HIGH RESOLUTION IMAGE RECONSTRUCTION IN SHAPE FROM FOCUS

R. R. Sahay and A. N. Rajagopalan

Image Processing and Computer Vision Lab
Department of Electrical Engineering, IIT Madras, Chennai
sahayiitg@yahoo.com, raju@ee.iitm.ac.in

ABSTRACT

In the Shape from Focus (SFF) method, a sequence of images of a 3D object is captured for computing its depth profile. However, it is useful in several applications to also derive a high resolution focused image of the 3D object. Given the space-variantly blurred frames and the depth map, we propose a method to optimally estimate a high resolution image of the object within the SFF framework.

Index Terms— Shape from focus, super-resolution, image enhancement, image sequence analysis, Cramer-Rao lower bound.

1. INTRODUCTION

Shape from focus is a method for estimating the structure of a 3D object. A sequence of images of the object is captured through a real aperture camera. Due to the finite depth of field, none of these observations will be in focus. The images suffer from aliasing, space-variant blurring and noise. Increasing the resolution of images in SFF is potentially useful in applications where one is interested in finer details on the surface of a 3D specimen. In the literature on SFF, thus far there has been no attempt to reconstruct a high resolution (HR) focused image of the 3D object. The resolution of the frames captured in SFF is limited by the resolution of the real aperture camera used. Super-resolution algorithms estimate an HR image from multiple low resolution (LR) frames by dealiasing and deblurring. Broadly, these algorithms can be classified into three categories; namely, motion-based, motion-free and learning-based. Motion-based algorithms [1, 2] exploit the relative motion between the camera and the scene which gives rise to sub-pixel displacements among the LR images, to obtain an HR image. Motion-free super-resolution [3] can be performed when there is no relative motion between the LR frames, by using the defocus cue. The classic paper of Papoulis [4], provides the theoretical foundation for motion-free super-resolution. Learning based algorithms [5, 6, 7] attempt to learn the statistical relationships between corresponding image regions in LR and HR images during the training phase and use these relationships to predict finer details for enlarging other LR images.

In SFF, space-variantly blurred observations are naturally captured as the 3D object is translated vertically. Since we know that blur can serve as a cue for super-resolution [3], our goal in this work is to exploit the defocus cue to extend the scope of traditional SFF to reconstruct a HR image of the underlying 3D object, given its LR depth map and the LR observations.

2. PROBLEM FORMULATION

In traditional SFF [8], a 3D object is placed on a translational stage which moves in a vertical direction in steps of Δd. A sequence of images is captured which are space-variantly blurred due to the 3D nature of the specimen and the finite aperture of the camera. The focus measure $F(x, y)$ at a point (x, y) in image I is computed using the sum-modified Laplacian (SML) operator [8]. The focus measure profile for the pixel at (x, y) is obtained by plotting the value of $F(x, y)$ computed for every image captured in the stack of observations. The final estimate of the depth at (x, y) is arrived at by using Gaussian interpolation of a few values near the peak value of the focus measure profile.

In this paper, we address the following problem: Given the depth map of the 3D object and the observations captured in a SFF scenario, can we super-resolve the focused image of the 3D object? In order to perform super-resolution, we need a model that relates the LR observations to the HR image. We assume p number of LR observations $\{y_m(i, j)\}$, each of size M x M which are decimated, blurred noisy versions of a single HR image $\{x(m, n)\}$ of size qM x qM. If $\mathbf{y}_m$ is the lexicographically arranged vector containing pixels from the m^{th} LR image of size M^2 x 1 and $\mathbf{x}$ is the lexicographically arranged vector containing pixels from the HR image of size q^2M^2 x 1 then they can be related in the following way [9]

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

where $\mathbf{H}_m$ is the blur matrix of size M^2 x M^2, $\mathbf{D}$ is the decimation matrix of size M^2 x q^2M^2 and $\mathbf{n}_m$ is zero mean noise vector of size M^2 x 1. The observation noise is assumed to be zero mean Gaussian with variance σ_η^2. In motion-free super-resolution, the degradation model that is typically used

is $\mathbf{y}_m = \mathbf{D}\mathbf{H}_m\mathbf{x} + \mathbf{n}_m$. However, in SFF the choice of degradation model given by eq. (1) is motivated by the fact that we have depth estimates at the same resolution as the LR observations.

To compute the blur matrix $\mathbf{H}_m$, we need to model the point spread function (PSF) of the real aperture camera. In the literature, the PSF is usually modeled by a 2D Gaussian function [10] $h(i,j) = \frac{1}{2\pi\sigma^2}\exp\frac{-(i^2+j^2)}{2\sigma^2}$ where σ is the blur parameter. Note that a 3D object induces space-variant blurring. The spatial distribution of the blur in relation to the depth at a point on the object can be given as

$$\sigma = \rho R v \left(\frac{1}{w_d} - \frac{1}{D}\right) \tag{2}$$

where w_d is the working distance of the camera, D is the distance of the object point from the lens, v is the distance from the lens to the image plane, R is the radius of the aperture of the lens and ρ is a camera constant.

Interestingly, it is possible to show that the relationship between the blur parameter at a particular point in one image in the stack and another image at the same point can be expressed as

$$\sigma_k = \sigma_0 + \rho R v \left(\frac{1}{D_0} - \frac{1}{D_0 \pm k\Delta d}\right) \tag{3}$$

where σ_k is the blur parameter at a particular point in the k^{th} frame, σ_0 is the blur parameter at the same point in the reference frame, Δd is the step size by which the stage is moved and D_0 is the distance of the object point from the lens when the stage was at the reference position. An appropriate calibration procedure must be adopted to find the value of the product $\rho R v$. Unlike in depth from defocus [3], only the stage is translated here and there is no need to change the lens settings. Hence, in SFF the value of $\rho R v$ remains constant during capture.

The blur matrix $\mathbf{H}_m$ in (1) can now be constructed from the depth map obtained using SFF and the relation between the blur parameter σ at each pixel across the stack of LR frames in (3). We assume here that the change in magnification is negligible across the stack.

3. SUPER-RESOLUTION

The problem of reconstructing the high resolution image $\mathbf{x}$ is an ill-posed inverse problem and some form of regularization is necessary. We propose to derive an optimal estimate of the HR image as the maximum a posteriori (MAP) estimate given by

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_p) \tag{4}$$

The MAP framework allows us to impose a priori constraints on the HR image. Since statistical models can encode contextual constraints in images in a natural way, we model the original HR image as a Markov random field (MRF) [11]. Specifically, we model it as a Gauss-Markov random field (GMRF).

Using the degradation model in (1) and assuming that the the noise processes are independent, we have

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[\sum_{m=1}^{p} \frac{\|\mathbf{y}_m - \mathbf{H}_m\mathbf{D}\mathbf{x}\|^2}{2\sigma_\eta^2} + \lambda \sum_{c\in C} (\mathbf{d}_c^T\mathbf{x})^2\right] \tag{5}$$

where $\mathbf{d}_c^T\mathbf{x}$ provides a measure of smoothness of the image by computing discrete approximations for first or second derivatives at each image pixel. We assume a first order MRF neighbourhood. Minimization by gradient descent is employed to find an estimate of $\mathbf{x}$. At the n^{th} iteration, the gradient of the cost function is given by

$$\text{grad}^{(n)} = \frac{1}{\sigma_\eta^2}\sum_{m=1}^{p} \mathbf{D}^T\mathbf{H}_m^T(\mathbf{H}_m\mathbf{D}\mathbf{x}^{(n)} - \mathbf{y}_m) + \lambda Q^{(n)} \tag{6}$$

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

Here, $\mathbf{D}^T$ spreads equally the LR pixel intensity value at corresponding pixel locations in the HR image. Matrix $\mathbf{H}_m$ is computed using the LR depth map obtained by the traditional SFF method [8] and the relation among the σ values across the stack of LR images given in (3). The regularization factor λ is typically tuned to derive the best estimate of $\mathbf{x}$. One can use a more complicated model such as the discontinuity adaptive MRF [11] but a reasonably low value of λ in the GMRF model will preserve edges well. Importantly, the GMRF is amenable to mathematical analysis. The estimate of the HR image at the $(n+1)^{th}$ iteration is obtained as $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \beta\,\text{grad}^{(n)}$ where β is the step size. The iterations continue until $\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\| <$ threshold.

The performance of the proposed method for superresolution can be bounded by analyzing the problem in a statistical framework using the Cramér-Rao bound. Without loss of generality, we derive relations for the 1-D case for mathematical convenience. Considering the degradation model given in (1), the log-likelihood function is given by

$$\log P(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_p|\mathbf{x}) = \sum_{m=1}^{p} \log\frac{1}{(2\pi\sigma_\eta^2)^{\frac{M}{2}}} - \sum_{m=1}^{p}\frac{\|\mathbf{y}_m - \mathbf{H}_m\mathbf{D}\mathbf{x}\|^2}{2\sigma_\eta^2} \tag{7}$$

The prior distribution of the HR one-dimensional signal, assuming a GMRF model is

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}|\frac{1}{\lambda}R_{\mathbf{x}}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{x}^T\left(\frac{R_{\mathbf{x}}}{\lambda}\right)^{-1}\mathbf{x}\right\} \tag{8}$$

where the length of the signal is N. The matrix $R_{\mathbf{x}}^{-1} = \mathbf{D}_r^T\mathbf{D}_r$ where $\mathbf{D}_r$ represents a one-step forward difference operator.

Matrix $\mathbf{D}_r^T\mathbf{D}_r$ can be approximated as a circulant matrix which is a Laplacian operator. Therefore, the posterior CR bound can be written as

$$J^{-1}(\mathbf{x}) = \left[\frac{1}{\sigma_\eta^2}\sum_{m=1}^{p}\mathbf{D}^T\mathbf{H}_m^T\mathbf{H}_m\mathbf{D} + \lambda\mathbf{D}_r^T\mathbf{D}_r\right]^{-1} \quad (9)$$

It is not possible to derive a closed form expression for the CRLB due to the space-variant nature of blurring in the LR observations. The term $\lambda\mathbf{D}_r^T\mathbf{D}_r$ improves invertibility of $J(\mathbf{x})$, an effect of regularization. The above expression can be generalized to the 2-D case in a straightforward manner. The bound gives a fundamental limit on the quality of estimates of the unknown HR image from LR space-variantly blurred, noisy images in SFF. The performance measure for evaluating the quality of the reconstructed HR signal is the average mean squared error $(\overline{MSE}) = \frac{1}{N}E\left[(\mathbf{x}-\hat{\mathbf{x}})^2\right]$, where $\hat{\mathbf{x}}$ is the estimated HR signal. The posterior CR bound for the error in the estimation of $\mathbf{x}$ and $\overline{MSE}$ are related as $\overline{MSE} \geq \frac{1}{N}\text{trace}(J^{-1}(\mathbf{x}))$.

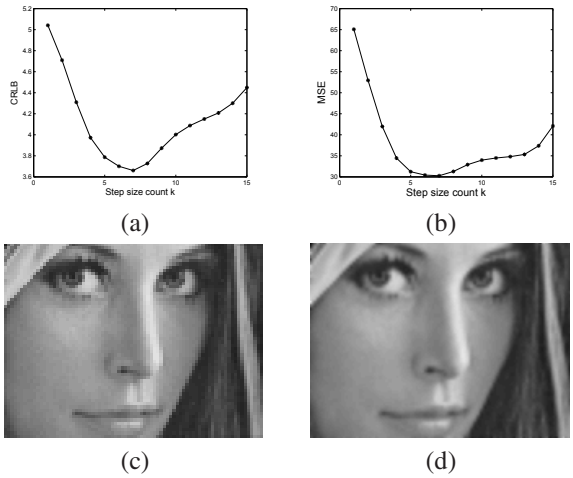

Fig. 1. (a) Plot of the variation of the CRLB with Δd.(b) Variation of the MSE for the Lena image. (c) One of the LR images. (d) HR image reconstructed using the proposed method for the optimal separation of 0.7mm (MSE = 30.26).

4. EXPERIMENTAL RESULTS

In this section, we first present experimental results for the synthetic case, assuming a hypothetical ramp object for which we simulated the SFF technique for a lens of objective 2.5x. The focused plane was assumed to be at a distance of 8.8mm from the lens. The translational stage was assumed to be at the reference plane, initially at a distance of 1.4mm from the focused plane. The height of the hypothetical ramp object was 0.5mm and the space-variant blur undergone by every point on it was computed. The size of this blur map was 60x80. Similarly, blur maps were computed for the upward movement of the stage by Δd = 0.1mm and 0.2mm, respectively, from the reference plane. The above procedure was repeated for 15 different values of Δd incremented in steps of 0.1mm. Thus, we obtained 15 sets of blur maps each of size 60x80. The corresponding CR bound given in (9) is plotted in Fig. 1 (a). This plot has a minimum at step size count k = 7, which corresponds to Δd = 0.7mm.

For simulating the above situation with actual images we took the Lena image, cropped it to size 120x160, decimated it by a factor of 2 and blurred it in a space-variant manner with each of the 15 different sets of blur maps obtained earlier. Thus, 15 sets of LR observations were generated. An HR image was estimated from each of these sets of LR observations using the proposed algorithm and the MSE was computed in each case. The MSE is plotted in Fig. 1 (b) for different values of Δd. It is interesting to note that the proposed algorithm indeed yields the best HR image for Δd = 0.7mm between the LR frames as predicted by the posterior CR bound plot in Fig. 1 (a). One of the LR frames used is shown in (c) and the super-resolved image corresponding to Δd = 0.7mm is shown in (d). If the blurring had been space-invariant, the proposed algorithm would yield the best HR image whenever one of the LR observations came close to the focused plane. Since, here we have space-variant blur it is difficult to give a physical interpretation for the CRLB minimum which corresponds to a particular distance of separation, Δd, between the LR frames.

Next, we present real results using the proposed method for estimating an HR image from a stack of LR, space-variantly blurred and degraded observations of a 3D object captured within the SFF setting. An LV-150 Nikon industrial microscope was used for imaging. The lens objective was 2.5x, for which the working distance w_d = 8.8mm, focal length f = 80mm and depth of field = 48.9μm. For this experiment, we chose λ = 0.001, σ_η^2 = 5 and the step size β = 1 for gradient descent. The upsampling factor q was chosen to be 2. The LR images captured were of size 100x135 pixels. A ring on which the face of a man is engraved was taken as the specimen 3D object and LR observations were captured by translating the stage in steps of Δd = 0.025mm. We used the traditional SFF method [8] to compute the LR depth map of the object. Seven LR frames were chosen from the stack and the corresponding blur maps were computed. The HR image was estimated by the proposed algorithm using the blur maps and the LR frames. In Fig. 2 (a) one of the LR images is shown. The blur map of the portion of the ring corresponding to this LR frame is shown in (b). The initial estimate for the proposed algorithm was the bilinearly interpolated LR image shown in (c). The super-resolved image obtained using our method is shown in (d). The facial features have been deblurred and come out clearly in (d) as compared to (c). The details of the dress on the shoulders and the beard also come out quite well.

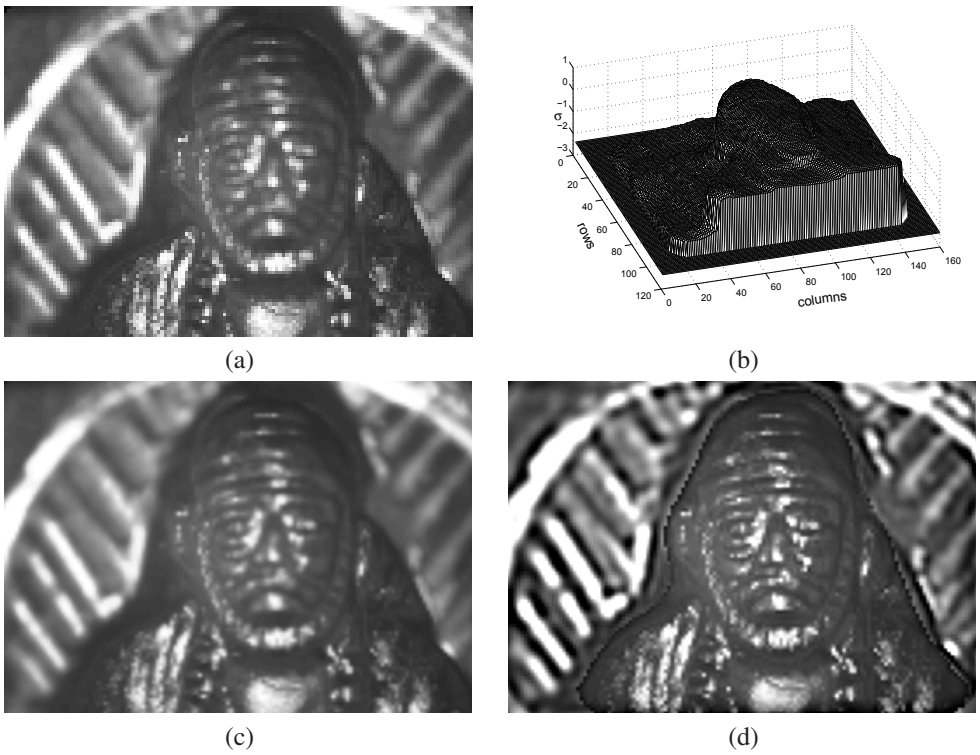

Fig. 2. (a) LR image of a portion of a ring. (b) The corresponding blur map. (c) HR image obtained by bilinear interpolation. (d) HR image obtained using the proposed method.

5. CONCLUSIONS

We proposed a method for obtaining a high resolution image of a 3D object in SFF, given its space-variantly blurred, noisy observations and the low resolution depth map. Using synthetic and real images, it was shown that the quality of the reconstructed high resolution image is quite good and can be potentially beneficial in many applications.

6. REFERENCES

[1] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, 1997.

[2] R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, 1996.

[3] 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.*, vol. 25, no. 9, pp. 1102–1117, 2003.

[4] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, , no. 11, pp. 652–654, 1977.

[5] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, 2002.

[6] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Intl. Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[7] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," *Proceedings of ACM SIGGRAPH*, pp. 341–346, 2002.

[8] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, 1994.

[9] N. K. Bose, M. K. Ng, and A. C. Yau, "A fast algorithm for image super-resolution from blurred observations," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 35726, 14 pages, 2006. doi:10.1155/ASP/2006/35726.

[10] S. Chaudhuri and A. N. Rajagopalan, "Depth from defocus: A real aperture imaging approach," *Springer-Verlag, New York*, 1999.

[11] S. Z. Li, "Markov random field modeling in computer vision," *Springer-Verlag, Tokyo*, 1995.