Inpainting in Shape from Focus: Taking a Cue from Motion Parallax

Rajiv R. Sahay
sahayiitm@gmail.com

A.N. Rajagopalan
raju@ee.iitm.ac.in

Image Processing and Computer Vision Lab
Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai 600 036, India

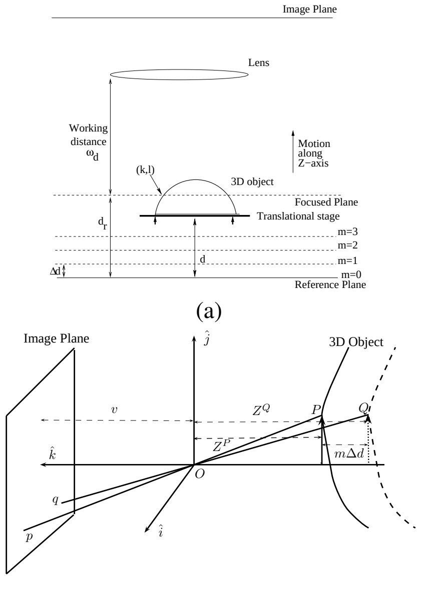
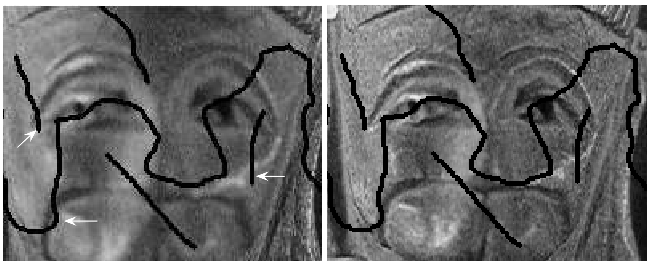


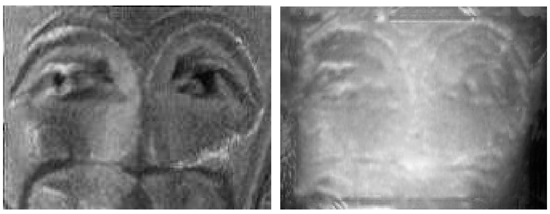
Figure 1: (a) Working principle of SFF. (b) Schematic showing mechanism of structure-dependent pixel motion in SFF.



(a) (b)

Figure 2: Wooden specimen depicting a face. (a, b) Two frames chosen from the scaled and defocused stack with missing data, simulating the effects of a camera having a damaged sensor.

Shape from focus (SFF) (Fig. 1 (a)) which uses a sequence of space-variantly defocused frames works under the constraint that there is 'no magnification' in the stack. In the presence of sensor damage and/or occlusions, there will be missing data in the observations and SFF cannot recover structure in those regions. In this paper, we investigate the effect of motion parallax in SFF and demonstrate the interesting possibility of how it can be judiciously used to jointly inpaint image and depth profiles. A Bayesian approach is adopted in which the inpainted depth profile and the focused image are modeled as separate Gauss-Markov random fields (MRFs). Their maximum a posteriori (MAP) estimates are obtained by minimizing a suitably derived objective function.



(a) (b)

Figure 3: (a) Inpainted focused image. (b) Grayscale image corresponding to the inpainted depth map.

To explain the mechanism of this structure-dependent pixel motion, let us assume that the focused image and the shape of the 3D specimen are known. As shown in Fig. 1 (b), we consider a specific point on the specimen which is moved relative to the pinhole camera. A point on the 3D object with world coordinates $P(X^P,Y^P,Z^P)$ moves to $Q(X^Q,Y^Q,Z^Q)$ along the Z-axis by a distance of $m\Delta d$ and away from the pinhole denoted by O. The distances of the points P and Q from the pinhole are Z^P and Z^Q, respectively. The point P is imaged at p on the image plane and has coordinates (x,y). Let this image be the reference plane. When the 3D object is moved away from the pinhole by an amount $m\Delta d$, the point Q is imaged at q with coordinates (x',y') on the image plane. The corresponding image is the m^{th} frame. Assuming that the size of the images is $M \times M$, according to basic perspective projection equations $x = \frac{vX^P}{Z^P}$, $x' = \frac{vX^Q}{Z^Q}$ and $y = \frac{vY^P}{Z^P}$, $y' = \frac{vY^Q}{Z^Q}$

It must be noted that the motion of the object relative to the pinhole is only along the Z-axis since the 3D specimen is translated away from or towards the camera along the optical axis. Hence, for the SFF scenario, we have $X^P = X^Q$, $Y^P = Y^Q$ and $Z^Q = Z^P + m\Delta d = w_d + \bar{d} + m\Delta d$, where v is the distance between the pinhole and the image plane, and $\frac{-M}{2} \leq x', y' \leq \frac{M}{2}$. Thus, it can be shown that $x' = \frac{x(w_d+\bar{d})}{(w_d+\bar{d})+m\Delta d}$, $y' = \frac{y(w_d+\bar{d})}{(w_d+\bar{d})+m\Delta d}$ Note that the pixel motion is a function of $\bar{d}$, the 3D structure of the scene. We will exploit this cue for inpainting.

Consider N frames, $\{y_m(i,j)\}$, $m = 0,1,\ldots,N-1$, each of size $M \times M$ from the stack. Assume that these are derived from a single focused image $\{x(i,j)\}$ of the 3D specimen. The scaled and defocused frames can be related to the focused image by the degradation model

$$\mathbf{y}_m^{\text{vis}} = \mathbf{O}_m[\mathbf{H}_m(\bar{\mathbf{d}})\mathbf{W}_m(\bar{\mathbf{d}})\mathbf{x} + \mathbf{n}_m] \qquad m = 0,\ldots,N-1 \tag{1}$$

where $\mathbf{y}_m^{\text{vis}}$ is the lexicographically arranged vector of size $M^2 \times 1$ derived from the visible regions in the m^{th} defocused and scaled frame, $\mathbf{W}_m$ is the matrix describing the motion of the pixels in the m^{th} frame, $\mathbf{H}_m$ is the blurring matrix for the m^{th} frame, $\mathbf{n}_m$ is the $M^2 \times 1$ Gaussian noise vector in the m^{th} observation and $\mathbf{O}_m$ is the operator which removes the missing/damaged regions and crops out the visible portions of the observations.

We seek to solve for the MAP estimate of $\bar{\mathbf{d}}$ and $\mathbf{x}$. Let us consider a set of p frames chosen from the stack of N observations. Assuming the noise process $\mathbf{n}_m'$s to be independent, the MAP estimates of $\bar{\mathbf{d}}$ and $\mathbf{x}$ can be obtained by minimizing the posterior energy function

$$U^P(\bar{\mathbf{d}},\mathbf{x}) = \sum_{m\in O} \frac{\|\mathbf{y}_m^{\text{vis}} - \mathbf{O}_m[\mathbf{H}_m(\bar{\mathbf{d}})\mathbf{W}_m(\bar{\mathbf{d}})\mathbf{x}]\|^2}{2\sigma_\eta^2} + \lambda_{\bar{\mathbf{d}}} \sum_{c\in C_{\bar{\mathbf{d}}}} V_c^{\bar{\mathbf{d}}}(\bar{\mathbf{d}}) + \lambda_{\mathbf{x}} \sum_{c\in C_{\mathbf{x}}} V_c^{\mathbf{x}}(\mathbf{x})$$

where $O = \{u_1, u_2, \ldots, u_p\}$, u_i is the frame number and σ_η^2 is the variance of the Gaussian noise.

We image a small wooden piece on which a face is carved. The captured images are space-variantly blurred and scaled. Note the locations indicated by white arrows in Fig. 2 (a). At these locations, in Fig. 2 (b) we can make the interesting observation that the position of the damaged regions is fixed but different image features are covered by the black pixels in this image. This is the cue which we exploit to fill-in the missing parts in both the focused image and the depth profile. The inpainted image using the proposed algorithm is shown in Fig. 3 (a). The values of the parameters chosen in the optimization procedure were $\lambda_{\bar{\mathbf{d}}} = 1 \times 10^8$ and $\lambda_{\mathbf{x}} = 0.05$, respectively. The reconstructed depth map is shown in Fig. 3 (b). We can see that the proposed method is able to reconstruct the depth information even in regions with missing observations. It is notable that various features features of the face like the eyes, the eyebrows, the nose and the straight edge below it are faithfully reconstructed both in the inpainted image (Fig. 3 (a)) and the inpainted depth map (Fig. 3 (b)), respectively.