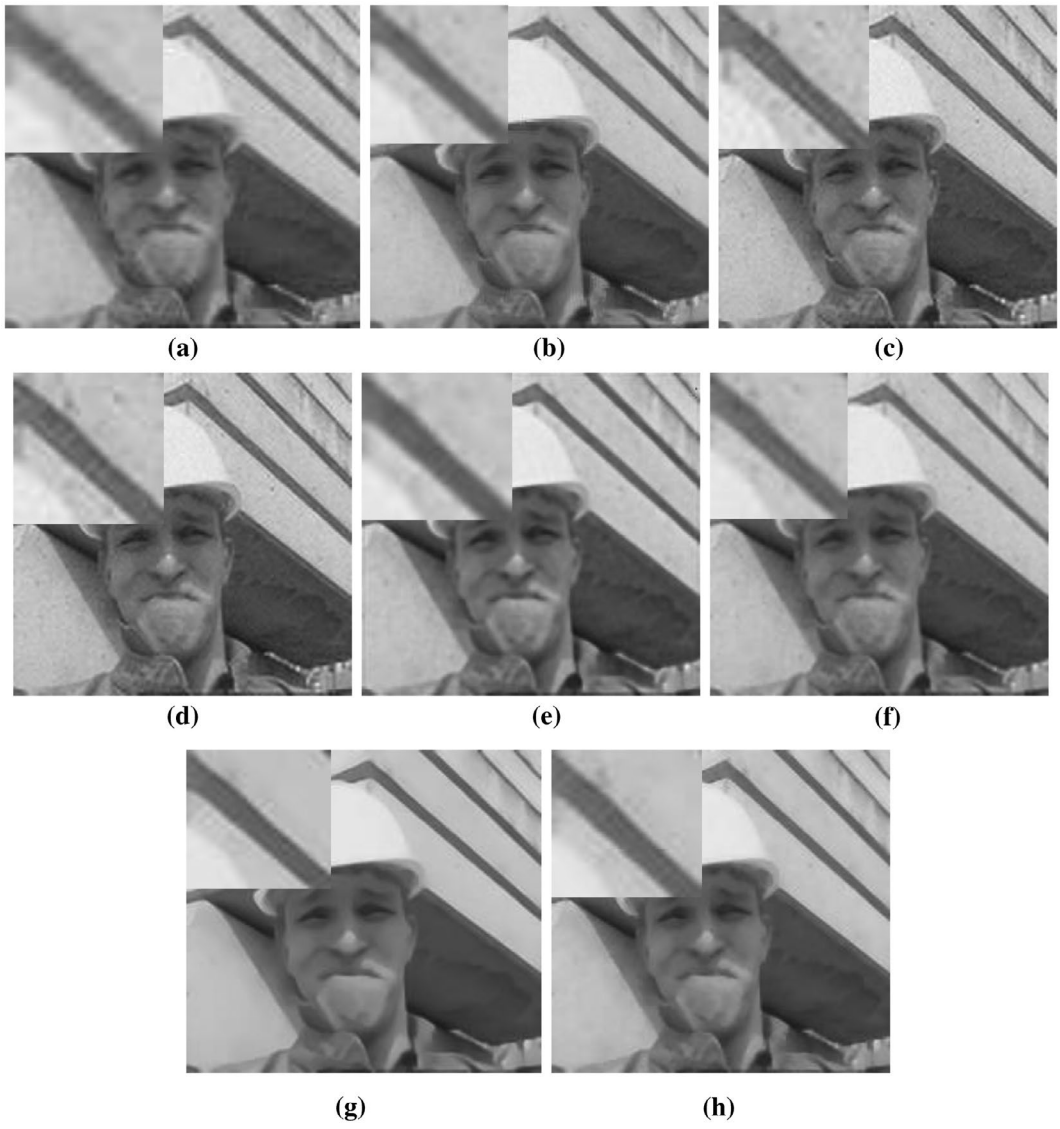

Figure 6. Foreman reconstructed results. (a) One of three successive frames of the video sequence. (b)NLM method, SNR = 2.5078, IE = 7.2116, AG = 4.8246, Operation time = 30.75 s. (c)SCSR method, SNR = 2.4171, IE = 7.2930, AG = 6.1950, Operation time = 62.0568 s. (d) BP-JDL method, SNR = 2.4220, IE = 7.2979, AG = 6.2618, Operation time = 119.8363 s. (e)AGMS method, SNR = 2.1597, IE = 7.2518, AG = 5.1207, Operation time = 1.3858 s. (f) Multiple SR method, SNR = 2.4142, IE = 7.2146, AG = 4.6796, Operation time = more than one hour. (g)The proposed method without regularization. (h)The proposed method, SNR = 2.5974, IE = 7.3004, AG = 6.2681, Operation time = 20.8112 s.

the objective quality of the reconstructed HR image. The operation times of each method are also given. When the values of the SNR, IE and AG are larger, the reconstructed image is clearer and contains richer information (*23*).

The experimental results for Castle are shown in Figure 5. The SNR, AG and IE demonstrate that the proposed algorithm can achieve better reconstructed results and recover more missing high-frequency information and clearer edge details in comparison with the other five methods. From the visual comparison results, Figure 5(h) with regularization is more natural than Figure 5(g) without regularization. We also notice that the proposed method produces a sharper image than the other methods. Among the detail results for the top right image in all methods, the HR image produced by our method has the least artefacts, indicating that the proposed method can better restore an HR image from LR images.

Our results for video image SR will show that this algorithm can reconstruct video images well. Therefore, in this section, we use the continuous three frames of video images as the input images to reconstruct the HR image. Experimental results are shown in Figure 6. It is observed that some visible artefacts are introduced in Figure 6(c), (d) and (f). In contrast, the proposed approach (Figure 6(h)) gives a clearer result without these artefacts. We also note that our method produces sharper edges than the

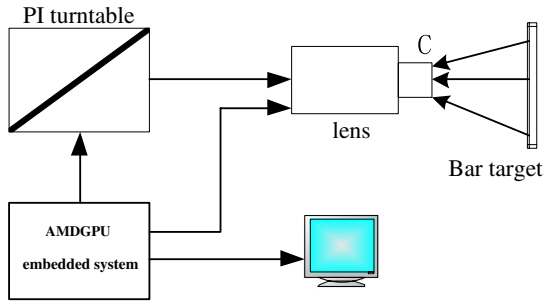

Figure 7. Camera system.

NLM and AGMS methods in Figure 6(b) and (e), respectively. Figure 6(h) is more natural in comparison to Figure 6(g). By comparing the SNR, AG and IE, we can draw the conclusion that the proposed method can produce better reconstructed HR image than other methods, and can adapt to SR reconstruction of video images.

To further assess the proposed method when adapted to real-scene SR reconstruction, we consider multiple real scenes from a camera system that we have assembled, and reconstruct one HR image in the embedded GPU system platform. As illustrated in Figure 7, the apparatus consists

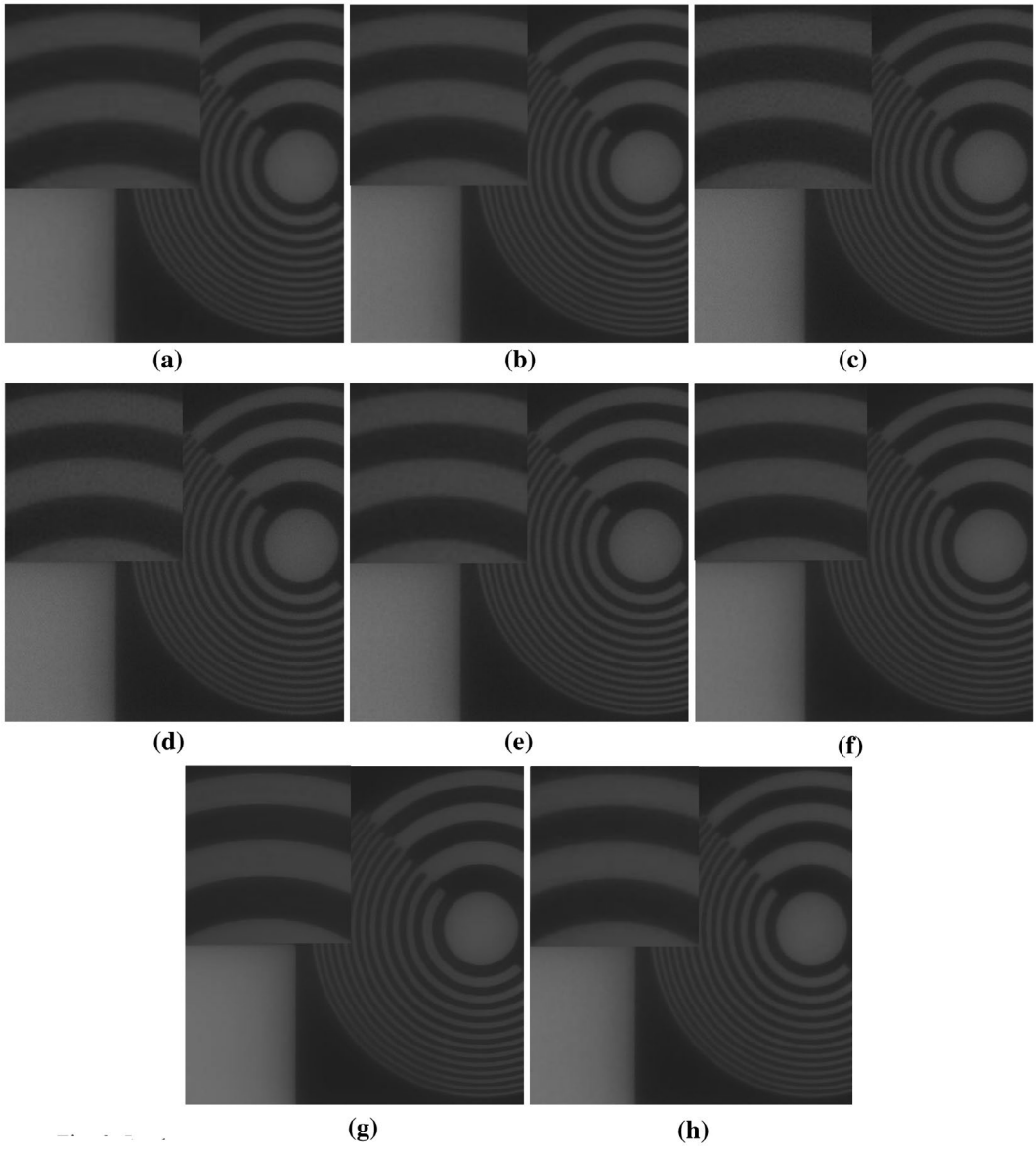

Figure 8. Real scene reconstructed results. (a) One of four real scenes. (b) NLM method, SNR = 6.7254, IE = 6.3151, AG = 1.1895, Operation time = 282.7817 s. (c) SCSR method, SNR = 6.7889, IE = 6.3411, AG = 1.7296, Operation time = 439.9516 s. (d) BP-JDL method, SNR = 6.8049, IE = 6.3409, AG = 1.7327, Operation time = 859.3259 s. (e) AGMS method, SNR = 6.2727, IE = 6.3080, AG = 1.2874, Operation time = 4.3652 s. (f) Multiple SR method, SNR = 7.1254, IE = 6.4141, AG = 1.8138, Operation time = more than one hour. (g) The proposed method without regularization. (h) The proposed method, SNR = 6.8283, IE = 6.8573, AG = 1.8912, Operation time = 48.5661 s.

of a micro-displacement lens, a PI turntable and an AMDGPU. By controlling the PI turntable, we can continuously receive multiple real images, such as that shown in Figure 8(a). Figure 8(b)–(f) are the reconstruction results from the NLM method, SCSR method, BP-JDL method, AGMS method and multiple SR method (*23*). Figure 8(g) and (h) are from the proposed method without regularization and with regularization (Section 3.4), respectively. We see that Figure 8(g) appears over-smoothed. From Figure 8(b)–(d) and (f), we see that these methods improve the resolution of LR images, but are prone to annoying images. The AGMS method can achieve better results than the SCSR and BP-JDL methods; however, its performance is obviously inferior to that of the proposed method, which produces sharper and clearer edges. Values of the SNR, IE and AG also demonstrate that our method achieves both sharper edges and details.

5. Conclusion

Based on the detailed discussion above, we have proven through both theory and experiment that the proposed algorithm can achieve better results than three other methods for multiple-frame degraded images, video images and real-time images. We propose the adaptive self-learning reconstruction method to reconstruct target HR block centre pixels in the K corresponding LR blocks. We then use the edge recovery method to obtain the missing high-frequency edge information. In Section 3.4, we adopt a new edge prior constraint to meet the global reconstruction constraint and get a final HR image. However, although the running time of the proposed algorithm is greatly improved, it still does not meet the real-time requirements. We need to further optimize the algorithm's procedures and make full use of the parallel real-time properties of the GPU. In the future, we plan to further speed up the proposed super-resolution and sharpness enhancement algorithm, and then to extend the proposed method to video super-resolution and enhancement in FPGA hardware.

Disclosure statement

The authors declare that there is no conflict of interests regarding the publication of this paper.

Funding

This work was supported by the project of the National Science Fund for Distinguished Young Scholars of China [grant number 60902067]; the key Science-Technology Project of Jilin province [grant number 11ZDGG001].

References

(1) Wen, D.S.; Liu, X.P.; Qiao, W.; Wang, H. Novel Subpixel Imaging System with Linear CCD Sensors. *Proc. SPIE* **2001**, *4563*, 116–122.
(2) Yang Wen-bo, W.B.; Zhu Ming, M.; Liu Zhi-ming, Z.M.; CHEN Dong-cheng, D.C. Super-resolution Reconstruction of Sub-Pixel Imaging Achieved by Three Linear Array Detectors. *Opt. Precis. Eng.* **2014**, *22* (8), 2247–2258.
(3) Xu, Z.P.; Zhai, L.P.; Ge, W.Q. CCD Geometric Super-resolution Method Based on Subpixel. *Opt. Precis. Eng.* **2008**, *16* (12), 2447–2453.
(4) Xu, L.M.; Fan, W.H.; Liu, J. High-resolution Reconstruction for Terahertz Imaging. *Appl. Opt.* **2014**, *53* (33), 7891–7897.
(5) Tomáš, L.; Pavel, K.; Svindrych, Z.; Benda, J. Three-dimensional Super-resolution Structured Illumination Microscopy with Maximum a posteriori Probability Image Estimation. *Opt. Express* **2014**, *22* (24), 29805–29817.
(6) Protter, M.; Elad, M.; Takeda, H.; Milanfar, P. Generalizing the Nonlocal-means to Super-resolution Reconstruction. *IEEE Trans. Image Process.* **2009**, *18* (1), 1958–1975.
(7) Takeda, H.; Milanfar, P.; Protter, M.; Elad, M. Super-resolution Without Explicit Subpixel Motion Estimation. *IEEE Trans. Image Process.* **2009**, *18* (9), 36–51.
(8) Yuan, Q.Q.; Zhang, L.; Shen, H.F. Multiframe Super-resolution Employing a Spatially Weighted Total Variation Model. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22* (3), 379–392.
(9) Yuan, Q.; Zhang, L.; Shen, H.F. Regional Spatially Adaptive Total Variation Super-resolution with Spatial Information Filtering and Clustering. *IEEE Trans. Image Process.* **2013**, *22* (6), 2327–2342.
(10) Jeong, S.C.; Kang, Y.; Song, B.C. Super-resolution Algorithm Using Noise Level Adaptive Dictionary. *IEEE 14th International symposium on consumer electronics*, Braunschweig, Germany, 2010.
(11) Zhang, L.P.; Yuan, Q.Q.; Shen, H.F.; Li, P.X. Multiframe Image Super-resolution Adapted With Local Spatial Information. *J. Opt. Soc. Am. A.* **2011**, *28* (3), 381–390.
(12) Gaurav, G.B.; Ajinkya, S.D.; Swarup, S.M. Image Super Resolution with Direct Mapping and De-noising. *IEEE 14th Emerging applications of information technology*, Kolkata, India, **2014**.
(13) Trinh, D.-H.; Luong, M.; Dibos, F.; Rocchisani, J. Novel Example-Based Method for Super-Resolution and Denoising of Medical Images. *IEEE Trans. Image Process.* **2014**, *23* (4), 1882–1895.
(14) Wang, L.F.; Xiang, S.M.; Meng, G.; Wu, H. Edge-directed Single-image Super-resolution via Adaptive Gradient Magnitude Self-interpolation. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23* (8), 1289–1299.
(15) Sun, J.; Xu, Z.B.; Shum, H.Y. Gradient Profile Prior and its Applications in Image Super-resolution and Enhancement. *IEEE Trans. Image Process.* **2011**, *20* (6), 1529–1541.
(16) Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image Super-resolution as Sparse Representation of Raw Image Patches. In *Proceedings of the 26th IEEE Conference on*

Computer Vision and Pattern Recognition, Anchorage, AK, June, 2008; pp 1–8.

(17) Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image Super-resolution via Sparse Representation. *IEEE Trans. Image Process.* **2010**, *19* (11), 2861–2873.

(18) He, L.; Qi, H.;Zaretzki, R. Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, Portland, OR, USA, June 2013; pp 345–352.

(19) Chang, H.; Yeung, D.-Y.; Xiong, Y. Super-resolution through Neighbor Embedding. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR '04), Washington, DC, USA, Vol. 1, 2004. pp 275–282.

(20) Zhang, K.; Gao, X.; Tao, D.; Li, X. Single Image Super-resolution with Multiscale Similarity Learning. *IEEE Transactions on Neural Networks and Learning Systems* **2013**, *24* (10), 1648–1659.

(21) Zhang, K.; Gao, X.; Li, X.; Tao, D. Partially Supervised Neighbor Embedding for Example-based Image Super-resolution. *IEEE J. Sel. Top. Signal Process.* **2011**, *5* (2), 230–239.

(22) Liu, Q.; Liu, L.; Wang, Y.; Zhang, Z. Locally Linear Embedding Based Example Learning for Pan-sharpening. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR '12)*, Tsukuba, Japan, 2012; pp 1928–1931.

(23) Mu, S.S.; Zhang, Y.; Jia, P.; Yang, X.; Qiu, X.F. Multiframe Superresolution Reconstruction Based on Self-learning Method. *Math. Prob. Eng.* **2015**, *2015*, 12 pages. Article ID 181864.

(24) Buades, A.; Col, B.; Morel, J.-M. A Non-local Algorithm for Image Denoising. *IEEE Conference CVPR*, San Diego, CA, USA, 2005.

(25) Zhang, K.; Gao, X.; Tao, D.; Li, X. Single Image Super-resolution with Non-local Means and Steering Kernel Regression. *IEEE Trans. Image Process.* **2012**, *21* (11), 4544–4556.