

Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera

Vinay P. Namboodiri

Subhasis Chaudhuri *

Department of Electrical Engineering
Indian Institute of Technology, Bombay
Powai, Mumbai 400076, India.

{vinaypn, sc}@ee.iitb.ac.in

Abstract

In this paper we investigate the challenging problem of recovering the depth layers in a scene from a single defocused observation. The problem is definitely solvable if there are multiple observations. In this paper we show that one can perceive the depth in the scene even from a single observation. We use the inhomogeneous reverse heat equation to obtain an estimate of the blur; thereby preserving the depth information characterized by the defocus. However, the reverse heat equation, due to its parabolic nature, is divergent. We stabilize the reverse heat equation by considering the gradient degeneration as an effective stopping criterion. The amount of (inverse) diffusion is actually a measure of relative depth. Because of ill-posedness we propose a graph-cuts based method for inferring the depth in the scene using the amount of diffusion as a data likelihood and a smoothness condition on the depth in the scene. The method is verified experimentally on a varied set of test cases.

1. Introduction

Shape from a single image is a task which has predominantly continued to elude computer vision researchers. While the community has gained considerable expertise to attack this problem when provided with multiple images, the same cannot be said when just a single image from an uncalibrated camera is provided. The task is quite daunting computationally, in spite of the ease with which the human system is able to achieve the same.

In this paper we present a method to perceive the depth layers from a single defocused image. The limited depth of field introduces a defocus blur in images captured with conventional lenses based on the range of depth variation in a scene. This artifact has been used in computer vision



Figure 1. A sample image of a scene captured with a low depth of field and the resultant depth map estimated using the proposed method without any additional information. Note that darker regions correspond to focused regions and lighter regions correspond to defocused regions.

for estimating depth in the scene when multiple defocused images are provided. Here we show that using this low-level cue it is still possible to perceive the scene structure using a single image to a fairly good extent of accuracy.

The method proposed in this paper uses a single defocused image of a scene taken with an uncalibrated real aperture camera having a low depth of field. We show that a surprisingly large amount of information of the 3D scene can be inferred based on just the defocus cue in a single image. In order to extract this information we face several challenges. The conventional methods for estimating depth from defocus ([4],[9]) have relied on multiple observations. The differences in blur among various observations are used as a cue for estimating the depth. However, it is more natural for a photographer to take a single image. There are in general a lot of beautiful images taken of natural scenes with a shallow depth of field, such as the image shown in fig.1(a). In this paper we explore for the first time as to what extent the blur can be estimated even with a single image. From fig.1(a), any human observer can infer that the flowering plant and the adjoining grass is closer to the camera than the waterfall. Can we teach a computer to infer the same? The depth map of the 3D scene shown in fig.1(b) demonstrates that it is indeed now possible to a certain extent using the method proposed in this paper.

The idea of using low-level cues for extracting a $2\frac{1}{2}$ D

*Financial assistances to VPN from Bharti Centre for Communication and SC under the J.C. Bose Fellowship are gratefully acknowledged.

sketch was proposed by Marr [17]. This forms a philosophical basis for our work. A successful approach for single image based structure recovery has been that by Criminisi *et al.* [5]. Here the authors have used projective geometry based techniques for computing the scene structure based on the prior knowledge of the vanishing line of a reference plane and a vanishing point. An interesting work has been that of obtaining 3D pop-up kind of structure from a single image based on learning appearance based models of geometric classes and using this information for obtaining cutouts ([12]). Another approach [20] based on similar lines uses images of scenes and their associated ground-truth depth maps. It discriminatively trains an MRF using multiscale local and global image features and uses it to predict the depth map as a function of the image. In our method we do not need such explicit prior knowledge or learning. There is a related work [22], where the authors do segmentation of images based on defocus cue by using the statistics of the wavelet coefficients, but it is highly feature dependent.

1.1. Related Work

The basic problem that has been addressed in the depth from defocus methodology (DFD) has been the measurement of the relative defocus between observations. Pentland [18] identified the problem of DFD as an estimation of linear space variant blur. The defocus parameter was recovered using the deconvolution in the frequency domain. However, the method depended on the availability of a perfectly focused image of the scene as one of the observations. Subbarao [21] proposed a more general method in which he removed the constraint of one image being formed with the pinhole aperture. Subsequently there have been several approaches in the frequency domain and in the spatial domain. An example of frequency based approach is [23]. In [23], the authors suggest the use of broadband rational filters. There has also been a substantial amount of work done to estimate depth from defocus where the image analysis is done in the spatial domain itself. One such approach has been that of modeling the problem in a statistical framework ([4], [19]). Extensive work has been done by the authors using this technique by modeling the depth and the image as separate Markov random fields (MRF). Further recent work in the spatial domain includes [6],[7], [8], [9], [16]. In [7], the authors pose the problem as one of reconstructing the shape and the radiance that minimize a measure of information divergence between blurred images. In [8], the authors have used the idea of diffusion for estimating the depth from defocus which is to some extent similar to our work. Here, however, we consider the reverse diffusion instead of using only the forward diffusion as done by the authors. In [16], the authors consider the issue of calibration for estimation of depth from defocus. In [6], the problem is explored by taking into account issues related to convexity,

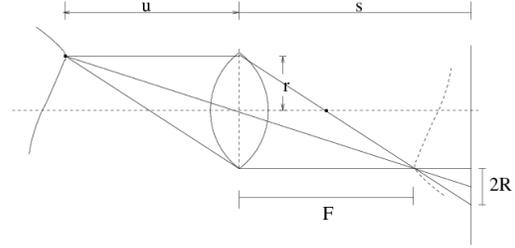


Figure 2. Illustration of image formation in a real aperture lens.

and using Bregman distances the authors provide means for estimating depth from defocus. However, while all the work done using defocus cue have been based on multiple observations, our work differs in the fact that we use a single observation with an uncalibrated camera for estimating the relative blur and thereby the relative depth in the scene.

2. Diffusion Based Modeling of Defocus

Consider the image formation process in a real aperture camera employing a thin lens [4]. When a point light source is in focus, all light rays that are radiated from the object point and intercepted by the lens converge at a point on the image plane. When the point is not in focus, its image on the image plane is no longer a point but a circular patch, as shown in fig.2. The point spread function (PSF) of the camera describes the image intensity caused by a single point light source. Geometric optics approximates the PSF to a circular disk. However, as discussed in the literature ([4],[18]), due to diffraction it will be approximately a circular blob with the brightness falling off gradually at the border rather than sharply. The resultant PSF has the general shape of a 2-D Gaussian function ([4],[18]). For an equifocal plane the resultant image formed is then given by

$$I(x) = \int f(y)h(x,y)dy, \quad (1)$$

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

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

where σ is a blurring parameter that is a function of depth at a given point.

It is known that, for a scene with constant depth the imaging model in eqn.(1) can be formulated in terms of the isotropic heat equation [13] given by

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} &= c \left(\frac{\partial^2 u(x,t)}{\partial x^2} \right) \\ u(x,0) &= f(x). \end{aligned} \quad (3)$$

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 Gaussian PSF evidently can be formulated in terms of the heat equation, since the Gaussian function is a fundamental solution of the heat equation. It is pertinent to note that even the cylindrical PSF can be modeled using heat equation. This has been shown by Guichard and Morel [11]. Thus, modeling the defocus process by the heat equation is indeed physically valid when we consider both Gaussian and cylindrical shapes for modeling the defocus PSF. The blurring parameter σ is related to the diffusion coefficient by the following relation [8]

$$\sigma^2 = \frac{tc}{\gamma} \quad (4)$$

where t is the time variable in the diffusion equation, c is the diffusion coefficient, and γ relates to the size of the blur radius in terms of pixel units which depends on the spatial resolution of the CCD array. In the depth from defocus problem, the depth in the scene varies over the image and hence the coefficient c and the time t together will be a function of relative depth at location x , i.e., it will vary over the image. This corresponds to a heat equation in an inhomogeneous medium.

3. Reverse Heat Equation

Given a defocused observation of a scene we would like to restore it using the reverse heat equation. The reverse heat equation is given as

$$\begin{aligned} \frac{\partial u}{\partial t} &= c \left(\frac{\partial^2 u(x, t)}{\partial x^2} \right) \\ u(x, \tau) &= I(x), \end{aligned} \quad (5)$$

where $I(x)$ is the blurred observation and c is the diffusion coefficient. This is identical to the heat equation given in eqn.(3), except for the initial condition given. Here we are given the blurred observation and we have to find the pin-hole equivalent image. That is we have to find the solution such that it satisfies the original image $u(x, 0) = f(x)$. This is achieved by reversing time in the heat equation and the resultant is

$$\frac{\partial u}{\partial t} = -c \left(\frac{\partial^2 u(x, t)}{\partial x^2} \right), \quad u(x, 0) = I(x). \quad (6)$$

The idea of using the reverse heat equation for restoring images was first proposed by Gabor in 1965 [15]. Recently the use of the reverse heat equation has been advocated by Buades *et al.* [3]. They propose the use of the reverse heat equation regularized by using the ‘‘non-local means’’ constraint. In this paper we use the reverse heat equation as

given in eqn.(6). The reason for not modifying the reverse heat equation as done by Buades *et al.* is that the relation between depth and diffusion coefficient and time, given by eqn.(4) is valid only for the heat equation. Hence, for an accurate depth estimation, the reverse heat equation should be used directly. The main problem faced while using the reverse heat equation is its divergent nature. Due to this the reverse heat equation remains stable for a short while and then degenerates very rapidly. The key to using the reverse heat equation is to have an effective stopping criterion that stops the reverse heat equation in its stable region.

Koenderink [13] had referred to the use of heat equation in the reverse direction indirectly in his work. He had observed that the notion of scale space in the reverse direction would be stable up to the initial condition and beyond that it would result in impulses being generated. Taking this into account we have devised a stopping criterion that would effectively stop the reverse heat equation at an appropriate time.

Consider the eqn.(5) where we are given a blurred observation $I(x)$ and we have to estimate the observation without blur i.e. $f(x)$ that was the initial condition. However, we do not know the value of the time either, i.e. we do not know how far in time should the reverse heat equation be carried out. An observation that can be used is that the eqn.(5) is valid only till time $t = 0$ and it breaks down if we go beyond this time. The breakdown of the heat equation is indicated by the degeneration of the gradient. Hence, the resultant formulation for reverse heat equation is

$$\frac{\partial u}{\partial t} = -\beta(x)c \left(\frac{\partial^2 u(x, t)}{\partial x^2} \right), \quad (7)$$

where $\beta(x)$ is given by

$$\beta(x) = \begin{cases} 1 & \text{if } |\nabla u - \overline{\nabla u}| < \theta \\ 0 & \text{else} \end{cases} \quad (8)$$

Here ∇u is the gradient of u and $\overline{\nabla u}$ is the average gradient in the neighborhood. The function $\beta(x)$ detects the degeneration of the gradient since the divergence of the gradient from the average gradient is an indicator of the degeneration of the gradient. The stopping time t of the reverse diffusion is then determined by the value of the constant θ . In our experiments we have used a small value of θ ranging from 0.2 to 0.4. The use of eqn.(7) results in an inhomogeneous stopping of the reverse heat equation based on the amount of defocus at a location. The relative depth in the scene is then given by

$$\hat{d}(x) = \int_0^{t(x)} c(x, t') dt'. \quad (9)$$

Here, $\hat{d}(x)$ is the approximate estimate of the depth at the location x and $t(x)$ is the reverse diffusion time at a location x . An estimate of $\hat{d}(x)$ for the scene in fig. 1 is shown in

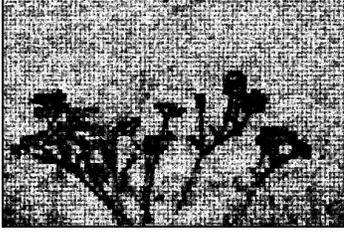


Figure 3. The integrated diffusion coefficient of the reverse heat equation for the scene shown in fig.1. The convention darker is the focused region (closer) is used throughout this paper.

fig. 3. It was demonstrated in [4] that the DFD method does not perform well in the absence of any regularization of the estimate. Hence, the depth estimate is further refined by modeling the depth as a Markov random field.

4. DFD Using Graph Cuts

Using the reverse heat equation coefficient we can obtain an approximate estimate of the depth in the scene using eqn.(9). However, in the regions which do not have texture, this depth estimate is not valid. Hence, we define a Markov random field (MRF) modeled on the relative depth (or blur) in the scene and solve it using the MAP-MRF framework [10].

The depth estimate obtained using the reverse heat equation is a measure of the disparity between the observed image and the restored image. Hence the depth estimate $\hat{d}(x)$ obtained in eqn.(9) is taken as an estimate of the observed blur \mathbf{D} , corresponding to the depth in the scene, i.e. \mathbf{D} is now the data term defined for a location x . We define a set of n discrete labels corresponding to different depths in the scene. $\mathcal{L} = \{l_1, \dots, l_n\}$. These labels are assigned over the image for the relative depth in the scene given by \mathbf{r} , one label r_x at each pixel x , that maximizes the posterior probability given by the Gibbs distribution

$$p(\mathbf{r}|\mathbf{D}) = \frac{p(\mathbf{D}|\mathbf{r})p(\mathbf{r})}{p(\mathbf{D})} = \frac{1}{Z_{\mathbf{r}}} \exp(-E(r)), \quad (10)$$

where $Z_{\mathbf{r}}$ is the normalizing constant (or partition function). The energy corresponding to a configuration r consists of a likelihood and a smoothness term as

$$E(\mathbf{r}) = \sum_{\mathbf{x}} \left(\phi(\mathbf{D}|r_x) + \sum_{\mathbf{y} \in \mathcal{N}_{\mathbf{x}}} \psi(r_x, r_y) \right) \quad (11)$$

The likelihood term $\phi(\mathbf{D}|r_x)$ is derived from the initial depth estimate \hat{d} and the smoothness term $\psi(r_x, r_y)$ is based on the prior on the depth in the scene. The neighborhood \mathcal{N}_x around x considered is the eight neighborhood around a pixel.

The prior in the scene $\psi(r_x, r_y)$ chosen to have the form

$$\psi(r_x, r_y) = \|(r_x - r_y)\|_2 \quad (12)$$

We do not explore the choice of the optimal energy function that can yield the best results in this paper.

An important issue here has been modeling the data likelihood term. Here we consider the depth term from the reverse heat equation around the edges and consider an equal likelihood for the data term where the edge is absent. Accordingly the data likelihood is given by

$$\phi(\mathbf{D}|\mathbf{r}_{\mathbf{x}}) = \begin{cases} (r_x - \hat{d}(x))^2 & \text{if } M(x) = 1 \\ \eta & \text{if } M(x) = 0 \end{cases} \quad (13)$$

Here η is the default data value which is uniform for all labels and is used in case the edge indicator function $M(x)$ indicates the absence of an edge. We use a binary valued Canny edge detector for the indicator function $M(x)$. Since the nearly homogeneous regions do not offer any information about the depth or the associated diffusion, we give more weight to the edge pixels. Accordingly the value of η is appropriately chosen.

We minimize eqn.(11), thereby maximizing the posterior probability using graph cuts ([1],[2]). 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 \mathcal{L} . The graph cut is solved using alpha expansion [2] which allows us to consider this method of using binary labels to minimize the cost over the entire set \mathcal{L} . The resulting energy function is

$$E(p_1, \dots, p_n) = \sum_{i < j} E^{i,j}(p_i, p_j). \quad (14)$$

Here p_1, p_2, \dots, p_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. These labels provide a discrete approximation of r and the corresponding minimization is same as minimization of $E(\mathbf{r})$ in eqn. 11. For an energy function of this form it has been proved by Kolmogorov and Zabih [14] 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:

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \quad (15)$$

which implies that the energy for two labels taking similar values should be less than the energy for the two labels taking different values. Since this is the case for the energies defined by us, we can find the desired configuration $\hat{\mathbf{r}}$ by minimizing eqn.(11).

5. Results

We now describe the experiments that we have performed using a variety of images. The first experiment was performed on a synthetic texture data set. Here we provide

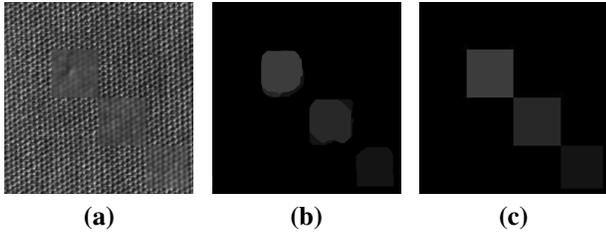


Figure 4. A texture map is modified with three blurred regions as shown in (a) and the resultant depth map estimated using the proposed method is shown in (b). The ground truth for this synthetic data set is shown in (c). Here we obtain accurate labels for 94% of the pixels.

a defocused input image fig. 4(a). This shows a texture image from the Brodatz texture database which is blurred with 3 different blur regions. The corresponding depth map estimated is shown in fig. 4(b) and the ground truth for the data is shown in fig. 4(c). The quantitative comparison of the estimated depth map with the ground truth showed that 94% of the pixels are labeled accurately.

We then tested our method on a general outdoor image in fig.1(a). This image is captured with a moderately low depth of field with the focus on the flowers in the foreground. The grass near the flowers is also in the foreground, whereas the waterfall is further back and the sloping hills towards the left are furthest away from the camera. All these details are captured appropriately in the depth map shown in fig.1(b). The result is obtained using 16 labels of depth with a gradient degeneration threshold θ in eqn.(8) of 0.2. The corresponding data likelihood term from the reverse heat equation is shown in fig.3. As can be seen from the recovered depth map, we are able to obtain a very good estimate of the relative structure in the scene just from the single image in fig.1(a).

We next consider an image taken from a sports scene shown in fig.5(a). As can be seen, here the player is in focus and the spectators are out of focus. This image is interesting due to the very low amount of texture present in the scene. The scene structure is estimated using the proposed technique and the result is shown in fig.5(b). The player is clearly seen in front and the spectators are seen in the background. Further details like the right arm of the player being in front, the face being a bit behind the body can also be perceived.

We next consider a data set with complex lighting conditions. The input image fig.6(a) shows a room with various artefacts. The figurines of puma are shows specular effects and there are also other diffuse reflectors in the scene. The relative depth map obtained using the proposed technique is shown in fig.6(b). This shows that the algorithm is able to estimate the relative layers of depth even in such challenging situations.



Figure 5. A sports scene shown in (a) with low texture is considered. Even in this challenging data set an appropriate depth map is obtained using the proposed method as seen in (b).

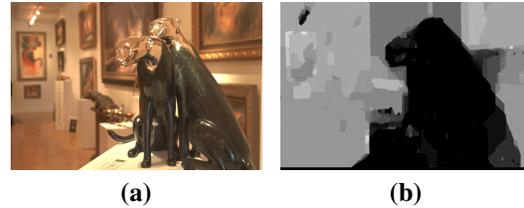


Figure 6. A scene with complex lighting conditions is shown in (a) and the resultant depth map is shown in (b).

5.1. Ambiguity in Depth Estimation Using Defocus

When we try to perceive the depth in the scene based on the defocus cue, there is an underlying assumption that all the objects are to one side of the defocus cone shown in fig. 2. A similar amount of defocus blur is generated on both sides of the defocus cone as is illustrated in fig.2. The same amount of blur is generated at the planes which are equidistant from the focus point F. Hence, it cannot be discerned whether the objects that are defocused are towards the front or back. This ambiguity is evident from our next experiment.

We consider two images of dolls (courtesy [8]) where the focus is interchanged between background and foreground objects. In the first case as seen in fig.7(a), the foreground is out of focus and the background is in focus. The corresponding recovered depth map obtained using the image is shown in fig.7(b). Note, that the depth map shown also captures the details like the right most doll has its front portion more in focus than the back and the depth variation around the hands is also reflected properly. Favaro *et al.* [8] have used the two images given in figures 7(a) and 8(a) to compute the depth. The result obtained by the method proposed by Favaro *et al.* is shown in fig.7(c) and those obtained by the proposed method (for the foreground defocused case for uniformity in comparison) is shown in fig.7(b). As can be seen in fig.7, the results obtained by the proposed method are definitely comparable and in some cases, as in around the right most doll, the depth map from the proposed method shows more detail. Note that, strictly speaking the depth maps cannot be compared as the proposed method uses just a single image.

When we consider the other image of the dolls where the foreground is in focus and the background is out of fo-

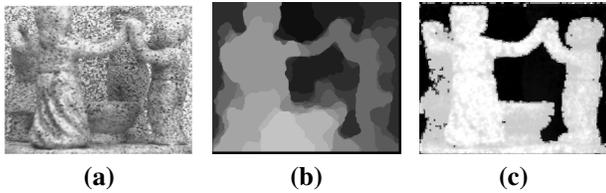


Figure 7. The dolls scene (courtesy [8]) with the foreground defocused as shown in (a). The depth map from the proposed method is shown in (b). The depth map obtained for the method proposed by Favaro *et al.* [8] using two images is shown in (c).

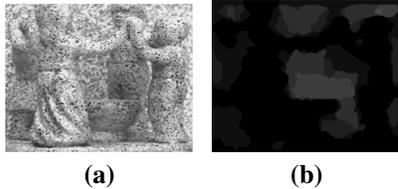


Figure 8. The dolls scene (courtesy [8]) with the background defocused as shown in (a). The depth map from the proposed method is shown in (b).

cus as seen in fig.8(a), then the resulting depth map shown in fig.8(b) will also be in the opposite direction. Here, the depth map shows the foreground portion in darker shade indicating that this region is in focus. Notwithstanding the above difference, a comparison of figs.7(b) and 8(b) shows that the recovered depth maps are mutually very consistent.

6. Conclusion

In this paper we demonstrate that it is indeed possible to recover the relative depth layers from a single image using the defocus cue. The reverse heat equation can be used for restoring the image in an inhomogeneous way based on the amount of defocus blur. The amount of reverse heat diffusion serves as a data likelihood and using this likelihood around the edges, a graph cuts based method is proposed to estimate the depth in the scene thereby enforcing regularization. We have demonstrated by experimentation on a variety of test cases of real data that the method consistently provides a correct perception of the scene structure. We would like to further explore the use of additional priors to improve the results further by considering the natural image statistics of the gradient in an image.

References

- [1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on PAMI*, 26(9):1124–1137, September 2004.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 20(12):1222–1239, November 2001.
- [3] A. Buades, B. Coll, and J-M. Morel. Image enhancement by non-local reverse heat equation. Technical Report 2006-22, Centre de Mathematiques et Leurs Applications, ENS Cachan, 2006.

- [4] S. Chaudhuri and A. N. Rajagopalan. *Depth From Defocus: A Real Aperture Imaging Approach*. Springer Verlag, New York, 1999.
- [5] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.
- [6] P. Favaro. Shape from focus and defocus: Convexity, quasiconvexity and defocus-invariant textures. In *Proc. IEEE International Conference of Computer Vision (ICCV)*, October 2007. Rio de Janeiro, Brazil.
- [7] P. Favaro, A. Mennucci, and S. Soatto. Observing shape from defocused images. *International Journal of Computer Vision*, 52(1):25–43, April 2003.
- [8] 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.
- [9] P. Favaro and S. Soatto. *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-Blur*. Springer-Verlag, London, 2007.
- [10] D.M. Greig, B.T. Porteous, and A.H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of Royal Statistical Societies, Series B*, 51(2):271–279, 1989.
- [11] F. Guichard and J-M. Morel. A note on two classical enhancement filters and their associated pde's. *International Journal of Computer Vision*, 52(2/3):153–160, 2003.
- [12] D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. *ACM Trans. Graph.*, 24(3):577–584, 2005. SIGGRAPH 2005 Conference Proceedings.
- [13] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [14] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on PAMI*, 26(2):147–159, February 2004.
- [15] M. Lindenbaum, M. Fischer, and A. Bruckstein. On gabor's contribution to image enhancement. *Pattern Recognition*, 27(1):1–8, 1994.
- [16] Y. Lou, P. Favaro, A.L. Bertozzi, and S. Soatto. Autocalibration and uncalibrated reconstruction of shape from defocus. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2007. Minnesota, US.
- [17] D. Marr. *Vision: A computational representation into the human representation and processing of Visual Information*. W.H. Freeman and Company, New York, 1982.
- [18] A. P. Pentland. A new sense for depth of field. *IEEE Trans. on PAMI*, 9(4):523–531, July 1987.
- [19] A.N. Rajagopalan and S. Chaudhuri. An mrf model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Trans. on PAMI*, 21:577–589, July 1999.
- [20] A. Saxena, S. H. Chung, and A. Y. Ng. Learning depth from single monocular images. In *Proc. of Conf. on Neural Information Processing Systems, NIPS 18*, December 2005.
- [21] M. Subbarao. Parallel depth recovery by changing camera aperture. In *Proceedings of International Conference on Computer Vision*, pages 149–155, 1988. Florida, US.
- [22] J.Z. Wang, J. Li, R.M. Gray, and G. Wiederhold. Unsupervised multiresolution segmentation for images with low depth of field. *IEEE Trans. on PAMI*, 23(1):85–90, January 2001.
- [23] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, May 1998.