

# A depth-based Multi-view Super-Resolution Method Using Image Fusion and Blind Deblurring

**Jun Fan, Xiangrong Zeng, Qizi Huangpeng, Yan Liu, Xin Long, Jing Feng, Jinglun Zhou**

College of Information System and Management, National University of Defense Technology,  
Changsha, China

[E-mail: fanjun891009@163.com; zengxrong@gmail.com]

Corresponding author: Xiangrong Zeng

*Received January 11, 2016; revised April 3, 2016; revised July 16, 2016; accepted August 28, 2016;  
published October 31, 2016*

---

## **Abstract**

Multi-view super-resolution (MVSR) aims to estimate a high-resolution (HR) image from a set of low-resolution (LR) images that are captured from different viewpoints (typically by different cameras). MVSR is usually applied in camera array imaging. Given that MVSR is an ill-posed problem and is typically computationally costly, we super-resolve multi-view LR images of the original scene via image fusion (IF) and blind deblurring (BD). First, we reformulate the MVSR problem into two easier problems: an IF problem and a BD problem. We further solve the IF problem on the premise of calculating the depth map of the desired image ahead, and then solve the BD problem, in which the optimization problems with respect to the desired image and with respect to the unknown blur are efficiently addressed by the alternating direction method of multipliers (ADMM). Our approach bridges the gap between MVSR and BD, taking advantages of existing BD methods to address MVSR. Thus, this approach is appropriate for camera array imaging because the blur kernel is typically unknown in practice. Corresponding experimental results using real and synthetic images demonstrate the effectiveness of the proposed method.

---

**Keywords:** Multi-view super-resolution, depth estimation, graph cuts, blind deblurring, ADMM

## 1. Introduction

The goal of MVSR is to estimate a HR image from a set of LR images captured from different viewpoints (typically by different cameras). Bennett Wilburn et al. [1] used multiple inexpensive cameras to approximate a video camera with a large synthetic aperture to obtain HR videos. Kartik Venkataraman et al. [2] constructed an ultra-thin high-performance monolithic camera array to acquire multi-view images, and then applied a two-stage MVSR construction to obtain the final HR image that corresponded to a selected reference camera. Guillem Carles et al. [3] realized super-resolution (SR) imaging using an array of 25 independent commercial off-the-shelf cameras.

Another problem of interest that is strongly related to MVSR is multi-frame super-resolution (MFSR) [4, 5], which aims to construct an HR image from several observed LR images. For example, MFSR has been extensively studied in video sequences [5, 6, 7, 8] in which the involved LR images have been captured using the same camera at different times. MVSR and MFSR share several common characteristics. First, both are ill-posed problems that can be addressed through regularization [5, 9, 10, 11]. Among the aforementioned articles, Qizi Huangpeng et al. [11] super-resolved multiple degraded LR frames of the original scene via multi-frame blind deblurring (MFDB) to address unknown blurring. Second, both approach have two steps: LR image registration (LRIR) and HR image reconstruction (HRIR). LRIR determines pixel correspondences among different input LR images, whereas HRIR reconstructs the desired HR image from input LR images based on the outcome of LRIR. The quality of the reconstructed HR image depends considerably on the accuracy of LRIR. The motions among LR images in MVSR are generally more complicated than those in MFSR. Thus, LRIR in MVSR is more difficult than that in MFSR, and consequently, MVSR is more difficult than MFSR.

Recently, many papers have been dedicated to cutting the edge of MVSR. T. Tung et al. [12] super-resolved input multi-view images to generate a complete 3D model for a single object. A super-resolution-free-viewpoint image synthesis (SR-FVS) method that used adaptive regularization for MVSR to address depth inaccuracies was proposed by Takahashi; this method simultaneously realized free-viewpoint image SR and free-viewpoint depth estimation [13]. Nakashima et al. [14] combined a learning-based SR method, namely, sparse coding SR (ScSR) [15], with an existing SR-FVS method [13] to improve the quality of the desired HR image of the target viewpoint.

SR reconstruction was divided into two stages in [2, 13, 14]: image fusion (IF), which was called image blend in [13, 14], and maximum-a-posteriori (MAP). In the MAP stage, however, [13] ignored the effect of blur, whereas [2, 14] assumed that blur was known. Consequently, all the energy functions constructed in the MAP stage had no regularization term for blur. In fact, the blur kernel was typically unknown in practical situations. In this study, we have adopted the two-stage SR framework in [2, 13, 14], and considered the effect of unknown blur (i.e., a regularization term for blur has been added into the energy function in the MAP stage). We have actually reformulated the MVSR problem into an IF problem and a blind deblurring (BD) problem. In the IF stage, we have introduced a reference LR image to improve the accuracy of depth estimation and the quality of the resulting IF image. In the BD stage, we have addressed the BD problem via alternating minimization in which each sub-problem is efficiently solved using the alternating direction method of multipliers (ADMM) [16, 17]. The proposed approach can estimate accurate depth maps and desired HR images from multi-view

input LR images; it is suitable for camera array imaging.

The remainder of the paper is organized as follows. In Section 2, we introduce a mathematical model that corresponds to the proposed MVSR method and illustrates its reformulation into the IF problem and the BD problem. The depth estimation method is described in Section 3. Section 4 presents the IF step, and Section 5 presents the BD step. Finally, Section 6 reports the experimental results.

## 2. Mathematical model

In multi-view settings, input images are usually captured from several cameras. We choose one as the reference camera, which is indexed as  $r$ , whereas the other cameras are indexed as  $1, 2, \dots, m$ . The HR image that we need to restore is a 2D projection of a 3D scene onto an HR grid of the selected reference camera. Let  $u$  be the desired, lexicographically ordered HR image that corresponds to the selected reference camera. The objective is to estimate  $u$  from the LR observations  $y_p, p=1, 2, \dots, m$  and  $y_r$  (which denotes the LR image captured by the reference camera). In accordance with [18, 19], we can define the forward imaging model that generates  $y_p$  as

$$y_p = DW_p H u + e_p, p = 1, 2, \dots, m ; \quad (1)$$

where the warping matrix  $W_p$  represents the displacement of the image captured from the  $p^{\text{th}}$  camera with respect to the reference camera. The matrix  $H$  denotes the total blur, which is unknown and assumed to be spatially invariant. The matrix  $D$  reflects the decimation step, whereas  $e_p$  denotes the imaging noise. Hereafter, we assume that  $H$  and  $D$  are identical for images captured by any cameras. For the LR image  $y_r$ ,  $W_r = I$ , then

$$y_r = D H u + e_r . \quad (2)$$

The desired HR image  $u$  can be also written as an HR image reconstruction process as follows:

$$u = F(u_1, u_2, \dots, u_m, u_r) + e , \quad (3)$$

where  $u_p = H^{-1} W_p^{-1} D^{-1} y_p (p = 1, 2, \dots, m)$ ,  $u_r = H^{-1} D^{-1} y_r$ ,  $e = F(n_1, n_2, \dots, n_m, n_r)$ ,  $F$  denotes a fusion process (which is introduced in Section 4), the operators  $H^{-1}$ ,  $W_p^{-1}$ , and  $D^{-1}$  are the inverse of the blurring matrix, warping matrix, and decimating matrix in the forward imaging process, respectively.  $n_p = -H^{-1} W_p^{-1} D^{-1} e_p (p = 1, 2, \dots, m)$  and  $n_r = H^{-1} D^{-1} e_r$ .

Given that the blurring matrix  $H$  is unknown, we blindly estimate the blur of the recovered SR image rather than assume that the blur is known. For simplicity, we separate the SR reconstruction process into two steps: IF and BD. The first step focuses on estimating the “blurry” HR image  $z$  as follows:

$$z = F(z_1, z_2, \dots, z_m, z_r) + s , \quad (4)$$

where  $z_p = H u_p = W_p^{-1} D^{-1} y_p (p = 1, 2, \dots, m)$ ,  $z_r = H u_r = D^{-1} y_r$ , and  $s = F(n_1, n_2, \dots, n_m, n_r)$ . Then, we obtain

$$\begin{aligned}
z &= FH(u_1, u_2, \dots, u_m, u_r) + s \\
&= KF(u_1, u_2, \dots, u_m, u_r) + s \\
&= K(F(u_1, u_2, \dots, u_m, u_r) + e - e) + s \\
&= Ku - Ke + s \\
&= Ku + n,
\end{aligned} \tag{5}$$

where  $K$  is the ultimate blur operator that should be estimated, and  $n = s - Ke$ .

The second step is to estimate the final HR image via blind deblurring of  $u$  from  $z$ , which minimizes

$$E(u, h) = \frac{\lambda}{2} \|Ku - z\|_2^2 + \rho(u) + \ell_\Omega(k), \tag{6}$$

where  $\lambda$  is a positive parameter,  $K$  is the convolution matrix constructed by the blur filter  $k$ , and  $\rho$  is a generalized total variation (GTV) regularizer given by

$$\rho(u) = \|D_x u\|_p^p + \|D_y u\|_p^p = \sum_i ([D_x u]_i^2 + [D_y u]_i^2)^{\frac{p}{2}}, \tag{7}$$

where  $D_x$  and  $D_y$  denote the derivative partial operators. Given that the distribution of gradients of natural images is more heavy-tailed than the Laplace distribution [20], we set  $0 \leq p \leq 1$ .  $\ell_\Omega$  is the indicator function of set  $\Omega$ , which is the probability simplex

$$\Omega = \{k : k \succeq 0, \|k\|_1 = 1\}. \tag{8}$$

In general, the SR method first finds pixel correspondence between non-reference images  $y_p, p = 1, 2, \dots, m$  and the reference image  $y_r$ . The process can be replaced with depth estimation if all the parameters of the cameras are known. Then, depth estimation is the premise of the IF step.

The flow diagram of the entire method is shown in Fig. 1. The steps of depth estimation, IF, and BD are detailed as follows.

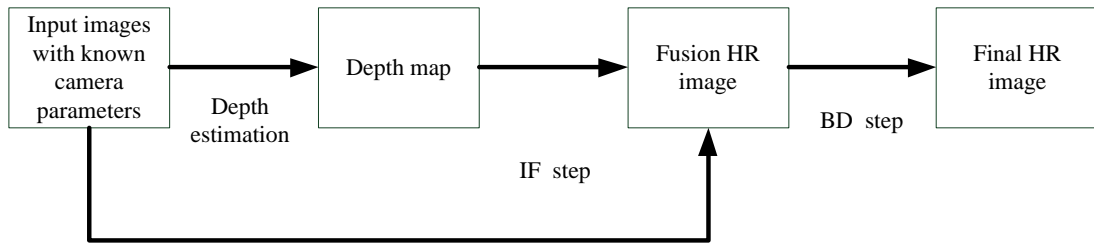


Fig. 1. Flowchart of the proposed method

### 3. Depth estimation

#### 3.1 Derivation of mapping function

We headline the mapping function first before performing depth estimation. In this study, the mapping function is equivalent to homography in multi-view geometry [21]. The mapping function  $P_{\alpha \rightarrow \beta}(q, d)$  is a function that maps a point  $q$  on camera  $\alpha$  onto camera  $\beta$  with a known depth value  $d$ . The mapping function is derived as follows.

Given the  $3 \times 4$  projection matrices for the two cameras,  $P_{(\alpha)}$  for camera  $\alpha$  and  $P_{(\beta)}$  for camera  $\beta$ , and a plane located at depth  $d$ :  $\pi^T X = 0$  with  $\pi = (0, 0, 1, -d)^T$  and  $X = (x, y, z, 1)^T$ , we first project an image point of camera  $\alpha$  onto the plane located at depth  $d$ , and then project it onto camera  $\beta$ . The homography induced by the plane is

$$\begin{pmatrix} u_\beta \\ v_\beta \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} P_{(\beta)} \\ \pi^T \end{pmatrix} \begin{pmatrix} P_{(\alpha)}^{-1} \\ \pi^T \end{pmatrix} \begin{pmatrix} u_\alpha \\ v_\alpha \\ 1 \\ 0 \end{pmatrix}, \quad (9)$$

where  $(u_\alpha, v_\alpha)$  is an image point of camera  $\alpha$ , and  $(u_\beta, v_\beta)$  is the corresponding point of camera  $\beta$ . When  $q = (u_\alpha, v_\alpha, 1)^T$ , the homography is equivalent to the mapping function  $P_{\alpha \rightarrow \beta}(q, d)$ , and the depth of point  $q$  is the precondition of the mapping function  $P_{\alpha \rightarrow \beta}(q, d)$ .

The steps to obtain the shift operator  $W_p$  ( $p = 1, 2, \dots, m$ ) are detailed as follows. For each pixel  $q$  in the reference camera,  $P_{r \rightarrow p}(q, d)$ , the mapping function from the reference camera to camera  $p$  should be determined. Then, the depth map  $G_u$  of the desired HR image  $u$  is required, which can be estimated by interpolating the depth map of the reference LR image  $y_r$ .

### 3.2 Depth estimation

The image registration process determines pixel correspondences for non-reference images with respect to the reference image. This process is equivalent to depth estimation if the parameters of the camera array are known. The reference image is typically selected as the base for which the depth value of each pixel is estimated.

Several discrete depths  $d$  are searched and distributed following a specific rule up to the minimum and maximum object depths. We quantize depth space into  $N$  levels as

$$\frac{1}{d_n} = \frac{1}{d_{\max}} + \frac{n-1/2}{N} \left( \frac{1}{d_{\min}} - \frac{1}{d_{\max}} \right) \quad (n = 1, 2, \dots, N), \quad (10)$$

where  $d_{\max}$  and  $d_{\min}$  are the minimum and maximum object depths.

The disparity estimation in [22] is translated into a multiple label energy minimization (MLEM) problem. Similarly, depth estimation in our study can also be represented as an MLEM problem. Let  $Q$  denote the set of pixels in the reference image  $y_r$ ;  $L$  be the set of depth levels  $\{1, 2, \dots, N\}$ , with the corresponding depth values of  $\{d_1, d_2, \dots, d_N\}$ ; and  $D(q)$  denotes the depth level of  $q$ . The problem of estimating the depth of each pixel in  $y_r$  is defined as follows.

Given the reference image and other non-reference images, the objective is to find the labeling  $f : Q \rightarrow L$  that assigns an appropriate depth level to each pixel in  $y_r$  such that the energy for labeling is minimized. The energy function for labeling is defined as

$$E(f) = E_{data} + E_{smoothness}, \quad (11)$$

where the data term is the sum of the data cost of each pixel in the reference image  $y_r$ . The data term measures multi-view intensity consistency.

$$E_{data} = \sum_{q \in C_r} A(q), \quad (12)$$

where  $A(q)$  is defined as

$$A(q) = \begin{cases} C(q, d(q)), & \text{if } C(q, d(q)) < \tau, \\ \tau, & \text{otherwise} \end{cases}, \quad (13)$$

where  $C(q, d(q))$  is the match cost for a pixel  $q$  in the reference image with the assigned depth value of  $d(q)$ , and  $\tau$  is a threshold for occluded pixels. In this application, we map the pixels from the reference image to all the other images by evaluating the match cost. We define  $C(q, d(q))$  as

$$C(q, d(q)) = \sum_{\{p\} \in C_{other}} \frac{1}{M} |y_p(P_{r \rightarrow p}(q, d(q))) - y_r(q)|^2, \quad (14)$$

where  $P_{r \rightarrow p}(q, d)$  is a function that maps a point  $q$  on the reference camera onto camera  $p$  with a known depth value  $d$ , and  $r$  represents the reference camera. A list  $C_{other}$  contains all the cameras except for the reference camera, and  $M$  indicates the number of cameras in the list.

The smoothness term measures the cost of assigning depth values to a pair of neighboring pixels. This term assumes that neighboring pixels should typically have similar depths, i.e.,

$$E_{smoothness} = \sum_{q \in C_r} \left( \sum_{p \in N_q} V(q, p) \right), \quad (15)$$

where  $(q, p)$  is a pair of neighboring pixels, and  $N_q$  is the neighbor of pixel  $q$ , such that  $\|p - q\|_1 = 1$ . Then, we define  $V(q, p)$  as

$$V(q, p) = \begin{cases} 0, & \text{if } |D(p) - D(q)| = 0 \\ \lambda_1, & \text{if } |D(p) - D(q)| = 1. \\ \lambda_2, & \text{if } |D(p) - D(q)| > 1 \end{cases} \quad (16)$$

As defined,  $D(p)$  and  $D(q)$  are the depth levels of pixels  $p$  and  $q$ , respectively, and  $D(p), D(q) \in \{1, 2, \dots, N\}$ .  $\lambda_1$  and  $\lambda_2$  are non-negative weights, where  $\lambda_1 \leq \lambda_2$ . In this study, we set  $\lambda_2 = 4\lambda_1$ .

To minimize energy in (11), we use an alpha–beta swap graph cuts algorithm [22, 23] to find the labeling for optimal depth. Then, for each pixel  $q$  in the reference camera, the best depth is detected.

#### 4. Image Fusion

The IF step aims to reconstruct the “blurry” image  $z$  with HR size from the LR observations  $y_p, p = 1, 2, \dots, m$  and  $y_r$ . Given

$$z = F(z_1, z_2, \dots, z_m, z_r) + s,$$

the first priority is to estimate  $z_p = W_p^{-1} D^{-1} y_p$  ( $p = 1, 2, \dots, m$ ) and  $z_r = D^{-1} y_r$ . When estimating operator  $W_p^{-1}$ , the HR image that corresponds to the  $p^{\text{th}}$  camera must be known. Then, we must first estimate  $m$  depth maps that correspond to images captured by  $m$  different cameras. We can use an equivalent operation (**Algorithm 1**) to realize  $z_p = W_p^{-1} D^{-1} y_p$  ( $p = 1, 2, \dots, m$ ). We only need estimate the depth map  $G_u$  of the HR image  $u$ .

##### Algorithm 1

1. We apply the method described in Section 3 to estimate the depth map  $G$  of the reference LR image  $y_r$ .
2. We calculate depth map  $G_u$  of the desired HR image  $u$  by interpolating  $G$ .
3. We use  $G_u$ , the projection matrix of the  $p^{\text{th}}$  camera, and the projection matrix of the reference camera with the theory in Section 3.1 to achieve the shift operator  $W_p$  ( $p = 1, 2, \dots, m$ ).
4. By using  $W_p$  ( $p = 1, 2, \dots, m$ ), we project each integer pixel point  $(x_{rh}, y_{rh})$  in  $u$  onto the HR grid of the  $p^{\text{th}}$  camera, and the corresponding position is indicated by  $(x_{ph}, y_{ph})$  ( $p = 1, 2, \dots, m$ ).
5. We translate every pixel  $(x_{ph}, y_{ph})$  in the HR grid of the  $p^{\text{th}}$  camera onto the LR grid of the  $p^{\text{th}}$  camera; the corresponding position is indicated by  $(x_{pl}, y_{pl})$ . In this study, we use  $factor = 2$  and  $(x_{pl}, y_{pl}) = \frac{(x_{ph}, y_{ph})}{factor} + 1$  ( $p = 1, 2, \dots, m$ ).
6. We use the LR observation  $y_p$  to interpolate each pixel  $(x_{pl}, y_{pl})$  and return these pixel values to their corresponding pixel positions in the reference HR image.

For  $y_r$ , we apply the bilinear interpolation method to realize  $z_r = D^{-1} y_r$ .

After estimating  $z_p$  ( $p = 1, 2, \dots, m$ ) and  $z_r$ , we can apply fusion operation  $F$  to calculate the “blurry” HR image  $z$ . For simplicity,  $F$  is set as an average operation that can statistically decrease noise, such that  $z$  can be approximated as

$$z = F(z_1, z_2, \dots, z_m, z_r). \quad (17)$$

In an average operation, we assign  $u_r$  as the reference and set a threshold  $t$ , and then estimate the  $z$  equal estimating pixel values of each pixel point  $q$  in  $z$ . For each pixel point  $q$  in  $z$ , we apply **Algorithm 2** to estimate its pixel value, which results in the “blurry” HR image  $z$ .

**Algorithm 2**

1. **Set**  $t > 0$ ,  $k = 1$ ,  $value = z_r(q)$
2. **for**  $p = 1$  to  $m$
3. **if**  $|z_p(q) - z_r(q)| < t$
4.  $value = value + z_p(q)$ ,  $k = k + 1$
5. **end**
6. **end**
7.  $value\_final = value/k$
8. **Return**  $value\_final$

## 5. Blind Deblurring

In this section, we blind estimate the desired HR image from the “blurry” HR image  $z$  obtained in the previous section.

### 5.1 Proposed algorithm framework

We obtain the following framework by alternatively minimizing (6) with respect to  $u$  and  $k$  while increasing parameter  $\lambda$ .

**Algorithm Proposed algorithmic framework**

1. **Input:** “blurry” HR image  $z$ ,  $\lambda$ , and  $\alpha > 1$ .
2. **Step I:** Blind estimation of blur filter  $k$  from  $u$  by alternatively looping over coarse-to-fine levels:
3. ▶Update the image estimate

$$\hat{u} \leftarrow \underset{u}{\operatorname{argmin}} \frac{\lambda}{2} \|\hat{K}u - z\|_2^2 + \rho(u), \quad (18)$$

where  $\hat{K}$  is the convolution matrix constructed using  $\hat{k}$  obtained from the following blur filter estimation.

4. ▶Update the blur filter estimate

$$\hat{k} \leftarrow \underset{k}{\operatorname{argmin}} \frac{\lambda}{2} \|\hat{U}k - z\|_2^2 + \ell_{\Omega}(k), \quad (19)$$

where  $\hat{U}$  is the convolution matrix constructed using  $\hat{u}$  obtained from the preceding image estimation.

5. ▶Increase parameter  $\lambda$

$$\lambda \rightarrow \alpha\lambda \quad (20)$$

6. **Step II:** Non-blind estimation of HR image  $u^*$  from  $z$  by solving (18) using final  $\hat{h}$  (obtained in Step I).



7. **Output:** HR image  $u^*$  and blur estimate  $\hat{h}$ .

Sub-problems (18) and (19) can be solved using many existing methods. In the next section, we show how these two sub-problems can be efficiently solved via ADMM.

## 5.2 ADMM optimization

Before proceeding, we first introduce ADMM [16, 17], which has been a popular tool for solving imaging inverse problems ([24] and the references therein). ADMM is suitable for addressing the general unconstrained minimization problem that comprises  $J$  sub-functions as follows:

$$\min_x \sum_{j=1}^J g_j(B^{(j)}x), \quad (21)$$

where  $B^{(j)}$  are the arbitrary matrices, and  $g_j$  are the functions. The ADMM for solving (21) presents the following form [24].

**Algorithm ADMM for solving (21)**

1. **Set**  $k = 0$ ,  $\beta > 0$ ,  $\varepsilon > 0$ ,  $v_0^{(1)}, v_0^{(2)}, \dots, v_0^{(J)}, d_0^{(1)}, d_0^{(2)}, \dots, d_0^{(J)}$

**Repeat**

2.  $r_k = \sum_{j=1}^J (B^{(j)})^T (v_k^{(j)} + d_k^{(j)})$

3.  $x_{k+1} = \left[ \sum_{j=1}^J (B^{(j)})^T B^{(j)} \right]^{-1} r_k$

**For**  $j = 1, 2, \dots, J$

4.  $v_{k+1}^{(j)} = \underset{x}{\operatorname{argmin}} \left( \frac{\tau}{2} \|x - (B^{(j)}x_{k+1} - d_k^{(j)})\|_2^2 + g_j(x) \right)$

5.  $d_{k+1}^{(j)} = d_k^{(j)} - (B^{(j)}x_{k+1} - v_{k+1}^{(j)})$

**End for**

6.  $k \leftarrow k + 1$

**until**  $\|x_{k+1} - x_k\|_2 / \|x_{k+1}\|_2 \leq \varepsilon$

Suppose  $v = B^{(j)}x_{k+1} - d_k^{(j)}$  following Almeida and Figueiredo [25]. Line 4 corresponds to the so-called Moreau proximity operator (MPO) as follows:

$$\operatorname{prox}_{\tau/g_j}(v) = \underset{x}{\operatorname{argmin}} \left( \frac{\tau}{2} \|x - v\|^2 + g_j(x) \right). \quad (22)$$

Then, we address sub-problems (18) and (19) using ADMM.

## 5.3 $u$ update using ADMM

Sub-problem (18) can be written in the form of (21), with

$$g_1(\cdot) = \frac{\lambda}{2} \|\cdot - z\|_2^2, g_2(\cdot) = g_3(\cdot) = \|\cdot\|_p^p, \quad (23)$$

$$B^{(1)} = \hat{K}, B^{(2)} = D_x, B^{(3)} = D_y. \quad (24)$$

Then, by solving (18) using ADMM, we yield the following algorithm.

**Algorithm ADMM for solving (18)**

1. **Set**  $k = 0$ ,  $\tau_1 > 0$ ,  $\varepsilon_1 > 0$ ,  $v_0^{(1)}, v_0^{(2)}, v_0^{(3)}, d_0^{(1)}, d_0^{(2)}, d_0^{(3)}$
- Repeat**
2.  $z_k^{(1)} = v_k^{(1)} + d_k^{(1)}$
3.  $z_k^{(2)} = v_k^{(2)} + d_k^{(2)}$
4.  $z_k^{(3)} = v_k^{(3)} + d_k^{(3)}$
5.  $r_k = \hat{K}^T z_k^{(1)} + D_x^T z_k^{(2)} + D_y^T z_k^{(3)}$
6.  $u_{k+1} = \left[ \hat{K}^T \hat{K} + D_x^T D_x + D_y^T D_y \right]^{-1} r_k$
7.  $v_{k+1}^{(1)} = \underset{x}{\operatorname{argmin}} \left( \frac{\tau_1}{2} \left\| x - (\hat{K}u_{k+1} - d_k^{(1)}) \right\|_2^2 + g_1(x) \right)$
8.  $d_{k+1}^{(1)} = d_k^{(1)} - (\hat{K}u_{k+1} - v_{k+1}^{(1)})$
9.  $v_{k+1}^{(2)} = \underset{x}{\operatorname{argmin}} \left( \frac{\tau_1}{2} \left\| x - (D_x u_{k+1} - d_k^{(2)}) \right\|_2^2 + g_2(x) \right)$
10.  $d_{k+1}^{(2)} = d_k^{(2)} - (D_x u_{k+1} - v_{k+1}^{(2)})$
11.  $v_{k+1}^{(3)} = \underset{x}{\operatorname{argmin}} \left( \frac{\tau_1}{2} \left\| x - (D_y u_{k+1} - d_k^{(3)}) \right\|_2^2 + g_3(x) \right)$
12.  $d_{k+1}^{(3)} = d_k^{(3)} - (D_y u_{k+1} - v_{k+1}^{(3)})$
13.  $k \leftarrow k + 1$
- until**  $\|u_{k+1} - u_k\|_2 / \|u_{k+1}\|_2 \leq \varepsilon_1$

In the preceding algorithm, line 6 is involved via the inversion of matrix  $\hat{K}^T \hat{K} + D_x^T D_x + D_y^T D_y$ , which is block-circulant. Thus, this matrix can be diagonalized via 2D discrete Fourier transform (DFT) with  $\mathcal{O}(n \log n)$  cost, and the inversion of the resulting diagonal matrix can be computed using  $\mathcal{O}(n)$  cost. Line 7 is the proximity operator of  $\tau_1/g_1$ , which can be obtained in a closed-form as follows:

$$v_{k+1}^{(1)} = \frac{\lambda z + \tau_1 (\hat{K}u_{k+1} - d_k^{(1)})}{\lambda + \tau_1}. \quad (25)$$

Lines 9 and 11 are the proximity operators of the  $\ell_p$  ( $0 \leq p \leq 1$ ) norm, and they have closed-form solutions for  $p \in \left\{ 0, \frac{1}{2}, \frac{2}{3}, 1, \frac{4}{3}, \frac{3}{2}, 2 \right\}$  [26]. For other general  $p$  values, no closed-form solution exists. However, such a solution can be precomputed numerically and used in the form of a lookup table as that in [20].

### 5.4 $k$ update using ADMM

Similarly, sub-problem (19) can be written in the form of (21), with

$$g_1(\cdot) = \frac{\lambda}{2} \|\cdot - z\|_2^2, g_2(\cdot) = \ell_{\Omega}(\cdot), \quad (26)$$

$$B^{(1)} = \hat{U}, B^{(2)} = I. \quad (27)$$

Then, by solving (19) using ADMM, we obtain the following algorithm.

**Algorithm ADMM for solving (19)**

1. **Set**  $k = 0$ ,  $\tau_2 > 0$ ,  $\varepsilon_2 > 0$ ,  $v_0^{(1)}, v_0^{(2)}, d_0^{(1)}, d_0^{(2)}$
- Repeat**
2.  $z_k^{(1)} = v_k^{(1)} + d_k^{(1)}$
3.  $z_k^{(2)} = v_k^{(2)} + d_k^{(2)}$
4.  $r_k = \hat{U}^T z_k^{(1)} + z_k^{(2)}$
5.  $k_{k+1} = [\hat{U}^T \hat{U} + I]^{-1} r_k$
6.  $v_{k+1}^{(1)} = \mathbf{argmin}_x \left( \frac{\tau_1}{2} \|x - (\hat{U}k_{k+1} - d_k^{(1)})\|_2^2 + g_1(x) \right)$
7.  $d_{k+1}^{(1)} = d_k^{(1)} - (\hat{U}v_{k+1}^{(1)} - v_{k+1}^{(1)})$
8.  $v_{k+1}^{(2)} = \mathbf{argmin}_x \left( \frac{\tau_1}{2} \|x - (k_{k+1} - d_k^{(2)})\|_2^2 + g_2(x) \right)$
9.  $d_{k+1}^{(2)} = d_k^{(2)} - (k_{k+1} - v_{k+1}^{(2)})$
10.  $k \leftarrow k + 1$
- until**  $\|k_{k+1} - k_k\|_2 / \|k_{k+1}\|_2 \leq \varepsilon_2$

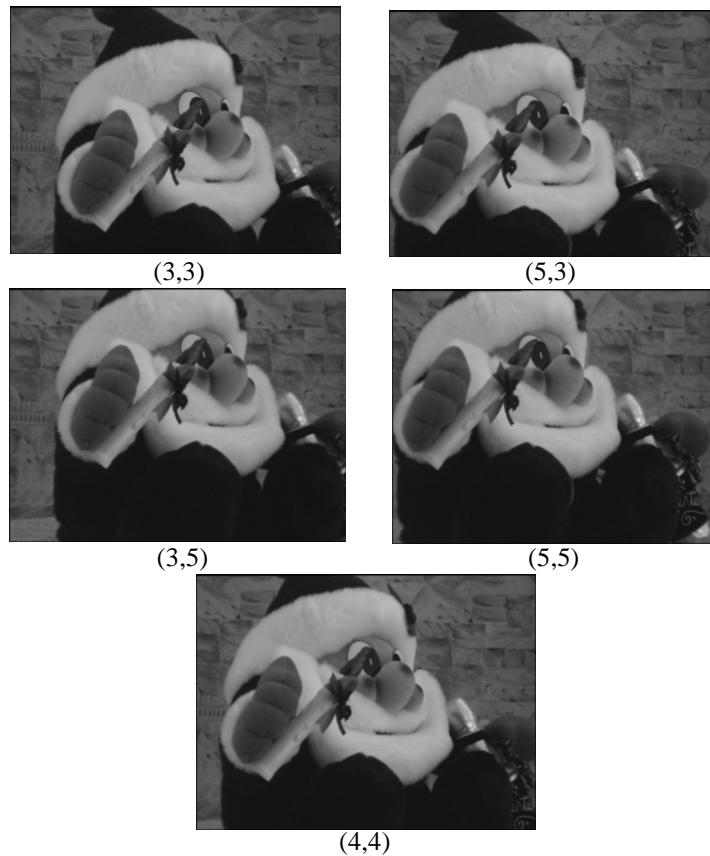
In line 5, matrix  $\hat{U}^T \hat{U} + I$  can also be diagonalized via DFT with  $\mathcal{O}(n \log n)$  cost. Line 6 can be evaluated in closed-form as (25). Line 8 is the projection onto the probability simplex  $\Omega$  in (8), which has been already addressed in [27].

## 6. Experiments

In this section, we report the detailed experimental results of the proposed method, and compare the results of the proposed method with those of the SR-FVS [13] and the ScSR + SR-FVS [14] methods. All the experiments were performed using MATLAB on a 64 bit Windows 8 personal computer with an Intel Core i7 3.6 GHz processor and 16 GB RAM. The setup of the proposed method is as follows:  $d_{max} = 1900$  mm,  $d_{min} = 300$  mm,  $N = 100$ ,  $\lambda_1 = 60$ ,  $\lambda_2 = 240$ ,  $\tau = 120$ , threshold  $t = 10$ ,  $\lambda = 1$ ,  $\alpha = 1.5$ ,  $\tau_1 = \tau_2 = 0.2$ ,  $p = 0.5$ , and  $\varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$ .

### 6.1 On non-blurred images of the doll data set

Five images from the doll data set (shown at the top of **Fig. 2**) that were obtained from the Multi-view Image Database<sup>1</sup> of the University of Tsukuba, Japan, were used as input in the experiments. The database is captured by a  $9 \times 9$  camera array. The top four input images are located at the corners of a square, i.e., (3,3), (5,3), (3,5), and (5,5), following the database notation. The size of the square was  $40 \times 40$  mm. The fifth image in **Fig. 2** is the reference LR image, which is located at the center of the square; this location is described as (4,4) using the database notation. The original images have  $640 \times 480$  pixels in color, and we only use their green channels. We reduce the images to  $320 \times 240$  pixels by downsampling for input. Output image size in this experiment is  $640 \times 480$ . The bottom image in **Fig. 2** is the original image in (4,4), which is used as the ground truth. Given the ground truth, we can use mean squared error (MSE), peak signal-to-noise ratio (PSNR), and the structural similarity (SSIM) index [28] to evaluate the SR results. The SSIM index is used to measure the similarity between two images. The resultant SSIM index is a decimal value between  $-1$  and  $1$ . A larger value indicates better result, and  $1$  is only achieved in the case of two identical images.



1. The Multi-view Image Database can be found on the website <http://www.image.esys.tsukuba.ac.jp/imagedb/>. The doll data set we used is called "Santa."



Ground truth

**Fig. 2.** Non-blurred input images and ground truth of the doll data set

**Fig. 3(a)** shows the depth map of the desired HR image, which is obtained by interpolating the depth map of the reference LR image. **Fig. 3(b)** shows the fusion HR image via the IF step. **Fig. 3(c)** shows the final HR image via the IF step and the BD step. The final HR image has more fine details than the fusion HR image. The results of the other SR methods are shown in **Fig. 4**.



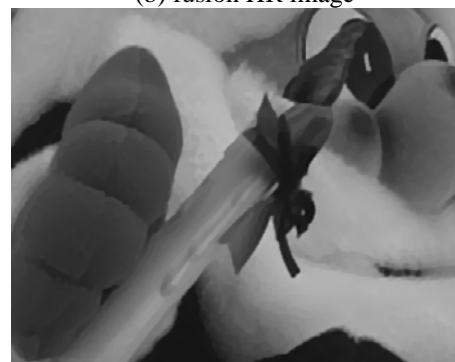
(a) depth map



(b) fusion HR image

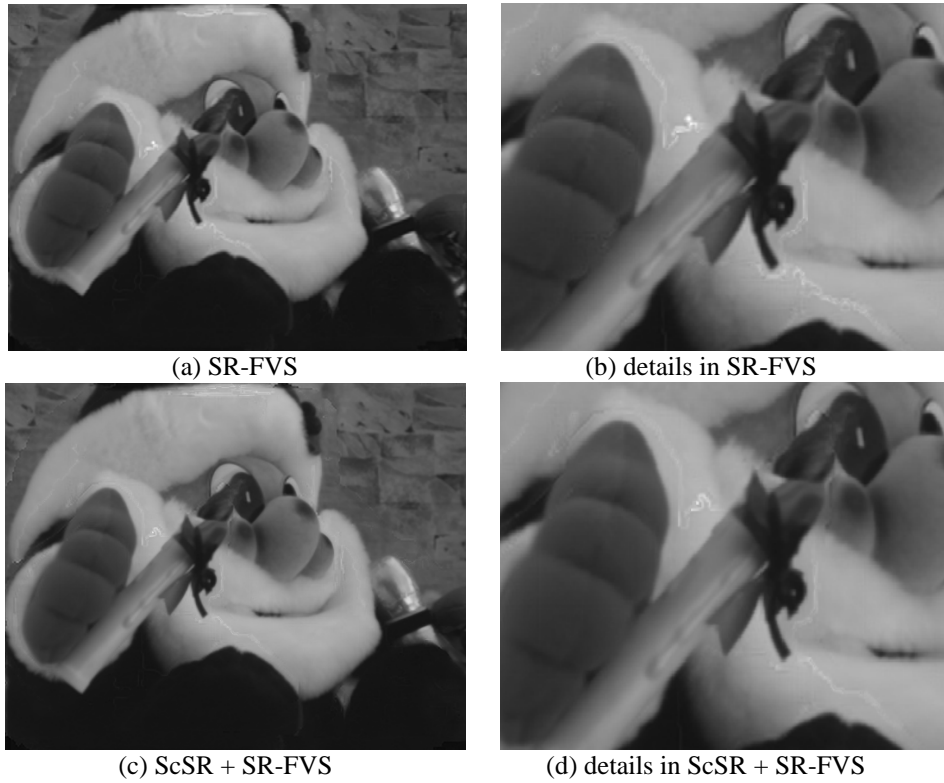


(c) final HR image



(d) details in final HR image

**Fig. 3.** Results based on the proposed method using non-blurred input images



**Fig. 4.** Results based on the other SR methods using non-blurred input images

The performance comparison among the proposed method, the SR-FVS method, and the ScSR + SR-FVS method on non-blurred images of the doll data set is presented in **Table 1**.

**Table 1.** Performance comparison on non-blurred images

	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
Proposed method	<b>14.12</b>	<b>36.63</b>	<b>0.9778</b>
SR-FVS	16.21	36.03	0.9088
ScSR+SR-FVS	15.62	36.19	0.9120

As shown in **Fig. 3**, **Fig. 4**, and **Table 1**, the proposed method exhibits outstanding performance in terms of MSE, PSNR, SSIM, and details compared with the other algorithms.

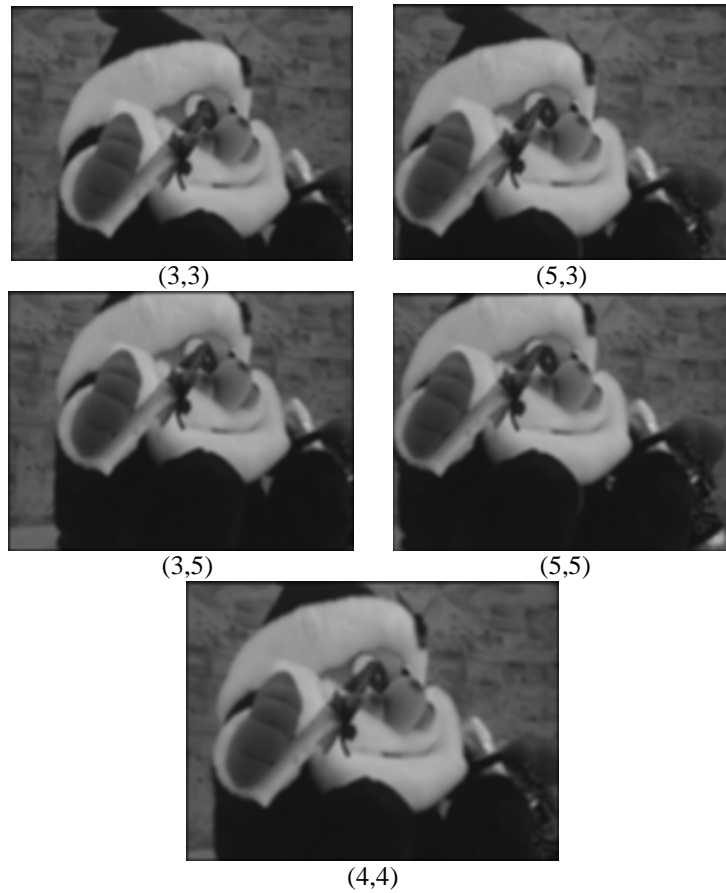
## 6.2 On blurred images of the doll data set

We construct three sequences of blurred images from five images of the doll data set. A  $5 \times 5$  uniform point spread function (PSF), a Gaussian PSF, and a motion PSF have been used for blurring and downsampling with the factor  $1/2$ . Finally, additive Gaussian noise with signal-to-noise ratio of 40 dB has been added to the LR images. The latter two PSFs are shown in **Fig. 5**. One of the sequences is shown in **Fig. 6**, in which the original image in (4,4) is selected as the ground truth.

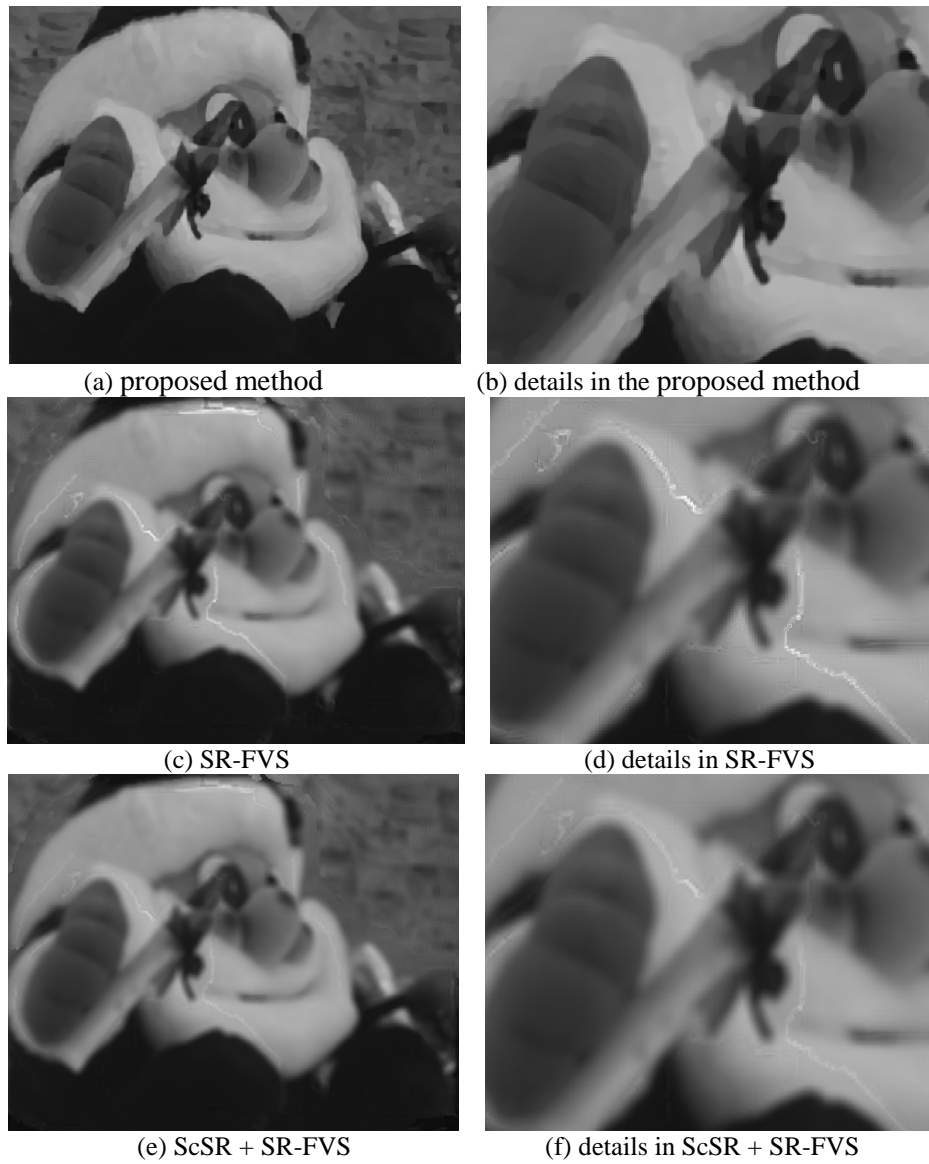
The SR results on the images blurred by the three PSFs are shown in **Figs. 7, 8, and 9**, respectively. The performance comparison of the three methods on three different blurred images from the doll data set are demonstrated in **Tables 2, 3, and 4**, respectively.



**Fig. 5.** Two PSFs used in the study



**Fig. 6.** Input images blurred via a uniform PSF

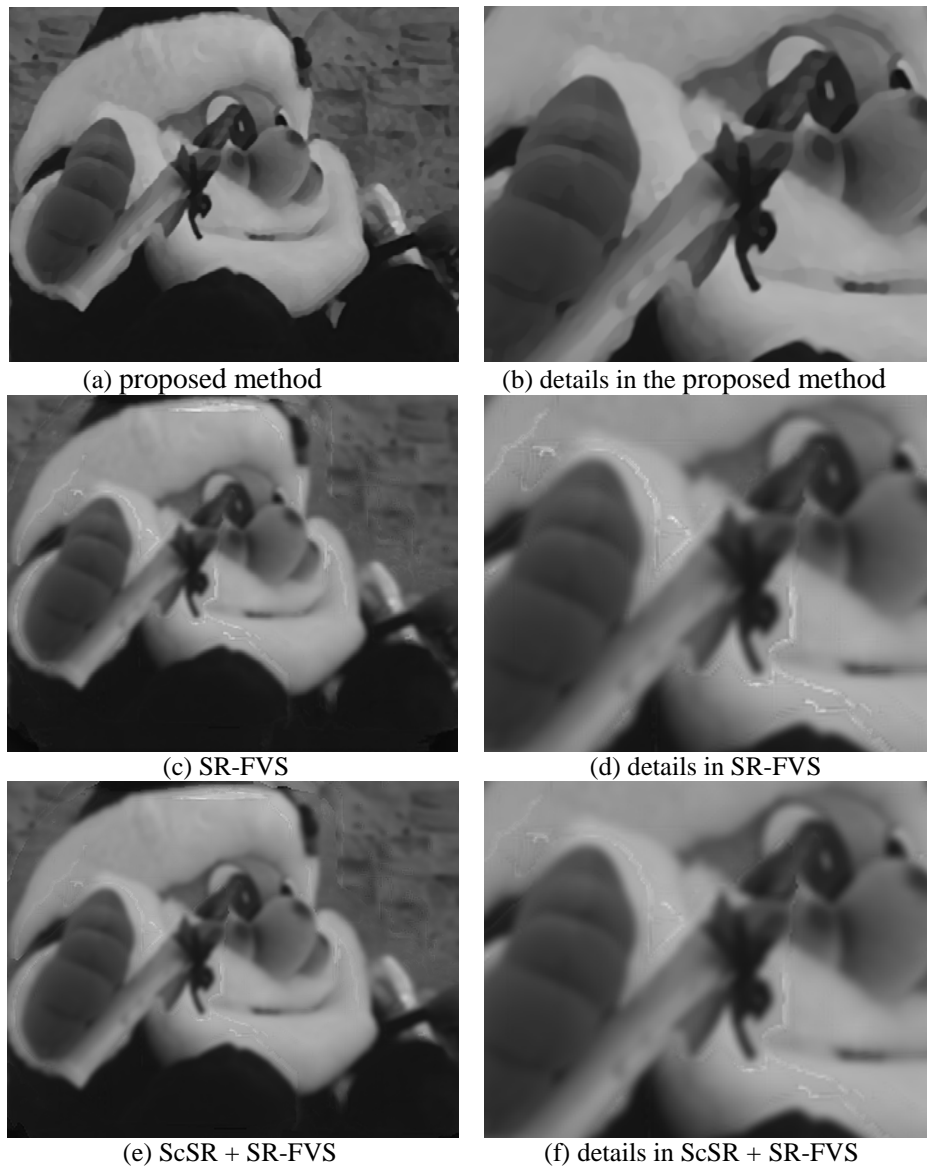


**Fig. 7.** Results using input images blurred via a uniform PSF

**Table 2.** Performance comparison on images blurred via a uniform PSF

	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
Proposed method	<b>21.29</b>	<b>34.85</b>	<b>0.9417</b>
SR-FVS	64.63	30.03	0.8440
ScSR+SR-FVS	63.82	30.08	0.8417

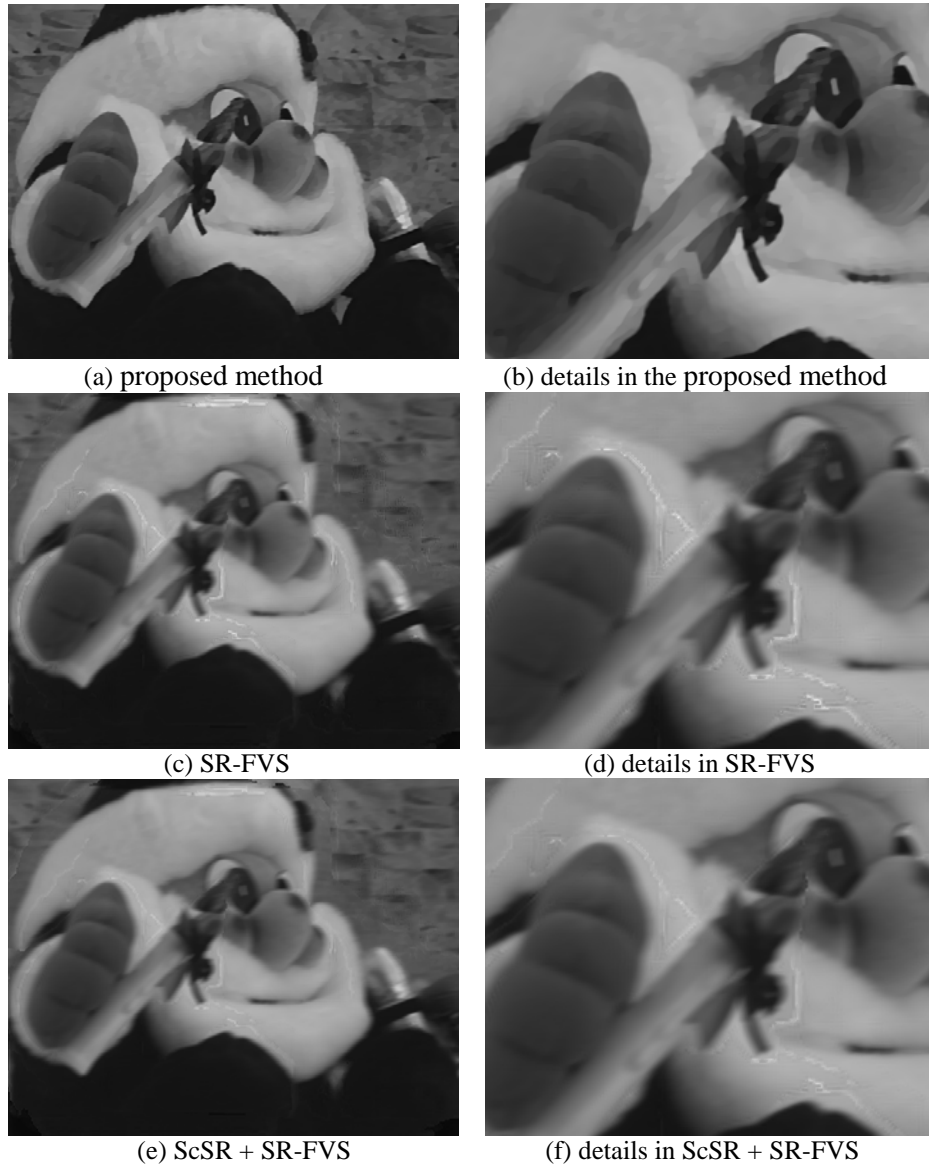




**Fig. 8.** Results using input images blurred via a Gaussian PSF

**Table 3.** Performance comparison on images blurred via a Gaussian PSF

	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
Proposed method	<b>26.66</b>	<b>33.87</b>	<b>0.9485</b>
SR-FVS	63.18	30.12	0.8514
ScSR+SR-FVS	57.71	30.52	0.8578



**Fig. 9.** Results using input images blurred via a motion PSF

**Table 4.** Performance comparison on images blurred via a motion PSF

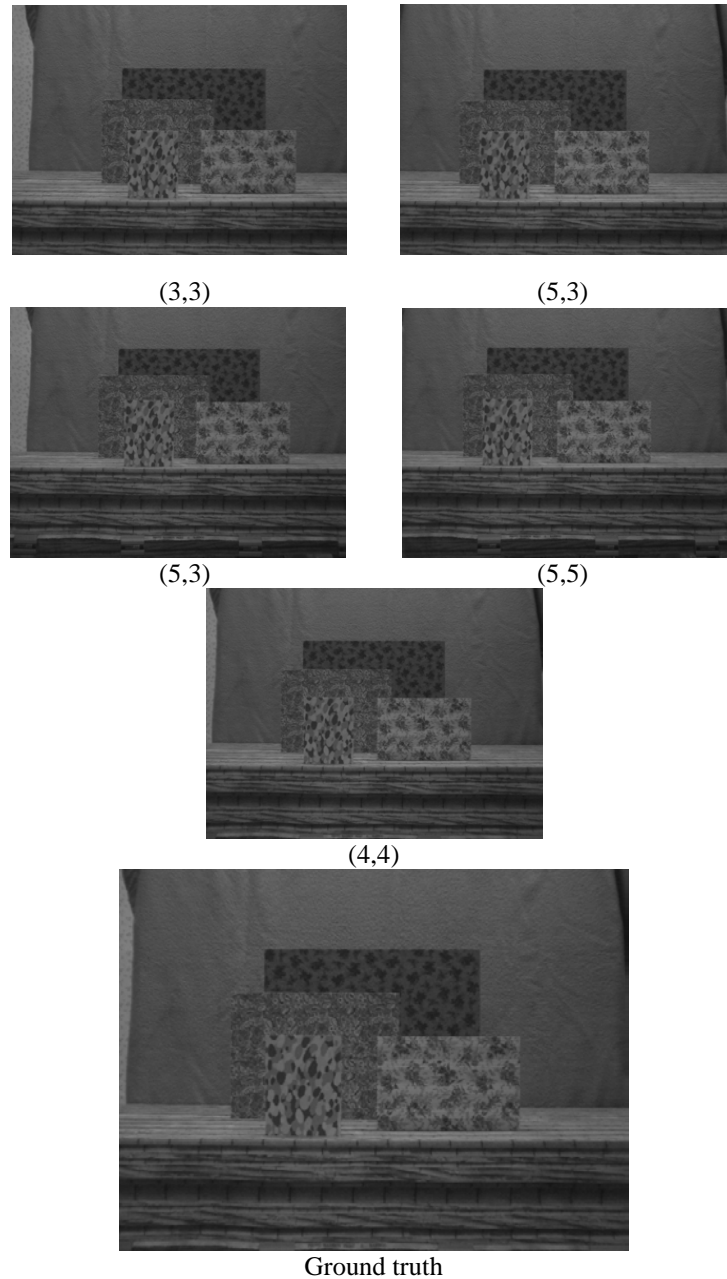
	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
Proposed method	<b>21.90</b>	<b>34.72</b>	<b>0.9641</b>
SR-FVS	54.34	30.78	0.8632
ScSR+SR-FVS	48.47	31.27	0.8694

As shown in **Figs. 7, 8, and 9**, as well as **Tables 2, 3, and 4**, the proposed method outperforms the other two methods in handling blurred images. In particular, the MSE of the proposed method is significantly smaller than those of the other methods.

### 6.3 On another image data set

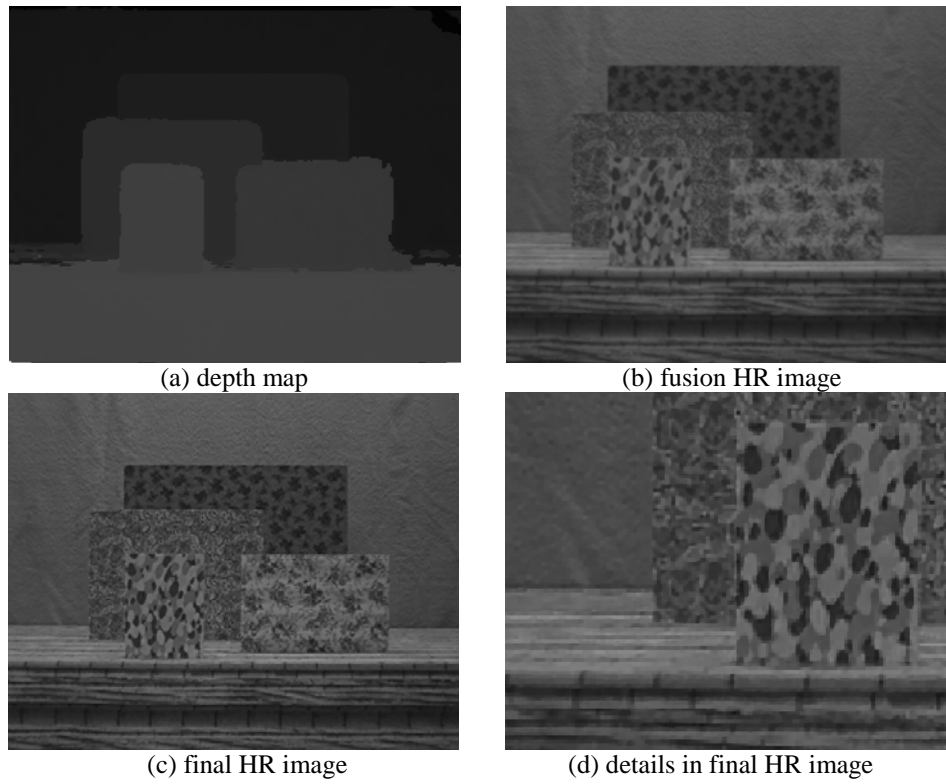
We also apply our method to a different data set, which is also included in the Multi-view Image Database of the University of Tsukuba, Japan. The database notations of the input

images are (3,3), (5,3), (3,5), (5,5), and (4,4). The input and output image sizes are  $320 \times 240$  and  $640 \times 480$ , respectively. The ground truth is the original image in (4,4), and its size is  $640 \times 480$ .

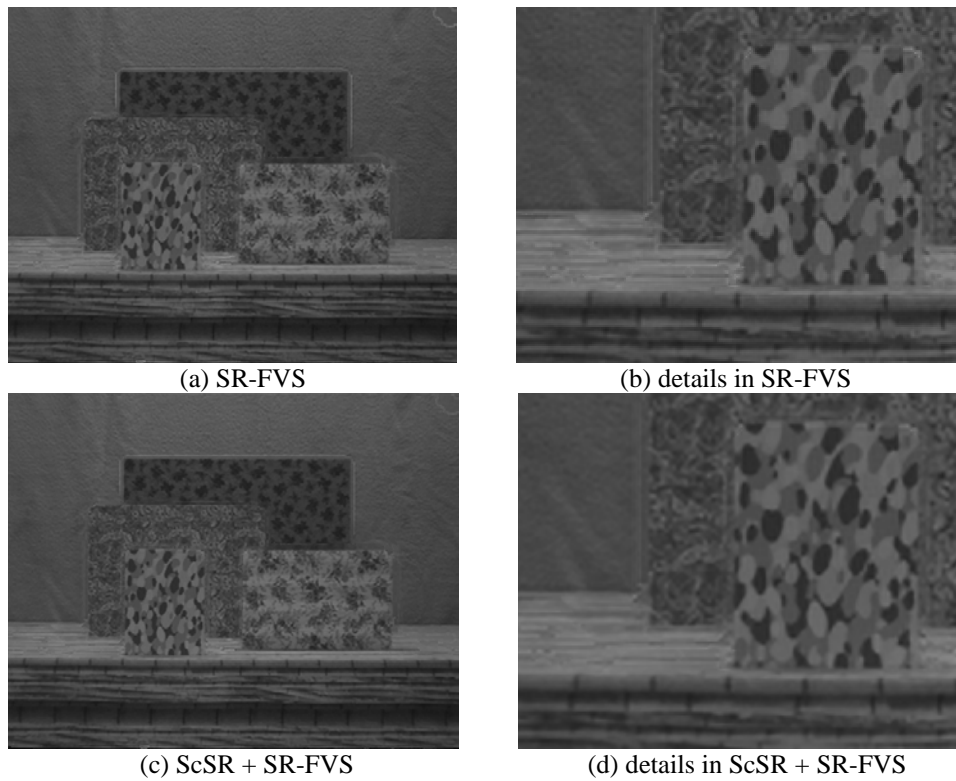


**Fig. 10.** Input images and ground truth of the board data set

**Fig. 11(a)** shows the depth map of the desired HR image, and **Fig. 11(b)** shows the fusion HR image via the IF step. **Fig. 3(c)** presents the final HR image via the IF step and the BD step. The SR results of other algorithms are shown in **Fig. 12**.



**Fig. 11.** Results based on the proposed method using input images of the board data set



**Fig. 12.** Results based on other SR methods using input images of the board data set

The comparisons of MSE, PSNR, and SSIM are shown in **Table 5**. Compared with the algorithms in [13, 14], the proposed algorithm achieves the smallest MSE, the highest PSNR, and the most significant SSIM, thereby indicating the competitiveness of the proposed method.

**Table 5.** Performance comparison on images of the board data set

	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
Proposed method	<b>32.76</b>	<b>32.98</b>	<b>0.9489</b>
SR-FVS	38.27	32.30	0.8585
ScSR+SR-FVS	33.92	32.83	0.8612

## 7. Conclusions

This study has proposed a depth-based MVSR approach, in which MVSR is addressed by solving an IF problem based on the depth map of the desired image and a BD problem using ADMM. Experiments on real and synthetic images demonstrate the effectiveness and competitiveness of the proposed method. The proposed method considers the effect of unknown blur (which typically occurs in camera array imaging) and bridges the gap between MVSR and BD, and thus, it is more suitable for camera array imaging than state-of-the-art MVSR methods. Our method also has several limitations. For example, since in the BD stage, we have to minimize the energy function (6) alternatively with respect to  $u$  and  $k$ , the optimization speed of our method is slower than most of state-of-the-art MVSR methods. In addition, we cannot guarantee that the final estimated HR image is overall optimal because our framework is divided into two stages. Future work should involve three aspects:

1. Accelerate the proposed method while preserving SR quality;
2. Propose an MVSR framework that jointly realizes IF and solves BD, such that the final estimated HR image is of higher quality;
3. Construct our own camera array system to verify above new ideas and do some further research on MVSR.

## Acknowledgment

This work is supported by the National Natural Science Foundation of China under project No. 61405252.

## References

- [1] B. Wilburn, N. Joshi, V. Vaish, et al, "High performance imaging using large camera arrays," *ACM Transactions on Graphics (TOG)*, vol.24, no.3, pp.765-776, July, 2005. [Article \(CrossRef Link\)](#).
- [2] K. Venkataraman, D. Lelescu, J. Duparré, et al, "PiCam: An ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics (TOG)*, vol.32, no.6, pp.2504-2507, November, 2013. [Article \(CrossRef Link\)](#).
- [3] C. Guillem, D. James, A. R. Harvey, "Super-resolution imaging using a camera array," *Optics letters*, vol.39, no.7, pp.1889-1892, April, 2014. [Article \(CrossRef Link\)](#).
- [4] T. Huang, R. Tsai, "Multi-frame image restoration and registration," *Advances in Computer Vision and Image Processing*, vol. 1, pp.317-339, January, 1984.

- [5] S. Farsiu, M. Dirk Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multi-frame Super-resolution," *IEEE Transactions on Image Process*, vol.13, no.10, pp.1327-1344, July, 2004. [Article \(CrossRef Link\)](#).
- [6] M. Crisani, S. C. Dong, V. Murino, and D. Pannullo, "Distilling information with super-resolution for video surveillance," in *Proc. of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks*, pp.2-11, January, 2004. [Article \(CrossRef Link\)](#).
- [7] F. Lin, C. Fookes, V. Chandran and S. Sridharan, "Investigation into optical flow super-resolution for surveillance applications," in *Proc. of the Australian Pattern Recognition Society Workshop on Digital Image Computing 2005*, pp.73-78, February, 2005. [Article \(CrossRef Link\)](#).
- [8] W. S. Yu, M. H. Wang, H. W. Chang, S. Q. Chen, "A Fast Kernel Regression Framework for Video Super-Resolution," *KSII Transactions on Internet & Information Systems*, vol.8, no.1, pp.232-248, January, 2014. [Article \(CrossRef Link\)](#).
- [9] D. Capel, "Image Mosaicing and Super-resolution," *Springer*, 2004. [Article \(CrossRef Link\)](#).
- [10] D. Capel and A. Zisserman, "Computer vision applied to super-resolution" *IEEE Signal Processing Magazine*, vol.20, no.3, pp.75-86, May, 2003. [Article \(CrossRef Link\)](#).
- [11] Qizi Huangpeng, Xiangrong Zeng, Quan Sun and Jun Fan, "Super-resolving blurry multiframe images through multiframe blind deblurring using ADMM," *Multimedia Tools and Applications*, pp.1-17, 2016. [Article \(CrossRef Link\)](#).
- [12] T. Tung, S. Nobuhara, and T. Matsuyama, "Simultaneous super-resolution and 3d video using graph-cuts," in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pp.1-8, 2008. [Article \(CrossRef Link\)](#).
- [13] K. Takahashi and T. Naemura, "Super-resolved free-viewpoint image synthesis based on view-dependent depth estimation," *IPSI Transactions on Computer Vision and Applications*, vol. 4, pp.134-148, January, 2012. [Article \(CrossRef Link\)](#).
- [14] R. Nakashima, K. Takahashi, and T. Naemura, "Super-resolved free-viewpoint image synthesis combined with sparse-representation-based super-resolution," in *Proc. of IEEE Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2013 Asia-Pacific, pp.1-6, 2013. [Article \(CrossRef Link\)](#).
- [15] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol.19, no.11, pp.2861-2873, November, 2010. [Article \(CrossRef Link\)](#).
- [16] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers and Mathematics with Applications*, vol.2, no.1, pp.17-40, 1976. [Article \(CrossRef Link\)](#).
- [17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol.3, no.1, pp.1-122, January, 2011. [Article \(CrossRef Link\)](#).
- [18] S. Lertrattanapanich, N. K. Bose, "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Trans. Image Process*, vol.11, no.12, pp.1427-1441, December, 2002. [Article \(CrossRef Link\)](#).
- [19] Z. Z. Wang, F. H. Qi, "On ambiguities in super-resolution modeling," *IEEE Signal Process. Letters*, vol.11, no.8, pp.678-681, August, 2004. [Article \(CrossRef Link\)](#).
- [20] D. Krishnan, and R. Fergus, "Fast image deconvolution using hyper-laplacian priors. In NIPS," in *Proc. of Advances in Neural Information Processing Systems*, pp.1033-1041, 2009. [Article \(CrossRef Link\)](#).
- [21] R. Hartley, A. Zisserman, "Multiple view geometry in computer vision," *Cambridge University Press*, 2000.
- [22] V. Kolmogorov, "Graph Based Algorithms for Scene Reconstruction from Two or More Views," PhD thesis, Cornell University, 2004.
- [23] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.26, no.9, pp.1124-1137, September, 2004. [Article \(CrossRef Link\)](#).

- [24] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol.20, no.3, pp.681-695, March, 2011. [Article \(CrossRef Link\)](#).
- [25] M. S. Almeida, F. Mario and M. A. T. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Transactions on Image Processing*, vol.22, no.8, pp.3074-3086, August, 2013. [Article \(CrossRef Link\)](#).
- [26] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no.4, pp.1168-1200, January, 2005. [Article \(CrossRef Link\)](#).
- [27] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the  $l_1$ -ball for learning in high dimensions," in *Proc. of 25th international conference on Machine learning*, pp.272-279, 2008. [Article \(CrossRef Link\)](#).
- [28] Z. Wang, A. C. Bovik, H. R. Sheikh, et al, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol.13, no.4, pp.600-612, April, 2004. [Article \(CrossRef Link\)](#).



**Jun Fan** received the B.S. degree in Mathematics from Wuhan University, Wuhan, Hubei, China, in 2011 and the M.S. degree in System Engineering from National University of Defense Technology, Changsha, Hunan, China, in 2013. Currently, he is working toward the Ph.D. degree in the College of Information System and Management, National University of Defense Technology. His current research interests include image super-resolution and image fusion.



**Xiangrong Zeng** received the B.S. degree in Management Engineering and M.S. degree in System Engineering from National University of Defense Technology, Changsha, Hunan, China, in 2008 and 2010, respectively, and received the Ph.D. degree in Electrical and Computer Engineering from Instituto Superior Tecnico, Universidade de Lisboa, Portugal, in 2015. He is currently an assistant professor in the College of Information System and Management, National University of Defense Technology, Changsha, China. His current research interests include computational photography and convex optimization methods in signal and image processing.



**Qizi Huangpeng** received the B.S. degree in Management engineering from National University of Defense Technology, Changsha, Hunan, China, in 2013. Currently, he is working toward the Ph.D. degree in the College of Information System and Management, National University of Defense Technology. His current research interests include pattern recognition and image super-resolution.



**Yan Liu** received the B.S. degree in Applied Mathematics from Northeastern University at Qinhuangdao, Qinhuangdao, Hebei, China, in 2013. Currently, he is working toward the M.S. degree in the College of Information System and Management, National University of Defense Technology. His current research interests include computer vision and image processing.



**Xin Long** received the B.S. degree in Management Science from Beijing Normal University, Beijing, China, in 2014. Currently, he is working toward the M.S. degree in the College of Information System and Management, National University of Defense Technology. His current research interests include computer vision and image processing.



**Jing Feng** received the B.S. and M.S. degrees in Applied Mathematics, and Ph.D. degree in Management Science and Engineering from National University of Defense Technology, Changsha, Hunan, China, in 1998, 2001 and 2004, respectively. She is currently an associate professor at the department of system engineering, College of Information System and Management, National University of Defense Technology, Changsha, China. Her current research interests include computational photography and image processing.



**Jinglun Zhou** received the B.S. degree in Applied Mathematics, M.S. degree in System Engineering and Ph.D. degree in Management Science and Engineering from National University of Defense Technology, Changsha, Hunan, China, in 1981, 1988 and 2000, respectively. He is currently a professor at the department of system engineering, College of Information System and Management, National University of Defense Technology, Changsha, China. His current research interests include computer vision and image processing.