A Multi-view Super-Resolution Method with Joint-optimization of Image Fusion and Blind Deblurring

Jun Fan, Yue Wu, Xiangrong Zeng, Qizi Huangpeng, Yan Liu, Xin Long, Jinglun Zhou

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

Changsha, China

[E-mail: fanjun891009@163.com]

*Corresponding author: Jun Fan

Received April 17, 2017; revised December 5, 2017; accepted January 8, 2018;

published May 31, 2018

Abstract

Multi-view super-resolution (MVSR) refers to the process of reconstructing a high-resolution (HR) image from a set of low-resolution (LR) images captured from different viewpoints typically by different cameras. These multi-view images are usually obtained by a camera array. In our previous work [1], we super-resolved multi-view LR images via image fusion (IF) and blind deblurring (BD). In this paper, we present a new MVSR method that jointly realizes IF and BD based on an integrated energy function optimization. First, we reformulate the MVSR problem into a multi-channel blind deblurring (MCBD) problem which is easier to be solved than the former. Then the depth map of the desired HR image is calculated. Finally, we solve the MCBD problem, in which the optimization problems with respect to the desired HR image and with respect to the unknown blur are efficiently addressed by the alternating direction method of multipliers (ADMM). Experiments on the Multi-view Image Database[1] of the University of Tsukuba and images captured by our own camera array system demonstrate the effectiveness of the proposed method.

Keywords: Multi-view super-resolution, Multi-channel blind deblurring, Alternating direction method of multipliers

1. The Multi-view Image Database can be found on the website http://www.image.esys.tsukuba.ac.jp/imagedb/. The dataset we used is called “board”.

1. Introduction

Multi-view super-resolution (MVSR) aims to estimate a high resolution (HR) image from multi-view low-resolution (LR) images, and it has been extensively investigated in recent years. An obvious benefit of MVSR is that it can achieve a HR image through several inexpensive cameras instead of an expensive one, such as in [2, 3, 4, 5], and camera arrays were designed based on MVSR. Specifically, Wilburn et al. [2] used multiple inexpensive cameras to approximate a video camera with a large synthetic aperture to obtain HR videos. Venkataraman et al. [3] designed Pelican Imaging Camera Array (PiCam) for mobile devices to acquire multi-view images and reconstruct an HR image through a post-capture super-resolution process. Carles et al. [4] presented a practical demonstration of super-resolution (SR) imaging using an array of 25 independent commercial off-the-shelf cameras, and then constructed a Multi-aperture (MA) imaging system employing a 3 × 3 array of free-form lenslets to realize SR imaging in [5].

Mudenagudi et al. [6] super-resolved the generated novel views of a 3D scene using multi-view images obtained from cameras in general positions. This method simultaneously realized MVSR and handled occlusion. A multi-view video SR approach was proposed by Najafi [7], in which the multi-view video SR problem is divided into LR frame fusion and deblurring. Meanwhile, a frame fusion scheme for multi-view video is proposed and exploiting multi-view sequence instead of single-view for frame fusion is studied in this thesis. Takahashi et al. [8] proposed a super-resolution-free-viewpoint image synthesis (SR-FVS) method that simultaneously realized free-viewpoint image SR and free-viewpoint depth estimation. Nakashima et al. [9] integrated a learning-based SR method, namely, sparse coding SR (ScSR) [10], into an existing SR-FVS method [8] to improve the quality of the desired HR image of the target viewpoint. Venkataraman et al. [3] constructed an ultra-thin high-performance monolithic camera array to acquire multi-view images; a two-stage (fusion SR stage and maximum-a-posteriori (MAP) SR stage) MVSR construction is then applied to obtain the final HR image that corresponded to a selected reference camera.

The preceding MVSR methods either ignored the effect of blur [8, 9] or assumed that blur was known [3, 6, 7]. Nevertheless, the blur kernel was typically unknown in practical situations. In our previous study [1], we reformulated the MVSR problem into an IF problem and a BD problem. This MVSR approach can accurately estimate depth maps and desired HR images from multi-view input LR images while handling unknown blur. Experiments in [1] were realized on multi-view image database of the University of Tsukuba.

We cannot guarantee that the final estimated HR image is overall optimal, because the entire MVSR reconstruction was divided into two stages in [3, 7, 8, 9]. In this paper, we derive the idea for handling blur in [1]; and generate an integrated MVSR framework that jointly realizes IF and solves BD based on an integrated energy function optimization. We

actually reformulate the MVSR problem into a special MCBD problem [11]. Furthermore, we efficiently address the MCBD problem via alternately optimizing the resulting sub-problems with respect to the unknown HR image and the unknown blur by using the ADMM [12, 13, 14, 15]. We also construct our own camera array system, and corresponding experimental results on the multi-view image dataset obtained by our system demonstrate the utilization and robustness of the presented method in real applications. We basically realized the prospect in our previous study [1].

The rest of the paper is organized as follows. In Section 2 we introduce the MVSR forward model, and the manner in which an MVSR problem is reformulated into an MCBD problem. Section 3 shows the basics of the MCBD problem solved by the ADMM and provides the details of the optimization process. Section 4 reports the experimental results and illustrates the performance of the presented method. Finally, a conclusion is shown in Section 5.

2. Problem Formulation

2.1 The observation model for MVSR

In multi-view settings, input images are usually captured by a camera array. Our camera array system consists of 16 low-cost cameras assembled into a plane array as shown in **Fig. 1**. Each individual camera employs an f/3.1 lens with a focal length of 12 mm. The resolution of each camera is 2592 × 1944 pixels, and the camera centers are separated by 45 mm.

Fig. 1. Our 4 × 4 camera array system

In the system, we choose one as the reference camera, which is indexed as r, while the other cameras are indexed as $1,2,...,m$. Let u be the desired, lexicographically ordered, HR image corresponding to the selected reference camera. The goal is to estimate u from the lexicographically ordered LR images $y_p, p=1,2,...,m$. In accordance with [16, 17], the forward imaging model generating each y_p can be modeled as:

$$y_p = DW_pHu + e_p, p=1,2,...,m, \tag{1}$$

where D denotes the downsampling matrix, and the warping matrix W_p represents the displacement of the image captured from the p^{th} camera with respect to the reference camera. The matrix H denotes the total blur, which is unknown and assumed as spatially-invariant, and e_p is the imaging noise. We assume that H and D are identical for all the cameras in the camera array system.

Let the size of each y_p be $MN \times 1$ (the size of the LR image is $M \times N$), the downsampling factor be k (in our paper $k=2$), the size of u be $k^2MN \times 1$ (the size of the HR image is $kM \times kN$), and the sizes of D, $W_p(p=1,2,...,m)$, H are $MN \times k^2MN$, $k^2MN \times k^2MN$ and $k^2MN \times k^2MN$, respectively.

The decimation matrix D is a simple matrix wherein only four element equal $1/4$ in each row. We note the decimation matrix of 1D signal as D_1, then for a 1D signal, $k=2$, $D_1 = \left[\frac{1}{2}\frac{1}{2}000...;00\frac{1}{2}\frac{1}{2}0...;0000\frac{1}{2}\frac{1}{2}...;...\right]$. Given that image is a kind of 2D signal, the decimation matrix D in the forward imaging model is:

$$D = D_1 \otimes D_1, \tag{2}$$

where $\otimes$ denotes the Kroneker product.

According to multi-view geometry [18] and SR-FVS [8], the warping matrices $W_p(p=1,2,...,m)$ can be estimated by using the depth map of the reference image and multi-view geometry knowledge. Herein, we use the depth estimation method in our previous study [1] to obtain the depth map of the reference image. The depth estimation problem is actually reformulated into a multiple label energy minimization problem, and the energy function for labeling is defined as

$$E(f) = \sum_{q \in C_r} A(q) + \sum_{q \in C_r} \left(\sum_{p \in N_q} V(q,p) \right), \tag{3}$$

where C_r represents the reference camera, (q,p) is a pair of neighboring pixels, and N_q is the neighbor of pixel q, such that $\|p-q\|_1 = 1$. We define $A(q)$ as

$$A(q)=\begin{cases} C(q,d(q)), & if \quad C(q,d(q))<\tau \\ \tau, & otherwise \end{cases}. \tag{4}$$

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

$$C\left(q,d\left(q\right)\right)=\sum_{\{p\}\in C_{other}}\frac{1}{M}\left|y_p\left(P_{r\to p}\left(q,d\left(q\right)\right)\right)-y_r\left(q\right)\right|^2, \tag{5}$$

where $P_{r\to p}(q,d)$ 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. The list C_{other} contains all the cameras except for the reference camera, and M indicates the number of cameras in the list.

Meanwhile, $V(q,p)$ in (3) is defined as

$$V\left(q,p\right)=\begin{cases} 0, if\left|D\left(p\right)-D\left(q\right)\right|=0 \\ \lambda_1, if\left|D\left(p\right)-D\left(q\right)\right|=1 \\ \lambda_2, if\left|D\left(p\right)-D\left(q\right)\right|>1 \end{cases}. \tag{6}$$

As defined, $D(p)$ and $D(q)$ are the depth levels of pixels p and q, respectively, and $D(p),D(q)\in\{1,2,...,N\}$. λ_1 and λ_2 are non-negative weights, wherein $\lambda_1\le\lambda_2$.

After estimating the depth map of the reference image, we refer to [8] in constructing the warping matrices $W_p(p=1,2,...,m)$ in our paper. The pseudo-code of constructing $W_p(p=1,2,...,m)$ is provided in Chapter 4.2 (Details of Transformation) in [8].

2.2 Basics of MCBD problem

We construct our optimization problem as follows

$$\min_{u,h}\frac{\lambda}{2m}\sum_{p=1}^{m}\left\|DW_pHu-y_p\right\|_2^2+\rho(u)+\ell_\Omega(h), \tag{7}$$

where λ is a positive parameter in the data, and the term H is the convolution matrix constructed by the blur filter h. Above $\rho(u)$ and $\ell_\Omega(h)$ are the regularization terms on image and blur, respectively. $\rho(u)$ is a generalized total variation (GTV) regularizer. In accordance with our previous work [19], we adopted a four direction version of GTV:

$$\rho(u)=\Phi\left(D_{xx}u,D_{xy}u,D_{yx}u,D_{yy}u\right)=\sum_i\left(\left|[D_{xx}u]_i\right|^q+\left|[D_{xy}u]_i\right|^q+\left|[D_{yx}u]_i\right|^q+\left|[D_{yy}u]_i\right|^q\right), \tag{8}$$

where D_{xx} and D_{yy} are the derivative partial operators performing derivatives with respect to x and y, respectively. Given that the distribution of gradients of natural images is more heavy-tailed than the Laplace distribution [20], we set $0 \le q \le 1$. Instead of only considering the image gradient in horizontal and vertical direction, (8) expands the image gradient into horizontal, vertical and two perpendicular diagonal directions. D_{xy} and D_{yx} are two perpendicular diagonal derivative operators. ℓ_Ω is the indicator function of set Ω, which is the probability simplex

$$\Omega = \left\{ k : k \succeq 0, \left\| k \right\|_1 = 1 \right\}. \tag{9}$$

From the preceding discussion, we then focus on the following MCBD problem:

$$\left(\hat{u}, \hat{h} \right) = \arg\min_{u,h} \frac{\lambda}{2m} \sum_{p=1}^{m} \left\| DW_p Hu - y_p \right\|_2^2 + \Phi\left(D_{xx}u, D_{xy}u, D_{yx}u, D_{yy}u \right) + \ell_\Omega(h), \tag{10}$$

which will be addressed by the ADMM.

3. Optimization of MCBD

The MCBD problem (10) can be solved by alternatively minimizing with respect to u and h. Each sub-problem can be efficiently solved by the ADMM.

3.1 Algorithm framework

We present the following framework by alternatively minimizing (10), which comprises two steps: minimization with respect to u and minimization with respect to h.

Algorithm *Proposed algorithmic framework*

1. **Input**: LR images captured by camera array $y_p, p = 1, 2, ..., m$, λ, and $\alpha > 1$.

2. **Step I**: Blind estimation of the blur filter h from u by alternatively looping over coarse-to-fine levels:

3. ▶Update the image estimate

$$\hat{u} \leftarrow \underset{u}{\mathbf{argmin}} \frac{\lambda}{2m} \sum_{p=1}^{m} \left\| DW_p \hat{H}u - y_p \right\|_2^2 + \Phi\left(D_{xx}u, D_{xy}u, D_{yx}u, D_{yy}u \right) \lambda, \tag{11}$$

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

4. ▶Update the blur filter estimate

$$\hat{h} \leftarrow \underset{h}{\mathbf{argmin}} \frac{\lambda}{2m} \sum_{p=1}^{m} \left\| DW_p \hat{U} h - y_p \right\|_2^2 + \ell_\Omega(h) \ \lambda, \tag{12}$$

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

5. ▶Increase parameter λ

$$\lambda \to \alpha\lambda \tag{13}$$

6. **Step II**: Non-blind estimation of HR image u^* from y_p, $p = 1, 2, \ldots, m$ by solving (11) using final $\hat{h}$ (obtained in Step I).

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

Sub-problems (11) and (12) can be efficiently solved via ADMM, which is suitable for addressing the general unconstrained minimization problem that comprises J sub-functions as follows:

$$\min_x \sum_{j=1}^{J} g_j \left(B^{(j)} x \right), \tag{14}$$

where $B^{(j)}$ are the arbitrary matrices, and g_j are the functions.

3.2 u update using ADMM

Sub-problem (11) can be written in the form of (14), with

$$g_p(\cdot) = \frac{\lambda}{2m} \left\| \cdot - y_p \right\|_2^2, \ p = 1, 2, \ldots, m, \tag{15}$$

$$g_{m+1}(\cdot) = g_{m+2}(\cdot) = g_{m+3}(\cdot) = g_{m+4}(\cdot) = \left\| \cdot \right\|_q^q, \tag{16}$$

$$B^{(p)} = DW_p \hat{H}, \ p = 1, 2, \ldots, m, \tag{17}$$

$$B^{(m+1)} = D_{xx}, B^{(m+2)} = D_{xy}, B^{(m+3)} = D_{yx}, B^{(m+4)} = D_{yy}. \tag{18}$$

By solving (11) using ADMM, we then derive the following algorithm.

Algorithm *ADMM for solving (11)*

1. **Set** $k=0$, $\tau_1>0$, $\varepsilon_1>0$, $v_0^{(p)}, p=1,2,...,m+4$, $d_0^{(p)}, p=1,2,...,m+4$

   Repeat

2. $z_k^{(p)}=v_k^{(p)}+d_k^{(p)}, p=1,2,...,m+4$

3. $r_k=\sum_{p=1}^{m}(DW_p\hat{H})^T z_k^{(p)}+D_{xx}^T z_k^{(m+1)}+D_{xy}^T z_k^{(m+2)}+D_{yx}^T z_k^{(m+3)}+D_{yy}^T z_k^{(m+4)}$

4. $u_{k+1}=\left[\sum_{P=1}^{m}(DW_p\hat{H})^T DW_p\hat{H}+D_{xx}^T D_{xx}+D_{xy}^T D_{xy}+D_{yx}^T D_{yx}+D_{yy}^T D_{yy}\right]^{-1} r_k$

5. $v_{k+1}^{(p)}=\underset{x}{\mathbf{argmin}}(\frac{\tau_1}{2}\left\|x-(DW_p\hat{H}u_{k+1}-d_k^{(p)})\right\|_2^2+g_p(x)), p=1,2,...,m,$

6. $d_{k+1}^{(p)}=d_k^{(p)}-(DW_p\hat{H}u_{k+1}-v_{k+1}^{(p)}), p=1,2,...,m,$

7. $v_{k+1}^{(m+1)}=\underset{x}{\mathbf{argmin}}(\frac{\tau_1}{2}\left\|x-(D_{xx}u_{k+1}-d_k^{(m+1)})\right\|_2^2+g_{m+1}(x))$

8. $d_{k+1}^{(m+1)}=d_k^{(m+1)}-(D_{xx}u_{k+1}-v_{k+1}^{(m+1)})$

9. $v_{k+1}^{(m+2)}=\underset{x}{\mathbf{argmin}}(\frac{\tau_1}{2}\left\|x-(D_{xy}u_{k+1}-d_k^{(m+2)})\right\|_2^2+g_{m+2}(x))$

10. $d_{k+1}^{(m+2)}=d_k^{(m+2)}-(D_{xy}u_{k+1}-v_{k+1}^{(m+2)})$

11. $v_{k+1}^{(m+3)}=\underset{x}{\mathbf{argmin}}(\frac{\tau_1}{2}\left\|x-(D_{yx}u_{k+1}-d_k^{(m+3)})\right\|_2^2+g_{m+3}(x))$

12. $d_{k+1}^{(m+3)}=d_k^{(m+3)}-(D_{yx}u_{k+1}-v_{k+1}^{(m+3)})$

13. $v_{k+1}^{(m+4)}=\underset{x}{\mathbf{argmin}}(\frac{\tau_1}{2}\left\|x-(D_{yy}u_{k+1}-d_k^{(m+4)})\right\|_2^2+g_{m+4}(x))$

14. $d_{k+1}^{(m+4)}=d_k^{(m+4)}-(D_{yy}u_{k+1}-v_{k+1}^{(m+4)})$

15. $k\leftarrow k+1$

   until $\left\|u_{k+1}-u_k\right\|_2/\left\|u_{k+1}\right\|_2\le\varepsilon_1$

Line 4 is involved in the preceding algorithm via the inversion of matrix $\sum_{P=1}^{m}(DW_p\hat{H})^T DW_p\hat{H}+D_{xx}^T D_{xx}+D_{xy}^T D_{xy}+D_{yx}^T D_{yx}+D_{yy}^T D_{yy}$, which is block-circulant. Thus it can be diagonalized by 2D discrete Fourier transform (DFT) with $O(N\log N)$ cost [14], and the

inversion of the resulting diagonal matrix can be computed with $O(N)$ cost. Suppose $v^{(p)} = DW_p\hat{H}u_{k+1} - d_k^{(p)}, p = 1,2,...,m$, following Almeida and Figueiredo [21], line 5 corresponds to the so-called Moreau proximity operator as follows:

$$prox_{\tau/g_p}(v^{(p)}) = \arg\min_x \left(\frac{\tau}{2}\|x - v^{(p)}\|^2 + g_p(x)\right), p = 1,2,...,m, \tag{19}$$

which can be obtained in a closed-form as follows:

$$v_{k+1}^{(p)} = \frac{\frac{\lambda}{m}y_p + \tau_1\left(DW_p\hat{H}u_{k+1} - d_k^{(p)}\right)}{\frac{\lambda}{m} + \tau_1}, p = 1,2,...,m. \tag{20}$$

Lines 9 and 11 are the proximity operators of the $\ell_q(0 \le q \le 1)$ norm, and they have closed-form solutions for $q \in \left\{0, \frac{1}{2}, \frac{2}{3}, 1, \frac{4}{3}, \frac{3}{2}, 2\right\}$ [22]. For other general q values, no closed-form solution exists. However, such a solution can be numerically pre-computed and used in the form of a lookup table as that in [20]. In our paper, we set $q = 0.3$

3.3 h update using ADMM

Similarly, sub-problem (12) can be written in the form of (14), with

$$g_p(\cdot) = \frac{\lambda}{2m}\|\cdot - y_p\|_2^2, p = 1,2,...,m, \tag{21}$$

$$g_{m+1}(\cdot) = \ell_\Omega(\cdot) \tag{22}$$

$$B^{(p)} = DW_p\hat{U}, p = 1,2,...,m. \tag{23}$$

$$B^{(m+1)} = I \tag{24}$$

By solving (12) using ADMM, we then obtain the following algorithm.

Algorithm *ADMM for solving (12)*

1. **Set** $k = 0$, $\tau_2 > 0$, $\varepsilon_2 > 0$, $v_0^{(p)}, p = 1,2,...,m+1$, $d_0^{(p)}, p = 1,2,...,m+1$

   Repeat

2. $z_k^{(p)} = v_k^{(p)} + d_k^{(p)}, p = 1,2,...,m+1$

3. $r_k = \sum_{p=1}^{m}(DW_p\hat{U})^T z_k^{(p)} + z_k^{(m+1)}$

4. $h_{k+1} = \left[\sum_{P=1}^{m} (DW_P \hat{U})^T DW_P \hat{U} + I \right]^{-1} r_k$

5. $v_{k+1}^{(p)} = \underset{x}{\mathbf{argmin}} \left(\frac{\tau_2}{2} \left\| x - (DW_P \hat{U} h_{k+1} - d_k^{(p)}) \right\|_2^2 + g_p(x) \right), p = 1,2,...,m,$

6. $d_{k+1}^{(p)} = d_k^{(p)} - (DW_p \hat{U} h_{k+1} - v_{k+1}^{(p)}), p = 1,2,...,m,$

7. $v_{k+1}^{(m+1)} = \underset{x}{\mathbf{argmin}} \left(\frac{\tau_2}{2} \left\| x - (h_{k+1} - d_k^{(m+1)}) \right\|_2^2 + g_{m+1}(x) \right)$

8. $d_{k+1}^{(m+1)} = d_k^{(m+1)} - (h_{k+1} - v_{k+1}^{(m+1)})$

9. $k \leftarrow k+1$

until $\left\| h_{k+1} - h_k \right\|_2 / \left\| h_{k+1} \right\|_2 \le \varepsilon_2$

Line 5 can be evaluated in a closed-form as (20). Line 7 is the projection onto the probability simplex Ω in (9), which has been already addressed in [23].

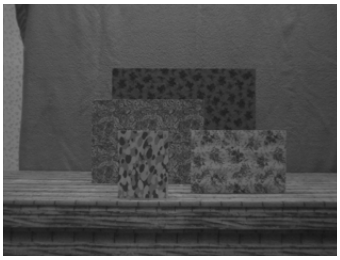
4. Experiments

In this section, we compare the proposed method with SR-FVS [8], SR-FVS + ScSR [9] and depth-based MVSR method [1]. 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. To demonstrate the effectiveness of the proposed method, we conducted experiments on two image database. One is the Multi-view Image Database of the University of Tsukuba, and the other is the image database obtained by our camera array system.

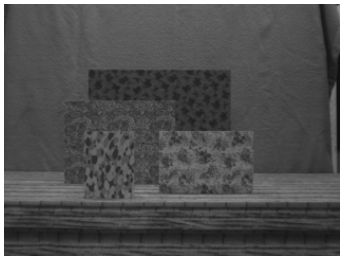
6.1 On images of the Multi-view Image Database

We first conducted experiment on images of the Multi-view Image Database of the University of Tsukuba. The setup of the proposed method for the Multi-view Image Database of the University of Tsukuba is as follows: d_{max} = 1900 mm, d_{min} = 300 mm, N = 100, λ_1 = 60, λ_2 = 240, τ = 120, λ = 1, α = 1.5, $\tau_1 = \tau_2$ = 0.2 , q = 0.3 and $\varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$ (in which d_{max} , d_{min} , N , λ_1 , λ_2 and τ are used for depth estimation). Five images of the board dataset (shown at the top of **Fig. 2**) in the Multi-view Image Database, were used as input in the experiment. The database is captured by a 9×9 camera array. The top four input images are indexed as (3,3), (5,3), (3,5), and (5,5), following the database notation. The fifth image in **Fig. 2** is the reference LR image, which

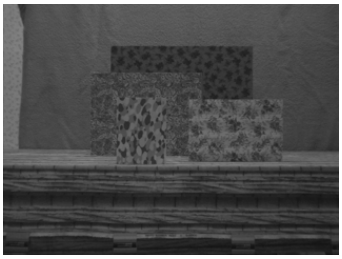
is located at the center of other four input images.; this location is described as (4,4) using the database notation. The original images have 640×480 pixels in color. We converted the images to gray-scale and reduced them to 320×240 pixels by downsampling for input. Output image size in this experiment is 640×480. The bottom image in **Fig. 2** is the original image in (4,4), which is used as the ground truth. Similar to what we conducted in [1], we use mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) index [24] 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.



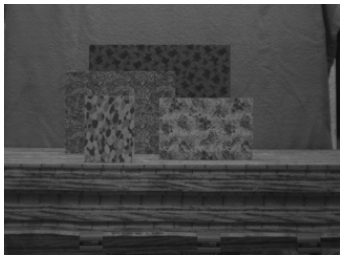
(3, 3)



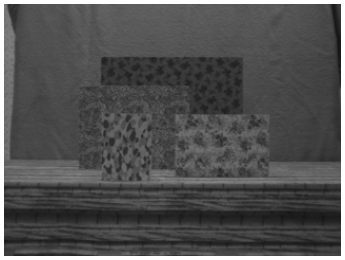
(5, 3)



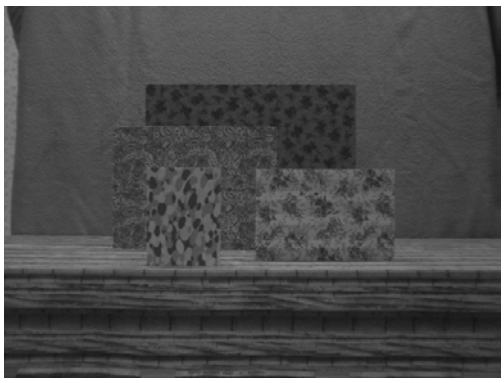
(3, 5)



(5, 5)



(4, 4)

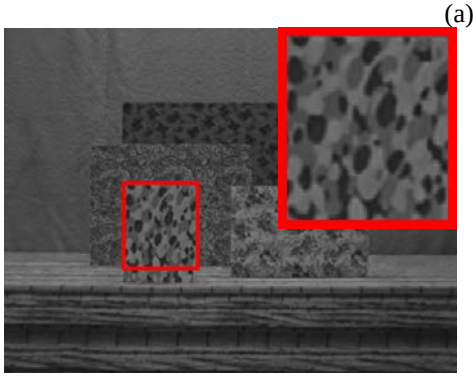


Ground truth

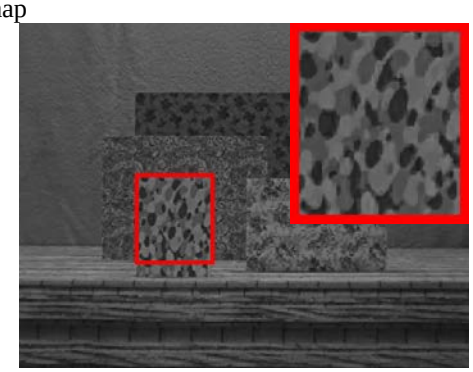
Fig. 2. Input images and ground truth of the board dataset

The depth map of the desired HR image is shown in **Fig. 3(a)**, and the LR reference image is presented in **Fig. 3(f)**. As can be seen from **Fig. 3(b)-(e)**, only the results of the proposed method and depth-based method obviously recover details on the board. To better assess the estimated results, the close-up images of board are shown in **Fig. 4**. From **Fig. 4**, we can find the reconstructed details of the depth-based method are over-deblurred.

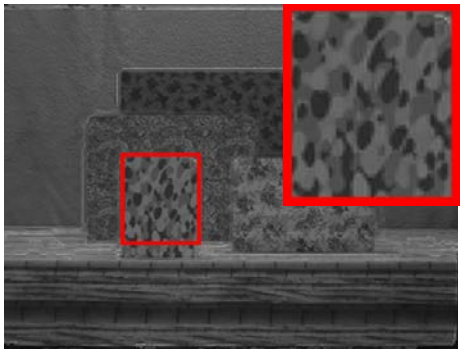
(a) Depth map



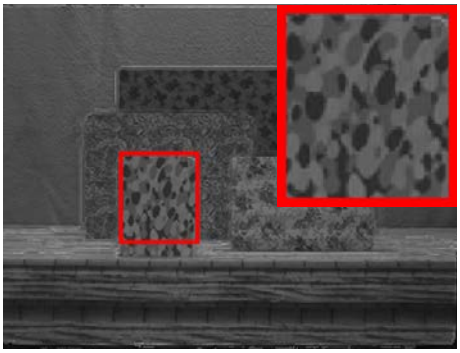
(b) Proposed method



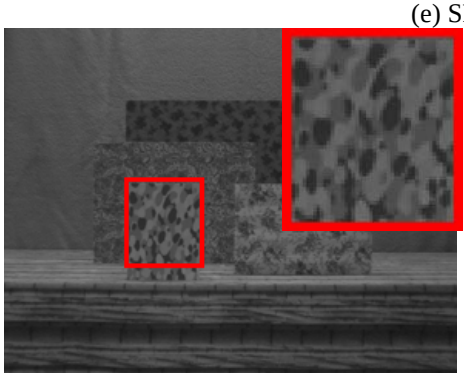
(c) Depth-based method [1]



(d) SR-FVS

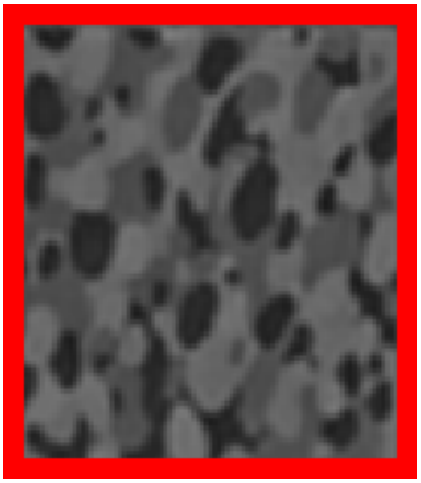


(e) SR-FVS+ScSR

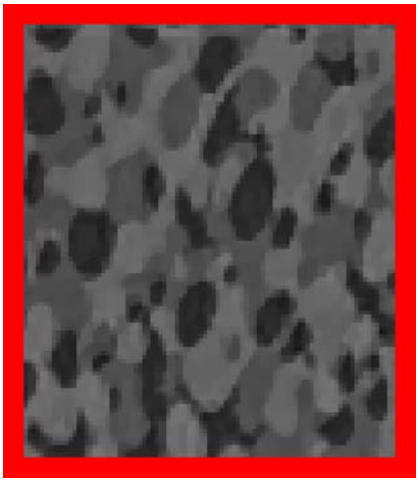


(f) LR reference image

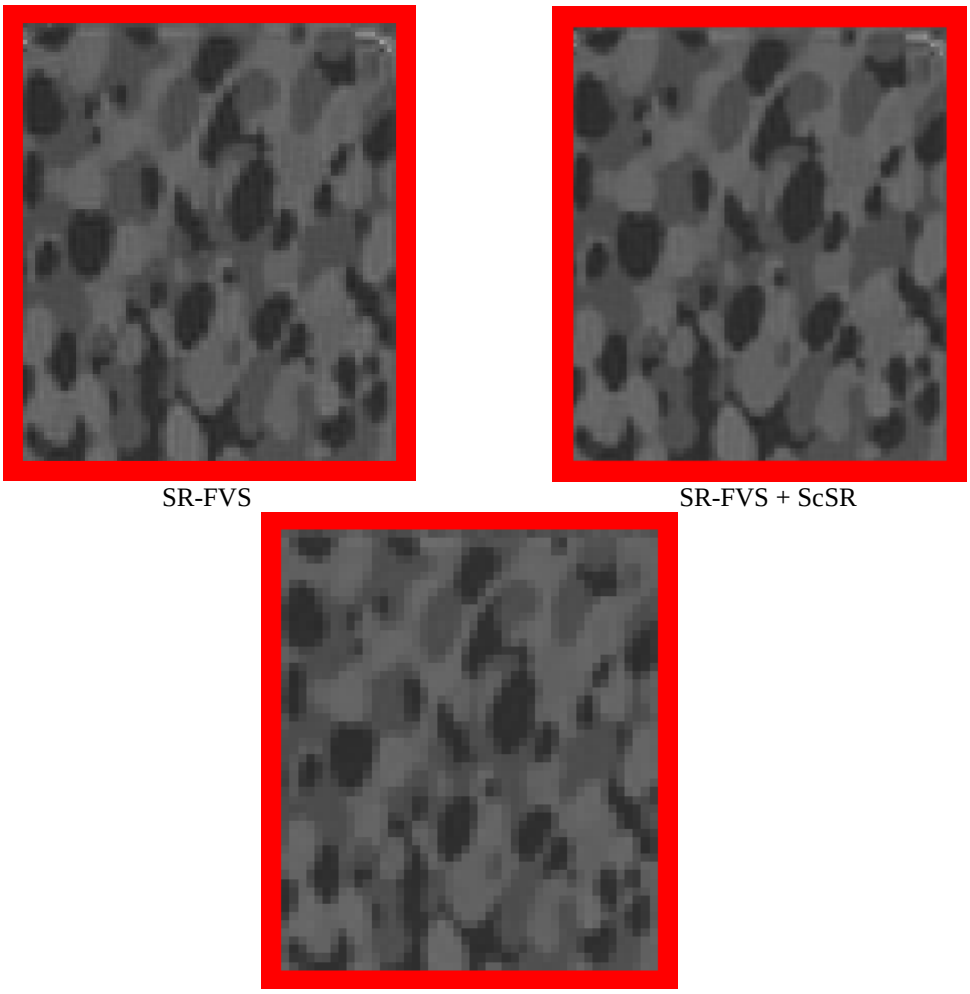
Fig. 3. Results based on the mentioned SR methods using board images



Proposed method



Depth-based method [1]



SR-FVS SR-FVS + ScSR

LR reference image (indexed as (2, 2))

Fig. 4. Close-up of details

A detailed performance comparison among the proposed method, SR-FVS, SR-FVS + ScSR and depth-based method on the board dataset is presented in **Table 1**. The proposed method achieves the smallest MSE, the highest PSNR, and the most significant SSIM. The increases of PSNR and SSIM of the proposed method to the most efficient method (SR-FVS) are 2.87% and 12.13%, respectively. Although our method is not the fastest, it produces better results than any of the other methods, thereby indicating the competitiveness of the proposed method.

Table 1. Performance comparison on images of the board dataset

	MSE	**PSNR**	**SSIM**	**Computational Time**
Proposed method	**30.51**	**33.29**	**0.9613**	45.78s
SR-FVS	37.71	32.36	0.8573	30.73s

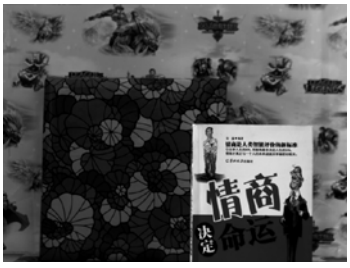
SR-FVS+ScSR	33.15	32.92	0.8637	283.62s
Depth-based method	31.76	33.11	0.9467	38.63s

6.2 On images of our image database

We also conducted experiments on our own images. Three datasets captured by our camera array system were used for comparing the proposed method with SR-FVS, SR-FVS + ScSR and depth-based method. The setup of the proposed method for our image database is as follows: d_{max} = 4000 mm, d_{min} = 1000 mm, N = 100, λ_1 = 60, λ_2 = 240, τ = 120, λ = 1, α = 1.5, $\tau_1 = \tau_2$ = 0.2, q = 0.3 and $\varepsilon_1 = \varepsilon_2 = 5\times10^{-4}$

6.2.1 On images of the book dataset

In the first experiment, five images of the book dataset (shown at the top of **Fig. 5**), which were obtained by our 4×4 camera array system, were used as inputs. The five images were captured by five cameras at the top left, as shown in **Fig. 6**. The top four input images are indexed as (1, 2), (2, 1), (2, 3), and (3, 2), following the database notation. The fifth image in **Fig. 5** is the reference LR image, which is located at the center of other four input images. This location is described as (2, 2) using the database notation. The original images have 2592 × 1944 pixels in RGB color. We converted the images to gray-scale and reduced them to 1296 × 972 and 648 × 486 pixels as the ground truth and the input, respectively. The bottom image in **Fig. 5** is the ground truth in (2, 2), with 1296 × 972 pixels.



(1, 2)

(2, 1)

(2, 3)

(3, 2)

(2, 2)

Ground truth

Fig. 5. Input images and ground truth of the book dataset

Fig. 6. Five cameras used in our experiments

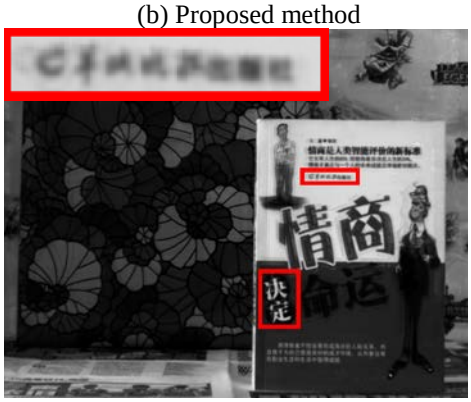
Fig. 7(a) shows the depth map of the desired HR image, whereas **Figs. 7(b)–(e)** demonstrate the results of the four mentioned reconstruction techniques, respectively. **Fig. 7(f)** presents

the input image indexed as (2, 2) (LR reference image). The close-ups of four reconstructed SR images are presented in **Fig. 8** for better visual comparison. Only the results of the proposed method and depth-based method obviously recover letters on the book. However, if we have a close-up view of **Fig. 8**, we can find that the reconstructed letters obtained by the depth-based method are excessively sharpened.

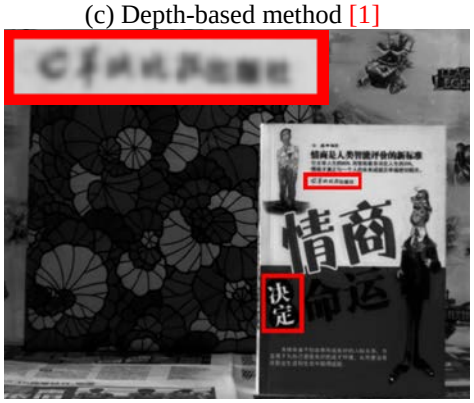
(a) Depth map

(b) Proposed method

(c) Depth-based method [1]



(d) SR-FVS



(e) SR-FVS + ScSR

(f) LR reference image (indexed as (2, 2))

Fig. 7. Results based on the mentioned SR methods using book images

Proposed method

Depth-based method [1]

SR-FVS

SR-FVS + ScSR

LR reference image (indexed as (2, 2))
Fig. 8. Close-up of details

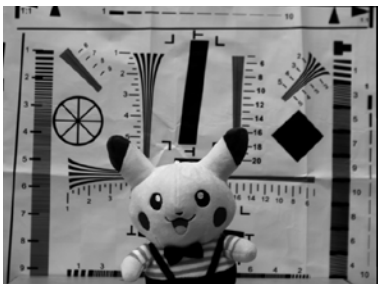
The calculated MSE, PSNR, SSIM and computational time are shown in **Table 2**. It is obvious that the proposed method consistently produces the best results among the four compared methods without much extra computational time.

Table 2. Performance comparison on images of the book dataset

	MSE	**PSNR**	**SSIM**	**Computational Time**
Proposed method	**16.41**	**35.98**	**0.9811**	195.63s
SR-FVS	21.25	34.85	0.9128	129.27s
SR-FVS+ScSR	21.07	34.89	0.9177	1050.59
Depth-based method	20.32	35.04	0.9238	165.89

6.2.2 On images of the doll dataset

Another experiment was conducted on the doll dataset. The database notations of the input images are (1, 2), (2, 1), (2, 3), (3, 2), and (2, 2). The sizes of input and output image are 648 × 486 and 1296 × 972, respectively. The ground truth is reduced from the original image in (2, 2), and its size is 1296 × 972.

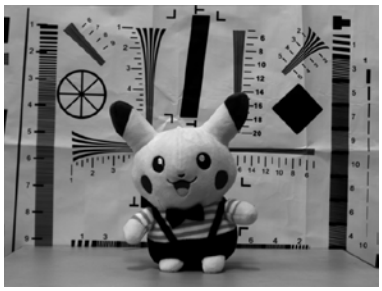


(1,2)

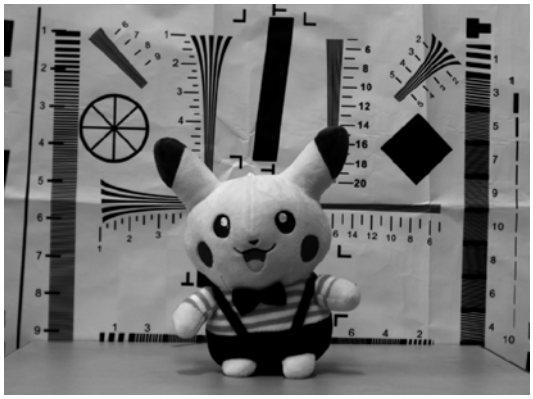
(2,1)

(2,3)

(3,2)



(2,2)



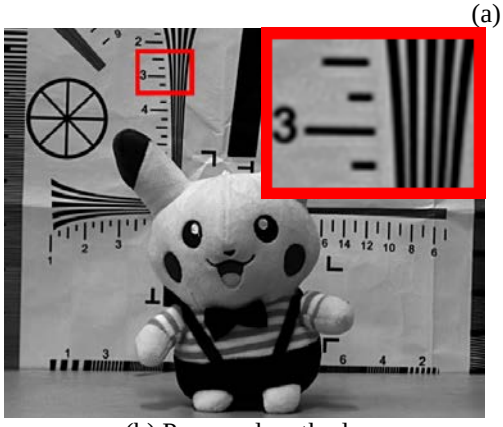
Ground truth

Fig. 9. Input images and ground truth of the doll dataset

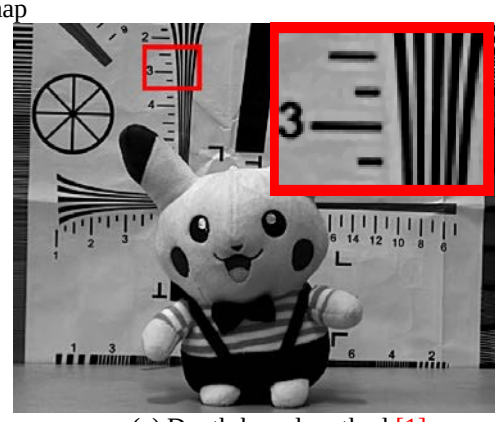
Fig. 10(a) presents the depth map of the desired HR image. **Figs. 10(b)–(e)** shows the results of the four mentioned reconstruction techniques, respectively. The LR reference image is

shown in **Fig. 10(f)**. We provided the close-up images of **Figs. 10(b)–(f)** in **Fig. 11** to better assess the estimated results. The proposed method obviously outperforms the other three methods and obtains a well-reconstructed HR image.

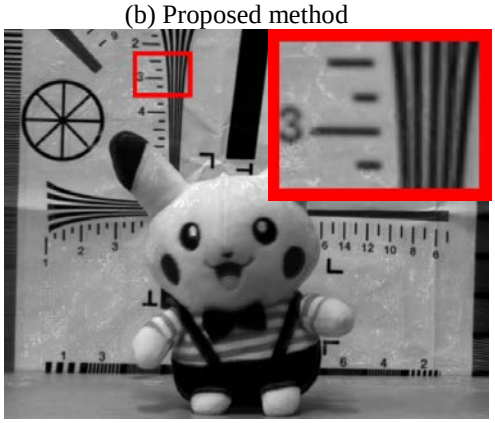
(a) Depthmap



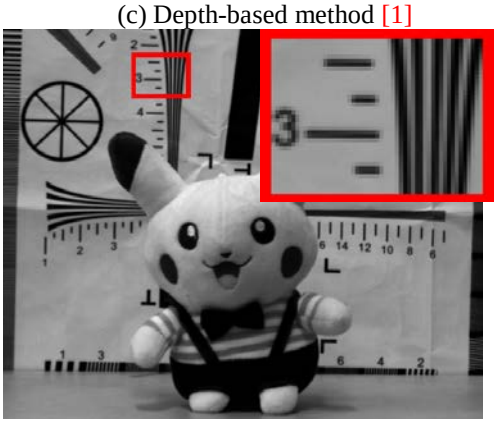
(b) Proposed method



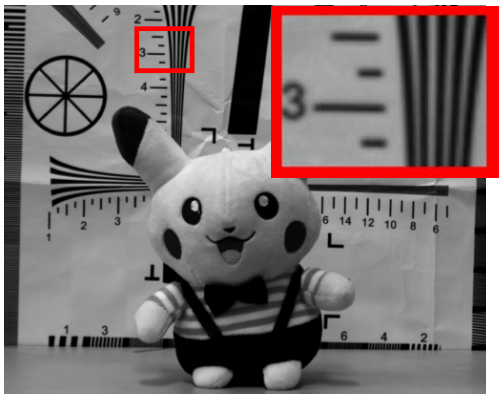
(c) Depth-based method [1]



(d) SR-FVS

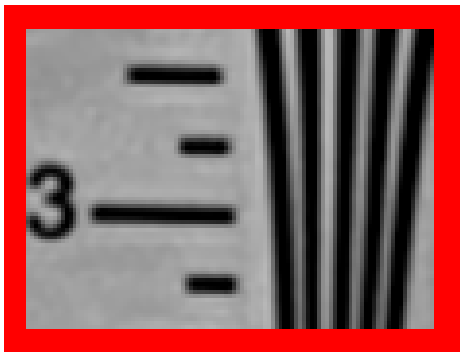


(e) SR-FVS+ScSR

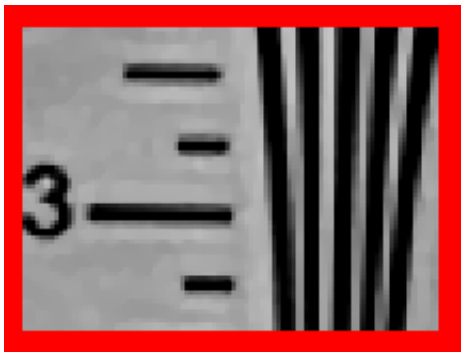


(f) LR reference image

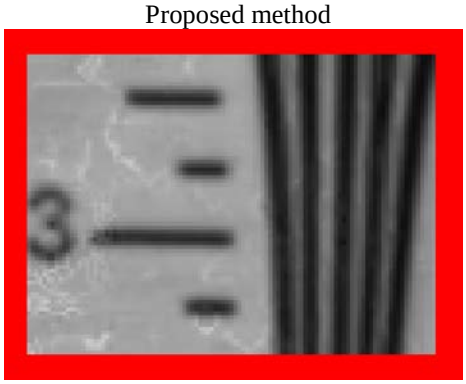
Fig. 10. Results based on the mentioned SR methods using doll images



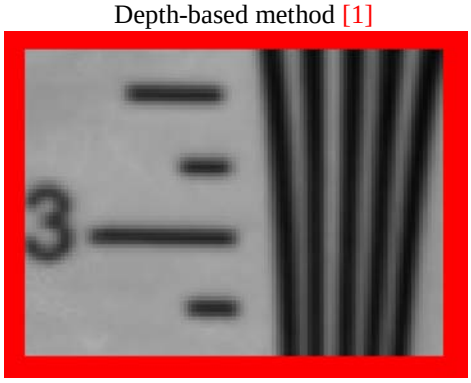
Proposed method



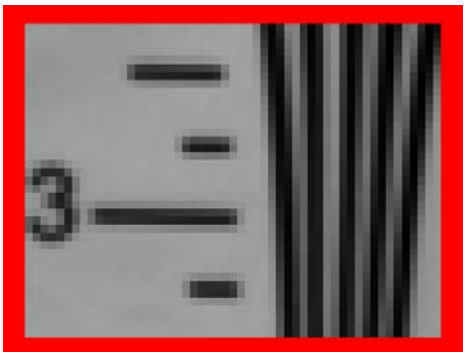
Depth-based method [1]



SR-FVS



SR-FVS + ScSR



LR reference image (indexed as (2, 2))

Fig. 11. Close-up of details

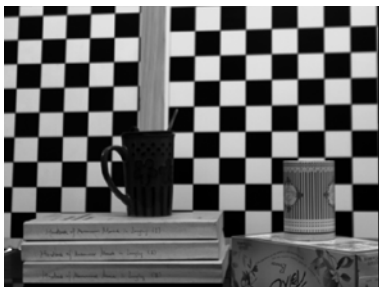
Comparisons of MSE, PSNR, SSIM and computational time are shown in **Table 3**. As shown in **Fig. 10** and **Fig. 11** as well as **Table 3**, the proposed method exhibits outstanding performance in terms of MSE, PSNR, SSIM, and details compared with the other methods, and its computational time is acceptable.

Table 3. Performance comparison on images of the doll dataset

	MSE	PSNR	SSIM	Computational Time
Proposed method	**31.78**	**33.11**	**0.9659**	197.15s
SR-FVS	34.27	32.78	0.9165	131.31s
SR-FVS + ScSR	33.13	32.93	0.9212	1123.05s
Depth-based method	35.88	32.58	0.9109	168.68s

6.2.3 On images of the cup dataset

The last experiment was conducted on our cup dataset. The input images are indexed as (1, 2), (2, 1), (2, 3), (3, 2), and (2, 2). The sizes of input and output image are 648 × 486 and 1296 × 972, respectively. The ground truth have 1296 × 972 pixels, and it is reduced from the original image in (2, 2).



(1,2)

(2,1)

(2,3)

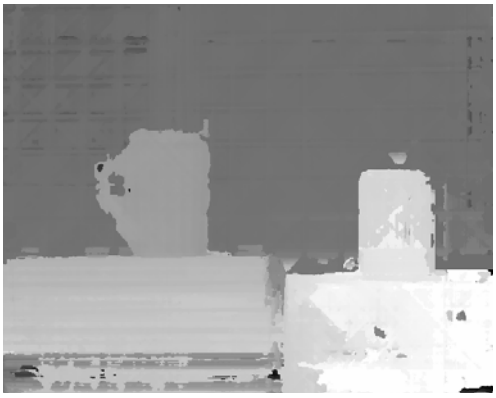
(3,2)

(2,2)

Ground truth

Fig. 12. Input images and ground truth of the cup dataset

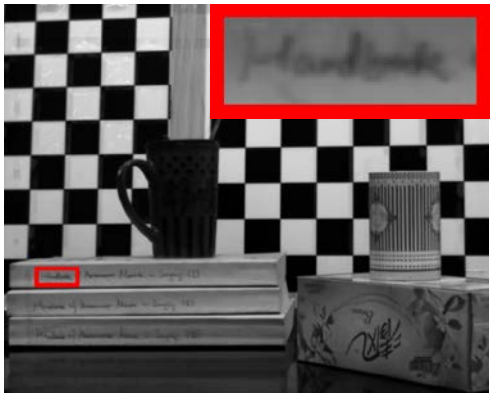
Fig. 13(a) shows the depth map of the desired HR image. **Figs. 13(b)–(e)** shows the results of the four mentioned reconstruction techniques, respectively. The LR reference image is presented in **Fig. 13(f)**. The close-up images of letters are shown in **Fig. 14**. It is obvious that the proposed method recovers the most detailed HR image.



(a) depth map

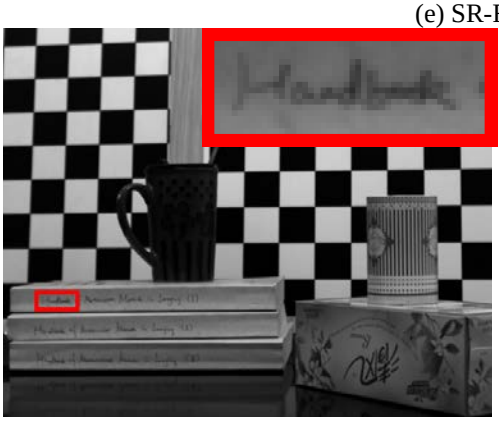
(b) Proposed method

(c) Depth-based method [1]



(d) SR-FVS

(e) SR-FVS+ScSR

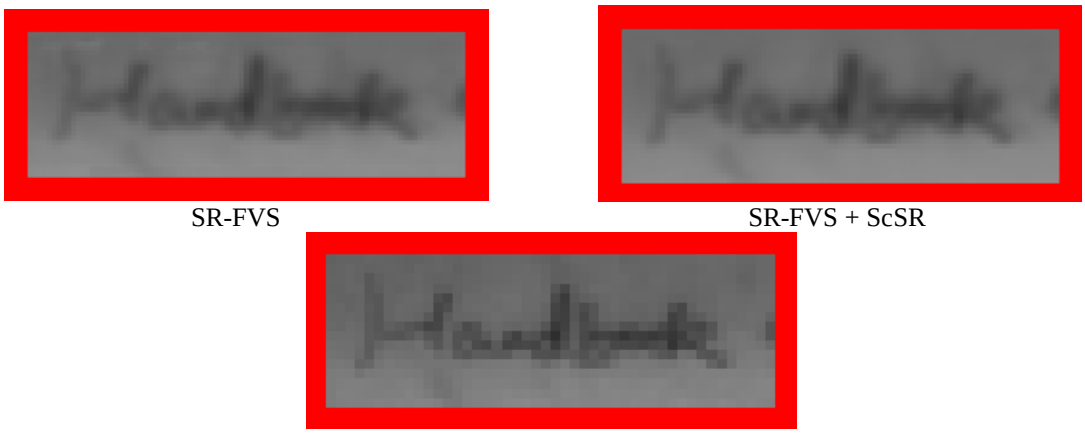


(f) LR reference image

Fig. 13. Results based on the mentioned SR methods using cup images

Proposed method

Depth-based method [1]



SR-FVS SR-FVS + ScSR

LR reference image (indexed as (2, 2))

Fig. 14. Close-up of details

Table 4 shows the MSE, PSNR, SSIM and computational time of the compared methods. It is clearly seen from the objective indexes in **Table 4** that the proposed method produces better MSE, PSNR and SSIM than any of the other methods with acceptable computational time.

Table 4. Performance comparison on images of the cup dataset

	MSE	**PSNR**	**SSIM**	**Computational Time**
Proposed method	**20.12**	**31.30**	**0.9640**	188.69s
SR-FVS	31.92	33.09	0.9472	125.89s
SR-FVS + ScSR	27.83	33.68	0.9474	1094.37s
Depth-based method	27.14	33.79	0.9509	161.13s

5. Conclusions

This study proposed a new MVSR method that addresses MVSR by solving an MCBD problem using the ADMM. We can jointly realize IF and BD based on an integrated energy function optimization. Essentially, the proposed method is an improved approach of our previous work [1], which means that it can also handle unknown blur typically occurring in camera array imaging. The experiments on the Multi-view Image Database of the University of Tsukuba and real images obtained by our camera array system demonstrate that the proposed method is favorably competitive to state-of-the-art MVSR methods. The proposed method also has several limitations. For example, the optimization speed of our method is slower than most state-of-the-art MVSR methods because we have to alternatively minimize and integrate the energy function (7) with respect to u and k. In addition, our reconstructed SR results are sensitive to estimated depth map; thus, a high-precision depth

map is necessary. Future work should involve two aspects:

1. Hasten the proposed method while preserving the SR quality;
2. Propose an MVSR framework that can handle depth inaccuracies;

Acknowledgment

This work was supported by NSFC (Grants No. 61602494 and 61405252) and NUDT (Grant ZK16-03-16).

References

[1] Jun Fan, Xiangrong Zeng, et al. "A depth-based Multi-view Super-Resolution Method Using Image Fusion and Blind Deblurring," *KSII Transactions on Internet & Information Systems*, vol. 10, no.10, pp. 5129-5152, 2016. Article (CrossRef Link).

[2] 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, 2005. Article (CrossRef Link).

[3] 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, 2013. Article (CrossRef Link).

[4] C. Guillem, D. James, A. R. Harvey, "Super-resolution imaging using a camera array," *Optics letters*, vol.39, no.7, pp.1889-1892, 2014. Article (CrossRef Link).

[5] C. Guillem, et al. "Compact multi-aperture imaging with high-angular-resolution," *Journal of the Optical Society of America A Optics Image Science & Vision*, vol.32, no. 3, pp. 411-419, 2015. Article (CrossRef Link).

[6] U Mudenagudi, et al. "Super Resolution of Images of 3D scenecs," in *Proc. of Computer Vision - Asian Conference on Computer Vision (ACCV) 2007*, pp.85-95, 2007. Article (CrossRef Link).

[7] S Najafi, "Single and Multi-view Video Super-resolution," Master's thesis, McMaster University, 2012.

[8] K. Takahashi and T. Naemura, "Super-resolved free-viewpoint image synthesis based on view-dependent depth estimation," *IPSJ Transactions on Computer Vision and Applications*, vol. 4, pp.134-148, 2012. Article (CrossRef Link).

[9] 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).

[10] 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, 2010. Article (CrossRef Link).

[11] F Sroubek, P Milanfar, et al. “Robust Multichannel Blind Deconvolution via Fast Alternating Minimization,” *IEEE Transactions on Image Processing*, vol.21, no.4, pp.1687-1700, 2011. Article (CrossRef Link).

[12] 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, pp.17-40, 1976. Article (CrossRef Link).

[13] 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, 2011. Article (CrossRef Link).

[14] M. V. Afonso, J. M. Bioucas-Dias, M. A. T. Figueiredo, “Fast Image Recovery Using Variable Splitting and Constrained Optimization,” *IEEE Transactions on Image Process*, vol. 19, no.9, pp. 2345-2356, 2010. Article (CrossRef Link).

[15] J. Kotera, F. Šroubek, P. Milanfar, “Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors,” *Computer Analysis of Images and Patterns 2013*, pp.59-66, 2013. Article (CrossRef Link).

[16] S. Lertrattanapanich, N. K. Bose, “High resolution image formation from low resolution frames using Delaunay triangulation,” *IEEE Trans. Image Process*, vol.11, pp.1427-1441, 2002. Article (CrossRef Link).

[17] Z. Z. Wang, F. H. Qi, “On ambiguities in super-resolution modeling,” *IEEE Signal Process. Letters*, vol.11, pp.678-681, 2004. Article (CrossRef Link).

[18] R. Hartley, A. Zisserman, “Multiple view geometry in computer vision,” *Cambridge University Press*, 2000.

[19] Qizi Huangpeng, Xiangrong Zeng, Quan Sun and Jun Fan, “Super-resolving blurry multiframe images through multiframe blind deblurring using ADMM,” *Multimedia Tools and Applications*. Article (CrossRef Link).

[20] D. Krishnan, and R. Fergus, “Fast image deconvolution using hyper-laplacian priors. In NIPS,” *Proceedings of Neural Information Processing Systems Blurred Lut Nr*, pp.1033-1041, 2009. Article (CrossRef Link).

[21] 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.22, pp.3074-3086, 2013. Article (CrossRef Link).

[22] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, pp.1168-1200, 2005. Article (CrossRef Link).

[23] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, “Efficient projections onto the l1-ball for learning in high dimensions,” in *Proc. of 25th international conference on Machine learning*, pp.272-279, 2008. Article (CrossRef Link).

[24] 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, 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.

Yue Wu received the B.S. degree in Computer Science and the M.S. degree in Management Science and Engineering from National University of Defense Technology, Changsha, Hunan, China, in 1997 and 2001, respectively. 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 processing.

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.

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.