Extension of the shape from focus method for reconstruction of high-resolution images

R. R. Sahay[1,*] and A. N. Rajagopalan[1]

[1]*Image Processing and Computer Vision Laboratory, Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600 036, India*
[*]*Corresponding author: sahayiitg@yahoo.com*

Received March 29, 2007; revised August 13, 2007; accepted September 24, 2007; posted September 28, 2007 (Doc. ID 81637); published October 30, 2007

Shape from focus (SFF) estimates the depth profile of a 3D object using a sequence of observations. Due to the finite depth of field of real aperture cameras and the 3D nature of the object, none of the observations is completely in focus. However, in many applications, it is important to examine finer image details of the underlying object in conjunction with the depth map. We propose an extension to the traditional SFF method to optimally estimate a high-resolution image of the 3D object, given the low-resolution observations and the depth map derived from traditional SFF. Using the observation stack, we show that it is possible to achieve significant improvement in resolution. We also analyze the special case of region of interest superresolution and show analytically that an optional interframe separation exists for which the quality of the estimated high-resolution image is the best. © 2007 Optical Society of America

OCIS codes: 100.6640, 100.3010.

1. INTRODUCTION

Shape from focus (SFF) is a technique for estimating the depth profile of a 3D object. It uses the degree of focus in an image as a cue to extract shape [1]. A sequence or stack of observations is captured with different portions of the object coming into focus in different images. A focus measure profile is computed for each pixel across all the observations in the stack. The depth of the corresponding point on the 3D object is found by Gaussian interpolation of a few values near the peak of the focus measure profile. Subbarao and Choi [2] fit a low-order polynomial to a few data points lying on the largest mode. In [3], a piecewise curved local window is proposed for computing the focus measure. Asif and Choi [4] use multilayer feed-forward neural networks for finding the shape. A dynamic programming-based optimization approach has been proposed in [5], which estimates the shape using the notion of a focused image surface (FIS). In [6], a method for finding the optimal illumination filter for a given optical setup and the exact number of images required for accurate shape recovery are described.

Within the SFF framework, wherein one is interested in estimating structure, it is important to examine whether one can also obtain a high-resolution (HR) focused image of the 3D surface from the low-resolution (LR) observations. Superresolution is useful in several vision tasks, such as automated inspection of industrial samples, printed circuit boards, silicon wafers, etc. In conjunction with the depth map, it enhances the ability to make better decisions regarding presence and localization of defects and in recognizing meaningful patterns in a sample. To our knowledge, no attempt has been made to obtain an HR image in SFF.

The resolution of the image captured by a camera is limited by the density of photodetectors in the sensor array. Images captured with an LR camera suffer from aliasing, blurring, and noise. The goal behind signal processing-based superresolution is to undo these effects and yield an HR image. Superresolution algorithms can be divided into three broad categories: motion based, motion free, and learning based. Motion-based [7–12] algorithms use the relative motion between the camera and the scene to derive an HR image from subpixel shifted LR images. The theoretical foundation for motion-free superresolution [13–16] is provided by the classic result of Papoulis [17] that a band-limited signal $x(t)$ can be uniquely reconstructed in terms of the samples of the responses of M linear time-invariant systems with input $x(t)$ and output sampled at $1/M$ the Nyquist rate. Learning-based [18–20] algorithms learn about the statistical relationships between LR and HR images during the training phase and use these relationships for superresolution.

In the SFF setting, defocusing occurs naturally as images of an object are captured by moving the object with respect to the camera. Since blur can be used as a cue for superresolution, we propose to extend the scope of traditional SFF to generate an HR focused image of the 3D object. The observation stack used by SFF and the depth map derived from it are taken as inputs. Note that the spatial resolution of the depth map is the same as that of the captured observations. By exploiting certain novel features of the mechanism of the process of blurring in SFF and the fact that many observations are available, we show that it is possible to achieve resolutions much beyond the factor "2", which is typically reported in works on motion-free superresolution using the depth from defocus (DFD) [13] technique.

In many applications, only a certain portion of the 3D object (and not the entire image) is required to be superresolved. This is also termed as region of interest (ROI)

superresolution. In such a situation, the blurring corresponding to the ROI can be treated as space-invariant (assuming that the magnitude of the undulations within the ROI is less than the depth of field of the camera). Given the stack of LR observations captured in the SFF setting, we address the following issue: If p LR frames are used to superresolve the ROI by a factor of q, does an optimal uniform interframe separation, which yields best performance exist? We investigate this problem in a statistical framework using the Cramér–Rao lower bound (CRLB). Further, we investigate the interesting possibility of how this result can be used in real situations.

The paper is organized as follows: Section 2 discusses the SFF technique and the process of image formation in a real aperture camera. The degradation model is presented in Section 3. A maximum *a posteriori*-Markov random field (MAP-MRF) framework for obtaining the HR image from LR observations is discussed in Section 4. Performance analysis of the proposed algorithm for ROI superresolution is given in Section 5. Experimental results are compiled in Section 6. Conclusions and directions for future work are outlined in Section 7.

2. SHAPE FROM FOCUS

The traditional shape from focus scheme [1] is depicted in Fig. 1. A 3D object is placed on a translational stage which moves in a vertical direction in finite steps of size Δd. At each step, an image is captured in which, barring a small portion of the object, other regions are defocused by different degrees, because a real aperture camera cannot bring all the points of a 3D object into focus at the same time. Thus, a stack of space-variantly defocused observations is obtained. The shape of the object is determined by searching for those frames in which the object points come in focus.

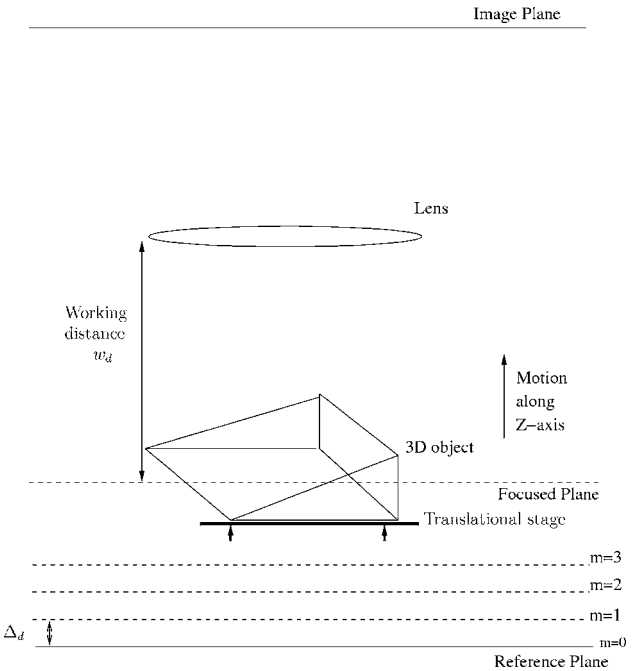

Fig. 1. Schematic of traditional SFF.

In traditional SFF [1], a focus measure profile is computed for each pixel across the image stack. The focus measure at a point (k,l) in an image I is computed using the sum-modified Laplacian operator (SML) as

$$F(k,l) = \sum_{m=k-W}^{k+W} \sum_{n=l-W}^{l+W} O_L(m,n), \quad \text{for } O_L(m,n) \geq T_1, \quad (1)$$

where T_1 is a threshold, $2W+1$ is the size of the window around the point (k,l), and O_L is the modified Laplacian defined in the discrete domain as $O_L(m,n)=|2I(m,n)-I(m-\delta,n)-I(m+\delta,n)|+|2I(m,n)-I(m,n-\delta)-I(m,n+\delta)|$, where δ is a variable spacing between the image pixels for computing the derivatives. The focus measure profile for a pixel at (k,l) is obtained by plotting the value of $F(k,l)$ computed at (k,l) in every image of the stack starting from the reference frame. An estimate of the depth at (k,l) is arrived by Gaussian interpolation of a few values near the peak of the focus measure profile.

According to the lens law, a point on the object at a distance D' from the lens will be in focus on the sensor plane, which is at a distance v from the lens plane, when D' is such that $1/f=1/D'+1/v$. Here f denotes the focal length. Points that are not at distance D' from the lens plane will appear blurred. The point spread function (PSF) of a camera is defined as the response of the camera to a point light source. Using ray tracing, a point light source at distance D from the lens plane will be imaged on the sensor plane as a circular disk called the circle of confusion with radius $r_b=Rv(1/f-1/v-1/D)$ where R is the radius of the aperture of the lens. Due to diffraction and lens aberrations, the PSF is best described by a circularly symmetric 2D Gaussian function [21]:

$$h(x,y) = \frac{1}{2\pi\sigma^2} \exp \frac{-(x^2+y^2)}{2\sigma^2}.$$

Here, σ is the blur parameter which is related to the blur radius as $\sigma=\rho r_b$, where ρ is camera constant. Several works [13,22–24] exist that validate the approximation of the true camera PSF by a 2D Gaussian. Note that the blurring undergone by a point on a 3D object is governed by the blur parameter σ, which is related to the depth D of that point as

$$\sigma = \rho R v \left(\frac{1}{f} - \frac{1}{v} - \frac{1}{D} \right). \quad (2)$$

When the 3D point satisfies the lens law, it will appear in perfect focus and the value of the blur parameter will be zero.

3. DEGRADATION MODEL

Suppose the image plane consists of $M \times M$ sensor elements. Let the LR image intensity values be denoted by $\{y(i,j)\}$, $(i,j)=0,1,\ldots,M-1$. If we superresolve by a factor q, the HR image x will be of size $qM \times qM$ where $\{x(m,n)\}$, $(m,n)=0,1,\ldots,qM-1$. The mapping of image intensity values from the original HR grid to the LR sensor grid can be described as

$$f(i,j) = \frac{1}{q^2} \sum_{m=qi}^{q(i+1)-1} \sum_{n=qj}^{q(j+1)-1} x(m,n). \quad (3)$$

The defocused image can be expressed as

$$g(i,j) = \sum_k \sum_l f(k,l) h(i,j;k,l), \quad (4)$$

where the space-variant function:

$$h(i,j;k,l) = \frac{1}{2\pi\sigma^2(k,l)} \exp\left(- \frac{(i-k)^2 + (j-l)^2}{2\sigma^2(k,l)} \right).$$

Observation noise is finally added to $g(i,j)$ to obtain the LR defocused image $y(i,j)$.

From the stack of observations available in SFF, suppose we pick p number of LR frames $\{y_m(i,j)\}$, $m = 1,2,\ldots,p$, each of size $M \times M$. These are decimated, blurred, and noisy versions of a single HR image $\{x(m,n)\}$ of size $qM \times qM$. If $\mathbf{y}_m$ is the lexicographically arranged vector containing pixels from the mth LR image of size $M^2 \times 1$, and $\mathbf{x}$ is the lexicographically arranged vector containing pixels from the original HR image of size $q^2M^2 \times 1$, then they can be related as [14,16],

$$\mathbf{y}_m = \mathbf{H}_m \mathbf{D} \mathbf{x} + \mathbf{n}_m, \quad m = 1, \ldots, p, \quad (5)$$

where $\mathbf{H}_m$ is a blur matrix of size $M^2 \times M^2$, $\mathbf{D}$ is the decimation matrix of size $M^2 \times q^2M^2$, and $\mathbf{n}_m$ is zero mean Gaussian noise vector of size $M^2 \times 1$ with variance σ_η^2. Another degradation model that is used in motion-free superresolution is $\mathbf{y}_m = \mathbf{D}\mathbf{H}_m\mathbf{x} + \mathbf{n}_m$. However, Bose *et al.* [16] have shown that the performance of these two observation models is comparable. In SFF, the choice of the degradation model given by Eq. (5) is motivated by the fact that the depth map is available at the same resolution as the observations.

The problem that we address is the extraction of the HR image $\mathbf{x}$ given the LR observations $\mathbf{y}_m$, $m = 1,2,\ldots,p$, and the depth map of the 3D object derived using traditional SFF.

When the 3D object is placed on the translational stage, which is at the reference plane as shown in Fig. 1, the blur induced by a point on the object in the reference frame is governed by blur parameter σ_0 which is given by

$$\sigma_0 = \rho R v \left(\frac{1}{w_d} - \frac{1}{D_0} \right), \quad (6)$$

where D_0 is the distance of the object point from the lens when the stage is at the reference position and w_d is the working distance of the camera, i.e., $1/w_d = 1/f - 1/v$. The stage is moved vertically by a distance of $m\Delta d$ to capture the mth LR frame. For the same point on the 3D object, the blurring induced in the mth frame can be expressed by the blur parameter σ_m which is given by

$$\sigma_m = \rho R v \left(\frac{1}{w_d} - \frac{1}{D_0 \pm m\Delta d} \right). \quad (7)$$

The change in magnification across the stack of LR observations is assumed to be negligible so that there are no errors due to registration. Eliminating the common term w_d from Eqs. (6) and (7), we get

$$\sigma_m = \sigma_0 + \rho R v \left(\frac{1}{D_0} - \frac{1}{D_0 \pm m\Delta d} \right). \quad (8)$$

Using the depth map computed by the SFF method and Eq. (6), the blur parameter σ_0 at every point in the reference image can be computed. The blur parameter σ_m at any point in the mth observation, $m = 1,2,\ldots,p$, can be determined with the knowledge of σ_0 at the same point in the reference frame using the relationship in Eq. (8). Note that the value of $\rho R v$ in Eq. (7) remains constant during the entire image capturing process.

For a magnification factor of q, the number of independent LR observations needed to estimate an HR image is at least q^2. Because in SFF we capture a large number of LR observations to compute depth, the advantage is that we can use these LR frames in the stack to super-resolve by higher factors.

4. MAP-MRF FRAMEWORK

In Eq. (5), the blur matrix $\mathbf{H}_m$ can be constructed using the relationship given by Eq. (7). The problem of finding $\mathbf{x}$ given the observations $\mathbf{y}_m$, $m = 1,2,\ldots,p$, and the blur matrix $\mathbf{H}_m$ in Eq. (5) is ill-posed. It is singular at worst and ill-conditioned at best. Regularization in the form of *a priori* constraints on the solution can be imposed by suitably modeling the HR focused image. We propose to compute the MAP estimate of the HR image. Using Bayes' rule we can write

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p)$$
$$= \arg\max_{\mathbf{x}} [\log P(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p|\mathbf{x}) + \log P(\mathbf{x})]. \quad (9)$$

MRF models capture information contained in the local intensity distribution of images in a natural way. Spatial correlation such as local dependencies among neighboring pixel values can be conveniently described using MRFs [25]. They can encode contextual constraints as well as provide *a prior* distribution with which to model the probability density function (PDF) of the original HR image.

The Markovian property of the MRF states that the probability of a pixel being assigned a particular gray value depends only on the image data in its neighborhood. The Hammersley-Clifford theorem [26] provides the all-important equivalence between MRF and the Gibbs random field (GRF). If the HR image $\mathbf{x}$ is modeled as a Gauss–Markov random field (GMRF) then

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left[- \sum_{c \in C} (\mathbf{d}_c^T \mathbf{x})^2 \right], \quad (10)$$

where Z is the partition function, c is a clique, C is the set of all cliques, $\mathbf{d}_c$ is an operation on clique c such that $\mathbf{d}_c^T\mathbf{x}$ provides a measure of smoothness of the image by computing discrete approximations of first (or second) derivatives at each image pixel. For a first-order neighborhood, we have

$$\sum_{c\in C}(\mathbf{d}_c^T\mathbf{x})^2=\sum_{i=1}^{qM}\sum_{j=1}^{qM}[(x(i,j)-x(i,j-1))^2+(x(i,j+1)-x(i,j))^2+(x(i+1,j)-x(i,j))^2+(x(i,j)-x(i-1,j))^2]. \tag{11}$$

Assuming the noise process $\mathbf{n}_m's$ to be independent in Eq. (5), and from Eqs. (9) and (11),

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}}\sum_{m=1}^{p}\frac{\|\mathbf{y}_m-\mathbf{H}_m\mathbf{D}\mathbf{x}\|^2}{2\sigma_\eta^2}+\lambda\sum_{i=1}^{qM}\sum_{j=1}^{qM}[(x(i,j)-x(i,j-1))^2+(x(i,j+1)-x(i,j))^2+(x(i+1,j)-x(i,j))^2+(x(i,j)-x(i-1,j))^2]. \tag{12}$$

The regularization factor λ plays a critical role in balancing the importance between the prior and the observations. The values of σ_η^2 and λ are typically tuned to derive the best estimate of the HR image $\mathbf{x}$. The number of LR observations p used to compute the cost function depends on the resolution factor. For example, to upsample by a factor of 2, we require four LR observations. But to upsample by a factor of 6, we would require 36 observations.

The cost function in Eq. (12) is convex since it is the sum of two convex functions. Hence, one can use simple gradient descent for optimization. At the nth iteration, the gradient of the cost function in Eq. (12) is given by

$$\text{grad}^{(n)}=\frac{1}{\sigma_\eta^2}\sum_{m=1}^{p}\mathbf{D}^T\mathbf{H}_m^T(\mathbf{H}_m\mathbf{D}\mathbf{x}^{(n)}-\mathbf{y}_m)+\lambda Q^{(n)}, \tag{13}$$

where

$$Q^{(n)}=\sum_{i=1}^{qM}\sum_{j=1}^{qM}2[4x^{(n)}(i,j)-x^{(n)}(i,j-1)-x^{(n)}(i,j+1)-x^{(n)}(i-1,j)-x^{(n)}(i+1,j)].$$

Operation $\mathbf{D}^T$ spreads the LR pixel intensity value equally at corresponding pixel locations in the HR image. Matrix $\mathbf{H}_m$ is obtained using relation Eq. (8) as described earlier. The estimate of the HR image in the $(n+1)$th iteration is given by

$$\mathbf{x}^{(n+1)}=\mathbf{x}^{(n)}-\beta\,\text{grad}^{(n)},$$

where β is the step size. The estimates of the HR image $\mathbf{x}$ are computed iteratively until $\|\mathbf{x}^{(n+1)}-\mathbf{x}^{(n)}\|<$ threshold. The initial estimate $\mathbf{x}^{(0)}$ can be chosen to be one of the LR observations upsampled using any standard interpolation technique.

5. REGION OF INTEREST SUPERRESOLUTION

It is often required to superresolve only a small portion corresponding to the ROI on the specimen. The blurring can be assumed to be space-invariant over this portion if the magnitude of the undulations in depth within the ROI is less than the depth of field of the camera. Using the proposed algorithm, we next seek to perform the task of ROI superresolution. Interestingly, the performance can be bounded by analyzing the problem in a statistical framework using the posterior CRLB.

Consider a scenario in which we have captured a stack of LR observations in SFF. We seek to superresolve the ROI by a factor of q using p LR frames with uniform interframe separation $\Delta d'$. In what follows, we show that an optimal value for $\Delta d'$ exists, which results in the best reconstruction of the HR image of the ROI.

Let $\mathbf{v}=[v_1,v_2,\ldots,v_N]^T$ be the realization of a random vector $\mathbf{V}$, which depends upon a random parameter vector $\boldsymbol{\Theta}=[\theta_1,\theta_2,\ldots,\theta_K]^T$ with prior distribution $P_{\boldsymbol{\Theta}}(\theta)$. The joint PDF $P_{\mathbf{V},\boldsymbol{\Theta}}(\mathbf{v},\theta)=P_{\mathbf{V}|\boldsymbol{\Theta}}(\mathbf{v}|\theta)P_{\boldsymbol{\Theta}}(\theta)$. Let $\hat{\boldsymbol{\Theta}}$ be an unbiased estimator of $\boldsymbol{\Theta}$. The covariance of the error in the estimate of $\boldsymbol{\Theta}$ is lower bounded as

$$E[(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})^T]\geq J^{-1}(\boldsymbol{\Theta}), \tag{14}$$

where $J(\boldsymbol{\Theta})$ is the Fisher information matrix, which is given by

$$J(\boldsymbol{\Theta})=-\left[E\left\{\frac{\partial^2\log P_{\mathbf{V}|\boldsymbol{\Theta}}(\mathbf{v}|\theta)}{\partial\boldsymbol{\Theta}^2}\right\}+E\left\{\frac{\partial^2\log P_{\boldsymbol{\Theta}}(\theta)}{\partial\boldsymbol{\Theta}^2}\right\}\right].$$

The left hand side of this inequality gives the error covariance matrix whose diagonal elements represent the mean squared error (MSE) of each component of $\boldsymbol{\Theta}$.

For mathematical convenience but without any loss of generality, let us consider a 1D case for ROI superresolution. The degradation model is given by

$$\mathbf{y}_m=\mathbf{H}_m\mathbf{D}\mathbf{x}+\mathbf{n}_m,\quad m=1,2,\ldots,p. \tag{15}$$

Here, $\mathbf{x}$ is the HR signal of size $N\times 1$, $\mathbf{D}$ is the decimation matrix of size $M\times N$, the blurring matrix $\mathbf{H}_m$ is of size $M\times M$, and $\mathbf{y}_m$ is the mth LR observation with *i.i.d.* zero mean Gaussian noise $\mathbf{n}_m$. Since the blur is assumed to be space invariant over the ROI, the blurring matrix $\mathbf{H}_m$ will be circulant.

Equation (15) can be transformed to the frequency domain through unitary discrete Fourier transform (DFT) matrices $\boldsymbol{\Phi}_\mathbf{H}$ of size $M\times M$ and $\boldsymbol{\Phi}_\mathbf{X}$ of size $N\times N$ as follows:

$$\begin{aligned}\boldsymbol{\Phi}_\mathbf{H}\mathbf{y}_m&=\boldsymbol{\Phi}_\mathbf{H}\mathbf{H}_m\mathbf{D}\mathbf{x}+\boldsymbol{\Phi}_\mathbf{H}\mathbf{n}_m\\&=\boldsymbol{\Phi}_\mathbf{H}\mathbf{H}_m\boldsymbol{\Phi}_\mathbf{H}^*\boldsymbol{\Phi}_\mathbf{H}\mathbf{D}\boldsymbol{\Phi}_\mathbf{X}^*\boldsymbol{\Phi}_\mathbf{X}\mathbf{x}+\boldsymbol{\Phi}_\mathbf{H}\mathbf{n}_m.\end{aligned} \tag{16}$$

Therefore,

$$\mathbf{Y}_m=\boldsymbol{\Lambda}_{\mathbf{H}_m}\boldsymbol{\Phi}_\mathbf{H}\mathbf{D}\boldsymbol{\Phi}_\mathbf{X}^*\mathbf{X}+\mathbf{N}_m=\boldsymbol{\Psi}_m\mathbf{X}+\mathbf{N}_m, \tag{17}$$

where $\mathbf{Y}_m$, $\mathbf{X}$, and $\mathbf{N}_m$ are the DFTs of $\mathbf{y}_m$, $\mathbf{x}$, and $\mathbf{n}_m$, respectively. Matrix $\boldsymbol{\Psi}_m=\boldsymbol{\Lambda}_{\mathbf{H}_m}\boldsymbol{\Phi}_\mathbf{H}\mathbf{D}\boldsymbol{\Phi}_\mathbf{X}^*$ where * represents complex conjugate. Matrix $\boldsymbol{\Lambda}_{\mathbf{H}_m}=\boldsymbol{\Phi}_\mathbf{H}\mathbf{H}_m\boldsymbol{\Phi}_\mathbf{H}^*$ is diagonal of size $M\times M$, and the diagonal elements are the DFT coefficients of the first column of $\mathbf{H}_m$ [27]. Equation (17) is the frequency domain equivalent of the spatial domain observation model of Eq. (15). Since the PDF of the noise $\mathbf{n}_m$ was assumed to be Gaussian with zero mean and variance σ_η^2, it remains a Gaussian in the frequency domain with the same mean and variance.Therefore, the log-likelihood function in the frequency domain becomes

$$\log P(\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_p|\mathbf{X})=\log\frac{1}{(2\pi\sigma_\eta^2)^{pM/2}}-\sum_{m=1}^{p}\frac{\|\mathbf{Y}_m-\boldsymbol{\Psi}_m\mathbf{X}\|^2}{2\sigma_\eta^2}.$$

The prior distribution of the HR signal in time domain, assuming a GMRF model, is

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\left|\frac{1}{\lambda}R_{\mathbf{x}}\right|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\left(\frac{R_{\mathbf{x}}}{\lambda}\right)^{-1}\mathbf{x}\right\}, \quad (18)$$

where the length of the HR signal is N. The matrix $R_{\mathbf{x}}^{-1} = \mathbf{D}_r^T\mathbf{D}_r$, where $\mathbf{D}_r$ is a one-step forward difference. Matrix $\mathbf{D}_r^T\mathbf{D}_r$ can be approximated as a circulant matrix which is the Laplacian and is related to the operator $\mathbf{d}_c$ discussed earlier as $\Sigma_{c\in C}(\mathbf{d}_c^T\mathbf{x})^2 = \mathbf{x}^T\mathbf{D}_r^T\mathbf{D}_r\mathbf{x}$. Applying the linear transformation $\mathbf{X} = \mathbf{\Phi}_{\mathbf{X}}\mathbf{x}$, the vector of DFT coefficients of the unknown HR 1D signal in the frequency domain is jointly Gaussian with covariance matrix $\tilde{R}_{\mathbf{X}} = \mathbf{\Phi}_{\mathbf{X}}R_{\mathbf{x}}\mathbf{\Phi}_{\mathbf{X}}^*$.

The posterior CRLB can be derived and shown to be

$$J^{-1}(\mathbf{X}) = \left[\frac{1}{2\sigma_\eta^2}\sum_{m=1}^{p}\mathbf{\Psi}_m^T\mathbf{\Psi}_m^* + \frac{\lambda}{2}\mathbf{\Phi}_{\mathbf{X}}^*\mathbf{D}_r^T\mathbf{D}_r\mathbf{\Phi}_{\mathbf{X}}\right]^{-1}. \quad (19)$$

Note that the prior term in Eq. (19) lends improved invertibility to $J(\mathbf{X})$, an effect of regularization. The motivation for analyzing the CRLB in the frequency domain is due to the following reasons. The PSF of the camera, which is modeled as a Gaussian in the spatial domain remains a Gaussian in the frequency domain. Hence, the diagonal elements of $\mathbf{\Lambda}_{\mathbf{H}_m}$ can be expressed in functional form. Matrices $\mathbf{\Phi}_{\mathbf{H}}$ and $\mathbf{\Phi}_{\mathbf{X}}$ are known unitary DFT matrices. Similarly, matrix $\mathbf{D}$ is also known. Therefore, the matrix $\mathbf{\Psi}_m$ can also be written in a closed form. Likewise, we observe that since $\mathbf{D}_r^T\mathbf{D}_r$ can be approximated as a circulant matrix and the matrix $\mathbf{\Phi}_{\mathbf{X}}$ is known, the second term in Eq. (19) can also be expressed in a functional form. This enables us to obtain a closed form expression for the posterior CRLB.

The CRLB expression can be generalized to the 2D case in a straightforward manner. The bound in Eq. (19) gives a fundamental limit on the quality of the estimated unknown HR image in ROI superresolution. The performance measure for evaluating the quality of the reconstructed HR signal is the average mean squared error $(\overline{MSE}) = (1/N)E[(\mathbf{x}-\hat{\mathbf{x}})^2]$ where $\hat{\mathbf{x}}$ is the estimated HR signal. This is reflected in the posterior CR bound for the error in the estimation of $\mathbf{X}$.

6. EXPERIMENTAL RESULTS

We first demonstrate the performance of the proposed algorithm for super-resolving the focused image of 3D objects in SFF. This is followed by results on ROI superresolution with performance analysis based on the CRLB. Both synthetic and real results are given.

An LV-150 Nikon industrial microscope was used for imaging. We remark here that microscopy is not the main focus of this paper. We use images taken through a microscope only for purpose of demonstration. The lens objective was 2.5×, the working distance $w_d = 8.8$ mm, focal length $f = 80$ mm, and the depth of field = 48.9 μm. The PSF of the camera was assumed to be Gaussian, which is a reasonable approximation as discussed in Section 2. The depth map of the object was constructed using the traditional SFF method [1] which yields D_0. Equations (6) and (8) are then used to compute the blur map corresponding to each LR image.

A. Focused Image Superresolution of Three-Dimensional Objects

In the first experiment, a metal object with an engraved floral design was used as the 3D specimen. A stack of 100 LR images of size 90×130 pixels of this object were captured by moving the translational stage of the microscope in steps of $\Delta d = 0.025$ mm. One of the LR images is shown in Fig. 2(a). To obtain the initial estimate of the HR image, the 70th LR frame was upsampled by a factor of 2 using bicubic interpolation. Four LR frames were chosen from the stack and the superresolved image obtained using the proposed algorithm (with $\lambda = 0.003$, $\sigma_\eta^2 = 5$, and $\beta = 1$) is shown in Fig. 2(b). Observe that we can resolve many details in the HR image. It is interesting to note that the region shown with an arrow in Fig. 2(b) and shaped like a flipped "7" can be confused with a sleeping "U" in the LR image of Fig. 2(a).

We next attempted superresolving the above specimen by a factor of 4 by choosing 16 LR observations from the stack. In Fig. 2(c), a portion cropped from one of the LR images is shown. To obtain the initial estimate for the HR image, we upsampled one of the 16 LR frames using bicubic interpolation. The proposed algorithm was then used with parameters $\lambda = 0.003$, $\sigma_\eta^2 = 5$, and $\beta = 1$, and the superresolved image is shown in Fig. 2(d). Note that the output is good, even for $q = 4$.

As another example of a 3D object, we used a metal ring which had the face of a man engraved on its surface. The object was placed on the translational stage of the microscope and the stage was moved in fixed steps of 0.025 mm to capture a stack of 150 frames. The size of the LR observation images was 100×135 pixels.

The blur map corresponding to the reference frame is given in Fig. 3(a) where we have plotted the absolute values of the blur parameters. In Fig. 3(b), one of the LR observations is shown. We attempted superresolution by a large factor, $q = 6$, and used 36 LR frames from the stack. With $\lambda = 0.05$, $\sigma_\eta^2 = 5$, and $\beta = 1$, the superresolved image obtained using the proposed algorithm is given in Fig. 3(c). It is interesting to observe that even at such a high magnification factor the details in the face are preserved. Importantly, there are no visible artifacts.

B. ROI Superresolution

In this subsection, the blur is assumed to be space-invariant over the ROI. Our objective here is to not only demonstrate ROI superresolution using the proposed algorithm, but also to investigate whether an optimal uniform interframe separation $\Delta d'$ exists between the p LR frames chosen from the stack for superresolving the ROI by a factor of q. The optimality is tested using the theoretical CRLB derived in Section 5.

We assumed a hypothetical planar ROI and simulated the SFF method for a lens objective of 2.5×. The focused plane was 8.8 mm below the lens, and the object was initially assumed to have been placed on the translational

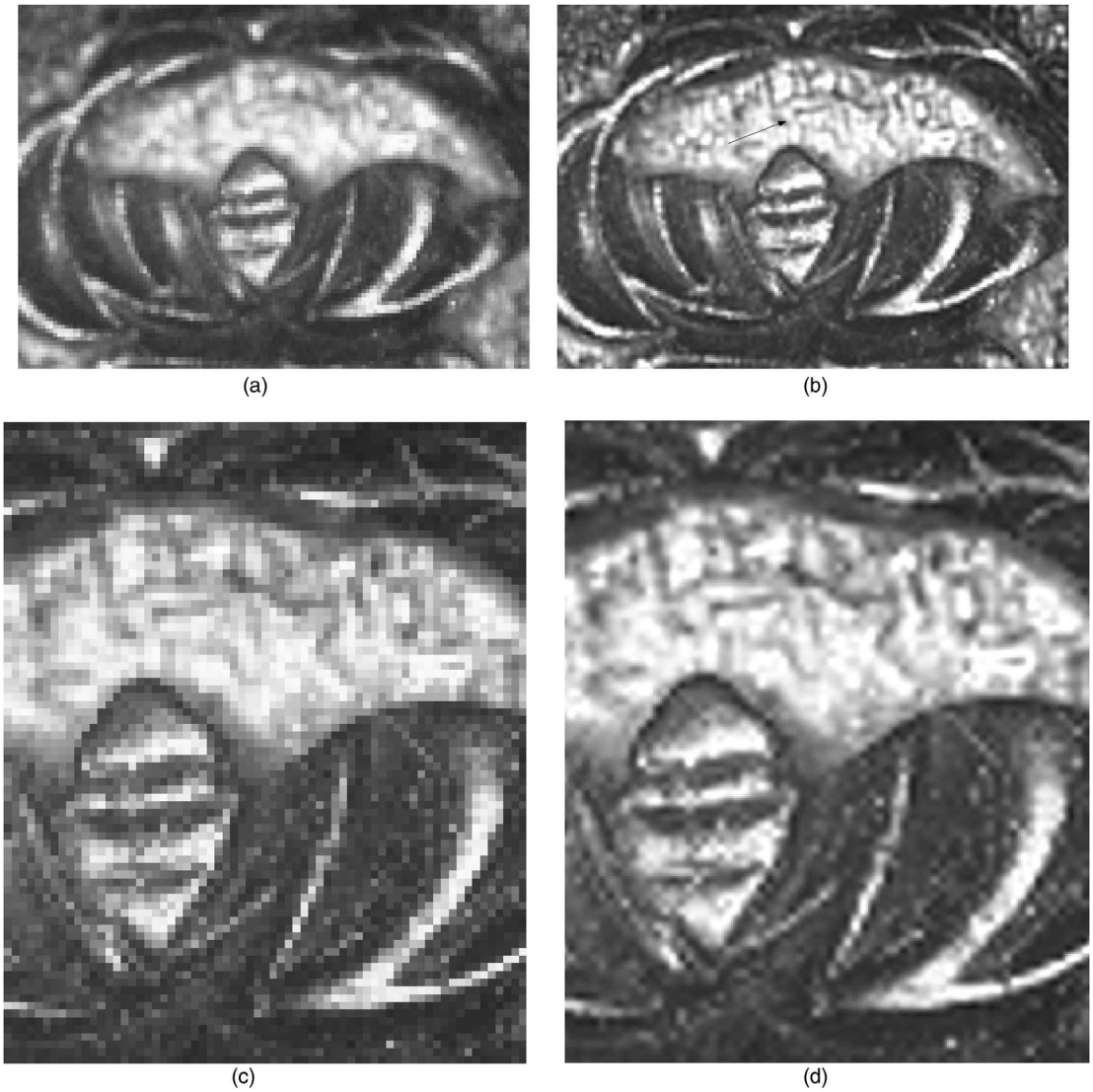

Fig. 2. (a) LR image of a portion of a metal object. (b) HR image obtained by using the proposed method for $q=2$. (c) Cropped region from the LR image. (d) Output of our method for $q=4$.

stage kept at the reference plane, which was 1.4 mm below the focused plane. At this position, the space-invariant blur σ_0 that would be induced by the ROI was computed. The interframe separation $\Delta d'$ was initially chosen as 0.1 mm. Blur parameters σ_1 and σ_2 for two more frames induced due to the upward movement of the stage by 0.1 and 0.2 mm, respectively, were then computed using Eq. (8). This procedure was repeated for 16 different values of $\Delta d'$ by incrementing it in steps of 0.1 mm to yield 16 different triplets of blur parameters $\{\sigma_0, \sigma_1, \sigma_2\}$. To test for optimality of $\Delta d'$,the posterior CR bound given by Eq. (19) was computed for each of these triplets, and this is plotted in Fig. 4(a). The plot gives the variation of CRLB with $\Delta d'$. The plot is seen to have two minima at $\Delta d'=0.7$ and 1.4 mm, respectively.

Next, we verified the theoretical CRLB plot of Fig. 4(a). For this purpose, we simulated the same situation as above, but now instead of using the CRLB expression [Eq. (19)], we experimented with images by computing the MSE between the original and the reconstructed HR image. The Lena image of size 152×152 was decimated by a factor of 2 and space-invariantly blurred with the same 16 triplets of blur parameters $\{\sigma_0, \sigma_1, \sigma_2\}$ computed earlier. Each triplet corresponds to a set of three LR observations captured with a certain $\Delta d'$. The proposed algorithm was then used to reconstruct the HR image corresponding to each of these triplets, and the MSE was computed for each case. The MSE averaged over 50 different trials is plotted in Fig. 4(b), which shows the variation of the MSE for different values of $\Delta d'$. It is interesting to observe from the plot that the proposed algorithm indeed reconstructs the best HR image for $\Delta d'=0.7$ and 1.4 mm, as predicted by the theoretical CRLB plot of Fig. 4(a).

The location of the minima can be explained as follows. As the observations are blurred and noisy, it is expected that in the case of ROI superresolution where the blurring is space-invariant, whenever one of the LR frames comes close to the focused plane, the degree of blur in that frame is the least and the proposed algorithm should yield a good estimate of the HR image. The corresponding uniform interframe separation between the LR frames is the optimal $\Delta d'$. We observe that whenever the proposed

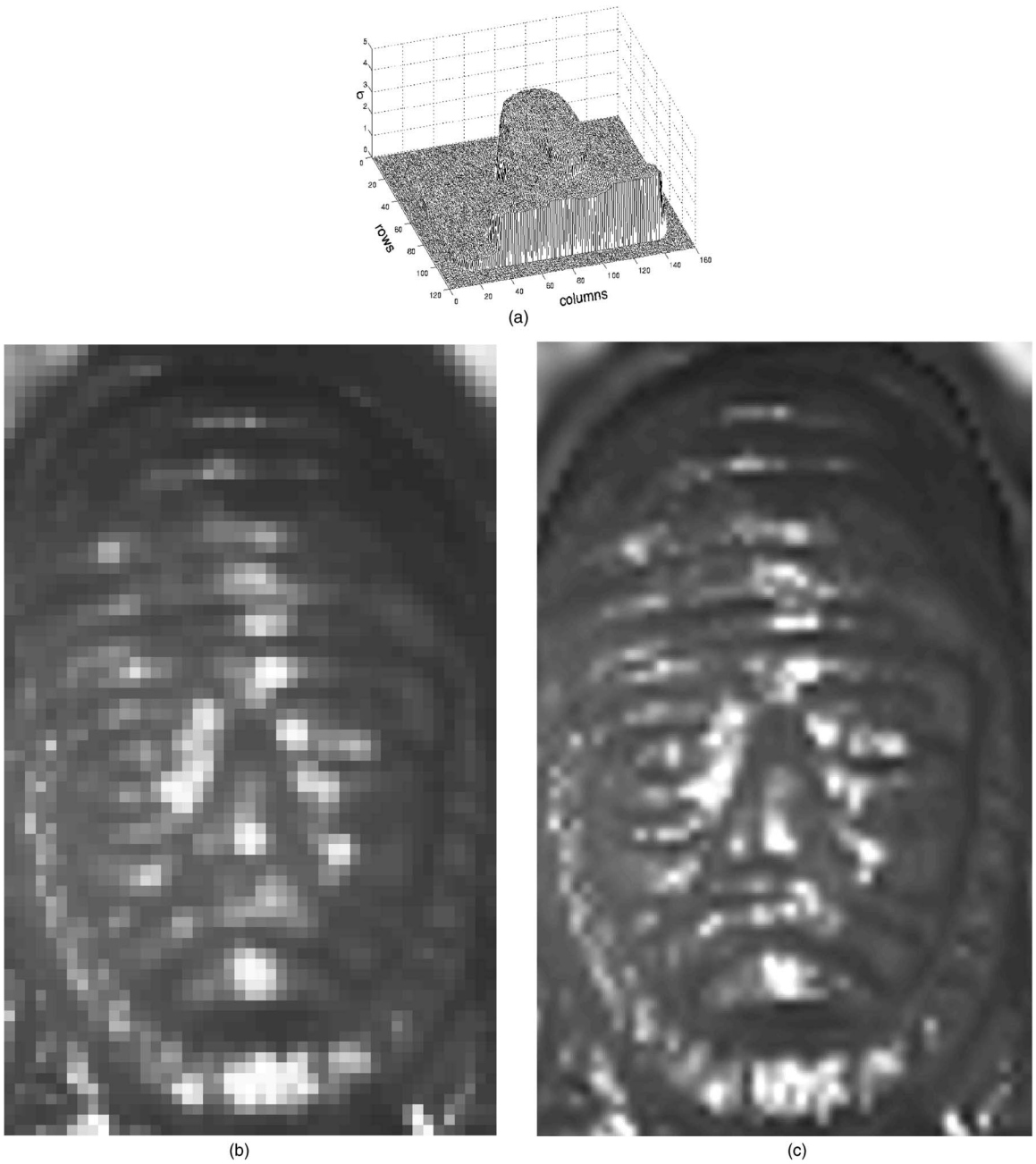

Fig. 3. (a) Blur map corresponding to the reference frame. (b) LR image of a portion of a ring. (c) HR image obtained using the proposed algorithm for $q=6$.

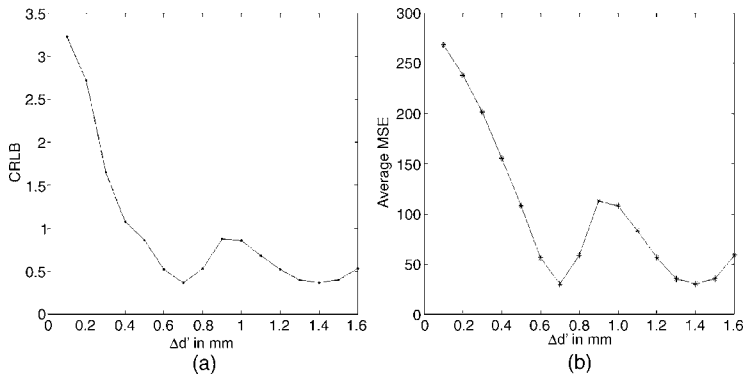

Fig. 4. (a) Plot showing the variation of the theoretical CRLB with $\Delta d'$. (b) Variation of MSE (averaged over 50 trials) with $\Delta d'$ for the Lena image.

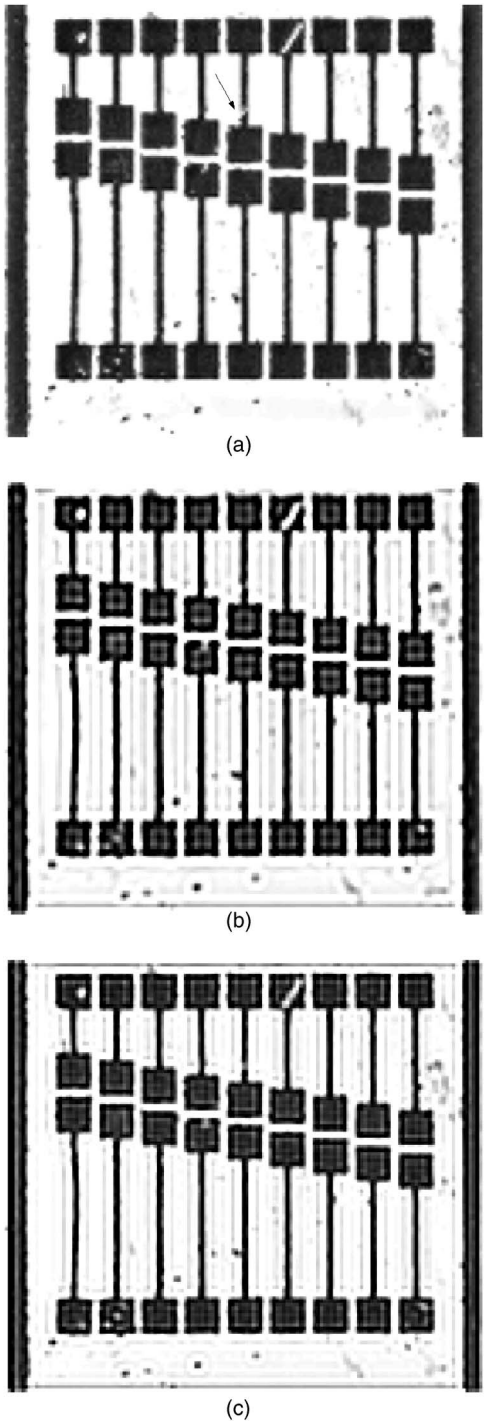

Fig. 5. Superresolved ROI of the MEMS sample obtained with (a) $\Delta d' = 0.6$ mm, (b) $\Delta d' = 0.4$ mm, and (c) $\Delta d' = 0.9$ mm.

algorithm yields the best HR image, one of the LR images is indeed almost focused because the value of σ_1 or σ_2 is very small for this situation.

We next demonstrate the utility of optimal $\Delta d'$ for real situations. We chose a fabricated sample, used for characterization of mechanical properties of the gold metal for microelectromechanical switches (MEMS). These switches find applications in 3G mobile phones, GPS, wireless LAN, etc. The challenge in successfully realizing the potential of MEMS lies in ensuring a high degree of reliability of operation as these are vulnerable to many failure mechanisms. Computer vision-based inspection can aid in the detection of defects.

We imaged a MEMS sample using the LV-150 microscope. The size of the ROI in the LR observations that we sought to superresolve (by a factor of 2) was 126×141 pixels. The ROI contained a mechanical defect, which manifested as a crack in one of the cantilever gold beams. To conform with the earlier synthetic experiment, 16 sets of three LR observations of the sample were captured for different values of $\Delta d'$. The space-invariant blur parameter σ_m was found for the ROI in each LR frame. For each value of $\Delta d'$, the three corresponding LR frames were used in the proposed algorithm to superresolve the ROI.

Typically, the position of the focused plane is not known *a priori*, but in SFF, since the depth map is computed using the observation stack, we know the frame in which the ROI would approximately come in focus. We observed that if the LR images are picked such that $\Delta d' = 0.6$ mm, then the quality of the reconstructed HR image [Fig. 5(a)] is quite good. Note that the discontinuity in one of the gold beams (shown by an arrow) is clearly visible in Fig. 5(a). Also, the edges of the gold beams and their supports are preserved well. The superresolved image obtained for $\Delta d' = 1.2$ mm was found to be equally good. A nonoptimal choice of $\Delta d' = 0.4$ mm yields the image shown in Fig. 5(b). Here, the crack is not visible and the quality of the HR image is poor. The HR image estimated using yet another nonoptimal $\Delta d' = 0.9$ mm and shown in Fig. 5(c) is again inferior compared to Fig. 5(a). We observe that the best HR images were obtained for $\Delta d' = 0.6$ and 1.2 mm, which are very close to the theoretically derived optimal values of $\Delta d'$.

7. CONCLUSIONS

We proposed an extension of the shape from focus method to superresolve the focused image of 3D objects. Using the observations in the SFF stack and the depth map of the object, we proposed a MAP-MRF technique to superresolve by magnification factors of 2 or greater. The method was validated with real data. The performance of ROI superresolution was analyzed in the frequency domain from a statistical perspective using the CRLB. It was shown that an optimal uniform interframe separation among the LR frames exists, which yields the best estimate of the HR image corresponding to the given ROI. Theoretical predictions were verified with experiments on both synthetic and real images.

REFERENCES

1. S. K. Nayar and Y. Nakagawa, "Shape from focus," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 824–831 (1994).
2. M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," IEEE Trans. Pattern Anal. Mach. Intell. **17**, 266–274 (1995).
3. J. Yun and T. S. Choi, "Accurate 3-D recovery using curved window focus measure," in *Proceedings of the International Conference on Image Processing* (IEEE, 1999), pp. 910–914.
4. M. Asif and T. S. Choi, "Shape from focus using multilayer feedforward neural networks," IEEE Trans. Image Process. **10**, 1670–1675 (2001).
5. M. B. Ahmad and T. S. Choi, "A heuristic approach for

finding best focused shape," IEEE Trans. Circuits Syst. Video Technol. **15**, 566–574 (2005).
6. M. Noguchi and S. K. Nayar, "Microscopic shape from focus using active illumination," in *Proceedings of the International Conference on Pattern Recognition* (IEEE, 1994), pp. 147–152.
7. T. S. Huang and R. Y. Tsai, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, T. Huang, ed. (JAI, 1984), Vol. 1, pp. 317–339.
8. H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," Comput. Vis. Graph. Image Process. **54**, 181–186 (1992).
9. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. Image Process. **5**, 996–1011 (1996).
10. M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy, and under sampled measured images," IEEE Trans. Image Process. **6**, 1646–1658 (1997).
11. M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," IEEE Trans. Image Process. **10**, 1187–1193 (2001).
12. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Trans. Image Process. **13**, 1327–1344 (2004).
13. D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1102–1117 (2003).
14. D. Rajan and S. Chaudhuri, "An MRF-based approach to generation of super-resolution images from blurred observations," J. Math. Imaging Vision **16**, 5–15 (2002).
15. A. N. Rajagopalan and V. P. Kiran, "Motion-free superresolution and the role of relative blur," J. Opt. Soc. Am. A **20**, 2022–2032 (2003).
16. N. K. Bose, M. K. Ng, and A. C. Yau, "A fast algorithm for image super-resolution from blurred observations," EURASIP J. Appl. Signal Process. **2006**, 35726 (2006).
17. A. Papoulis, "Generalized sampling expansion," IEEE Trans. Circuits Syst. **24**, 652–654 (1977).
18. W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," Int. J. Comput. Vis. **40**, 25–47 (2000).
19. A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," in *Proceedings of ACM SIGGRAPH 2001* (ACM, 2001), pp. 327–340.
20. S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Trans. Pattern Anal. Mach. Intell. **24**, 1167–1183 (2002).
21. A. Pentland, "A new sense for depth of field," IEEE Trans. Pattern Anal. Mach. Intell. **9**, 523–531 (1987).
22. M. Born and E. Wolf, *Principles of Optics* (Pergamon, 1965).
23. M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Proceedings of the International Conference on Computer Vision* (IEEE, 1988), pp. 149–155.
24. S. Chaudhuri and A. N. Rajagopalan, *Depth from Defocus: A Real Aperture Imaging Approach* (Springer-Verlag, 1999).
25. S. Z. Li, *Markov Random Field Modeling in Computer Vision* (Springer-Verlag, 1995).
26. J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. R. Stat. Soc. Ser. B (Methodol.) **36**, 192–236 (1974).
27. A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, 1989).