DEPTH ESTIMATION USING FUSION OF DEFOCUS AND STEREO

*Uma Mudenagudi and Subhasis Chaudhuri**

Abstract

In this paper we propose a new method for estimating depth using a fusion of defocus and stereo. It avoids the correspondence problem of stereo. Main advantage of this algorithm is simultaneous recovery of depth and image restoration. Restoration of image degraded by linear shift variant blur, by itself, is a challenging task. The depth (blur or disparity) in the scene and the intensity process in the focused image are individually modeled as Markov random fields (MRF). It avoids the windowing of data and allows incorporation of multiple observations in the estimation procedure. The performance of the proposed method is evaluated for different set of images. The accuracy of depth estimation and the restored image are improved compared to the individual depth from defocus or the stereo method.

1 INTRODUCTION

In recent years, an important area of research in computer vision has been the recovery of 3D information about a scene from its 2D images. In the case of human vision, there is also the concept of binocular fusion, which is when stereoscopically presented image appears a single entity. Julesz[?] showed that random dot stereograms provide a cue for disparity even when each image does not provide any high level cue for disparity. Pentland [?] reported that the gradient of focus inherent in biological and most of the optical system is actually a useful source of depth information. Binocular stereo matching is in general ambiguous if the matching is evaluated independently at each point purely by using image properties. All stereo matching algorithms examine the candidate matches by calculating how much support they receive from their local neighborhood. Marr and Poggio[?] proposed a cooperative stereo algorithm based on a multi resolution framework. Barnard and Thompson[?] proposed a feature based (statistical approach) iterative algorithm for correspondence problem. Large number of papers have appeared in the literature on stereo analysis and review of them can be found in [?].

Let us now look at the literature on depth recovery from defocused images. Rajagopalan and Chaudhuri proposed new methods, based on block shift-variant blur model[?] that incorporates the interaction of blur among neighboring subregions. Space variant approach for depth recovery using a space-frequency representation framework are given in [?],[?]. They have also proposed a method [?] of estimating SV blur as well as focused image of the scene from two defocused images. In this method, both focused image and blur are modeled as separate MRFs and their MAP estimates are obtained using SA.

Computationly efficient methods are available in the literature for each of them. Kanade and Okutomi[?] have given new stereo matching algorithm with an adaptive window, the size of the

* *The authors are with Department of Electrical Engineering, Indian Institute of Technology-Bombay, Mumbai - 400076, India. E-mail:zahid@ee.iitb.ernet.in*

window is selected by evaluating the local variation of the intensity and the disparity. In [?], nonlinear diffusion is used to estimate the window size. But accuracy of estimates in depth from defocus(DFD) are inferior to that of stereo based methods, while the stereo gives a sparse depth map and the setting up the correspondence is a difficult task. In this method we fuse these two methods to estimate the depth information for an improved accuracy. Tsai and others[?] proposed a scheme of integrating stereo and defocus. But they have used rough depth estimates obtained from defocus as a guideline for the stereo algorithm. Yoav and others[?] have given analysis of defocus vs stereo and responses of DFD stereo and motion to the perturbations. As we know in stereo the disparity is directly related to depth. In defocus blur parameter is σ is also directly related to the depth. Hence disparity, a function of σ, is known in terms of lens settings and base line distance. This information is used to fuse the two methods, there by getting the advantages of both the methods. In the proposed method, given four images of a scene, ie, two defocused stereo pairs of images, we estimate the focused image of the scene and a dense depth (blur or disparity) map, which is space variant, using an MAP-MRF approach. Computational problem for the MAP-MRF is solved using simulated annealing(SA).

In section 2 detailed theory for the proposed method is developed. Section 3 describes the different parameters involved in the iterative algorithm. The performance of the algorithm is validated by a set of results in section 3. Main conclusions are given in section 4.

2 FUSION OF DEFOCUS AND STEREO

In this proposed method we are simultaneously estimating blur/disparity and restoring one of the focused image of the scene in the stereo pair (say the left image). Estimating the other stereo pair is trivial once we know the disparity. As in the most literature we assume the epipolar line constraint, so that disparity is only in the y direction. The left image is given by

$$f_R = f_L(x, y + d_r(x, y)) + w(x, y)$$

where d_r is the disparity associated with the stereo pair and w is white Gaussian noise. The basic structure of the method is given in figure-1.

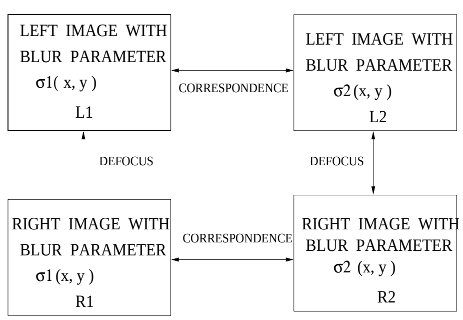

Figure 1: Basic structure of depth from defocused stereo

Let us denote by, L1 = left image with $\sigma_1(x, y)$ as a blur parameter, L2 = left image with $\sigma_2(x, y)$ as a blur parameter, R1 = stereo pair of L1 with same blur parameter $\sigma_1(x, y)$, R2 = stereo pair of L2 with same blur parameter $\sigma_2(x, y)$. For the DFD camera setup, we also have

$$\sigma_2(x, y) = \alpha\sigma_1(x, y) + \beta \tag{1}$$

Where α and β are known constants that depend on camera settings. The relative blur between the two defocused images is estimated using the intensity information by assuming an appropriate model for the optical transfer function. Usually a Gaussian model is assumed. Though the Gaussian blur is of infinite extent, a finite spatial extent approximation ($\pm 3\sigma$ pixels) is assumed for Gaussian blurring windows. We note that blurring PSF $\sigma_i(x, y)$, i=1,2 is space varying and it is directly related to the depth in the scene for a fixed camera setting. Similarly $\sigma_i(x, y)$ is directly related to the scene disparity $d_r(x, y)$ through a known relationship. The problem is addressed under the framework of MAP-MRF approach. Computation based on simulated annealing is carried out for simultaneous recovery of depth estimates and the focused image. The utility of MRF lies in its ability to capture local dependencies and its equivalence to the Gibbs random fields(GRF). Space variant blur parameter which is related to depth is modeled as an MRF. The local property of MRF leads to an algorithm which can be implemented in a local and parallel manner. Let S denote the random field corresponding to space variant (SV) blur parameter $S_{ij} = \sigma_1(i, j)$, and G denote the random field corresponding to observed image while $/F$ denote the random field corresponding to the focused image (intensity process). Assume that S can take P possible levels and F can take M possible levels. S is statistical independent to both F and noise field W. The noise field is assumed to be white Gaussian with zero mean and variance σ_w^2. The relation between focused image and defocused image is governed by the observation model given by

$$g = Hf + W$$

where g, f and w represent lexicographical ordering of $g(i, j)$, $f(i, j)$ and $w(i, j)$ respectively. H is the blur matrix corresponding to SV blurring function h(i,j;m,n). Since it is SV H does not possess the nice property of having a block toeplitz structure. The above problem of recovering f given g is ill posed and may not yield unique solution, unless additional constraints like smoothness are added to restrict the solution space. Since S is modeled as MRF, we can write

$$P(S = s) = \frac{1}{z} exp\{-\sum_{c \subset C} V_c(s)\} \tag{2}$$

Where z is a partition function.

A posteriori probability distribution of S is given by $P(S = s/G = g)$. Using Bayes' rule, the problem of estimating the SV blur parameter can then be posed as following MAP problem.

$$max \quad \frac{P(G = g/S = s)P(S = s)}{P(G = g)}$$

since S and F are statistically independent of each other, for each image g fixed , $P(G = g/S = s)$ is Gibbs distribution with energy function $U^P(s)$ can be assumed to be

$$U^P(s) = \lambda_s \int (s_x^2 + s_y^2) dxdy + \frac{||g - Hf||^2}{2\sigma_w^2}$$

Extending above analysis for four observed images g_{l_1}, g_{l_2}, g_{r_1} and g_{r_2} with random fields G_{L_1}, G_{L_2}, G_{R_1} and G_{R_2} respectively. The observation model are

$$g_{l_k} = H_k f_L + w_k, \quad g_{r_k} = H_k(d_r) f_R + w_k.$$

As both S and F_L are modeled as MRFs we can write

$$P(S = s) = \frac{1}{z^s} exp\{-U^s(s)\},$$

$$P(F=f)=\frac{1}{z^f}exp\{-U^f(f)\}$$

The terms $U^s(.)$ and $U^f(.)$ correspond to the energy functions of the space-variant blurring process and intensity processes respectively. Given realization of s the blurring function $h_1(.)$ is known and hence the matrix H_1 is known. Moreover, $h_2(.)$ is also determined by $\sigma_{ij_2}=\alpha\sigma_{ij_1}+\beta$. Now, given the four observed images the posteriori conditional joint probability of S and F is given by Bayes' rule,

$$P(S=s,F=f/G_{L_1}=g_{l_1},...,G_{R_2}=g_{r_2})=\frac{P(S=s)P(F=f)P(G_{L_1}=g_{l_1},..,G_{R_2}=g_{r_2}/S=s,F=f)}{P(G_{L_1}=g_{l_1},...,G_{R_2}=g_{r_2})}$$

Since S and F_L are assumed to be statistically independent, we pose the problem of simultaneous space-variant blur estimation and image restoration as the following MAP problem.

$$\max_{s,f}\ P(G_{L_1}=g_{l_1},.../S=s,F=f)P(S=s,F=f)$$

For fixed observations with an appropriate model (first order smoothness), one can show that the posterior energy function is given by

$$\begin{aligned}U^P(s,f_L)=&\frac{||g_{L1}-H_1f_L||}{2\sigma_w^2}+\frac{||g_{L2}-H_2f_L||}{2\sigma_w^2}\\&+\frac{||g_{R1}-H_1f_R||}{2\sigma_w^2}+\frac{||g_{R2}-H_2f_R||}{2\sigma_w^2}\\&+\int[\lambda_s(s_x^2+s_y^2)+\lambda_f(f_x^2+f_y^2)]dxdy\\&+\lambda_{st}||g_{R1}-g_{L1}(x+d)||\\&+\lambda_{st}||g_{R2}-g_{L2}(x+d)||\end{aligned}\tag{3}$$

From the above analysis computing MAP estimates is equivalent to minimizing the posteriori energy function. Smoothness constraints on the estimates of space-variant blur parameter and the intensity process are encoded in the potential function. In order to preserve the discontinuities in the blurring process and the focused image of the scene, line fields are also incorporated into the energy function. The horizontal and vertical line fields corresponding to the blurring process and binary intensity process are denoted by l_{ij}^s, v_{ij}^s, l_{ij}^f and v_{ij}^f respectively. The posteriori energy function to be minimized is defined including line fields as $U^P(s,f_L,l_{ij}^s,v_{ij}^s,l_{ij}^f,v_{ij}^f)$, where the smoothness term in equation-3 can be replaced by

$$\begin{aligned}&\sum_{i,j}\lambda_s[(s_{i,j}-s_{i,j-1})^2(1-v_{i,j}^s)\\&+(s_{i,j+1}-s_{i,j})^2(1-v_{i,j+1}^s)+(s_{i,j}-s_{i-1,j})^2(1-l_{i,j}^s)\\&+(s_{i+1,j}-s_{i,j1})^2(1-l_{i+1,j}^s)]\\&+\sum_{i,j}\lambda_f[(f_{i,j}-f_{i,j-1})^2(1-v_{i,j}^f)\\&+(f_{i,j+1}-f_{i,j})^2(1-v_{i,j+1}^f)+(f_{i,j}-f_{i-1,j})^2(1-l_{i,j}^f)\\&+(f_{i+1,j}-f_{i,j1})^2(1-l_{i+1,j}^f)]\\&+\gamma_s[l_{i,j}^s+l_{i+1,j}^s+v_{i,j}^s+v_{i,j+1}^s]\\&+\gamma_f[l_{i,j}^f+l_{i+1,j}^f+v_{i,j}^f+v_{i,j+1}^f]\end{aligned}$$

Simulated annealing algorithm is used to obtain the MAP estimates of the SV blur parameter and focused image simultaneously. The temperature variable is introduced in the objective function . Annealing cum cooling schedule are carried out at each iteration with linear cooling. Since the random fields associated with SV blur and image are assumed to be statistically independent, the values of blur at every point(σ_{ij}) and f_{ij} are changed independently. The parameters of MRF model are chosen in an ad hoc way. The initial estimates of blur are obtained from Subbarao's window based method[**?**]. The posteriori energy function is, in general non-convex, algorithms based on steepest descent are prown to get trapped in a local minima. Hence we chose simulated annealing (SA) algorithm for minimizing the posteriori energy function. It is important to note that the locality property of the posteriori distribution is what enables us to successfully employ the SA algorithm.

3 RESULTS

In this section, we present the performance of the proposed method in estimating the space variant blur and restoring the image. Results of experimentation are presented on a random dot pattern and a corridor image. The number of discrete levels for SV blur was chosen as 32. For the intensity process, 256 levels were used which is the same as the CCD dynamic range. Defocused versions of random dot pattern were first generated such that $\sigma_{i,j2} = 0.5\sigma_{i,j1}$. Figure-2 shows the four defocused stereo pair of images. Window based method of Subbarao is used to estimate the SV blur parameter from the noisy defocused images (size of window 8x8 pixels). Figure-2 (2g) shows the initial estimates of the blur.The *rms* value of the error in the estimate of the blur is 0.55. These estimates are used to initialize the proposed scheme. The values of various parameters used in SA algorithm were $T_0 = 10.0$, $\lambda_s = 5000.0$, $\lambda_f = 0.005$, $\lambda_{st} = 0.01$, $\gamma_s = 10.0$, $\gamma_f = 15.0$, $\delta = 0.9$, $\theta s = 0.4$, $\theta_f = 25.0$, $\sigma s = 0.1$, $\sigma f = 6.0$, annealing iterations=200, metropolis iterations=100. Where T_0 is the initial temperature, θ_s and θ_f are threshold for deciding an edge in the blur and image respectively. ${\sigma_s}^2$ and ${\sigma_f}^2$ are variances with which new samples are estimated. The estimated SV blur and restored image are shown in figure-2 (2f) and (2h) respectively. The value of *rms* error reduced to 0.12. From the figure it is seen that the blur is well captured even at the edges. It is important to note that using the proposed method we have been able to perform simultaneous SV image restoration.
The algorithm is now tested on a corridor image in which the ceiling has less spectral content than the floor. The original fully focused image is shown in figure-3 (3e) and actual depth map is shown in fig-3(3f). From fig-3(3g), which is restored image using defocus alone, the estimates are poor at maximum blur which does not have enough spectral content. Results were improved with the proposed scheme since it fuses stereo also. From fig-3(3i) which shows the estimates of SV blur using defocus method. Again the estimates are poor where there is a less spectral content. Estimates at the ceiling of the corridor were poor since it is homogeneous region without any spectral content. we can see the improvement whenever the image does not contain enough variations where the defocus method fails (like near the ball). The *rms* error is reduced from 0.78 to 0.34. The method can be easily extended to having multiple observations.

4 CONCLUSIONS

We have proposed a new method of fusing the DFD and the stereo based methods to improve the accuracy of the depth estimation. The method uses the advantages of both DFD and stereo. The rms error in the estimates of space varying blur is reduced compared to DFD alone. One can simultaneously restore the image of the scene also. The recovered depth map is dense and no separate interpolation or feature matching is required. Currently we are looking at ways to speed up the computation.

Figure 2: Left pair of defocused images with $\sigma_1 = 2\sigma_2$.

Figure 3: Right pair of defocused images with $\sigma_1 = 2\sigma_2$.

Figure 4: Focused Image and Reconstructed image.