REGULARIZED DEPTH FROM DEFOCUS

Vinay P. Namboodiri [1], *Subhasis Chaudhuri* [2], *Sunil Hadap*[3]

[1] ESAT-PSI/VISICS, KU Leuven, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium.
[2]Department of Electrical Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai, India.
[3]Advanced Technology Labs, Adobe Systems Incorporated, San Jose, California, USA
vinay.namboodiri@esat.kuleuven.be, sc@ee.iitb.ac.in, sunilhadap@acm.org

ABSTRACT

In the area of depth estimation from images an interesting approach has been structure recovery from defocus cue. Towards this end, there have been a number of approaches [4, 6]. Here we propose a technique to estimate the regularized depth from defocus using diffusion. The coefficient of the diffusion equation is modeled using a pair-wise Markov random field (MRF) ensuring spatial regularization to enhance the robustness of the depth estimated. This framework is solved efficiently using a graph-cuts based techniques. The MRF representation is enhanced by incorporating a smoothness prior that is obtained from a graph based segmentation of the input images. The method is demonstrated on a number of data sets and its performance is compared with state of the art techniques.

Index Terms— Focus, Defocus, Depth from Defocus, MAP-MRF, Graph-Cuts,

1. INTRODUCTION

The problem that is addressed in this paper is one of depth estimation from defocused images. Depth estimation from images has been one of the well studied problems in computer vision. One of the methods used for depth estimation is based on the use of defocus cue. Here, one uses the optical properties of cameras whereby due to the real aperture, an observation of a real scene is blurred by a defocus blur proportional to the depth in the scene. This is illustrated in Fig. 1. When the point is not in focus, its image on the image plane is no longer a point but a circular patch of radius σ that defines the amount of defocus associated with the depth of the point in the scene. It can be shown that [4]

$$\sigma = \kappa r v (\frac{1}{F} - \frac{1}{v} - \frac{1}{Z}) \quad (1)$$

where r is the radius of the aperture, v is the lens-to-image plane distance, F is the focal length of the lens, Z is the depth at that point and κ is a camera constant that depends on the sampling resolution on the image plane. From the eqn.(1) we note that $C = (r, F, v)$ defines the camera parameters each of which may be changed to effect a different amount of defocus blur for a fixed depth.

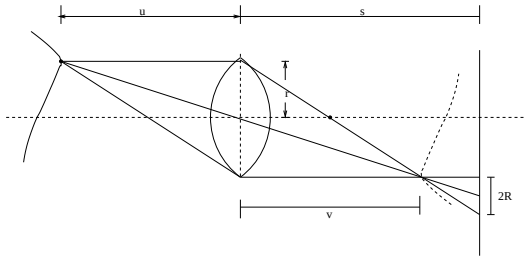


Fig. 1. Illustration of image formation in a convex lens.

There has been considerable research done towards using this cue to estimate depth [4, 6]. The approach used here is based on the modeling of defocus blur as a diffusion process [5, 13, 14]. This method was explored first by Favaro *et al.* where they used a linear diffusion process to estimate depth in the scene. Subsequently, the use of linear diffusion was used in the spectral domain [13]. The problem was also addressed using stochastically perturbed diffusion [14]. However, in [13, 14], regularization was not incorporated. In [5], the authors used $\mathcal{L}_2$ regularization that results in overly smooth images. In this paper we address the shortcomings of previous approaches and propose a Markov random field representation to estimate the diffusion coefficient.

1.1. Need for robust regularization

The problem of depth from defocus is an ill-posed problem because, in the absence of texture the depth in the scene cannot be estimated. Thus it becomes an ill-posed problem in the Hadamard sense, because in these areas the depth estimate cannot be obtained uniquely. A common approach adopted is to therefore regularize the solution by considering the solution in the neighborhood or by adopting some assumption of smoothness of the solution. The earlier approaches were based on usage of Tikhonov regularization. There a regularization term would be added to the minimization term. The regularization term would specify the form of the solution based on $\mathcal{L}_2$ smoothness of the result which

could then be solved by the calculus of variation approach using Euler-Lagrange equations [9]. However this approach results in overly smooth solutions. An approach made towards solving this problem is by using total variation based regularization [16]. A more principled approach is by using energy minimization using the discrete optimization framework of graph-cuts as proposed by Boykov *et al.* [3]. This approach can be mathematically formulated as an approach towards exact maximum a posteriori (MAP) estimation of a Markov random field (MRF) [8].

2. RELATED WORK

The earliest works towards the use of graph-cuts in image processing has been towards denoising of images where Greig *et al.* [8] used the Ford-Fulkerson idea of Graph-Cuts towards solving the problem of denoising of images by solving it in a MAP-MRF framework proposed earlier by Besag [1]. The use of MAP-MRF towards solving the problem of stereo was proposed by Roy and Cox [15]. An important contribution was by Boykov *et al.* [3] who demonstrated a fast approximate energy minimization technique for solving computer vision problems by using the idea of α expansion and α swap. A theoretical understanding of the energy functions that can be minimized using graph cuts was done by Kolmogorov and Zabih [12]. Further work done by Kolmogorov and Zabih showed effective use of graph cuts for computation of depth from stereo in the presence of occlusion [11]. While subsequently, graph-cuts has been used in many computer vision problems, the usual application has been based on the disparity in intensity values. In our problem we use graph cuts in order to compute the amount of defocus blur at each location in the image and this cannot be directly computed from the pixel intensities.

The MAP-MRF framework has been used in depth from defocus quite successfully by Chaudhuri and Rajagopalan [4]. They have used the Wigner-Ville distribution based representation for computing the relative blur which is then estimated using the MAP-MRF framework. They have also shown that it is possible to simultaneously compute depth and restore the image. The main drawback in their method was the use of simulated annealing for solving the MAP-MRF framework which is computationally prohibitively expensive. In [5], Favaro *et al.* consider the estimation of diffusion coefficient using gradient descent with $\mathcal{L}_2$ regularization. However, as mentioned earlier $\mathcal{L}_2$ regularization results in overly smooth results. The use of graph-cut allows seamless incorporation of robust regularization functions like the Huber function and total variation.

3. REPRESENTATION OF DEFOCUS CUE

The defocusing process can be modeled as

$$I(x) = \int f(\tau) h(x, y) dy, \tag{2}$$

where we adopt x to denote the 2D space co-ordinates in an image, $f(x)$ is the focused image of the scene and h is the space-varying PSF. Here $h(x)$ is given by a circularly symmetric 2-D Gaussian function

$$h(x) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right), \tag{3}$$

where σ is a function of depth at a given point and its relationship to the depth in the scene is given by eqn.(1). It is quite well-known that, for a scene with constant depth the imaging model in eqn(2) can be formulated in terms of the isotropic heat equation [10] given by

$$\frac{\partial u(x;t)}{\partial t} = c\left(\triangle u(x;t)\right), \quad u(x,0) = f(x)$$

where $\triangle u$ is the Laplacian operator. Here the solution $u(x,t)$ taken at a specific time $t = \tau$ plays the role of an image $I(x) = u(x,\tau)$ and $f(x)$ corresponds to the initial condition, i.e. the pin-hole equivalent observation of the scene. Note that we have used $u(x,t)$ to represent the evolution of heat everywhere in the paper. The blurring parameter σ is related to the diffusion coefficient by the following relation [5]

$$\sigma^2 = \frac{2tc}{\gamma} \tag{4}$$

where t is the time variable in the diffusion equation, c is the diffusion coefficient, and γ is a proportionality constant relating the blur radius to the spread (σ) of the blur kernel that can be determined via initial calibration. In the depth from defocus problem, the depth in the scene varies over the image and hence the constant c will actually be $c(x)$, i.e., it will vary over the image. This corresponds to a heat equation in an inhomogeneous medium. Now, given two images $I_1(x)$ and $I_2(x)$ one can estimate the diffusion coefficient such that

$$\frac{\partial u(x;t)}{\partial t} = c\left(\triangle u(x;t)\right), \quad u(x,t_1) = I_1(x).$$

Here, without loss of generality the initial condition is taken to be $I_1(x)$ and the equation is evolved to estimate the diffusion coefficient such that $u(x,t_2) = I_2(x)$. This is done by estimating the relative blurring variance σ_r^2 that blurs image $I_1(x)$ to equate $I_2(x)$.

Here we estimate the value of σ_r. Let w_i denote the label or σ_r value of pixel i in an image $w = (w_1, \ldots, w_n)$, then a Bayesian formulation specifies an *a priori* distribution $p(w)$ over all allowable images. Here $p(w)$ is assumed to be a Markov random field (MRF). Let w^* denote the unknown true σ labels corresponding to the scene. Here we have $z = (z_1, \ldots, z_n)$ denotes the observed values of w^*. The observed values are obtained by convolving a particular location with a label. The likelihood $l(z|w)$ of any image w is combined with $p(w)$ in accordance with Bayes' theorem to form an *a posteriori* distribution $p(w|z) \propto l(z|w)p(w)$. The maximum a posteriori (MAP) estimate of w^* is the label $\hat{w}$ that maximizes $p(w|z)$

The values $z_1, \ldots, z_n$ are assumed to be conditionally independent given w. Maximizing $p(w|z)$ is equivalent to minimizing the following term

$$E(w) = \sum_i \left(\phi(z|w_i) + \sum_{j \in \mathcal{N}} \psi(w_i, w_j) \right) \quad (5)$$

Here, the first term ϕ is the data likelihood and the second term ψ is the interaction potential determined by the prior. The data likelihood is estimated using a Euclidean distance measure between the destination image and the source image blurred by a label w_i. The interaction potential is given by

$$\psi(w_i, w_j) = \beta(i,j) M(w_i, w_j). \quad (6)$$

Here $M(w_i, w_j)$ is a robust error term between the two labels w_i, w_j. In the experiments the truncated linear term was used after experimental comparison. The term $\beta(i,j)$ incorporates the prior obtained by a segmentation of the input image using a graph-based segmentation described in [7] and is given by

$$\beta(i,j) = \begin{cases} 1 \text{ if segment } i = \text{segment } j \\ 0 \text{ else .} \end{cases} \quad (7)$$

This energy function can be minimized using graph cuts as discussed in the next section. An advantage of this formulation is the symmetric nature in which the value of σ can be estimated. In [14] and in the approach by Favaro *et al.* [5], preprocessing of images had to be done to ensure that the diffusion was always carried out in the forward direction only. Here, since the label for σ_r is being estimated we can equally assume positive and negative labels, wherein positive labels imply blurring of I_1 to obtain I_2 and negative labels imply vice-versa. This method thus simplifies the problem of requiring pre-processing since the labels are estimated with regularization.

4. GRAPH-CUTS FOR SOLVING MAP-MRF FRAMEWORK

We minimize eqn.(5), thereby maximizing the posterior probability using graph cuts ([2],[3]). The graph cut finds the cut with the minimum cost separating terminal vertices, called the source and sink. Here, the terminal vertices are assigned the presence and absence of a discrete label from w_i. The graph cut is solved using alpha expansion [3] which allows us to consider this method of using binary labels to minimize the cost over the entire set w.

The resulting energy function is a energy function of binary variables of the form

$$E(w_1, \cdots, w_n) = \sum_{i<j} E^{i,j}(w_i, w_j). \quad (8)$$

Here $w_1, w_2, \cdot, w_n$, correspond to vertices in the graph and each represents a binary variable where they are either connected to the sink or to the source. For an energy function of this form it has been proved by Kolmogorov and Zabih [12], that the function can be minimized provided that it is regular, i.e. minimization is possible if and only if each term of the energy function satisfies the following condition:

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \quad (9)$$

which implies that the energy for two labels taking similar values should be less than the energy for the two labels taking different values. In this case the labels denote the σ values and we can have a metric defined over σ. Hence, it would satisfy the above condition and we can therefore minimize the resultant energy function $E(w)$. In the next section we present the results using the proposed method.

5. EXPERIMENTAL RESULTS

We evaluate the method with synthetic and real image data sets and compare the results obtained with some of the latest techniques. The method compares well with these methods.

5.1. Simulated Data

Fig.2 shows a test data where a standard texture map from the Brodatz texture database has been blurred to create blocks of varying depths using Gaussian blur with variances 0.8, 1.6 and 3.8 respectively. Figures 2(a,b) show that there are three distinct layers of depth in the simulated observations. There exists a gap of 3 pixels between the blurred regions. However, due to the convex assumption, the depth map obtained by the method proposed in [5] results in the regions being connected as can be seen in Fig. 2(c). Fig. 2(d) shows the corresponding estimated depth map obtained from the technique proposed in [14]. Here, the three regions can be seen separately, however, the result is noisy due to absence of regularization. In fig. 2(e) the result obtained from the proposed technique is shown where the depth in the different regions is seen separately and the result is also smoother due to regularization. The brighter areas correspond to regions closer to the camera. The accuracy is confirmed against the expected depth map.

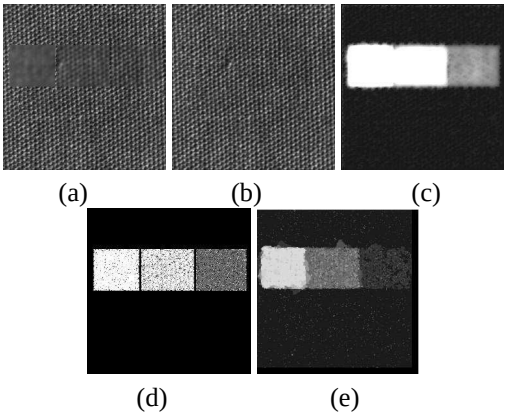
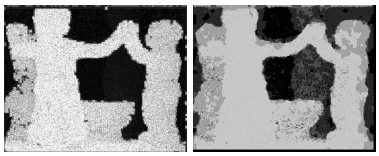


Fig. 2. Here (a,b) show a standard texture with high spectral details, synthetically blurred assuming three different layers of depth. (c) shows the resulting recovered structure from the method of Favaro *et al.*[5]. (d) shows the result obtained by stochastic technique [14] and (e) shows the result using the proposed method.

(a) (b) (c)



(d) (e)

Fig. 3. Here (a,b) are two real data sets showing a few dolls at different depths(Images courtesy [5]). (c) shows the resultant depth map for the method by Favaro *et al* [5]. (d) shows the resultant depth map obtained by the stochastic depth from defocus method [14] and (e) shows the resultant obtained by the proposed regularized depth from defocus method.

5.2. Real data

The first real data set used for evaluation is the "dolls" data set [5]. The scene depicts a few dolls situated at various depths. The dolls are focused at different depths in the scene with the focal plane shifting from foreground to the background. The result obtained by the linear diffusion method explained in [5] can be seen in 3(c). Here the authors have used $\mathcal{L}_2$ regularization. The result obtained by stochastically perturbed depth from defocus method is shown in 3(d). Here no regularization has been used. The result obtained by using regularized depth from defocus using graph-cuts is shown in 3(e). It can be seen that the result obtained by the proposed technique is more improved as compared to the other techniques. The regularization used definitely improves the depth-map obtained. We now test our method on a more challenging real image data set which has a distinct shadows and a non-textured background. Fig. 4(a) shows the image where the near vegetables are in focus and fig. 4(b) shows the scene where the far vegetables are in focus while the near vegetables are defocused. The result obtained by using the proposed method is shown in fig. 4(c). The resultant depth map in this challenging data set clearly shows the different vegetables and we are able to correctly estimate the depth.

6. CONCLUSION

We have seen the need for regularization and have provided a principled method for regularizing the diffusion coefficient estimated using a Markov random field framework which is solved by an efficient graph-cut based method. The results demonstrate that use of regularization indeed helps in obtaining a more reliable estimate of the depth in the scene.

(a) (b) (c)

Fig. 4. Here (a,b) are two real data sets showing a few vegetables at different depths. (c) shows the result obtained by the proposed method.

7. REFERENCES

[1] J. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B(Methodological)*, 48(3):259–302, 1986.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on PAMI*, 26(9):1124–1137, September 2004.

[3] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE transactions on PAMI*, 20(12):1222–1239, November 2001.

[4] S. Chaudhuri and A. N. Rajagopalan. *Depth From Defocus: A Real Aperture Imaging Approach*. Springer Verlag, New York, 1999.

[5] P. Favaro, S. Osher, S. Soatto, and L. Vese. 3d shape from anisotropic diffusion. In *Proceedings of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 179–186, 2003. Madison, Wisconsin, USA.

[6] P. Favaro and S. Soatto. *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-Blur*. Springer-Verlag, London, 2007.

[7] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):1–26, September 2004.

[8] D.M. Greig, B.T. Porteous, and A.H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Societies, Series B*, 51(2):271–279, 1989.

[9] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, Massachusetts, 1986.

[10] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[11] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. In *IEEE International Conference on Computer Vision*, volume 2, pages 508–515, 2001.

[12] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE transactions on PAMI*, 26(2):147–159, February 2004.

[13] V.P. Namboodiri and S. Chaudhuri. On defocus, diffusion and depth estimation. *Pattern Recognition Letters*, 28(3):311–319, February 2007.

[14] V.P. Namboodiri and S. Chaudhuri. Shape recovery using stochastic heat flow. In *Proc. of British Machine Vision Conference (BMVC)*, September 2007. Warwick, UK.

[15] S. Roy and I.J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *ICCV*, pages 492–502, 1998.

[16] F. Park T. Chan, S. Esedoglu and A. Yip. Recent developments in total variation image restoration. In O. Faugeras N. Paragios, Y. Chen, editor, *Handbook of Mathematical Models in Computer Vision*, pages 17–30. Springer Verlag, New York, 2005.