Depth From Defocus in Presence of Partial Self Occlusion

Sundeep Singh Bhasin and Subhasis Chaudhuri
Department of Electrical Engineering
Indian Institute of Technology-Bombay
Mumbai, India-400076
bhasin@ee.iitb.ac.in
sc@ee.iitb.ac.in

Abstract

Contrary to the normal belief we show that self occlusion is present in any real aperture image and we present a method on how we can take care of the occlusion while recovering the depth using the defocus as the cue. The space-variant blur is modeled as an MRF and the MAP estimates are obtained for both the depth map and the everywhere focused intensity image. The blur kernel is adjusted in the regions where occlusion is present, particularly at the regions of discontinuities in the scene. The performance of the proposed algorithm is tested over synthetic data and the estimates are found to be better than the earlier schemes where such subtle effects were ignored.

1. Introduction

Depth from defocus (DFD) is a ranging technique where the defocus blur is used as a cue to recover depth of the given scene. Two images of a scene are enough to recover the complete depth map. However it has been shown that the result improves as the image set increases [1]. The images are taken with two different but known camera settings, and the depth map can be recovered from the relative blur in two images. Researches until now were convinced that the occlusion problem common in stereo does not take place in DFD methods. However, it was shown in [2] that such affect is also present in the DFD. But no effort was made in suggesting a possible solution in alleviating this effect. An attempt has been made in this paper to account for such an effect by incorporating change into the blur kernel dynamically while recovering depth using the MAP-MRF approach.

Pentland [1], [3] identified the problem of DFD as a linear space variant estimation of the blur. The defocus parameter was recovered using the deconvolution in the frequency domain. However, the method depended on a perfectly focused image of the scene for results. Subbarao [4] proposed a more general method in which he removed the constraint of one image being formed with the pinhole aperture. A Gaussian defocus operator was assumed and the blur was recovered in the frequency domain through inverse filtering assuming local shift invariance of the blur. However, the estimates of blur are sensitive to the presence of noise in the observations. Ens and Lawrence [5] formulate the problem as a regularized deconvolution problem. The regularization is with respect to the shape of the point spread function (PSF). They propose a matrix based approach to DFD and argue that this can lead to a greater accuracy compared to inverse filtering. Some work has also been performed in the area of active depth from defocus. The idea of active DFD was proposed by Pentland et al. [6]. The method can provide range estimates in homogeneous regions by projecting a known pattern of light. Nayar et al. [7] describe a real time range focus sensor that uses an optimized illumination pattern to obtain accurate and high resolution depth maps from both textured and textureless surfaces. However, active ranging techniques find limited applications in practical situations.

The DFD problem is, in general, a problem of space variant (SV) blur identification. When the depth varies continuously the space frequency representation (SFR) can be used to yield reasonably accurate estimates of the blur [8]. As the change in depth in the scene is usually gradual, this constraint can be used to improve the depth estimates. The SFR based approach is amenable to incorporation of smoothness constraints on the blur parameter. This regularized approach is known to yield better results as shown by Rajagopalan and Chaudhuri [9]. The above mentioned methods can recover the depth map quite well but they do not perform any SV restoration of the defocused input images. A method has been proposed in [10] that models the depth and the focused image intensity processes as separate MRFs and recover both these fields using a MAP estimator.
The concept of modeling depth/image as an MRF is quite well known in the literature [11]. The utility of the MRF model is well established in the DFD framework [12].

All published works on DFD assume that there is no self occlusion in the input images. This assumption is not valid even when the true depth map in the scene is convex shaped. However as explained in section 3, self occlusion may take place in any real aperture imaging system whenever there is a depth discontinuity. Precise conditions under which the self occlusion takes place have been presented in section 3. Unlike in the pin-hole camera where the self occlusion is defined in terms of visibility of a point in the scene at the optical center of the camera as per the geometric optics, in real aperture imaging (i.e. a lens with finite aperture) the occlusion could be partial or total. In case of total self occlusion, a point is not visible in either of the two input images and hence the depth at the point cannot be obtained. Partial self occlusion means that the blur kernel due to the depth related to the defocus is partially obstructed (resulting in a truncated blur kernel) in one or both of the input images and hence the shape of the blur kernel is a function of the amount of the self occlusion. It may be noted that the size of the blur kernel is a function of the depth as shown in section 2. Hence a DFD system has to estimate both the size and the shape of the space varying blur. All prior research work deal with estimating the size of the blur kernel, while we present a method where in the truncation of the blur kernel is estimated iteratively during the blur estimation, thus refining the accuracy of the estimate by incorporating the effects of partial self occlusion.

The remainder of the paper is structured as follows, in section 2, we provide an overview of the theory of the DFD system and the image formation process. In section 3, analysis of the problem of self occlusion in the real aperture imaging is described. Section 4 deals with the proposed framework to recover the estimates of depth in presence of self occlusion. In section 5, we present the results of experiment performed on the simulated observations from self occluded images in comparison to the observation where such effects are ignored. In section 6, we conclude with discussions on our future work.

2. Depth from Defocus

The theory of DFD is essentially based on the fact that in the real aperture imaging only one of the planes is focused, i.e. only one depth in the whole scene can satisfy the thin lens relationship \( \frac{1}{v} + \frac{1}{v_0} = \frac{1}{f_1} \) where \( D, v \) and \( f_1 \) are the object distance, the image plane to lens distance and the focal length of the lens, respectively. When any point on a given plane at distance \( D \) is not focused, its image is a point but a circular patch [4] of radius \( r_b \) given by

\[
\sigma = \rho r_b \nu \left( \frac{1}{F_1} - \frac{1}{v_0} - \frac{1}{D} \right), \quad (1)
\]

where the lens aperture is \( r_0 \), lens to image plane distance is \( v_0 \). For a discrete image captured by a CCD array, the blur radius \( \sigma \) is given \( \sigma = \rho r_b \nu \). The phenomenon of depth related defocus is illustrated in Figure 1. For two different lens settings, we get

\[
\sigma_k = \rho r_k \nu \left( \frac{1}{F_k} - \frac{1}{v_k} - \frac{1}{D} \right), \quad k = 1, 2 \quad (2)
\]

By eliminating \( D \) from the Eqn(1), it can be shown that \( \sigma_2 = \alpha \sigma_1 + \beta \), where \( \alpha \) and \( \beta \) are known constants that depend only on the camera settings [4]. For a diffraction limited lens system, the point spread function (PSF) of the camera may be approximately modeled [4] as a circularly symmetric 2-D Gaussian function given by

\[
h(i, j) = \frac{1}{\pi \sigma^2} \exp\left( -\frac{(i^2 + j^2)}{2\sigma^2} \right)
\]

Since the PSF, specified by the spread factor \( \sigma \) is a function of depth \( D \), the variation in depth thus renders the blurring process shift variant. The intensity at the \((i, j)\)th pixel in the defocused image can then be obtained as a space variant convolution of the focused image.

\[
g(i, j) = \sum_m \sum_n f(m, n) \ h(i, j; m, n) \quad (3)
\]

where \( f() \) is the pin-hole focused image and \( h(\cdot, \cdot; i, j) \) is the space varying PSF at pixel \((i, j)\). Given two observations of a same scene \( f(i, j) \) and assuming the depth to be locally constant, the relative blur in two observations can be

![Figure 1. Geometry of the image formation in a real aperture camera.](image-url)
obtained from Eqn(3) (see [4] for details). Now using the Eqn(2), the depth can be estimated locally.

3. Analysis of Self Occlusion in DFD

In all previous work on DFD it has been assumed that the problem of self occlusion is absent. The observation that the binocular methods are prone to the missing parts problem is a consequence of the non-zero baseline associated with the lenses. If the angle subtended at a point in the scene by the optical centers of the two lenses is small, it reduces the number of points that will be visible to a part of the lens while being occluded at another part. In a monocular system like the DFD, this angle is zero and hence it was widely believed that there is no self occlusion. This observation would have been true had the lens been a pin-hole one. In DFD the lens has a finite aperture and thus there is a blur kernel associated with each value of a depth for a particular lens setting. It is this non zero blur kernel that gives rise to self occlusion in real aperture imaging. In fact, the defocus point-spread could be larger than that for a small baseline stereo and hence the chances of occlusion phenomenon is higher for DFD than the stereo [2]. However, in stereo if one of the rays is blocked it results in the matching problem, in contrast the DFD relies on a continuum of rays, thus allowing estimation although with an error (partial self occlusion).

We now explain the occlusion phenomenon in DFD using an example of step discontinuity in the scene. This is illustrated in Figure 2. Let us assume the simple case that there are two planes lying perpendicular to the optical axis of the lens and the discontinuity occurs exactly along the axis of the lens Figure 2. The two depths, as shown in Figure 2 are $D_1$ and $D_2$ respectively. From Figure 1 we notice that for any point in the scene, all rays emanating within the solid angle subtended by the aperture of the lens are collected by the thin lens and the image is formed on the other side. Compare this for a point lying between point A and B in Figure 2. Some of the emitted light rays within this solid angle now get blocked due to the discontinuity at the point A. Thus the corresponding blur kernel, as illustrated in the Figure 2, is no longer circular but it has the shape of the truncated disk. The kernel becomes a semi-circle for point A and slowly it acquires the shape of a complete circle as we move to the point B. Hence we observe that there is partial self occlusion in DFD. Beyond the point B, the image is free of such self occlusions.

It is interesting to note that if, instead of a horizontal step discontinuity BACD, one had an oblique discontinuity like BA'CD, the points lying within the region A to A' contribute to the self occlusion phenomenon even though they would not be visible on the image plane. For the given configuration in Figure 2 it is easy to show that

$$BA' = AA' = r_0 \frac{D_2 - D_1}{D_1}$$

Similarly for any point $P$ in within the region BA having the distance $x$ from the optical axis, the amount of self-occlusion $r_{eff}$ as shown in Figure 2 is given by

$$r_{eff} = (\frac{r_0}{r} - 1) \frac{D_1}{D_2 - D_1} x = \frac{r_0}{r_0} \frac{D_1}{D_2 - D_1} x$$

The above relationship must be used while redefining the blur kernel due to self occlusion in the DFD method. However, the quantities $r_0$, $D_1$, $D_2$ are not known apriori as depth (or blur) is what needs to be estimated in DFD. Hence an iterative scheme must be developed to handle the self occlusion. This method is discussed in section 4.2. It should be noted here that the truncation of the blur kernel depends on the orientation of the depth discontinuity. Similarly, if the object has a corner/wedge type of discontinuity, the corresponding truncation in the blur kernel will also be wedge shaped.

We end this section with a lemma, proof of which is quite trivial.

**Lemma:** Convexity of an object (or scene) is not a sufficient condition for not introducing any self-occlusion.

4. Recovery of Depth and Intensity maps

In this section, we address the DFD problem in the context of MRF based modeling. A detailed discussion on MRF can be found in [12]. With two defocused images and the knowledge that occlusion does affect the observation one
4.1. Basic Framework

The DFD problem, in general, is ill-posed [4] and may not yield a unique solution unless additional constraints are introduced to restrict the solution space. Both the scene intensity and the SV blur can be modeled as separate MRFs. Smoothness constraints are then imposed on both SV blur parameter and the image intensity. Such a smoothness constraint also serves to improve the depth estimates. Given two defocused images simultaneous recovery of depth and the restoration of the focused image is then posed as a MAP problem. The estimation process is schematically represented in Figure 3. The solution set becomes mathematically tractable because of the parametric form of the blur problem. The estimation process is schematically represented in Figure 3. The solution set becomes mathematically tractable because of the parametric form of the blur-}

Figure 3. Block schematic of the MAP-MRF based method for the recovery of depth and the image restoration.

has to develop an algorithm to recover the SV blur and to restore the focused pin-hole image. In the next subsection we present the basic framework for blur recovery and extend it to include the self occlusion effects subsequently.

Let $g_k(i, j) = f(m, n) h_k(i, j; m, n) + w_k(i, j)$, $k = 1, 2$

where $g_1()$ and $g_2()$ are the two defocused images of the scene and $f()$ is the original focused image of the scene which is not known. The space varying blurring functions $h_1()$ and $h_2()$ are assumed to be Gaussian with the spread parameters given by $\sigma_k$, $k = 1, 2$, such that $\sigma_2 = \alpha \sigma_1 + \beta$, where $\alpha$ and $\beta$ are related in terms of the camera parameters. However the blur parameter $\sigma_k$ is, in general, space varying and is directly related to the depth variations in the scene. The observation noise fields, denoted by $w_i$ are assumed to zero mean, independent white Gaussian processes with variances $\sigma_w^2$.

Let $S$ denote the random field corresponding to the space-variant blur parameter $\sigma_k(m, n)$ while $F$ denote the random field corresponding to the intensity image $f(m, n)$ over the $N \times N$ lattice of sites $L = \{(i, j) : 1 \leq i, j \leq N \}$. We assume that $S$ and $F$ can take $P$ and $M$ levels, respectively. We also assume that the $S$ and $F$ are statistically independent of each other as well as with $w_1$ and $w_2$. Let $G_1$ and $G_2$ denote the random fields corresponding to the observed images.

In matrix notation, $q_n$ which involves linear space variant filtering can be written as

$$g_k = H_k f + w_k, k = 1, 2,$$

where the vectors $g_k, f$ and $w_k$ represent lexicographical ordering of $g_k(i, j), h(i, j)$ and $w_k(i, j)$. $H_k$ is the blur matrix corresponding to the space-variant blurring function $h_k(i, j; m, n)$. Unlike in the shift-invariant case, the blur matrix $H_k$ does not possess the nice property of having a block-Toeplitz structure.

Since $S$ and $F$ are both modeled as MRFs we have $P[S = s] = \frac{1}{Z_S} e^{-U(s)}$ and $P[F = f] = \frac{1}{Z_F} e^{-U(f)}$ where, $U(s)$ and $U(f)$ are the energy functions [13] associated with the $S$ and $F$. They are given by $U(s) = \sum_{c \in C} V_c(s)$ and $U(f) = \sum_{c \in C} V_c(f)$ where $V_c(s)$ and $V_c(f)$ are the potential functions associated with each clique $c$ while $C_s$ and $C_f$ denote the set of all cliques corresponding to the SV and $F$, respectively. The terms $Z_s$ and $Z_f$ correspond to the partition functions of $S$ and $F$. Given a realization of $S$, the blurring function $h_1()$ is known and hence the matrices $H_1$ and $H_2$ in Eqn 7 are both known.

Now, given the observations $g_1$ and $g_2$, the a posteriori conditional probability of $S$ and $F$ is given by $P[S = s, F = f | G_1 = g_1, G_2 = g_2]$. Using the Bayes' rule, and the fact that the $S$ and $F$ are assumed to be statistically independent, the problem of simultaneous space-variant blur identification and image restoration can then be posed as the following MAP estimation problem:

$$\max_{s,f} \frac{P[G_1 = g_1, G_2 = g_2 | S = s, F = f] P[S = s] P[F = f]}{P[G_1 = g_1, G_2 = g_2]}$$

The posterior distribution must have a neighborhood structure in order to have a practical implementable computational algorithm. It has been shown in the literature that the posterior distribution, indeed, corresponds to a reasonable MRF neighborhood structure [12]. The MAP estimation is equivalent to minimizing the posterior energy function. Smoothness constraints on the estimates of the SV blur and the intensity process can be encoded separately in the potential function. In order to preserve the discontinuities in the blurring process and in the focused image of the scene, appropriate line fields are incorporated into the energy function as suggested in [13].

4.2. Incorporating Self Occlusion Effect

Initially we obtain the estimate of the blur field as discussed in section 4.1. Then we perform an edge detection
Figure 4. (a) and (b) set of defocused images due to a step discontinuity ignoring the self occlusion effects.

Figure 5. (a) and (b) set of self occluded defocused images due to a step discontinuity.

on this blur map and locate the points where there is a sharp discontinuity using a suitable threshold. From section 3, we know that the effect of the self occlusion is confined to a region close to the detected edge points. From the estimated value of the blur, and using the knowledge of the camera parameters and Eqn(5), the blur kernel $H_k$ in Eqn(7) can be modified and the intensity fields are once again obtained. The region in which the updates of the estimates are to be restricted is given by Eqn(4). Based on this revised estimates of the blur map, we can have a better estimate of the location and the amount of depth discontinuity at various points in scene. The above process may be repeated till no further refinements are possible.

The MAP estimates of the above problem can be solved using any optimisation technique. The modifications are incorporated into the same to include the effect of self occlusion. The description of the algorithm has been given in [12]. The modifications to the algorithm can be succinctly given as

Step1: Obtain initial MAP estimates of image $f^0(i,j)$ and blur fields $\sigma^0(i,j)$. Set $n = 0$

Step2: Locate discontinuities in $\sigma^n(i,j)$ through the edge detection

Step3: Obtain $r^0_{eff}(i,j)$ using the Eqns(2) and (5).

Step4: Update the estimates $f^{n+1}(i,j)$ and $\sigma^{n+1}(i,j)$ and set $n = n + 1$. Goto Step2 until convergence.

5. Experimental Results

In this section, we present the results of the performance of the proposed algorithm in Section 4.2 in estimating the space-variant blur and obtaining the pin-hole restored image. The space variant discontinuous blur is generated using the following parameters in Figure 2, $D_1 = 6.5cm$, $D_2 = 30cm$, $r_0 = 3.0cm$, $F_1 = 3.0cm$ and the camera settings are changed by varying the image plane-lens distance $\nu_1 = 6.0cm$ and $\nu_2 = 7.2cm$ to obtain two observations. The value of $\rho$ is fixed to 1.0. The value of the blur parameter can be obtained using Eqn (2). A binary random dot pattern is blurred assuming the circular blur kernel for the given blur parameter. The resulting observations are illustrated in Fig 4. The images obtained by including the effects of self occlusion, resulting in the change in blur kernel, are illustrated in Fig 5.

It is interesting to note the differences in pictures given in Figures 4 and 5. Although the left to right transition in the blur is quite apparent in both the figures due to the sharp discontinuity, the brightness remains quite constant in Fig 4. In Fig 5, the average brightness drops down significantly as one moves near the depth discontinuity. This appears more like a dark band to the right of the discontinuity. This is due to the partial self occlusion as a significant amount of rays get occluded resulting in the loss of energy received by the detector. The pictures given in the Figure 5 represent the images captured with a real aperture camera (the aperture used in this case is deliberately chosen to be larger to highlight the effect of self occlusion) for two different lens-to-image plane distances. These images are used in the simulation to recover the depth map using the proposed DFD method. The estimated SV blur when no occlusion is considered is given in Fig. 6. The estimates are accurate except at the discontinuity and the $rms$ error was found to be 0.0075. However, when occlusion effects are taken into consideration then the error at the discontinuity decreases considerably, which is reflected in by a decrease in $rms$ error of 0.0070. The resulting depth maps are shown in Fig 7. The regions where the occlusion effects are prominent, i.e. close to the discontinuity the results are closer to the actual values. In order to demonstrate the improvement in estimating the depth, instead of computing the $rms$ error in the entire scene, we compute the error within a band of ±4 pixels about the line of the discontinuity as beyond this region the estimates would be nearly identical in both the method. The corresponding values of the $rms$ error is 0.0142 and 0.0080 respectively. This substantiate our claim that the accuracy of the DFD method can be improved by taking the effect of self occlusion in account.
6. Conclusions

In this paper we have shown that partial self occlusion does exist in DFD methods due to the finite aperture of the camera. We have obtained a mathematical relationship to quantify the amount of self occlusion. Further, we have shown that even though a particular point is not visible to the camera, it may still partially occlude neighboring points. An iterative MAP estimation technique is proposed to estimate both the depth and the focused (pin-hole) image of the scene in presence of such partial occlusions. The depth and the intensity maps have been modeled as separate MRFs in this study. Inclusion of the effect of occlusion results in an improved accuracy of the depth map. The discontinuity is captured much more smoothly in occlusion included images. Computationally the scheme is slightly more expensive because of iterative approach. Currently we are looking into ways of handling arbitrary types of discontinuities.

References


