Accurate Structure Recovery via Weighted Nuclear Norm: A Low Rank Approach to Shape-from-Focus

Prashanth Kumar G.[1] and Rajiv Ranjan Sahay[2]
Computational Vision Laboratory,
[1]Advanced Technology Development Centre,
[2]Department of Electrical Engineering,
[1,2]Indian Institute of Technology Kharagpur
[1]prashanth02.b@gmail.com, [2]rajivsahay@gmail.com

Abstract

In recent years, weighted nuclear norm minimization (WNNM) approach has been attracting much interest in computer vision and machine learning. Due to the ability of WNNM to preserve large-scale sharp discontinuities and small-scale fine details more effectively, we propose to use it as a regularizer to recover the 3D structure using shape-from-focus (SFF). Initially, we estimate the All-in-focus image and subsequently 3D structure is recovered using space-variantly blurred observations from the SFF stack. Since estimation of 3D shape is a severely ill-posed problem, we use weighted nuclear norm as a regularizer in the proposed algorithm. Finally, the estimated shape profile is post-processed to compensate for the effect of specular reflections in the observations on shape reconstruction. We conducted several experiments on various synthetic and real-world datasets and our results confirm that the proposed method outperforms other state-of-the-art techniques.

1. Introduction

Patch-based techniques [1, 2, 3] have become more popular in computer vision and machine learning in recent years. An image exhibits self-similarity property [1], but pixel based methods [4, 5] can not capture the redundancy of small patches inside the same image. However, patch based methods [1, 2, 3] have yielded excellent results in many image applications exploiting the self-similarity property of an image. The matrix constructed by stacking self-similar non-local patches together in an image has low rank [6, 7]. As the data in many practical problems possessing intrinsic low rank structure, low rank matrix approximation (LRMA) has achieved great success in various computer vision applications. Low rank minimization methods reconstruct the data matrix by imposing an additional rank constraint upon the estimated matrix. Since direct rank minimization is NP hard, the problem is generally approached by minimizing the nuclear norm of the estimated matrix, which is a convex relaxation for minimizing the matrix rank [6]. This methodology is named as nuclear norm minimization (NNM) and the nuclear norm of a matrix is defined as the sum of its singular values. The main limitation of NNM is that, it treats all singular values equally and shrinks each of them with the same threshold. The large singular values of a matrix constructed by stacking similar patches deliver the major edge and texture information and therefore we should shrink the larger singular values less while the smaller ones should be shrunk more. Clearly, traditional NNM model is not flexible enough to handle such issues. To improve the flexibility of NNM, weighted nuclear norm method is proposed in [7, 8], in which different weights are assigned to all singular values by penalizing significant values by smaller weights and smaller values by larger weights. Due to the ability of WNN to preserve large-scale sharp edges and small-scale fine details more effectively, we propose to use WNN as a regularizer to recover the 3D structure of the 3D objects using shape-from-focus (SFF).

Shape-from-focus [9] is a technique which uses degree of focus as a cue to estimate the 3D structure or shape of the objects in the scene. In this technique, one captures multiple images (focal stack) by translating the 3D object relative to the camera. The captured images are space-variantly blurred due to finite depth-of-field of the camera. The structure of the object is estimated by computing the degree of focus of each point of the object in the stack of captured images. Estimation of 3D shape from the blurred images is an ill-posed problem and hence prior knowledge of the unknown is required to regularize the solution. In the proposed method we use low rank prior [10, 7] and formulate an appropriate objective function using the WNN regularizer. We resort to non-negative Garrote thresholding of the singular values and estimate the shape of 3D specimen using the split-Bregman framework [11, 12] for optimization.

Since, in WNNM the singular values are penalized, choice of good shrinkage operator is important. The disadvantage of soft thresholding is that the soft shrinkage estimates tend to have bigger bias, due to the shrinkage of large coefficients. Due to the discontinuities of the shrinkage function, the hard shrinkage estimates tend to have bigger variance and can be unstable becoming sensitive to small changes in the data. As a good compromise, the non-negative Garrote threshold (NNGT) shrinkage function [13, 14] is used in the proposed approach to reduce the bias.

Estimating 3D structure of metallic objects is challenging if there exist specular reflections in the observations which is a common problem in SFF. As per our knowledge, the existing techniques in the literature of SFF [9, 15, 16, 17, 18, 19, 20] didn't consider the effect of specularity into account for shape recovery. Since SFF computes the degree of focus in images to estimate structure erroneous estimates are obtained at specular regions as image intensities are saturated.

Once the 3D shape has been estimated using the proposed method, we apply inpainting as a post-processing step after localizing the specular regions in the corresponding focused image. We divide the all-in-focus image into super-pixels using the technique in [21] and use k-means clustering [22] to separate regions of specular reflections on the low-textured metallic 3D specimen. Subsequently, a binary mask is obtained which identifies the locations of specularities. Using this binary mask we inpaint the erroneous depth estimates by WNNM-NNGT algorithm.

The rest of the paper organized as follows. A brief literature survey of SFF and WNNM is given in section 2. A simplified introduction to low rank minimization with weighted nuclear norm is provided in section 3. Formation of image in SFF and formulation of the problem is described in section 4. Details of the proposed algorithm are presented in section 5. The results of the proposed algorithm are compiled in section 6. Finally, section 7 provides the concluding remarks for our work.

2. Related Works

Extracting 3D shape or depth from 2D images is a fundamental problem in computer vision. The popular methods are, stereo based technique [23, 24, 25], which measures disparities between a pair of images of the same scene taken from two different viewpoints to recover depth. Structure from motion (SfM) [26, 27] computes the correspondences between images to obtain the 2D motion field, which in turn used to recover the 3D motion and the depth. The depth-from-defocus (DFD) [28, 29, 30] is another technique which uses two defocused blurred images of the same scene captured using single camera with different focal and aperture settings [31, 32].

Shape-from-focus (SFF) [9] is another popular method to estimate the shape of 3D object by measuring the degree of focus [17, 18, 19, 20] from the set blurred images. SFF method uses a single camera to capture a sequence of space-variantly blurred images. The degree of focus in the stack of blurred images is exploited to arrive at an estimate of the shape of the object. Authors in [18] showed that, the accuracy of SFF method is known to be superior than DFD technique. A multilayer feed-forward neural networks approach to estimate shape in SFF is used in [33]. The authors in [15] recover shape by maximizing the focus measure in the 3D image volume. In [16], the authors extended the recovery of shape using SFF by considering the relative defocus blur as a cue. The authors in [34], reconstructed the high resolution (HR) shape profile of the 3D object by modeling it as an independent Markov random fields. An MRF-based approach was used in [17] for extracting shape of smooth and low textured objects using Iterative conditional mode (ICM). Even though their method was robust to scene texture, the optimization steps take longer time to run which is an inherent problem in MRF based approaches. A discontinuity-adaptive Markov random field (DAMRF) prior to estimate the structure is proposed in [35], the non-convex objective function is solved using graduated non-convexity (GNC) algorithm. In [36], the authors estimated shape of the object using total variation (TV) prior. A new method is proposed in [37] that extends the capability of SFF taking parallax into an account to estimate the depth profile of 3D objects in the presence of structure-dependent pixel motion using simulated annealing. Since low rank methods attracted significant interest in recent years and become state-of-the-art techniques in many computer vision problems. nonlocal low-rank regularization approach is proposed in [38] to recover the image in compressive sensing. Authors in [10] combined low rank prior and TV prior to deblur the image under Gaussian noise and salt-and-pepper noise and achieved significant improvement over the state-of-the-art deblurring methods. Video denoising problem based on NNM is presented in [39]. The authors in [7, 8] proposed a technique for image denoising using WNNM which was outperformed on several state-of-the-art techniques in image denoising. Due to the ability of low rank approaches to recover images accurately, we proposed an algorithm to recover the shape of the 3D object using WNN regularization from the set of space-variantly blurred images.

3. Low Rank Minimization with Weighted Nuclear Norm

As mentioned in [6], minimization of nuclear norm is the tightest convex relaxation of the original rank minimization problem. Given a data matrix $\mathcal{Y} \in \mathbb{R}^{m \times k}$, the goal of NNM is to find a matrix $\mathcal{X} \in \mathbb{R}^{m\times k}$ of rank r which satisfies the following objective function,

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2 + \lambda\|\mathcal{X}\|_* \quad (1)$$

where λ is a positive constant, $\|\cdot\|_*$ is nuclear norm (NN) of a matrix $\mathcal{X}$ and first term is the Frobenious norm of data fidelity term. In [40], it is shown that the low rank matrix can be perfectly recovered from the degraded/corrupted data matrix with high probability by solving the NNM problem. It is shown in [41], that NNM problem can be easily solved by imposing soft-thresholding operation. The solution of Eq. (1) is given by

$$\hat{\mathcal{X}} = \mathbf{U}\mathbf{S}_\lambda(\mathbf{\Sigma})\mathbf{V}^T \quad (2)$$

where $\mathcal{Y} = \mathbf{U}(\mathbf{\Sigma})\mathbf{V}^T$ is the SVD of $\mathcal{Y}$ and $\mathbf{S}_\lambda(\mathbf{\Sigma})$ is the soft-thresholding operator function on diagonal matrix $\mathbf{\Sigma}$ with parameter λ and is defined as

$$\mathbf{S}_\lambda(\mathbf{\Sigma}) = max(\mathbf{\Sigma} - \lambda, 0) \quad (3)$$

The main drawback of nuclear norm regularization is that, it penalizes all singular values equally. As pointed out in [42], for various machine vision tasks, the large singular values of the data matrix are often much more important than the smaller ones as they are related to the principal components of the data. The larger singular values of a matrix constructed from the similar patches of an image deliver the major edges and texture information. Hence, the weighted nuclear norm proposed in [7, 8], assigns different weights for each singular value. WNNM uses smaller weight value to penalize the larger singular values and larger weight value to penalize smaller singular values.

The weighted nuclear norm of a matrix $\mathcal{X}$ is defined as a weighted sum of its singular values:

$$\|\mathcal{X}\|_{\mathbf{w},*} = \sum_{i=1}^{n} w_i\sigma_i(\mathcal{X}) \quad (4)$$

where $\sigma_1(\mathcal{X}) \geq \sigma_2(\mathcal{X}) \geq \ldots.\sigma_n(\mathcal{X}) \geq 0$, the weight vector $\mathbf{w} = [w_1, w_2, \ldots w_n]$ and $w_i \geq 0$.

With WNN regularization, the low rank minimization problem studied in [7, 8] is

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2 + \|\mathcal{X}\|_{\mathbf{w},*} \quad (5)$$

The global optimal solution of Eq. (5) under the order constraints $0 \leq w_1 \leq w_2 \leq \ldots. \leq w_n$ as proved in [43] is given by

$$\hat{\mathcal{X}} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (6)$$

where

$$D = \begin{pmatrix} diag(d_1, d_2, ..., d_n) \\ 0 \end{pmatrix}$$

and $d_i = \max(\sigma_i - w_i, 0)$, $i = 1, 2, ..., n$. Further, if all the nonzero singular values of $\mathcal{Y}$ are distinct, then $\hat{\mathcal{X}}$ is the unique optimal solution.

For more general case, [42] showed that $(d_1, d_2, ..., d_n)$ is the solution of the following convex optimization problem:

$$\begin{aligned} &\min_{d_1,..,d_n} \sum_{i=1}^{n} \frac{1}{2}(d_i - \sigma_i)^2 + w_i d_i, \\ &s.t. \qquad d_1 \geq d_2 \geq \ldots \geq d_n \geq 0 \end{aligned} \quad (7)$$

and they also verified that the globally optimal solution of Eq. (7) has a closed form when the weights satisfy $0 \leq w_1 \leq w_2 \leq \ldots.. \leq w_n$.

Given a patch x_j located at j in the image, we can construct a matrix $\mathcal{X}_j$ by stacking all patches similar to patch x_j into columns of the matrix and hence all column vectors in $\mathcal{X}_j$ have similar image structures. This means that $\mathcal{X}_j$ is a low rank matrix. Then the estimate of $\mathcal{X}_j$ can be obtained by solving the following minimization problem:

$$\hat{\mathcal{X}}_j = \arg\min_{\mathcal{X}_j} \frac{1}{2}\|\mathcal{Y}_j - \mathcal{X}_j\|_F^2 + \|\mathcal{X}_j\|_{\mathbf{w},*} \quad (8)$$

Next, the latent image can be estimated via aggregating all the estimated patches. The setting of the weights $\mathbf{w}$ is very important in WNNM method. As defined in [7, 8], the weight for the i^{th} singular value of $\mathcal{X}_j$ is

$$w_i^j = \frac{c\sqrt{N}_{sp}}{\sigma_i(\mathcal{X}_j) + \varepsilon} \quad (9)$$

where $c > 0$ is a constant, N_{sp} is the number of similar patches in $\mathcal{Y}_j$. Thus the latent image can be obtained via singular value shrinkage. However, contrary to the hard shrinkage operator, the soft shrinkage operator given in Eq. (3) tends to have smaller variance but bigger bias because of shrinking all the entries. As a good compromise, the non-negative Garrote threshold shrinkage function was proposed in [13, 14]. We modified the WNNM problem to WNNM-NNGT by replacing the soft shrinkage function with non-negative Garrote thresholding function to shrink the singular values. The non-negative Garrote shrinkage function used in the proposed algorithm is

$$\mathcal{S}_\tau^{GT}(z) = \begin{cases} z - \frac{\tau^2}{z}, & \text{if} \quad |z| > \tau \\ 0, & \text{if} \quad |z| < \tau \end{cases} \quad (10)$$

where $\tau > 0$ is a threshold value, $\mathcal{S}^{GT}$ is Garrote thresholding function and z is a value to be thresholded. We later show in our experiments that convergence of WNNM-NNGT is superior than WNNM.

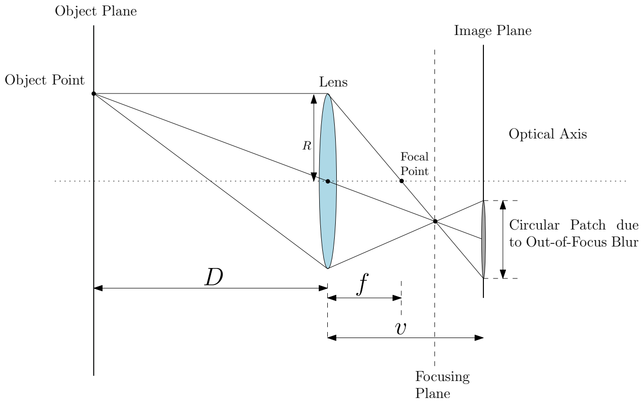


Figure 1. Formation of blurred image due to defocus blur [28].

4. Image Formation in SFF

Space-variantly blurred observations are captured in shape-from-focus scheme [9, 17, 36], by placing a 3D object on a translational stage which moves in vertical direction towards as away from the camera in finite steps of size Δd. Since the camera has a finite depth-of-field, only those points on the object which lie inside the depth-of-field are focused and other points are blurred. Initial position of the translational stage is termed as the reference plane. The plane passing through all those points of the 3D object which are perfectly focused on the sensor plane is known as focused plane. Working distance w_d is the separation between focused plane and the lens. The focal length of the lens is denoted by f and v is the distance between lens and the image plane. As the translational stage moves in steps of Δd from the reference plane, an image is captured at each step to obtain the stack of blurred observations. Since, a real aperture camera with a finite depth-of-field cannot bring all the points of a 3D object into focus at the same time, only the points (k, l) on the focused plane which satisfy lens law are perfectly focused and the other regions are defocused at different degrees. The shape of the object is determined from the captured stack of blurred observations by searching for those frames in which the object points come in focus.

4.1. Finite Aperture Camera

The point spread function (PSF), is the response of a camera to a point light source. As shown in Fig. 1, a point light source at a distance D from the lens will be imaged as a circular disc on the image plane with radius r_b given by

$$r_b = Rv(\frac{1}{f} - \frac{1}{v} - \frac{1}{D}) \tag{11}$$

where R is the radius of aperture of the lens.

Considering lens aberrations and diffraction, the PSF is approximated by a 2D circularly symmetric Gaussian function [44]. The standard deviation of such a PSF is given by

$$\sigma = \rho r_b = \rho Rv(\frac{1}{f} - \frac{1}{v} - \frac{1}{D}) \tag{12}$$

where ρ is a camera constant.

The stack of observations $\{y_m(i, j)\}; m = 0, 1, ., N - 1$ are captured, by moving the translating stage upwards in steps of Δd [36]. For the m-th frame the blur parameter $\sigma_m(i, j)$ at the point (i, j) can be expressed as

$$\sigma_m(i, j) = \rho Rv(\frac{1}{\omega_d} - \frac{1}{\omega_d + m\Delta d - \bar{d}(i, j)}) \tag{13}$$

Here $\bar{d}(i, j)$ is the distance by which the stage must be moved to bring the point (i, j) into focus [36]. The blur parameter $\sigma_m(i, j)$ becomes zero, when the point (i, j) is perfectly in focus, which happens when $m\Delta d = \bar{d}(i, j)$. We know from Eq. (13) that $\sigma_m(i, j)$ is related to the depth $\bar{d}(i, j)$ of the scene. The 2D Gaussian PSF $h_m(i, j)$ can be approximated to span the rectangle defined by $(i - 3\sigma_m(i, j), j - 3\sigma_m(i, j))$ to $(i + 3\sigma_m(i, j), j + 3\sigma_m(i, j))$ centered at (i, j) in the m^{th} observation. The PSF $h_m(i, j; k, l)$ of the camera used in the SFF setup is modeled as a 2D gaussian function.

$$h_m(i, j; k, l) = \frac{1}{2\pi\sigma_m^2(i, j)} \exp(-\frac{(i - k)^2 + (j - l)^2}{2\sigma_m^2(i, j)}) \tag{14}$$

where the standard deviation of the Gaussian function $\{\sigma_m(i, j)\} \in \mathbb{R}^{N \times N}$ is the space varying blur parameter at location (i, j) in the m^{th} observation.

From Eq. (12) and (13), we observe that σ is dependent on the depth of the object in the scene and hence blurring in the image is space-variant which can be expressed as

$$g_m(i, j) = \sum_k \sum_l x(k, l) h_m(i, j; k, l) \tag{15}$$

where $\{x(k, l)\}$ is the pixel at (k, l) in the focused image x and the space variant blurring kernel $h_m(i, j; k, l)$, which span the rectangle defined by $(k - 3\sigma_m(k, l), l - 3\sigma_m(k, l))$ to $(k + 3\sigma_m(k, l), l + 3\sigma_m(k, l))$ centered at (k, l) can be expressed as

$$h_m(i, j; k, l) = \frac{1}{2\pi\sigma_m^2(k, l)} \exp(-\frac{(i - k)^2 + (j - l)^2}{2\sigma_m^2(k, l)}) \tag{16}$$

The formation of the m^{th} observation is modeled as

$$y_m(i, j) = g_m(i, j) + n_m(i, j) \tag{17}$$

where $n_m(i, j)$ is Gaussian noise.

4.2. Problem Formulation

The formation of space variant blurred image $y_m \in \mathbb{R}^{N \times N}$, can be modeled as

$$y_m(i, j) = \sum_k \sum_l x(k, l) h_m(i, j; k, l) + n_m(i, j) \tag{18}$$

where $m \in \{0, ..., (M-1)\}$ is the index of the observations in the stack. Here $\{x(k,l)\} \in \mathbb{R}^{N \times N}$ is the pixel at (k,l) in the focused image x. Eq. (18) can also be written as

$$y_m(i,j) = \sum_k \sum_l x(k,l) h_{p_m}(i,j;k,l) + n_m(i,j) \quad (19)$$

where

$$\begin{aligned} h_{p_m}(i,j;k,l) &= \frac{1}{Z_m(k,l)} \exp\left(-\frac{(i-k)^2+(j-l)^2}{2\sigma_m^2(k,l)}\right) \\ \text{and,} \quad Z_m(k,l) &= \sum_k \sum_l \exp\left(-\frac{(i-k)^2+(j-l)^2}{2\sigma_m^2(k,l)}\right) \end{aligned} \quad (20)$$

is the p^{th} blurring kernel in the m^{th} observation at pixel position (i,j) in the image. The formation of a stack of space variantly blurred images can be modeled as

$$\mathbf{y}_m = \mathbf{H}_m(\bar{\mathbf{d}}) \cdot \mathbf{x} + \boldsymbol{\eta} \quad (21)$$

Let $\bar{\mathbf{d}}, \mathbf{x}, \boldsymbol{\sigma}_m, \mathbf{y}_m, \boldsymbol{\eta} \in \mathbb{R}^{N^2 \times 1}$ be the vectors formed by lexicographically arranging matrices $\bar{d}(i,j)$, $x(i,j)$, $\sigma_m(i,j)$, $y_m(i,j)$ and $\eta(i,j)$, respectively. The matrix $\mathbf{H}_m$ represents the operation of structure-dependent blurring on the focused image $\mathbf{x}$. $\boldsymbol{\eta}$ is assumed to be an additive white zero mean Gaussian noise model. Since we are interested in estimating shape profile $\bar{\mathbf{d}}$, so Eq. (21) can be rewritten as

$$\mathbf{y}_m = \mathbf{X}\mathbf{h}_{p_m} + \boldsymbol{\eta} \quad (22)$$

where $\mathbf{X}$ is a sparse matrix and contains only elements from $\mathbf{x}$. The vector $\mathbf{h}_{p_m}$ is formed by stacking together all PSFs. The variable $\mathbf{h}_{p_m}$ is a function of $\boldsymbol{\sigma}_m$ which is a function of depth of the scene $\bar{\mathbf{d}}$. As the value of $\bar{\mathbf{d}}$ varies at each point so do $\boldsymbol{\sigma}_m$ and $\mathbf{h}_{p_m}$. Since the kernel size depends on $\sigma_{m_{max}}$, let us assume a constant kernel size $n \times n$ at all the points where $n = 6\sigma_{m_{max}} + 1$. So the size of the vector $\mathbf{h}_{p_m}$ will be $n^2N^2 \times 1$. Sparse matrix $\mathbf{X} \in \mathbb{R}^{N^2 \times n^2N^2}$ is constructed from $\mathbf{x}$. The matrix $\mathbf{X}$ and vector $\mathbf{h}_{p_m}$ are formed such that outputs of Eq. (19) and Eq. (22) are identical.

5. The Proposed Algorithm

The problem of estimation of 3D shape from the stack of space-variantly blurred 2D images captured in SFF is typically ill-posed. Regularization in the form of priori constraints on the solution has to be imposed to estimate the unknown depth map. In the proposed method we use WNN regularization [7, 8] to estimate the shape profile $\bar{\mathbf{d}}$.

5.1. Estimation of Shape Profile

Since $\bar{\mathbf{d}}, \mathbf{x}, \boldsymbol{\sigma}_m, \mathbf{y}_m \in \mathbb{R}^{N^2 \times 1}$ are vectors and as proved in [45], Frobenious norm is equivalent to ℓ_2 norm, the least square estimate of $\bar{\mathbf{d}}$ with low rank regularization constraint can be modeled as

$$\hat{\bar{\mathbf{d}}} = \arg\min_{\bar{\mathbf{d}}} \frac{1}{2} \sum_{m=1}^{q} \|\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}(\bar{\mathbf{d}})\|_2^2 + \sum_{j \in \mathcal{P}} \|\mathbf{R}_j \bar{\mathbf{d}}\|_{\mathbf{w},*} \quad (23)$$

where $\mathcal{P}$ denotes the set of indices of all reference patches of $\bar{\mathbf{d}}$. The operator $\mathbf{R}_j$ firstly collects all the patches similar to the reference patch located at j^{th} position in $\bar{\mathbf{d}}$, and then stacks those patches into a matrix which is assumed to be of low rank. Solving the model in Eq. (23) is not straightforward and therefore we introduce an auxiliary variable $\mathbf{u}$ to split the model into easily solvable sub-problems. The constrained objective function is thus given by

$$\begin{aligned} \hat{\bar{\mathbf{d}}} = \arg\min_{\bar{\mathbf{d}}} \frac{1}{2} \sum_{m=1}^{q} \|\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}(\bar{\mathbf{d}})\|_2^2 + \sum_{j \in \mathcal{P}} \|\mathbf{R}_j \mathbf{u}\|_{\mathbf{w},*} \\ \text{s.t.} \quad \mathbf{u} = \bar{\mathbf{d}} \end{aligned} \quad (24)$$

To solve Eq. (24), we convert it into an unconstrained objective function by adding a quadratic penalty term to it. The split-Bregman method introduced in [11] is used to solve the unconstrained problem and is given by

$$\begin{aligned} (\hat{\bar{\mathbf{d}}}^{l+1}, \mathbf{u}^{l+1}) = \arg\min_{\bar{\mathbf{d}},\mathbf{u}} \frac{1}{2} \sum_{m=1}^{q} \|\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}(\bar{\mathbf{d}}^l)\|_2^2 \\ + \sum_{j \in \mathcal{P}} \|\mathbf{R}_j \mathbf{u}^l\|_{\mathbf{w},*} + \frac{\lambda}{2} \|\bar{\mathbf{d}}^l - \mathbf{u}^l - \mathbf{b}^l\|_2^2 \end{aligned} \quad (25)$$

and

$$\mathbf{b}^{l+1} = \mathbf{b}^l + (\mathbf{u}^{l+1} - \bar{\mathbf{d}}^{l+1}) \quad (26)$$

where l is the iteration number and $\lambda > 0$ is a constant that balances between data fidelity and the quadratic penalty term in order to control the quality of the estimated $\bar{\mathbf{d}}$. According to split-Bregman method [11], we can split the Eq. (25) into two separate subproblems:

$$\begin{aligned} \hat{\bar{\mathbf{d}}}^{l+1} = \arg\min_{\bar{\mathbf{d}}} \frac{1}{2} \sum_{m=1}^{q} \|\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}(\bar{\mathbf{d}}^l)\|_2^2 \\ + \frac{\lambda}{2} \|\bar{\mathbf{d}}^l - \mathbf{u}^l - \mathbf{b}^l\|_2^2 \end{aligned} \quad (27)$$

$$\mathbf{u}^{l+1} = \arg\min_{\mathbf{u}} \sum_{j \in \mathcal{P}} \|\mathbf{R}_j \mathbf{u}^l\|_{\mathbf{w},*} + \frac{\lambda}{2} \|\bar{\mathbf{d}}^{l+1} - \mathbf{u}^l - \mathbf{b}^l\|_2^2 \quad (28)$$

and

$$\mathbf{b}^{l+1} = \mathbf{b}^l + (\mathbf{u}^{l+1} - \bar{\mathbf{d}}^{l+1}) \quad (29)$$

where l is the iteration number, Eq. (27) to Eq. (29) constitute the proposed algorithm to estimate the 3D shape profile $\bar{\mathbf{d}}$ using a few observations from the stack of blurred observations. The shape profile $\bar{\mathbf{d}}$ can be obtained by alternatively minimizing the Eq. (27) with respect to $\bar{\mathbf{d}}$ and Eq. (28) with respect to $\mathbf{u}$ while updating the dual variable $\mathbf{b}$ according to Eq. (29).

Since the Eq. (27) contains all differentiable quadratic terms, its least square solution can be obtained easily using gradient based approach. The update equation for $\bar{\mathbf{d}}$ is hence given by

$$\bar{\mathbf{d}}^{t+1} = \bar{\mathbf{d}}^{t} - \delta_{\bar{\mathbf{d}}}\left(\sum_{m=1}^{q} \frac{\partial e_m}{\partial \bar{\mathbf{d}}} + \frac{\lambda}{2}\frac{\partial}{\partial \bar{\mathbf{d}}}\|\bar{\mathbf{d}}^t - \mathbf{u}^t - \mathbf{b}^t\|_2^2 \right) \tag{30}$$

where $\delta_{\bar{\mathbf{d}}}$ is the step size, t is iteration number of gradient descent algorithm for estimating $\bar{\mathbf{d}}$ and e_m represents the data fidelity term given by

$$e_m = \frac{1}{2}\|\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}(\bar{\mathbf{d}})\|_2^2 \tag{31}$$

From Eq. (13) and Eq. (20), it is clear that, the kernel $\mathbf{h}_{p_m}$ at a point (i,j) in an image depends on the $\sigma_m(i,j)$, which in turn depends on the depth $\bar{\mathbf{d}}(i,j)$ at that point. Hence the partial derivative of data fidelity term $\frac{\partial e_m}{\partial \bar{\mathbf{d}}}$ can be written using chain rule as

$$\frac{\partial e_m}{\partial \bar{\mathbf{d}}} = \frac{\partial \boldsymbol{\sigma}_m}{\partial \bar{\mathbf{d}}} \cdot \frac{\partial \mathbf{h}_{p_m}}{\partial \boldsymbol{\sigma}_m} \cdot \frac{\partial e_m}{\partial \mathbf{h}_{p_m}} \tag{32}$$

We now provide the details for each term in the RHS of the Eq. (32). According to Eq. (13)

$$\sigma_m(i,j) = \rho R v \left(\frac{1}{w_d} - \frac{1}{w_d + m\Delta d - \bar{\mathbf{d}}(i,j)} \right) \tag{33}$$

lexicographically arranging $\{\sigma_m(i,j)\}$ into a vector $\boldsymbol{\sigma}_m$, and differentiating with respect to $\bar{\mathbf{d}}$, we get

$$\begin{aligned} \frac{\partial \boldsymbol{\sigma}_m}{\partial \bar{\mathbf{d}}} &= \frac{\partial}{\partial \bar{\mathbf{d}}} \rho R v \left(\frac{1}{w_d} - \frac{1}{w_d + m\Delta d - \bar{\mathbf{d}}(i,j)} \right) \\ &= -\frac{\rho R v}{(w_d + m\Delta d - \bar{\mathbf{d}}(i,j))^2} \end{aligned} \tag{34}$$

Differentiating the Eq. (20) with respect to $\boldsymbol{\sigma}_m$ we get

$$\begin{aligned} \frac{\partial \mathbf{h}_{p_m}}{\partial \boldsymbol{\sigma}_m} =& \frac{\mathbf{h}_{p_m}(i,j;k,l)}{\sigma^3(k,l)} \Bigg(d(i,j;k,l) \\ & - \sum_k \sum_l \mathbf{h}_{p_m}(i,j;k,l) d(i,j;k,l) \Bigg) \end{aligned} \tag{35}$$

where $d(i,j;k,l) = (i-k)^2 + (j-l)^2$. Differentiating Eq. (31) with respect to $\mathbf{h}_{p_m}$ we get,

$$\begin{aligned} \frac{\partial e_m}{\partial \mathbf{h}_{p_m}} &= \frac{1}{2}\frac{\partial}{\partial \mathbf{h}_{p_m}} \big(\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}\big)^T \big(\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}\big) \\ &= \frac{1}{2}\frac{\partial}{\partial \mathbf{h}_{p_m}} \big(\mathbf{y}_m^T - \mathbf{h}_{p_m}^T \mathbf{X}^T\big)\big(\mathbf{y}_m - \mathbf{X}\mathbf{h}_{p_m}\big) \\ &= \mathbf{X}^T\big(\mathbf{X}\mathbf{h}_{p_m} - \mathbf{y}_m\big) \end{aligned} \tag{36}$$

Gradient of the quadratic penalty term is given by

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial \bar{\mathbf{d}}}\|\bar{\mathbf{d}} - \mathbf{u} - \mathbf{b}\|_2^2 &= \frac{1}{2}\frac{\partial}{\partial \bar{\mathbf{d}}}(\bar{\mathbf{d}}^T - \mathbf{u}^T - \mathbf{b}^T)(\bar{\mathbf{d}} - \mathbf{u} - \mathbf{b}) \\ &= [\bar{\mathbf{d}} - (\mathbf{u} + \mathbf{b})] \end{aligned} \tag{37}$$

The update equation Eq. (30) can now be written as

$$\bar{\mathbf{d}}^{t+1} = \bar{\mathbf{d}}^{t} - \delta_{\bar{\mathbf{d}}}\left(\sum_{m=1}^{q} \frac{\partial e_m}{\partial \bar{\mathbf{d}}} + \lambda\big[\bar{\mathbf{d}}^t - (\mathbf{u}^t + \mathbf{b}^t)\big] \right) \tag{38}$$

Solution to Eq. (28) can be obtained in the same way as in [7, 8]. To obtain the optimal solution of $\mathbf{u}$, we can minimize Eq. (28) with respect to each group of similar patches $\mathbf{R}_j\mathbf{u}$ in the estimated structure $\bar{\mathbf{d}}$, respectively, and then aggregate all the estimated patches to derive the final solution of $\mathbf{u}$. For each group of similar patches $\mathbf{R}_j\mathbf{u}$, we have the following minimization problem

$$\mathbf{R}_j\mathbf{u}^{l+1} = \|\mathbf{R}_j\mathbf{u}^l\|_{\mathbf{w},\star} + \frac{\lambda}{2}\|\mathbf{R}_j\mathbf{u}^l - (\mathbf{R}_j\bar{\mathbf{d}}^{l+1} - \mathbf{R}_j\mathbf{b}^l)\|_2^2 \tag{39}$$

The solution of the Eq. (39) with WNNM-NNGT is given by

$$\mathbf{R}_j\mathbf{u}^{l+1} = \mathbf{U}\begin{pmatrix} diag(d_1, d_2, ..., d_n) \\ 0 \end{pmatrix}\mathbf{V}^T \tag{40}$$

where

$$\big(\mathbf{R}_j\bar{\mathbf{d}}^{l+1} - \mathbf{R}_j\mathbf{b}^l\big) = \mathbf{U}\begin{pmatrix} diag(\sigma_{\bar{d}b,1}, \sigma_{\bar{d}b,2}, ., \sigma_{\bar{d}b,n}) \\ 0 \end{pmatrix}\mathbf{V}^T \tag{41}$$

$d_i = \max\left(\left(\sigma_{\bar{d}b,i} - \frac{w_i^2}{\sigma_{\bar{d}b,i}}\right), 0\right)$ for $i = 1, 2, ..., n$, with weights $w_i = \frac{c\sqrt{N_{sp}}}{\sigma_i(\mathbf{R}_j\mathbf{u}^{l+1}) + \varepsilon}$. After finding the solutions $\hat{\bar{\mathbf{d}}}^{l+1}$ and $\mathbf{u}^{l+1}$ alternatively, the dual variable $\mathbf{b}^{l+1}$ is updated according to Eq. (29).

5.2. Specular Region Segmentation and Shape Recovery

Existence of specular reflections in observations of the SFF stack results in erroneous estimate of 3D structure at that region. We use the all-in-focus image of 3D specimen to localize the specular regions wherein estimates of 3D shape is erroneous. We divided the entire focus image into super-pixels using the method proposed in [21]. The obtained group of super-pixels are separated into two clusters corresponding to specular and non-specular regions using k-means clustering algorithm [22]. Using the located specular regions we generate a binary mask which is used to repair the estimated erroneous shape profile. We inpainted the estimated shape profile with the aid of the generated binary mask using the proposed WNNM-NNGT approach. Let $\bar{\mathbf{d}}_e$ be the estimated erroneous shape profile, $\mathbf{S}$ be an operator which acts as a binary mask identifying the specular regions, the proposed objective function used to estimate the inpainted shape $\hat{\bar{\mathbf{d}}}_n$ is given by

$$\hat{\bar{\mathbf{d}}}_n = \arg\min_{\bar{\mathbf{d}}_n} \frac{1}{2}\|\bar{\mathbf{d}}_e - \mathbf{S}\cdot\bar{\mathbf{d}}_n\|_2^2 + \sum_{j\in\mathcal{P}} \|\mathbf{R}_j\bar{\mathbf{d}}_n\|_{\mathbf{w},*} \quad (42)$$

where first term in Eq. (42) represents data fidelity term and second term is the WNN prior for shape profile inpainting. The inpainted shape profile $\hat{\bar{\mathbf{d}}}_n$ of Eq. (42) is obtained using split-Bregman method in the same fashion as outlined in section 5.1.

5.3. Estimation of All-in-focus Image

Although, one can ideally solve for the focused image in the optimization framework proposed here, in order to avoid excessive computational expense we use the simple method given in [9, 17] using sum-modified-Laplacian (SML) focus measure operator. The procedure is as follows: when the 3D object is imaged, a particular point on the 3D specimen gets focused if it cuts the focusing plane. At that point, the blur parameter σ is zero. For a particular pixel (i, j), we compute the SML focus measure, using a window around the point in every frame of the captured stack. The frame for which the focus measure profile peaks, is the frame wherein the pixel (i, j) is in best focus. For every point on the 3D specimen, we can choose the frame in which it comes into focus, and pick the corresponding pixel intensity for the focused image $\mathbf{x}$ from that frame. Following this procedure for every pixel (i, j) we estimate the entire focused image $\mathbf{x}$. Even though the estimate so obtained is an approximation to the actual focused image, this is sufficient to estimate the shape profile using the proposed approach.

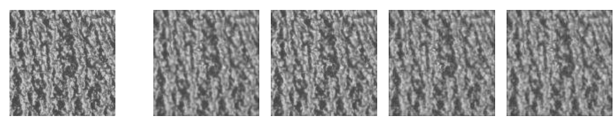
6. Experimental Results

In this section we demonstrate the results of our approach using low rank prior. Initially, we conducted synthetic experiments using the proposed method, wherein, known depth maps are used to generate the space-variantly blurred observations synthetically with the 'calf' texture from the Brodatz album [46] chosen as focused image. We also conducted experiments on real-world blurred observations of 3D objects captured using the Nikon LV 150 industrial microscope for imaging. The value of working distance $w_d = 8.8$mm, focal length $f = 80$mm and the depth-of-field $= 48.9\mu$m. The PSF of the camera was assumed to be circularly symmetric Gaussian [44]. In synthetic experiments, we assumed random initial estimate of the depth generated using 'randn' function of Matlab. In real-world experiments, we derived the initial estimate of the structure and the focused image from the stack of observations using the SFF technique proposed in [9]. Using the estimated focused image and few space-variantly blurred observations, the depth $\bar{\mathbf{d}}$ is estimated by minimizing the proposed objective function using split-Bregman algorithm.

We ran a fixed number of iterations of conjugate gradient algorithm [47, 48] using adaptive step size with line-search [49] to minimize Eq. (27). We alternatively minimized Eq. (28) using WNNM-NNGT at each Bregman iteration and subsequently updated the dual variable using Eq. (29). Parameter λ is empirically chosen to control the quality of the estimated output. We compare our algorithm with several state-of-the-art SFF techniques [9, 15, 16, 17, 36] in the recent literature.

6.1. Synthetic Data

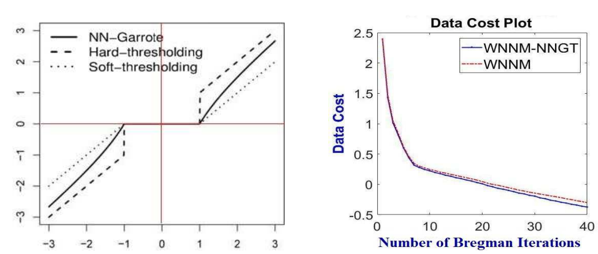
Initially, we used a known depth map of a face (Red-Indian) and the Art depth map of Middlebury dataset [50] to generate space-variantly blurred observations. We simulated the translation of the object in steps of $\Delta d = 25\mu$m relative to the camera and obtained the observation stack.



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

Figure 2. (a) Focused image. (b) to (e) Blurred images corresponding to translation steps $m = 140, 175, 185$ and 200, respectively.

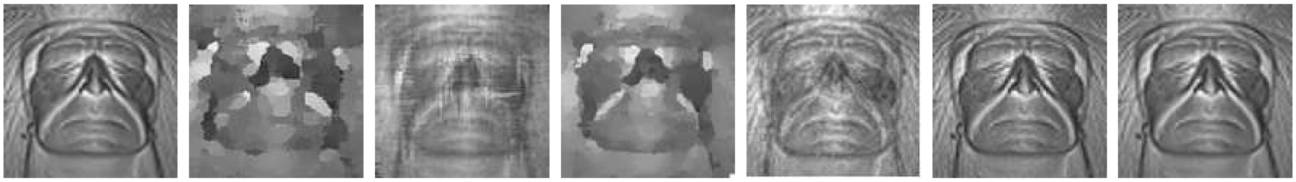
In the first synthetic experiment we used a depth map of a face (Red-Indian) to generate the blurred observations of size 120×120 pixels according to Eq. (22) with translation step $m = [140, 170, 185 \text{ and } 200]$ which are shown in Fig. 2 (b) to Fig. 2 (e). The corresponding focused image is given in Fig. 2 (a).



(a) (b)

Figure 3. (a) Singular value thresholding functions. (b) Data cost of WNNM and WNNM-NNGT for 3D structure estimation.

Firstly, we compared the performance of the proposed WNNM-NNGT algorithm using non-negative Garrote thresholding with WNNM using soft thresholding for



(a) (b) (c) (d) (e) (f) (g)

Figure 4. (a) Ground truth shape profile. (b)-(g): Estimated shape profile using methods in [9], [15], [16], [17], [36] and the proposed algorithm, respectively.

synthetic case of face (Red-Indian) depth map. Fig. 3 (a) shows the difference between hard thresholding, soft thresholding and Garrote thresholding functions indicated using plots with dashed, dotted and solid lines, respectively. Fig. 3 (b) shows the plot of data cost versus number of Bregman iterations. From Fig. 3 (b), we observe that convergence rate of the proposed WNNM-NNGT algorithm for depth estimation is better than WNNM algorithm.

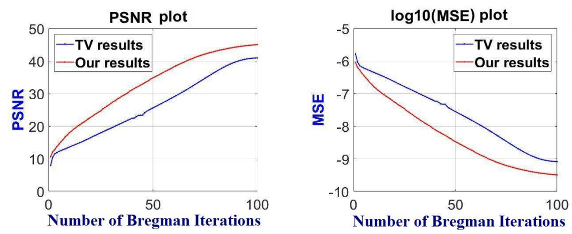
The value of λ in Eq. (38) is found empirically and fixed as $\lambda = 1 \times 10^6$ in our all synthetic experiments. Fig. 4 shows the comparison of the results of the proposed approach for synthetic face depth map (Red-Indian) with other state-of-the-art SFF techniques. In Fig. 4 (g) we observe that our approach has successfully reconstructed most of the details such as nose, eyes, lips, texture of the feathers, even the philtrum and left ear-ring.

Fig. 5 shows in detail the superiority of the proposed algorithm over the method in [36] for face depth map (Red-Indian). Fig. 5 (a) shows the ground truth depth map (b) and (c) represents the results obtained using [36] and the proposed approach, respectively. In Figs. 5 (b) and (c), it is difficult to visually discern the superiority of our approach over the method in [36]. Therefore, we estimate the residual map by taking absolute difference between the estimated and the ground truth depth maps which are shown in Figs. 5 (d) and (e), respectively.

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

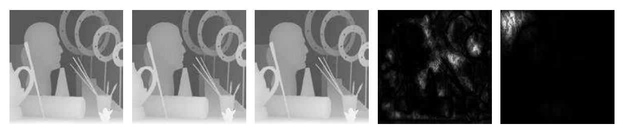
Figure 5. (a) Ground truth depth profile of face (Red-Indian). (b),(c) Depth profile estimated using [36] and the proposed algorithm, respectively. (d),(e) Residual maps corresponding to (b),(c).

The PSNR values of the estimated results are calculated $PSNR = 10\log\left(\frac{\bar{\mathbf{d}}^2_{max}}{MSE}\right)$, where $\bar{\mathbf{d}}_{max} = 3200\mu$m is the maximum value of depth $\bar{\mathbf{d}}$. Variation of PSNR in dB and mean square error (MSE) with respect to number of Bregman iterations is shown in Figs. 6 (a) and 6 (b), respectively. We observe that the PSNR of depth profile estimated using low rank prior is better than the state-of-the-art method proposed in [36].



(a) (b)

Figure 6. (a) Plot of PSNR values and (b) Mean square error values versus number of Bregman iterations, respectively.



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

Figure 7. (a) Ground truth depth profile of the Art depth map taken from Middlebury data set [50]. (b), (c) Depth profile estimated using [36] and the proposed algorithm, respectively. (d), (e) Residual map corresponding to (b), (c).

In the second synthetic experiment, we consider the Art depth map taken from Middlebury dataset [50] to generate the stack of blurred observations of size 140×140 pixels by considering translation step $m = [20, 35, 185 \text{ and } 200]$. Fig. 7 (a) shows the ground truth depth map. The results obtained using [36] and the proposed approach are shown in Figs. 7 (b) and (c), respectively. The residual maps obtained by taking absolute difference between the estimated and the ground truth depth map and are shown in Figs. 7 (d) and (e), respectively which show the superiority of our technique.

In Table 1. we provide a quantitative comparison of the proposed algorithm with [36] in terms of MSE, PSNR and SSIM parameters corresponding to face (Red-Indian) depth map and Art depth map.

6.2. Real-world Data Sets

In the first real-case experiment, we used a coin as the specimen, with the head of a lion engraved on it. A stack of images of size 227×197 pixels are captured using Nikon LV 150 industrial microscope by moving the translational

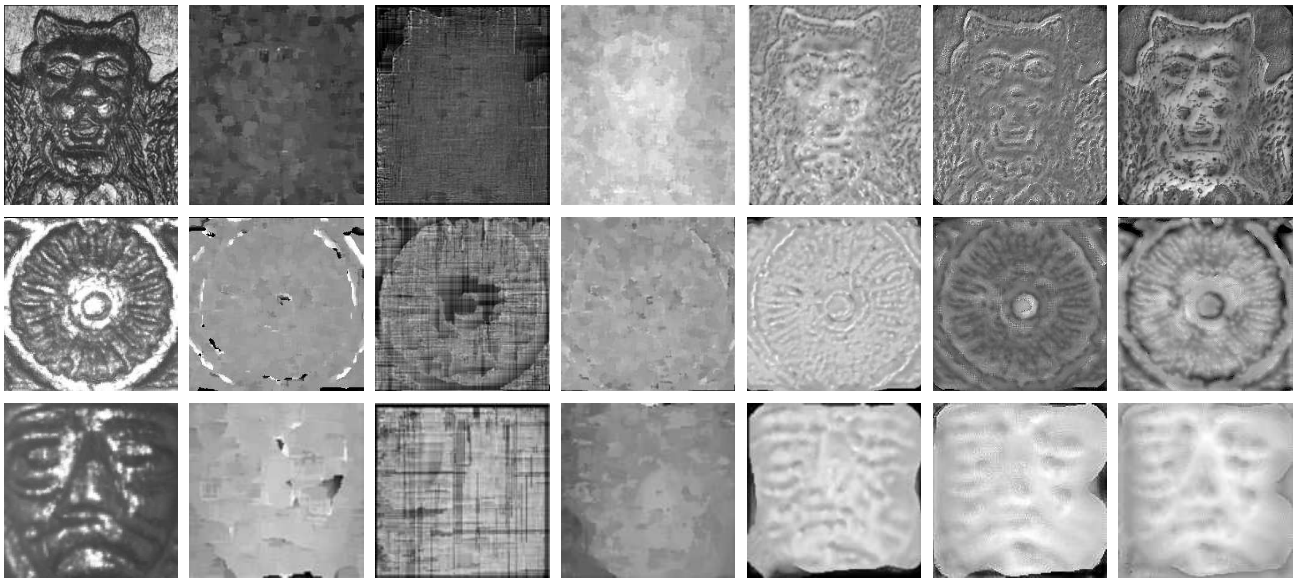

(a) (b) (c) (d) (e) (f) (g)

Figure 8. (a) Focus image. (b)-(g): Estimated shape profile using methods in [9], [15], [16], [17], [36] and the proposed algorithm, respectively.

Table 1. Quantitative evaluation of our and the method in [36].

	Methods	**MSE**	**PSNR in dB**	**SSIM**
Face depth	**TV** [36]	8.226e-10	40.951	0.9999
	Ours	3.202e-10	45.048	1.0000
Art depth	**TV** [36]	4.658e-9	33.420	0.9999
	Ours	6.907e-10	41.710	1.0000

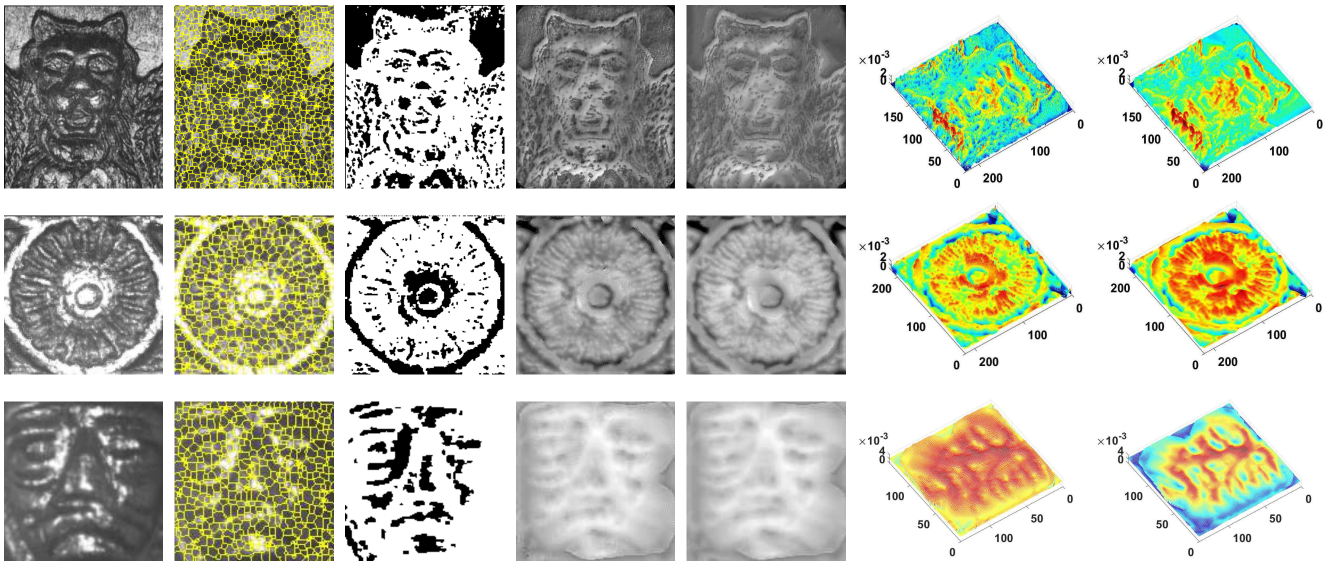
stage in-step of $\Delta d = 25\mu$m. The estimated focused image is shown in the first row of Fig. 8 (a). The initial estimate of the 3D profile for real-world cases is obtained using SFF method as given in [9] after median filtering using a 25×25 kernel. We used two images corresponding to $m = [10, 20]$ from the captured stack to estimate the structure. An empirical value of $\lambda = 1 \times 10^9$ is used in our experiments to estimate the structure. Figs. 8 (b) to (g) shows the recovered shape using the state-of-the-art techniques in [9, 15, 16, 17, 36] and the proposed approach, respectively. As compared to other state-of-the-art techniques, our algorithm successfully reconstructed the finer details such as eyes, nostrils, hair on the lion's mane, etc.

For the second experiment using a real 3D object, we imaged a small portion of a coin depicting a wheel. A stack of images of size 219×219 pixels are captured using the same industrial microscope as in the previous case by moving the translational stage in-step of $\Delta d = 25\mu$m. The estimated focused image is shown in second row of Fig. 8 (a). Using the estimated focus image and four blurred images from the stack corresponding to $m = [30, 40, 70, 80]$ and $\lambda = 5 \times 10^9$, we estimated the structure. Recovered shape profile using the state-of-the-art techniques in [9, 15, 16, 17, 36] and the proposed approach are shown Figs. 8 (b) to (g), respectively. As compared to other state-of-the-art techniques our approach is more successful in reconstructing fine variations like spokes of the wheel.

In the third experiment, we used a metal ring as a 3D specimen which had the face of a man engraved on its surface. Using fixed steps of $\Delta d = 25\mu$m we captured the image stack with each observation of size 140×140 pixels. Using estimated focus image shown in third row of Fig. 8 (a) and the observations corresponding to $m = [85, 95, 130, 140]$, we estimated the shape profile by choosing an empirical value of $\lambda = 1 \times 10^8$. In Figs. 8 (b) to (g) we present the results obtained using the methods in [9, 15, 16, 17, 36] and the proposed approach, respectively.

6.3. Structure Recovery at Specular Regions

To recover the structure at specular regions, we divide the entire focused image into super-pixels first using the technique in [21]. A binary mask corresponding to specular regions is generated by clustering the group of super-pixels into specular and non-specular regions with the aid of k-means clustering algorithm [22]. We inpainted the erroneous shape profile at specular regions by minimizing the objective function in Eq. (42). The rows of Fig. 9 (a) represent the focused images of the three real-word datasets, Fig. 9 (b) depicts the estimated super-pixels of the focused images and the generated binary masks are shown in Fig. 9 (c), respectively. Erroneous shape profile and the inpainted structure are given in columns of Fig. 9 (d) and (e). Columns (f) and (g) of Fig. 9, show 3D plots of estimated



(a) (b) (c) (d) (e) (f) (g)

Figure 9. Result of depth inpainting at specular regions. (a) Focused image. (b) Estimated super-pixels of focused image using [21]. (c) Estimated binary mask using k-means clustering. (d) Erroneous 3D structure. (e) Inpainted 3D structure. (f) 3D plot of erroneous structure. (g) 3D plot of inpainted depth profile.

and inpainted shape profiles, respectively.

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

Figure 10. (a) Erroneous shape profile. (b) Inpainted structure using [51]. (c) Recovered shape profile using [52]. (d) Results of the proposed method.

We compared the performance of the proposed inpainting approach with other state-of-the-art inpainting techniques in [51, 52]. The estimated shape profile erroneous at specular regions is shown in Fig. 10 (a). Inpainted depth maps obtained using the methods in [51, 52] and the proposed algorithm are shown in Figs. 10 (b), (c) and (d), respectively. From Fig. 10 (d), we observed that the proposed algorithm for inpainting the erroneous depth map performs superior than the other state-of-the-art inpainting techniques [51, 52].

Throughout our experiments, we assumed an additive white, zero mean, Gaussian noise (AWGN) model. This is a realistic mathematical model for our imaging conditions since adequate bright light illuminates the 3D specimen as it is translated under the lens of the microscope. Only when physical constraints such as low-power light source, short exposure time affect imaging the major source of noise would be strongly signal-dependent. Consequently, it would be more reasonable to model noise as Poisson-distributed.

7. Conclusion

We proposed an algorithm to estimate the structure of 3D objects using low rank prior in SFF. Since, 3D structure estimation is an ill-posed inverse problem, we used WNN as a regularizer and optimized the proposed objective function using split-Bregman technique. We adopted a new singular value thresholding scheme, WNNM-NNGT, instead of soft-thresholding to achieve superior convergence. We have addressed the problem of specular reflections which is inevitable in SFF and recovered the inpainted structure.

Acknowledgements: The authors are especially grateful to Department of Science and Technology (DST), India for providing the funding support for the proposed work. We thank Mr. Sankaraganesh Jonna, PhD research scholar, associated with Computational Vision laboratory for useful discussions which benefited the proposed work.

References

[1] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in *Proceedings of IEEE International Conference on Computer Vision*, 1999, pp. 1033–1038.

[2] M. Elad, M. A. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” *Proceedings of the IEEE Special Issues on Applications of Sparse Representation and Compressive Sensing*, vol. 98, no. 6, pp. 972–982, 2010.

[3] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 60–65.

[4] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 165–168, 1980.

[5] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[6] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.

[7] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[8] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *International Journal of Computer Vision*, vol. 121, no. 2, pp. 183–208, 2017.

[9] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[10] L. Ma, L. Xu, and T. Zeng, "Low rank prior and total variation regularization for image deblurring," *Journal of Scientific Computing*, vol. 70, no. 3, pp. 1336–1357, 2017.

[11] T. Goldstein and S. Osher, "The split bregman method for$\ell 1$-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[12] J.-F. Cai, S. Osher, and Z. Shen, "Split bregman methods and frame based image restoration," *Multiscale Modeling & Simulation*, vol. 8, no. 2, pp. 337–369, 2009.

[13] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[14] H.-Y. Gao, "Wavelet shrinkage denoising using the nonnegative garrote," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 469–488, 1998.

[15] M. B. Ahmad and T.-S. Choi, "Shape from focus using optimization technique," in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, 2006, pp. 493–496.

[16] K. Pradeep and A. Rajagopalan, "Improving shape from focus using defocus cue," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1920–1925, 2007.

[17] R. Sahay and A. Rajagopalan, "Shape extraction of low-textured objects in video microscopy," *Journal of Microscopy*, vol. 245, no. 3, pp. 252–264, 2012.

[18] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 266–274, 1995.

[19] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493–500, 2007.

[20] J. Yun and T.-S. Choi, "Accurate 3-d shape recovery using curved window focus measure," in *Proceedings of IEEE International Conference on Image Processing*, 1999, pp. 910–914.

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[22] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[23] U. R. Dhond and J. K. Aggarwal, "Structure from stereo-a review," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489–1510, 1989.

[24] S. T. Barnard and M. A. Fischler, "Computational stereo," *ACM Computing Surveys (CSUR)*, vol. 14, no. 4, pp. 553–572, 1982.

[25] S. W. Hasinoff and K. N. Kutulakos, "Confocal stereo," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 82–104, 2009.

[26] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[27] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 557–564.

[28] S. Chaudhuri and A. N. Rajagopalan, "Depth recovery from defocused images," in *Depth From Defocus: A Real Aperture Imaging Approach*. Springer, 1999, pp. 14–27.

[29] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 406–417, 2005.

[30] P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 518–531, 2008.

[31] M. Subbarao, "Parallel depth recovery by changing camera parameters." in *Proceedings of IEEE International Conference on Computer Vision*, 1988, pp. 149–155.

[32] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 70, 2007.

[33] M. Asif, T.-S. Choi *et al.*, "Shape from focus using multilayer feedforward neural networks," *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1670–1675, 2001.

[34] R. Sahay and A. Rajagopalan, “Harnessing defocus blur to recover high-resolution information in shape-from-focus technique,” *IET Computer Vision*, vol. 2, no. 2, pp. 50–59, 2008.

[35] K. Ramnath and A. N. Rajagopalan, “Discontinuity-adaptive shape from focus using a non-convex prior.” in *DAGM-Symposium*. Springer, 2009, pp. 181–190.

[36] D. Kumar CH, V. HI, and R. R. Sahay, “Shape-from-focus using total variation prior and split bregman algorithm,” in *Proceedings of Indian Conference on Computer Vision Graphics and Image Processing*, 2014, p. 80.

[37] R. R. Sahay and A. N. Rajagopalan, “Dealing with parallax in shape-from-focus,” *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 558–569, 2011.

[38] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, “Compressive sensing via nonlocal low-rank regularization,” *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.

[39] H. Ji, C. Liu, Z. Shen, and Y. Xu, “Robust video denoising using low rank matrix completion,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1791–1798.

[40] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, p. 717, 2009.

[41] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[42] Q. Xie, D. Meng, S. Gu, L. Zhang, W. Zuo, X. Feng, and Z. Xu, “On the optimal solution of weighted nuclear norm minimization,” *CoRR*, vol. abs/1405.6012, 2014.

[43] K. Chen, H. Dong, and K.-S. Chan, “Reduced rank regression via adaptive nuclear norm penalization,” *Biometrika*, vol. 100, no. 4, pp. 901–920, 2013.

[44] A. P. Pentland, “A new sense for depth of field,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 523–531, 1987.

[45] Z. Zha, X. Zhang, Q. Wang, Y. Bai, and L. Tang, “Analyzing the weighted nuclear norm minimization and nuclear norm minimization based on group sparse representation,” *arXiv preprint arXiv:1702.04463*, 2017.

[46] P. Brodatz, *Textures: a photographic album for artists and designers*. Dover Pubns, 1966.

[47] Y.-H. Dai and Y. Yuan, “A nonlinear conjugate gradient method with a strong global convergence property,” *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999.

[48] W. W. H and H. Zhang, “A survey of nonlinear conjugate gradient methods,” *Pacific journal of Optimization*, vol. 2, no. 1, pp. 35–58, 2006.

[49] W. W. Hager and H. Zhang, “A new conjugate gradient method with guaranteed descent and an efficient line search,” *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 170–192, 2005.

[50] H. Hirschmuller and D. Scharstein, “Evaluation of cost functions for stereo matching,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[51] A. Criminisi, P. Pérez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[52] K. H. Jin and J. C. Ye, “Annihilating filter-based low-rank hankel matrix approach for image inpainting,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3498–3511, 2015.