Photometric Stereo Under Blurred Observations

M.V. Joshi and Subhasis Chaudhuri
Department of Electrical Engineering
Indian Institute of Technology - Bombay
Powai, Mumbai-400076. India
{mvjoshi, sc}@ee.iitb.ac.in

Abstract

In this paper we address the problem of simultaneous estimation of structure and restoration of images from blurred photometric measurements. Given the blurred observations of a static scene captured with a stationary camera, under different illuminant directions, we obtain the structure represented by the surface gradients and the albedo and also perform blind image restoration. The surface gradients and the albedo are modeled as separate Markov random fields (MRF) and a suitable regularization scheme is used to estimate the different fields as well as the blur parameter.

1. Introduction

Researchers traditionally treat the shape from shading problem without considering the blur introduced by the camera. However, when one captures the images with a camera, the degradation in the form of blur and noise is often present in these observed images. It is natural that the variations in image intensity due to camera blur affects the estimates of the surface shape. Thus, the estimated shape differs from the true shape in spite of possibly having the knowledge of the true surface reflectance model. This limits the applicability of these techniques in 3D computer vision problems. It is to be mentioned here that all the existing approaches in the literature assume a pinhole model that inherently assumes that there is no camera blur during observation. However the blur could happen due to a variety of reasons such as improper focus setting or camera jitter. This motivates us to restore the image as well, while recovering the structure. The problem can then be stated as follows: given a set of blurred observations of a static scene taken with different light source positions, obtain the true depth map and the albedo of the surface as well as restore the images for different light source directions. Since the camera blur is not known, in addition, we estimate the blur point spread function (PSF) which caused the degradation. In this paper we assume a point light source illumination with known source directions and an orthographic projection. Due to above, the problem can be classified as a joint blind restoration and surface recovery problem. Since such a problem is inherently ill-posed, we need suitable regularization of all the fields to be estimated, i.e., surface gradients as well as the albedo.

Researchers in computer vision have attempted to use the shading information to recover the 3D shape. Horn was one of the first researchers to study this problem by casting it as a solution to second order partial differential equations [5]. Shape from shading (SFS) problem is typically solved using four different approaches. These approaches include the regularization approach, the propagation approach, the local approach and the linear approach. Most of the traditional SFS algorithms assume that the surface has constant albedo values, but the photometric stereo (PS) does not. The idea of PS was initially formulated by Woodham [13] and later applied by others [7, 11]. Some of the recent approaches to PS include a neural network based method for a rotational object with a non uniform reflectance factor [8], and integrating the SFS with the PS in order to improve the performance of shape recovery [10]. The general approaches for image restoration include both stochastic and deterministic methods. For a comprehensive survey of various digital image restoration techniques the reader is referred to [1]. A plethora of methods have also been proposed to solve the problem of blind image deconvolution [9]. Recently, Candela *et al.* used local spectral inversion of a linearized total variation model for denoising and deblurring [3]. As discussed above, the researchers have treated the shape estimation and restoration problems separately. Also, for shape estimation using the shading cue, the blur introduced by the camera is never considered. We demonstrate in this paper that both the shape estimation and restoration problems can be handled jointly in a unified framework.

2. Problem Definition

Consider a scene illuminated with different light source positions where both the object and the camera positions are stationary. We capture the images with a large distance between the object and the camera, thus making a reasonable assumption of orthographic projection and neglect the depth related perspective distortions. The light source is assumed to be a distant point light. Now given an ensemble of images captured with different light source positions, using the theory of photometric stereo we can express the intensity of the image at a point using the image irradiance equation as

$$f(x,y) = \rho(x,y)\mathcal{R}(p(x,y),q(x,y)) = \rho(x,y)\hat{n}(x,y).\hat{s}, \quad (1)$$

where $\hat{n}$ is the unit surface normal at a point on the object surface, $\hat{s}$ is the unit vector defining the light source direction and ρ is the albedo or the surface reflectance of the surface. The surface gradients $(p,\ q)$ are used to specify the unit surface normal. $\mathcal{R}$ is called the reflectance function. Here, we concentrate on the Lambertian model in our study, but the method can be expanded to other reflectance models also. In practice, one uses more than three observations to estimate the $p,\ q$ and the albedo due to inconsistency in measurements. The solution to equation (1) using the different measurements gives the true surface gradients and the albedo in the least squares sense only when we do not consider the camera blur. However, due to improper focus setting or camera blur the observations are often blurred. Thus, considering the effect of blur the observed image can be expressed as

$$g(x,y) = h(x,y) * f(x,y) + \eta(x,y), \quad (2)$$

where $h(x,y)$ represents the two-dimensional point spread function (PSF) of the imaging system, and $\eta(x,y)$ is an additive noise introduced by the system. Considering k light source positions, and using the vector/matrix notations, equation (2) can be expressed as

$$\mathbf{g}_m = \mathbf{H}(\sigma)\mathbf{f}_m(\rho,p,q) + \boldsymbol{\eta}_m, \quad m = 1,\cdots,k \quad (3)$$

where $\mathbf{H}(\sigma)$ represents the PSF with σ representing the blur parameter, and $\mathbf{f}_m$ is the true focused image for the m^{th} light source position, which is a function of the surface gradients and the albedo as seen from equation (1). In this paper we assume that the blur is due to the camera out-of-focus which can then be modeled by a pillbox blur or by a Gaussian PSF characterized by a single parameter σ. $\boldsymbol{\eta}_m$ is the noise vector which is a zero mean i.i.d process. Our problem now is to estimate the blur parameter σ, the albedo $\boldsymbol{\rho}$, the surface gradients $\mathbf{p}$ and $\mathbf{q}$, and also to perform blind image restoration given the observations $\mathbf{g}_m, m = 1,\cdots,k$.

3. Simultaneous Estimation of Structure and Blind Restoration

As we are using a regularization based approach for simultaneous estimation of different parameter fields ($\mathbf{p},\ \mathbf{q},$ and $\boldsymbol{\rho}$), we need to use suitable priors for the fields to be estimated. The MRF provides a convenient and consistent way of modeling context dependent entities such as image pixels, depth of the object and other spatially correlated features. Let Z be a random field over an arbitrary $N \times N$ lattice of sites $L = \{(i,j)|0 \leq i,j \leq N-1\}$. From the Hammersley-Clifford theorem [2] which proves the equivalence of an MRF and a Gibbs random field (GRF), we have $P(Z=z) = \frac{1}{Z_p}e^{-U(z)}$ where z is a realization of Z, Z_p is the partition function given by $\sum_z e^{-U(z)}$ and $U(z)$ is the energy function given by $U(z) = \sum_{c\in\mathcal{C}} V_c(z)$. $V_c(z)$ denotes the potential function of clique c and $\mathcal{C}$ is the set of all cliques.

In order to impose the spatial correlation, we consider pair wise cliques on a first order neighborhood and impose a quadratic cost which is a function of finite difference approximations of the first order derivative at each pixel location, i.e.,

$$\begin{aligned}\sum_{c\in\mathcal{C}} V_c(\mathbf{z}) &= \mu \sum_{k,l=1}^{N-2} [(z_{k,l} - z_{k,l-1})^2 + (z_{k,l} - z_{k-1,l})^2 \\ &+ (z_{k,l+1} - z_{k,l})^2 + (z_{k+1,l} - z_{k,l})^2] \\ &= U(\mathbf{z}), \quad (4)\end{aligned}$$

where μ represents the penalty for departure from the smoothness in $\mathbf{z}$. The prior model as defined above was used for each of the different fields $\mathbf{p},\ \mathbf{q},$ and $\boldsymbol{\rho}$. Theses priors are used in conjunction with the image formation model given in equation (3).

We obtain k observations of a static scene by varying the direction of the point light source. It is assumed that the directions are known. We also assume that the reflectance model is known. We introduce the context dependencies in the estimated fields by modeling them as separate MRFs. Thus the corresponding priors are $U(\mathbf{p})$, $U(\mathbf{q})$, and $U(\boldsymbol{\rho})$. Considering the brightness constraint term and the smoothness term for regularizing the solution, the final cost function can then be expressed as

$$\epsilon = \sum_{m=1}^{k} \|\mathbf{g}_m - \mathbf{H}(\sigma)\mathbf{f_m}(\rho,p,q)\|^2 + U(\mathbf{p}) + \mathbf{U}(\mathbf{q}) + \mathbf{U}(\boldsymbol{\rho}). \quad (5)$$

This cost function is convex and can be minimized using a gradient descent approach. The blur parameter and the structure (along with the images for different light source directions) are then estimated in an alternative way by keeping the blur parameter constant and updating the structure and vice-versa. It should be noted here that we are not performing image deconvolution which is highly ill-posed and often leads to numerical instability.

In order to do the simultaneous blur estimate along with the structure and image restoration, we must first estimate the amount by which an image is blurred. When the images are captured with a camera, the blur phenomenon could occur due to various reasons even when the camera is stationary. Considering that the unknown blur is due to the effect of improper focusing, it can be modeled by a Gaussian PSF, when we need to estimate the blur parameter σ (standard deviation) that determines the severity of the blur.

We estimate the blur by using a simple approach as suggested by Subbarao [12]. Since the blur is mostly due to camera defocus, the PSF can be easily parameterized by single parameter σ (see [4] for details). Hence the PSF estimation problem simplifies drastically. Let $g(x,y)$ represent the blurred image while $f(x,y)$ is the true focused image. Then $g(x,y)$ can be expressed in terms of $f(x,y)$ by a simple convolution operation as

$$g(x,y) = h(x,y;\sigma) * f(x,y). \quad (6)$$

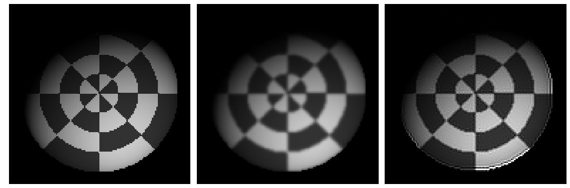
Using the fact that $h(x,y,\sigma)$ is Gaussian, and taking the Fourier Transform on both sides, we can easily derive the following equation for the estimated blur as

$$\sigma^2 = \frac{1}{A}\int\int_R \frac{-2}{w_x^2+w_y^2}\log\frac{G(w_x,w_y)}{F(w_x,w_y)}dw_x dw_y, \quad (7)$$

where R is a small region in the frequency domain and A is the area of R. Measuring the Fourier transform at a single spectral point (w_x, w_y) is, in principle, sufficient to obtain the value of σ^2, but a more robust estimate can be obtained by taking the average over a small area in the frequency domain.

The blur estimation technique as discussed above gives the estimate of blur only when the true focused image $f(x,y)$ and its blurred version $g(x,y)$ are available. But our problem is to estimate the blur given the blurred observations only, since the true focussed images for different light source positions are unknown. Only the reflectance model is known. We now suggest how the blur can be estimated from the given data itself.

Using the photometric stereo we obtain the least squares estimates of the fields $\mathbf{p}$, $\mathbf{q}$ and $\boldsymbol{\rho}$ from the observations disregarding the blurring effect. The optimization is carried out using the initial estimates of fields and an initial value of $\sigma^{(0)}$ by minimizing the cost function in equation (5) for $\mathbf{p}$, $\mathbf{q}$, $\boldsymbol{\rho}$ keeping $\sigma^{(0)}$ constant. The new estimates of fields $\mathbf{p}^{(n)}$, $\mathbf{q}^{(n)}$, $\boldsymbol{\rho}^{(n)}$ are then used in image irradiance equation (1) along with the source directions to get the estimates of the images at different light source positions. We then obtain the new estimate of blur $\sigma^{(n)}$ by using equation (7) holding $\mathbf{p}^{(n)}$, $\mathbf{q}^{(n)}$, $\boldsymbol{\rho}^{(n)}$ constant. Here the blurs are calculated between the observed images $\mathbf{g_1}, \mathbf{g_2}, \cdots, \mathbf{g_k}$ and the image estimates $\hat{\mathbf{f}}_1^{(n)}, \hat{\mathbf{f}}_2^{(n)}, \cdots, \hat{\mathbf{f}}_k^{(n)}$ and the average of these estimated blurs is used as the updated one. This new value of $\sigma^{(n)}$ obtained is then used again in the optimization to update the fields $\mathbf{p}$, $\mathbf{q}$, $\boldsymbol{\rho}$. In effect the estimation of blur parameter and the different fields are carried out alternately until the convergence is obtained for the estimated σ. The blur thus obtained is the final estimated one. The corresponding gradient fields are then used to calculate the depth map as given in [6]. The mask size chosen for the PSF should be sufficiently large compared to the value of σ. We have carried out extensive experiments under varying initial conditions and different measurement sets and we did not experience any difficulty in convergence. We also experimented on simulated data sets when the observation noise is quite high and the amount of defocus blur is large. Under such taxing circumstances we found the estimate of the blur parameter σ to be a bit underestimated.



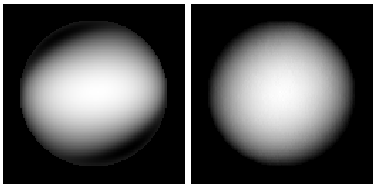
(a) (b) (c)

Figure 1: (a) Synthesized checker-board sphere image. (b) Observation using a Gaussian blur $\sigma = 1$ corresponding to Figure in (a). (c) Restored image.

4. Results and Discussions

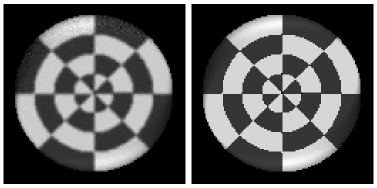
We now present the results for the proposed approach. First, we consider the experiments using the synthetically generated images. For this experiment, we generated a set of images of a spherical surface for different source positions. The sphere had a checker-board patterned albedo (see Figure 1). The obtained images are then blurred by using a Gaussian blur mask of size 7×7 with $\sigma = 1$. The final estimated σ for this experiment is 1.0352. Figure 1(c) shows the efficacy of our algorithm for the estimation of true synthesized image from its blurred version (see Figure 1(b)). The restored image is quite comparable to the true synthetic image displayed in Figure 1(a). Of the eight images generated with different light source positions, we show a single image with source position $p_s = 0.45$, $q_s = 0.80$. We see that the boundary curve on each segment in the estimated image are sharper when compared to the blurred image indicating the restoration of high frequency details. The estimated depth map shown as an intensity variation in Figure 2(b) is also quite correct as the intensity is highest at the center and decreases as we move away from it which definitely reflects the shape of a hemisphere. The shape distortion seen in the depth map of Figure 2(a) as obtained from a standard PS method clearly indicates the loss of depth information. Finally, we show the estimated and the true albedo maps in Figures 3(a, b). The restored albedo map is also quite comparable with the true one. In all our experiments the value of μ is chosen as 0.01 for the estimation of all the three fields namely, $\mathbf{p}$, $\mathbf{q}$ and the $\boldsymbol{\rho}$.

Next, we consider an experiment using the real images. We take images of a stuffed doll 'Jodu' for eight different positions of the light source. No attempt was



(a) (b)

Figure 2: (a) Depth map for blurred checker-board images. (b) Restored depth map.



(a) (b)

Figure 3: (a) Estimated albedo. (b) True albedo obtained from the synthesized images.

made to bring the object (Jodu) in to focus and hence, as seen in Figure 4 the observations are partly blurred. The blurred observations are then used to derive the fields $\mathbf{p}^{(0)}$, $\mathbf{q}^{(0)}$ and the $\boldsymbol{\rho}^{(0)}$ which are used as the initial estimates for our algorithm. An initial estimate of $\sigma^{(0)} = 0.7$ computed over a mask size 9×9 was used in order to estimate the different fields and to restore the images iteratively. After every 100 iterations in gradient descent operation for updating the surface gradients and the albedo (from which the images for different light source positions are calculated) a new value of σ is estimated and is used again to refine the fields and the images. The final estimated σ was found to be 1.0576. The results of the experiment are illustrated with the following Figures. Two of the eight blurred observations with source positions $p_s = -0.8389$, $q_s = -0.7193$ and $p_s = -0.3639$, $q_s = -0.5865$ are shown in Figures 4(a, b). The restored Jodu images for the same are displayed in Figures 5(a, b). As can be seen restored images are much sharper. Although we did not observe the perceptual difference in the estimated depth map and the depth map due to blurred observations, there was an improvement in the estimated depth map in terms of MSE (mean squared error). The MSE between the depth map due to focused observations captured by keeping the aperture very small and the depth map due to the blurred observations was found to be 0.0069, where as it was 0.0031 for the proposed method. This clearly indicates the improvement in the depth map estimation using the proposed algorithm. Similar conclusions can be drawn for the albedo estimate where the MSEs were 0.031 and 0.026, respectively.

We have presented a new approach for the simultaneous estimation of structure and image recovery along with the blur parameter estimation from blurred photometric observations. We do obtain an improved accuracy using the proposed approach.

(a) (b)

Figure 4: Observed Jodu images with the camera defocus.

(a) (b)

Figure 5: Restored Jodu images.

References

[1] M. R. Banham and A. K. Katsaggelos. Digital Image Restoration. *IEEE Signal Processing Magazine*, 14(2):24–41, 1997.

[2] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of Royal Statistical Society, Series B*, 36:192–236, 1974.

[3] V. F. Candela, A. Marquina, and S. Serna. A Local Spectral Inversion of a Linearized TV Model for Denoising and Deblurring. *IEEE Trans. on Image Processing*, 12(7):808–816, 2003.

[4] S. Chaudhuri and A. N. Rajagopalan. *Depth From Defocus: A Real Aperture Imaging Approach*. Springer-Verlog, NewYork, 1999.

[5] B. K. P. Horn. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Massachussetts Inst. of Technology, 1970.

[6] B. K. P. Horn. *Robot Vision*. MIT Press, 1986.

[7] K. Ikeuchi. Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method. *IEEE Trans. on Pattern analysis and Machine Intelligence*, 3(6):661–669, 1981.

[8] Y. Iwahori, R. J. Woodham, Y. Watanabe, and A. Iwata. Self Calibration and Neural Network Implementation of Photometric Stereo. In *Proc. International Conf. on Pattern Recognition*, pages 359–362, Quebec, Canada, 2002.

[9] D. Kundur and D. Hatzinakos. Blind Image Deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.

[10] U. Sakarya and I. Erkmen. An Improved Method of Photometric Stereo Using Local Shape from Shading. *Image and Vision Computing*, 21(11):941–954, 2003.

[11] W. M. Silver. Determining Shape and Reflectance Using Multiple Images. Master's thesis, Department of Electrical and Computer Science, MIT, Cambridge, 1980.

[12] M. Subbarao. Parallel Depth Recovery by Changing Camera Parameters. In *Proc. international Conference on Computer Vision*, pages 149–155, Florida, USA, 1988.

[13] R. J. Woodham. Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings. Technical Report 457, MIT AI Lab, June 1978.