Novel diffusion based techniques for depth estimation and image restoration from defocused images

A thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

by

Vinay P. Namboodiri
(Roll No. 02407703)

Under the guidance of
Prof. Subhasis Chaudhuri

Department of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY–BOMBAY
2008

Thesis Approval

The thesis entitled

Novel diffusion based techniques for depth estimation and image restoration from defocused images

by

Vinay P. Namboodiri
(Roll No. 02407703)

is approved for the degree of

Doctor of Philosophy

Examiner	Examiner
Guide	Chairman

Date: _____

Place: _____

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, INDIA

CERTIFICATE OF COURSE WORK

This is to certify that **Vinay P. Namboodiri** (Roll No. 02407703) was admitted to the candidacy of Ph.D. degree Jan. 2003, after successfully completing all the courses required for the Ph.D. programme. The details of the course work done are given below.

S.No	Course Code	Course Name	Credits
1	EE 678	Wavelets	6
2	EE 702	Computer Vision	6
3	EES 801	Seminar	4
		Total Credits	**16**

IIT Bombay

Date: Dy. Registrar (Academic)

Abstract

An intrinsic property of real aperture based imaging is the blurring of an observation due to defocus. There are two major aspects related to the defocus blur present in the image. The first aspect is based on use of the defocus blur for estimating the depth in the scene. The other aspect relates to restoration of the image. This problem manifests itself as a challenging blind, space varying deconvolution problem. In this thesis we explore diffusion based methods for depth estimation and image restoration from defocused observations.

We are given two observations of a scene that are taken with different camera parameters and are defocused to different extents. We use the idea of defocus morphing in the spectral domain to define a defocus space of observations from the two observations given. The defocus morphing technique is obtained from the use of linear diffusion equation. Based on the defocus space we estimate depth in the scene and the pin-hole observation. The framework proposed demonstrates the equivalence between depth from defocus and depth from focus based techniques. Since the depth in the scene is varying, one has to use a local spectral morphing procedure using a windowed (short-term) Fourier transform. The windowing procedure introduces artifacts in the defocus morphing procedure. We therefore consider the defocus blur by evolving the heat equation in the spatial domain.

The defocus blur can be estimated by evolving the diffusion equation in the spatial domain. This implies that the defocus blur kernel is Gaussian. However, on account of self occlusion and aperture artifacts there are deformations of the Gaussian point spread function. We therefore propose a stochastically perturbed diffusion model that implicitly handles the departure from the Gaussian assumption. We use stochastic level sets for propagating the stochastically perturbed diffusion and thereby estimating depth in the scene. An additional advantage of using this technique is that a global minimization procedure is adopted instead of a convex gradient descent technique. Thus problems of local minima are avoided in this technique.

The models of linear diffusion and stochastically perturbed diffusion, however, do not

incorporate spatial regularization effectively. We address this issue and use a pairwise Markov random field (MRF) representation for estimating the diffusion coefficient in the scene. We use a graph-cuts based method for obtaining the maximum a posteriori (MAP) solution of the MRF. By using the neighborhood information we obtain a regularized solution which is more accurate.

We then consider the problem of restoration of blurred images. Towards this end we propose the use of a non-parametric method based on reverse mean shift. The reverse mean shift procedure is divergent. We therefore use a stopping criterion based on cluster separation. Due to the non-parametric nature of the formulation the reverse mean shift procedure can be used for blind, space-varying deblurring. However, this cannot be used to handle deblurring in the presence of noise. We therefore propose a generalized mean shift method which combines the forward and reverse mean shift procedure. The switching between the forward and the reverse mean shift procedures is done based on correlation criterion defined on the residual and original observation. As a result, where there exist strong edges that are blurred, reverse mean shift procedure operates and in areas that are textureless the forward mean shift is active and so we are able to achieve both denoising and deblurring. However due to the opposite nature of deblurring and denoising procedures, the result obtained is somewhat constrained.

Consider an image which is blurred using a Gaussian kernel. In this case, the non-parametric method due to its non-linear nature would not obtain the exact deblurred image. It would obtain only an approximate solution for the deblurring. The formation of defocused image can be better modeled using the heat equation. Therefore a better solution for deblurring an image is obtained by using the reverse heat equation. Since the reverse heat equation is an unstable procedure we propose a technique for geometrically stabilizing the reverse heat equation. Though the non-parametric technique for deblurring performs better restoration, the stabilized reverse heat equation demonstrates how even in the linear case one can obtain acceptable deblurring performance.

We then consider the problem of depth estimation from a single defocused image. While, it is possible to estimate depth using two defocused images, the estimation of depth from a single defocused image is considered to be impossible. We show that it is still possible to perceive the depth in the image using only a single defocused observation. Since defocus can be modeled using the heat equation we use the reverse heat equation to formulate a technique for depth estimation from a single defocused image. The data constraint is derived by considering

the reverse heat diffusion coefficient along the edges and the problem is solved in the MAP-MRF framework using graph cuts. The results obtained demonstrate that it is indeed possible to obtain depth from a single defocused observation. This is an important contribution since it underlines the fact that defocus as a cue is indeed a strong cue.

Contents

List of Tables

List of Figures

Notations

I_1, I_2	Defocused observations with two different camera parameters.
I_0	Pin-hole equivalent image of a scene.
c	Diffusion coefficient.
σ	Radius of blur function corresponding to variance of the point spread function.
r	Radius of aperture of the lens.
F	Focal length of the lens.
v	Lens to image plane distance.
Z	Depth of the scene from the lens.
ρ	Camera constant that depends on the sampling resolution on the image plane.
α	Amount of defocus morphing.
$u(x, y, t)$	Diffusion variable with $u(x, y, 0)$ denoting the initial observation I_1.
∇u	Gradient of u given by $[\partial u/\partial x \ \partial u/\partial y]^T$
$\triangle u$	Laplacian of u given by $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$
$h(x, y)$	Point spread function typically given by the Gaussian function with the radius of the blur function σ^2 as the variance.
$h(x, y, x', y')$	Space varying point spread function.
$W(t)$	Brownian motion in time.
$\kappa(x, y, t)$	Curvature of $u(x, y, t)$.

n(x,y,t) Normal of $u(x, y, t)$.

$d(x, y)$ Depth in the scene at each point (x, y).

G Bandwidth matrix that determines the window width at each data point.

g Scalar value that specifies window width when the bandwidth matrix is assumed to be identical at all data points.

$k(x)$ Bounded kernel function with compact support used in kernel density estimator.

w_i Label of depth at a point.

$\phi(D|w_x)$ Data likelihood function of the depth at a location x being given the label w_x.

$\psi(w_x, w_y)$ Pairwise interaction potential between labels w_x and w_y.

Chapter 1

Introduction

The use of computer vision is increasingly pervading everyday life with applications in entertainment, security systems, industrial automation, medical imaging, biometrics and information extraction. The increasing use of digital cameras in everyday life poses fresh challenges and interesting aspects that need to be addressed. There are a few fundamental problems that are always relevant in the field and form the basis for the development of various applications. They are problems like structure recovery from images, image enhancement, object recognition, classification and tracking, rendering of novel images. In this thesis we address the first two problems, that of structure recovery from images and image enhancement.

1.1 Depth Estimation

Given a scene, humans have a tremendous ability for discerning the geometric relationship among objects in the scene. While exact distance maps may be difficult to perceive, it is definitely possible for a person to be able to estimate the relative nature of geometry of objects in the scene in terms of ordering of the objects. In doing so humans make use of various cues, both geometric and photometric. Based on these cues there has been a substantial amount of work done on building systems for estimating depth. The geometric cues used are based on principles of stereo vision and motion. The photometric cues are those based on shading, texture and focus/defocus. The basic principle underlying the use of cues for shape estimation from images is that the depth in the scene is encoded implicitly based on those cues and by using these cues it is possible to analyze the images and obtain the depth.

In depth estimation by geometric cues it is the geometric relationship among various ob-

jects in the scene that encode the depth. In stereo based vision two images of a scene are taken from two different viewpoints (i.e. cameras placed at two separate locations). There is a relative disparity between the locations of the objects in the images. This disparity is related to the depth in the scene. Hence, there is larger disparity for objects closer to the camera and less disparity in the locations of the objects in the images that are far away from the camera. Based on this principle it is possible to obtain the depth in the scene. It is based on this two-view principle that humans predominantly perceive depth. Structure from motion follows a similar principle where a moving object is captured and the relative movement offers an estimate of depth in the scene. While geometry offers a strong cue, it is possible to obtain depth by considering photometric cues as well.

In depth estimation by photometric cues one uses the depth based variations in shading, texture and focus. In shape from shading, given an image of a scene with a known light source and given a reflectance model one uses the variation of shading to estimate the shape in the scene. In shape from texture, instead of assuming that the reflectance of the scene is known or constant, one assumes that the deformation of individual texture elements in the scene is due to projective deformation caused by the variation in orientation and shape in the scene and this is used to recover the depth in the scene. These photometric cues due to the assumptions are fairly limiting in nature. An additional cue that is available is in terms of the defocus generated by the lens. This photometric cue is more widely usable as it is affected by the camera parameters and does not impose limiting assumptions on the scene. Hence, in this thesis we explore the use of this cue in depth estimation.

The fundamental principles of optics affect much of what we see or capture using a camera. The human or physical lens deflects the light rays and the observation is captured on the imaging plane or retina. The observation captured is determined by the geometric properties of the lens. The objects that are around the focal plane are captured sharply and those not in focus are defocused in the observation. Hence by estimating the amount of defocus blur it is possible to estimate the depth in the scene. This is an innovative use of the limited depth of field of the camera. While, conventional depth estimation methods require multiple cameras or necessitate human motion, depth from focus/defocus methods require only a single camera but with multiple camera parameter settings. This is particularly applicable in scenarios like occular medical imaging and industrial automation in constrained environments. An interesting application is for depth estimation of intensely textured objects like hair and artifacts where stereo matching

would lead to multiple spurious matches. There are also scenarios where the baseline available for stereo is very narrow and in such scenarios changing the lens parameters are more feasible.

In the area of depth from focus/defocus there have been two main approaches. The first approach is depth from focus where many images are obtained by changing the lens parameters and in which each part of the scene imaged is in focus in at least one of the observations. Then based on the lens to image plane distance and the thin lens equation the depth in the scene is estimated. Another approach is depth from defocus where the defocus blur in the scene is estimated between two (or more) observations and the relative change in defocus blur is used to estimate the depth in the scene.

1.2 Image Enhancement

While we have seen that the presence of blur can be used as an advantage in terms of obtaining the depth map, it is often an unwanted artifact in usual imaging scenarios. Due to real apertures present in physical imaging systems, if an object is not in the focal plane then the resultant observation is blurred. This blur would be space invariant if the objects are at the same depth and space variant (which is the more usual case) if the objects are at different depths. The deblurring of images is an ill-posed problem. This is because the process of blurring is filtering an image by low-pass filtering. Hence in order to undo the effects of the filtering one has to use inverse filtering which is an unstable process resulting in noise boosting and high frequency distortions. In the general case there would also be noise present in a blurred image. Hence the general problem of image restoration would be to deblur and denoise a given image. The problem is formulated as

$$Y(x,y) = \int\int U(t,\tau)h(x,y;t,\tau)dtd\tau + N(x,y). \tag{1.1}$$

Here $Y(x,y)$ is the observation, $U(x,y)$ is the original undegraded image, $h(x,y;t,\tau)$ is the space varying blurring kernel and $N(x,y)$ is the noise. The noise $N(x,y)$ is often assumed to be additive white Gaussian noise.

While eqn.(1.1) is the general formulation of the restoration problem there are several specific cases that are explored in the literature. The first case is the case where the point spread function $h(x,y)$ is known. In this case the problem is one of deconvolution. A more challenging case is where the point spread function is not known. This problem is termed as

blind deconvolution. In most cases the kernel is assumed to be shift invariant. The problem of blind space varying deconvolution is a very difficult problem where one has to restore the image without knowing the kernel. The problem of deblurring is usually solved in the absence of noise. The presence of noise makes the deblurring process harder because the process of deblurring increases the high frequency information and the process of denoising reduces the high frequency information. Hence the joint deblurring and denoising problem is a very challenging problem. We discuss further in chapter 2 the approaches used in the literature for addressing the problem and we propose new techniques in the thesis for solving the general problem of image restoration. We now discuss the actual model of image formation and defocus generation.

1.3 A Real Aperture Imaging Model

Consider the image formation process in a real aperture camera employing a thin lens [19]. 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. This is illustrated in Fig. 1.1. 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 [19]

$$\sigma = \rho r v(\frac{1}{F} - \frac{1}{v} - \frac{1}{Z}) \tag{1.2}$$

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 ρ is a camera constant that depends on the sampling resolution on the image plane. From the eqn.(1.2) 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.

1.4 Diffusion based modeling of defocus blur

The process of blurring an image can be obtained as a convolution with a point spread function. The point spread function used to model defocus blur is the Gaussian kernel. This model of defocus blur enables us to model the defocus blurring process in terms of the heat equation. Koenderink [62] had shown that the Gaussian convolution model that is used in the context of

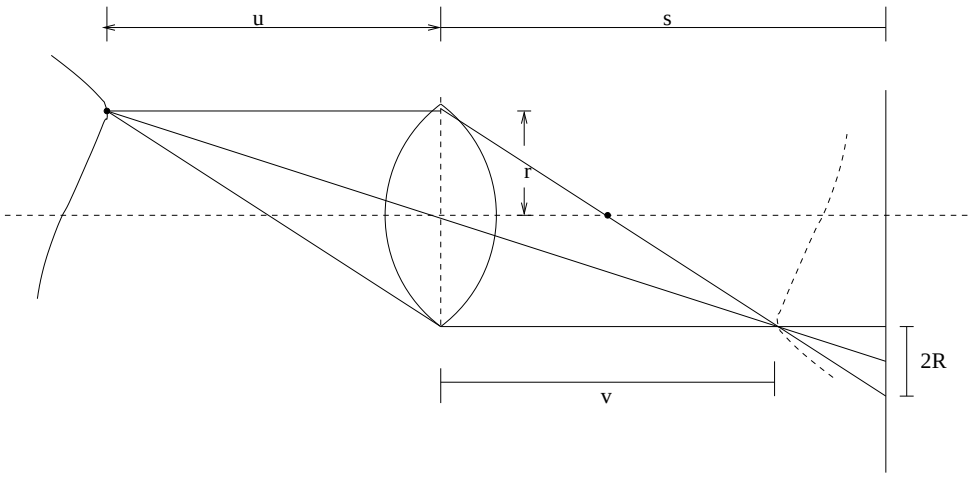


Figure 1.1: Illustration of image formation in a convex lens.

scale space is equivalent to diffusing an image using the heat equation given by

$$\frac{\partial u(x,y,t)}{\partial t} = c\triangle u(x,y,t) \tag{1.3}$$

where $u(x,y,t)$ represents the intensity of an image $I(x,y)$ and $\triangle u$ is the Laplacian operator. This procedure can be modeled by the heat equation. Here, the amount of blurring of an image is given by the coefficient of diffusion c. Since, in general the depth in the scene varies, the diffusion would also correspond to diffusion in an inhomogeneous medium with the differing diffusion coefficient corresponding to the various depths in the scene, i.e. instead of a space invariant diffusion coefficient c we have a space varying diffusion coefficient $c(x,y)$. In this thesis, we propose various methods for estimating the diffusion coefficient in order to estimate the depth in the scene. Given two images of a scene, there are two approaches for estimating the depth in the scene. In the first approach we diffuse an image from the more focused image to the defocused image (it is possible to distinguish these using simple techniques). In doing so, the diffusion coefficient can be estimated. The second approach uses reverse diffusion to go from a focused image to its pin-hole equivalent observation. This is a difficult technique due to the unstable nature of the reverse diffusion procedure. In our thesis we pursue both these techniques for estimating the inhomogeneous diffusion coefficient by successfully solving the various challenges posed in using both these techniques. The details pertaining to each of these are discussed in later chapters.

The problem of space varying blurring of an image is also considered from the restoration perspective. The space varying nature of blur is quite challenging especially when the model of the blurring procedure is not known. Using the diffusion formulation we propose inverse

diffusion procedures which reverse the blurring procedures and use specific stopping criterion in order to obtain the appropriate deblurred observations. While this procedure suffices for deblurring an image in the absence of noise, it fails when the blurred observation is also corrupted with noise. We consider these cases by using mixed diffusion. In textureless regions which contain mainly noise, we use forward diffusion and in regions which have strong edges which are blurred we use reverse diffusion. This approach enables us to solve the challenging problem of blind space varying deconvolution, including the case of deconvolution in the presence of noise.

1.5 Contributions of the thesis

In this thesis we address two related but specific and challenging problems, the first being depth estimation from defocused images and the second being blind space varying deconvolution. We formulate a principled approach towards addressing both these problems using an underlying framework based on the heat equation, i.e. diffusion formulation. While, there have been specific instances of works which have used a similar formulation, there has been no sustained effort at addressing the problems in their entirety as has been done here. In this thesis, we have explored spectral as well as spatial formulations of defocus modeled using the diffusion equation. The deterministic and stochastic approaches have both been used for evolution of the heat equation. We have studied both forward and reverse diffusion procedures and have also considered the issue of stopping criterion which is important in the reverse diffusion procedure. Since, in real images there are several regions that are textureless, we require effective regularization to solve the problem. This has been considered by using a MAP-MRF framework. While there have been several methods which use multiple images for depth estimation from defocused image, there has been no method for depth estimation from a single defocused image. Here we show that depth can be perceived even from a single defocused image. Summarizing, the various contributions of the thesis are as follows:

- We have used the linear diffusion principle for depth estimation from two (or more) defocused images. The use of linear diffusion was done by generating a defocus space of images using the defocus morphing principle. The process of morphing is done locally using the windowed Fourier transform. In the spectral domain the defocus morphing technique explains how an in between observation can be synthesized from two other ob-

servations using interpolation in the frequency domain of the signals where the spectral coefficients are raised to the interpolating factor (in a manner similar to root filtering). We also show that an extended defocus space can be generated where one can extrapolate the defocus blur beyond the range of defocus blur in the two observations. The extrapolation is bounded on one end by the pin-hole observation. The other end eventually converges to the average signal value. We use a band-pass filter to detect the occurrence of the pin-hole observation. We use the detection of the synthesized pin-hole observation as a means for estimating the depth in the image. As a by-product we also obtain the pin-hole equivalent observation. The method illustrates the equivalence of the depth from focus and depth from defocus methodologies. The drawback of this approach is windowing effect, ill-posed nature of extrapolation especially in textureless regions and limited accuracy.

- The drawbacks of the spectral approach can be addressed by considering the diffusion equation in the spatial domain itself. This was shown by Favaro *et al.* [32] in their work on depth estimation using diffusion. There were however two drawbacks that were not addressed. The first was that there are specific cases of self occlusion and aperture aberrations which result in departure from the Gaussian assumption of the point spread function. The second point is that in their method Favaro *et al.* have used gradient descent scheme which allows only convex minimization of the function. However, since in general the structure of the scene is not convex, it is possible that the local minimization approach provides convex approximate results. These points are addressed by formulating a stochastically perturbed diffusion scheme. The stochastic perturbation implicitly handles the departure from the Gaussian assumption. The stochastically perturbed diffusion is evolved using a stochastic level set technique which ensures that the non-convex nature is handled appropriately. This method is shown to be quite useful especially in handling depth estimation of complex objects like hair and fur. However, this method estimates the depth at each pixel independently without incorporating the neighborhood information. As a result, in textureless areas the depth estimated is ambiguous.

- In order to solve the problem of non-unique depth estimates in textureless scenes, it is required to incorporate spatial regularization. Towards this end, we use a Markov random field (MRF) framework for representing the diffusion coefficient. Here, the Markov random field network is connected pairwise to form a grid that takes into account the spa-

tial neighborhood relationship. The solution of pairwise MRF is done using graph cuts framework. The α expansion algorithm is used for estimating the variance of the point spread function which corresponds to estimating the diffusion coefficient and thereby the depth in the scene. This representation also allows a symmetric procedure for estimating the diffusion coefficient, i.e. it is not required to preprocess images to classify the focused and defocused regions in an image. In this framework it is possible to use robust regularization functions, including truncated functions of the Huber kind and total variation regularization. The resulting depth estimates are more accurate.

- The methods considered so far used the diffusion equation in the forward direction. An interesting aspect is to consider the reverse diffusion equation. While, the forward diffusion blurs an image the reverse diffusion would ideally deblur an image. However, due to the ill posed nature of the diffusion equation it is not possible to directly use the reverse diffusion. We therefore consider an approximation using the gradient based cluster separation. The forward counterpart is commonly known as the mean shift technique and has been demonstrated for denoising an image. We propose the reverse mean shift technique for deblurring an image. Since reverse mean shift is divergent we suggest an appropriate stopping criterion based on cluster separation factor to terminate the divergent procedure in order to provide the deblurred solution. While this technique can be used to achieve blind space varying deconvolution in the absence of noise, it cannot be used in the presence of noise. This is because due to the opposing nature of deblurring and denoising, the deblurring procedure accentuates the noise in the observation. To address this we suggest the use of mixed diffusion. The switching between forward and reverse mean shift techniques is done locally using an optimal criterion. The criterion correlates the residual with the image to decide whether forward or reverse mean shift technique is to be used and it achieves the desired result of denoising and deblurring of the image.

- While the generalized mean shift technique approximates the deblurring of an image, it is not exactly equivalent to the reverse heat equation. We next consider the problem of stabilizing the reverse heat equation itself. Instead of using non-linear schemes to approximate the linear reverse heat equation, we show here that it is possible to geometrically stabilize the reverse heat equation, thereby solving the problem in the linear case itself. The heat equation comprises of a tangential term and a normal term. The normal term of

the heat equation contributes more to the blur in an image than the tangential term. This factor suggests that the differential evolution of the normal and tangential term might be beneficial. We therefore damp the evolution of the normal term. As a result at the pin-hole equivalent observation point the tangential term degenerates whereas the normal term would not have degenerated. We therefore use the degeneration of the tangential term as a stopping criterion for halting the reverse heat equation and obtain the deblurred result.

- The success in the use of reverse heat equation for deblurring of an image allows us to consider the extremely challenging problem of depth estimation from a single defocused observation. There have been other techniques that have considered the depth estimation from defocus for multiple images, but depth estimation from a single defocused image was so far not thought to be possible. Indeed, depth estimation from a single image itself has mostly eluded vision researchers except for a very few efforts which either use learning or specific information like epipolar points (which require some human intervention). We show that using the diffusion coefficient obtained using the reverse heat equation as a data likelihood it is possible to perceive the relative depth in the scene. The diffusion coefficient is valid mostly near the edges and it is not valid in textureless regions. Hence we frame the problem in an MRF framework with the data likelihood weighed with the edge information. This framework is solved efficiently using graph cuts and as a result the relative depth in the scene is obtained. This clearly establishes that defocus is a strong cue.

1.6 Organization of the thesis

In the next chapter we consider the related work done till date in detail. We especially review the literature related to depth estimation from defocus/focus and image deblurring problems. The challenges that these problems present are evident from the various techniques that have been proposed towards solving these problems. In chapter 3 we describe the use of linear diffusion principle for simultaneously estimating the depth and obtaining the pin-hole equivalent image using the defocus morphing principle. In chapter 4 we present the use of stochastically perturbed diffusion for depth estimation from two (or more) defocused images. In chapter 5 we discuss the recovery of regularized depth from multiple defocused images using graph cuts.

While in chapter 3 we used a reverse diffusion technique to estimate the depth, in chapters 4 and 5 we use only the forward diffusion technique. We again consider the reverse diffusion principle, but now consider its application for deblurring an image in chapter 6. This chapter also discusses solution of the general problem of deblurring and denoising of an image. The technique used in this chapter is a non-parametric one based on the mean shift procedure. In chapter 7 we propose a geometrically stabilized reverse heat equation for image restoration. In chapter 8 we culminate with the significant result of the thesis where we present a technique for depth estimation from a single defocused image using the reverse heat equation in a MAP-MRF framework. We finally conclude in chapter 9 where we also consider the possibilities for future work.

Chapter 2

Literature Review

In this chapter the literature related to depth estimation is first reviewed and then the related work done in image enhancement is considered. The literature reviewed here gives an overview of the various methods used towards solving these problems. We discuss specific methods very closely related to the contributions of the thesis in the corresponding chapters.

2.1 Depth estimation

The primary consideration in depth estimation from images has been mostly on the stereo cue and the related cue of structure from motion. These cues are geometric in nature and there has been substantial work done as evident from the works by Ma *et al.*[74], Hartley and Zisserman [49], Faugeras [29] and Pollefeys *et al.* [92, 93, 73]. While geometric cues have been extensively researched by the vision community, there are photometric cues that are also relevant for depth estimation. These cues include shading and defocus. While stereoscopic cues rely on two (or more) cameras, the photometric cues can be used to infer depth using a single camera but with different parameter settings. These cues explain the basis on which people who have defect in a single eye are still able to perceive depth. Hence they are relevant biologically and also practically as there are cases where only a single camera can be used or where the baseline distances are large that makes the stereoscopic depth perception erroneous. In this thesis we consider the cue of defocus blur for estimating shape in the scene. The methods for shape from texture are discussed by Forsyth and Ponce [38] and recent work has been done by Forsyth [37]. Shape from shading techniques are discussed by Horn [55] and by Forsyth and Ponce [38]. Recent work in shape from shading has been done by Prados and Faugeras [94]. We now consider

the defocus blur based literature in detail.

2.1.1 Depth From Defocus (DFD)

The basic problem addressed in the depth from defocus methodology is the measurement of the relative defocus between observations. Broadly speaking, the approaches have been based on active and passive techniques.

Active Techniques

The active method for depth from defocus was suggested by Girod and Scherock [43]. This was further refined by Pentland *et al.* [89]. In active depth from defocus, a known pattern of light is projected onto the scene. The optical path of the projected pattern is maintained same as that of the observing camera using an optical beam splitter. The structured light source projects a pattern on the scene while the camera records it. Since the original projected light source is known, the defocus blur introduced by the depth in the scene can be measured against the original pattern and the blur introduced due to defocus and thereby depth in the scene is measured. This basic technique for active depth from defocus was refined by Noguchi and Nayar [85] for microscopic shape from defocus and Nayar *et al.* [83] and they proposed a real-time focus range sensor. In these works the illumination pattern to be used for depth estimation is analyzed in detail. They performed careful Fourier analysis of the various aspects of the defocus, illumination pattern and the focus operator used for measuring the focus, and thereby optimized the illumination pattern. Further, based on the illumination optimization and the need for dense depth estimation, the focus operator is tuned to enhance the depth estimation. These refinements were used to develop a robust real time focus range sensor [83]. In [1] the authors consider the integration of stereo disparity, focus and vergence cue to build an active vision system for surface reconstruction. The performance analysis of such a system was carried out in [25] by the authors. There has been very recent work done by Moreno-Noguer *et al.* [76] where the earlier work done on active depth from defocus is revisited. Here they use this method to demonstrate active refocusing of images and videos. In this work a dot-pattern is used instead of a stripe pattern so that the dot-pattern projected can be easily removed while refocusing the image. The authors also use segmentation techniques for obtaining a dense depth map. A similar work has been done by Zhang and Nayar [132]. Here they use the projector for

projecting a shifting light pattern and a set of images of the scene is taken. The depth of a pixel is computed by analyzing the temporal variation of the brightness of the pixel due to defocus. While these works demonstrate the efficacy of active depth from defocus, the main drawback with the methodology of active illumination based methods is that they require elaborate set-ups and can be done only in carefully controlled environments with calibrated illumination conditions. These requirements preclude the use of these techniques in outdoor environments and natural conditions.

Passive Techniques

There has been considerable research done on the passive method of depth from defocus. In the passive method, two or more observations of a scene are taken with different camera parameters. Based on the difference in defocus blur between the two observations, the depth in the scene is estimated. This method for estimating depth from defocus was introduced by Pentland [88]. He suggested the use of a half mirror for splitting the observation and then passing the observation through two lenses with varying apertures. Then for each image one could use a patch based Fourier transform and based on the ratio of the windowed Fourier transforms one could obtain the relative defocus. If one of the images is focused then one could obtain the depth estimate from the relative blur. An additional insight that was presented in this work is that one could use step edges and obtain a crude estimate of the blur function at certain sharp edges. He further explored the use of defocus cue by humans for estimating depth in terms of human perception related experiments justifying the use of defocus cue as a depth cue similar to motion or stereopsis. Around the same time Grossman [46] presented a method for depth estimation based on defocus along similar lines. The edge detection was done using Marr-Hildreth edge operator and the blur was measured using a simple metric based on the width of the edge. The method was tested using simple structured objects with well-defined edges. These methods were subsequently extended by Subbarao. In [120] Subbarao and Gurumoorthy considered the method of blur estimation from blurred edges in a manner similar to that done earlier by Pentland [88] and Grossman [46]. They proposed a closed form solution for estimating the blur parameters and extended the earlier assumption of the point spread function (PSF) from being Gaussian to the more general case of being rotationally symmetric. In [118], Subbarao considered the usual case of depth estimation from defocus by changing the camera parameters. The contributions were in making the earlier method by Pentland [88] more robust and in allowing changing of

more than one camera parameters simultaneously. The methods based on estimation of blur from step edges were too restrictive and could not be widely used. The approach for depth from defocus based on using two or more images from different camera parameters was further explored. In the earlier cases, ratio of windowed Fourier transforms were taken for estimating the depth in the scene. This approach was refined by Gokstorp [44]. In the work a local frequency representation was adopted wherein local estimates of instantaneous frequency, amplitude and phase were computed. For obtaining the representation a set of complex-valued Gabor filters are used where a large set of such filters are used each tuned to a different frequency. The final estimate of blur is computed by averaging the result from the various filters. A similar method was also proposed by Xiong and Shafer [130]. They proposed moment filters to address the problem of tuning the individual frequency components. The property of moment filters in that it is a polynomial approximation with the order of the polynomial being tuned for obtaining a large number of narrow band filters. Due to the recursiveness in frequency and spatial domain moment filters can be implemented efficiently. While the methods based on narrow band filters are attractive, typically it is computationally intensive and to a large extent the accuracy is a function of averaging over a large number of such filters. A different approach was adopted by Watanabe and Nayar [124]. Here the authors suggest the use of broadband rational filters. The technique uses a normalized ratio of near and far images, that are then convolved with a small number of broadband rational filters that are invariant to the texture. The method provided impressive results. However, a shortcoming of the method is its dependence on the assumption, that, the normalized ratio of the near and far focused images is linear in nature. This assumption is valid only when the amount of defocus between the near and far images is very small and will not work for a general class of defocused images.

We now consider methods based on the spatial domain. In [27], the authors suggest a method for estimating depth from defocus using a matrix based approach. They propose the estimation of the transfer function from the near image to the far image using a regularization based approach which is regularized by a circulant matrix that detects the smoothness of a function and a diagonal matrix that penalizes the transfer function for having nonzero tails. This equation is solved iteratively. They show that their method performs better than previously proposed inverse filtering solutions. An approach for depth from defocus using spatial deconvolution was proposed by Subbarao and Surya [121]. This is achieved by using an operation termed S-Transform. The method assumes a cubic polynomial form of an image and based on

this model derive a formula wherein convolution and deconvolution can be expressed in terms of a simple Laplacian operator modulated by the second moment of the PSF. The resultant depth maps obtained were however not very accurate. A related work by Zious and Deschenes [133] proposes a technique based on image decomposition using the Hermite polynomial basis. The resultant computation suggests the explicit use of higher-order polynomial fitting. The blur is computed from from image derivatives with the base case being similar to the S-Transform proposed by Subbarao and Surya [121].

There has been substantial work done by Rajagopalan and Chaudhuri towards solving the problem of depth from defocus [19]. They first explored a block shift variant circulant blur model for estimating the relative blur between two defocused images in the Fourier domain [98],[99]. The contribution here explicitly considered the contribution from neighboring blocks. The authors then considered a space-frequency representation using complex spectrogram and Wigner-Ville distribution [99]. They used these representations to estimate the relative blur. The authors then incorporated a smoothness constraint that incorporates the blur information in the neighborhood and solves the problem using calculus of variation [100]. The authors then used a maximum likelihood (ML) estimator for recovering the depth from two defocused images of a scene [95]. They also analyzed the effect of the degree of relative blurring on the accuracy of the estimate of depth and they calculated the optimal camera setting for recovering depth using the Cramer-Rao bound. They then improved their ML estimator by proposing a method for recursive computation of the likelihood function [96]. They then modeled the relative blur parameter as a Markov random field (MRF) to represent the local dependencies of the depth [97]. Further on they showed how one could recover the depth as well as estimate the restored image by modeling both as separate Markov random fields and estimating the parameters using simulated annealing [102].

Lately the problem of depth from defocus has been actively explored by Favaro and Soatto [35]. We now consider the specific approaches used by Favaro and Soatto in solving the problem of depth from defocus. In [34] the authors consider two specific cases, one where the form of the PSF is known and the other when the form of the PSF is unknown. When the form of the PSF is known the authors use a least squares solution in solving the projection between finite and infinite dimensional Hilbert spaces and is achieved by estimating a set of orthogonal operators. They use functional singular value decomposition for estimating the operators. The values are truncated beyond a certain singular value and this results in regularization being in-

corporated into the solution. If the form of the PSF is not known then they first learn the set of projection operators from blurred images. They then use the learnt set of projection operators to estimate the shape in the least square sense as done previously. In [31], the authors analyze the role of radiance, defocus and observability in a rigorous manner. In the previous approach [34], the least squares solution does not enforce the condition that the image is positive. In [31] the authors ensure that all the quantities are non-negative. They pose the problem of estimating depth from defocus as one of reconstructing the shape and the radiance that minimizes a measure of information divergence between blurred images instead of least squares. For solving this they use an iterative scheme that reduces the cost function with convergence to local minimum based on Euler-Lagrange equation. Recently, in [32] Favaro *et al.*, have used the idea of diffusion for estimating depth from defocus which is very similar to our work. However, there are a few shortcomings in the basic approach described by them as it does not address the problem of self occlusion appropriately. Further they adopt a convex gradient descent approach and the resultant depth estimate obtained is only a local minima approximation of the true scene depth. Very recently in [72], the authors consider the issue of calibration for estimation of depth from defocus and in [30], they explore the problem by taking into account issues related to convexity and using Bregman distances the authors provide means for estimating depth from defocus. Lately there has been a work by Hasinoff and Kutulakos [51] where they use a layer based restoration framework and are able to incorporate high dynamic range (HDR), focus and noise in a single framework and use alternating minimization to estimate each of these while holding the rest constant. These are addressed by us in the thesis. Besides this we are able to greatly extend the work by achieving depth estimation from a single defocused image. These points are expounded further in the thesis.

Most of the methods in depth from defocus literature assumed that the observations do not suffer from occlusion. However it was brought out in [112] that occlusion can occur in depth from defocus techniques as well. Occlusion effects have also been studied in [81]. The handling of occlusion effects in depth computation was addressed in [8] and [33].

2.1.2 Depth From Focus (DFF)

There are a number of papers in the literature which address the problem of obtaining depth information from focus. This includes work by Nayar and Nakagawa [82], and work by Subbarao and Choi [119]. The basic method followed has been to obtain different focus levels by adjust-

ing the camera parameters, i.e. either the lens to image plane distance v, the focal length f or the aperture radius r (cf. 1.1). The methods involve obtaining many observations for the various camera parameters and estimating the focus measure using various criterion functions. Krotkov [66] has experimentally evaluated several such criteria including the Laplacian and Teningrad operators. In [80] the authors discuss a method in which the blur is evaluated from the intensity change along corresponding pixels in the multi-focus images instead of using window-based blur estimation operators. Recently Hasinoff and Kutulakos [50] have shown that for very high resolution images the depth from focus can be seen to be reduced to color comparison with regions of an aperture-focus image representation for each pixel. The fundamental weakness of the DFF method is, however, the time required for image acquisition. In practice about ten or even more images are required to estimate the depth of a scene for a reasonable level of accuracy.

2.2 Image Restoration

The problems central to the task of image restoration are image denoising and image deblurring. In the thesis we address the problem of image deblurring both with and without noise. In the general sense, where the blur function is not known, the problem of image deblurring is one of blind deconvolution. The methods adopted towards solving these problem include regularized inverse filtering techniques, Bayesian methods, partial differential equations based techniques, frequency and wavelet based techniques and restoration using edge preserving regularization techniques. Extensive review of these techniques are presented in [9], [5], [67], [122], [4]. We now discuss only some of the related representative works. We review the PDE based literature in more detail as they are closely related to the work researched in the thesis. There exists a substantial number of works based on other methodologies like frequency based and learning based techniques, however, we do not undertake a comprehensive review of all these methods.

One of the commonly used techniques has been inverse filtering based on Wiener filtering. An approach that suggests an iterative Wiener filter specifically updating the covariance estimates is communicated by Hillery and Chin [52]. Due to the ill-posed nature of the problem there have been several approaches that adopt iterative regularized techniques. In [60] the authors propose an iterative constrained least squares formulation that converges using the method of successive approximations. The regularization methods proposed usually assume smooth

properties and do not preserve edges. An approach was proposed by Chan and Wong [17] which used total variation regularization for estimating the point spread function and also deblurring the image. The use of total variation allowed preservation of edges. Other approaches have proposed use of wavelets for regularizing image deconvolution [78], [84]. In [36] an EM algorithm is proposed for wavelet based image deconvolution. In [117], the authors propose a technique for combining curvelets and wavelets for image deconvolution.

The methods described so far solve the problem of blind deconvolution when the point spread function (PSF) is assumed to be the same throughout the image. The literature is relatively sparse when one considers the problem of spatially varying PSF. Rajagopalan and Chaudhuri [101] have solved the corresponding restoration problem using a Markov random field (MRF) prior. However, their method is strictly speaking, non-blind as they assume availability of multiple observations. The other methods that can be used for space varying deconvolution are those based on use of partial differential equations. A classical approach for this problem is that of using hyperbolic partial differential equations termed as shock filters. The shock filter was first proposed by Kramer and Bruckner [65]. It is based on the idea of using a dilation process near a maximum and an erosion process around a minimum. The decision whether a pixel belongs to the influence zone of a maximum or a minimum is made on the basis of its Laplacian. The term *shock filter* was first introduced by Rudin [108]. The experimental shock filter by Rudin was based on a modification of the nonlinear Burgers' equation. This model was further improved by Osher and Rudin in [87], where the total variation preserving computational approach and the theoretical basis for the same was developed. Further, a modification was suggested by Alvarez and Mazorra [2] where they incorporated a smoothing kernel in the model. The relationship of these methods to the Kramer-Bruckner filter became evident later ([47], [111])

The recent work in this field includes work by Kimmel *et al.* [61], Weickert [126], Gilboa *et al.* ([40], [42], [41]) and Remaki and Cheriet [104]. In [61], Kimmel *et al.* have developed a shock filter based on a geometric framework and the inverse diffusion is carried out along the edge. In [126], Weickert describes a coherence enhancing shock filter where the shock filter is steered with the orientation information. In [40] by Gilboa *et al.*, the authors have modified the diffusion coefficient in the Perona-Malik formulation [90] and they use a diffusion coefficient which switches adaptively between forward and backward diffusion process. In [42], Gilboa *et al.* extend the work done in [40] and define a triple-well potential based diffusion process which

is an energy minimizer flow aimed at reducing oscillations among three low energy states. In [41], Gilboa *et al.* suggest complex shock filters based on the complex diffusion process where the diffusion coefficient lies in the complex domain. In [104], the authors consider the problem of shock filters in the framework of generalized functions and propose shock filters where the speed of shock propagation is also controlled. In the thesis we propose two new techniques for image restoration. They are compared with these techniques and have shown comparable or better performance compared to these techniques.

In this chapter we have considered the literature related to depth estimation and image restoration in considerable detail. We now consider these problems in detail and propose diffusion based techniques that solve these problems specifically for defocused images.

Chapter 3

Linear Diffusion

In this chapter we establish the use of the diffusion process for modeling the blurring process. The idea of diffusion has been one of the important methodologies in the field of computer vision. It stems largely from the idea of modeling the image (observation) generation process using the heat equation. The pioneering work was done by Witkin in [129] where he proposed a scale space for images based on smoothing of images using a Gaussian kernel. Koenderink in [62] proved that this was equivalent to solving the heat equation. This approach has subsequently been widely used in low level vision tasks like smoothing, segmentation and edge detection.

Here we discuss how the linear diffusion principle can be used for depth estimation based on defocus as the cue. This was first explored by us in [79]. In depth estimation using defocus as the cue, the basic principle is to use the characteristics of the imaging system. There have been two methodologies in the literature, one is to obtain depth from focus [66] and the other to obtain depth from defocus [19].

In the procedure for obtaining depth information from focus, a sequence of images of a scene is obtained by continuously varying the distance between the lens and the image detector [119]. The corresponding fully focused observation is locally estimated from the sequence of images. A measure of image sharpness is used to decide whether the point is in focus or not. From the fully focused image point the distance of the corresponding object point is calculated using the standard thin lens equation

$$\frac{1}{F} = \frac{1}{Z} + \frac{1}{v} \tag{3.1}$$

where F is the focal length, Z is the distance of the object from the principal plane and v is the

distance of the focused image from the lens plane. For a good ranging accuracy, one is required to sample the observation space densely by changing v slowly.

Consider a planar scene with the plane perpendicular to the optical axis. When the image plane is kept at the focal point then all the points from the planar scene are captured exactly in focus. However, when the image plane is not in focus, the observation is blurred by a circular patch as can be seen from fig. (1.1). This is called the equifocal observation. In general there is variation in depth in the scene. Hence the resultant observation is not equifocal. The blur in the observation thus varies spatially. In depth from defocus, given two images of a scene recorded with different camera settings, one obtains an estimate of the blur at each point [19]. Subsequently, by using the estimate of the blur, one can recover the depth information in the scene with the knowledge of the lens parameters.

In this chapter we show that, given two observations obtained by two sets of lens parameters as is commonly employed in depth from defocus, we can generate the entire set of images in the *defocus space* of the input images using the diffusion equation. The defocus space of a particular scene refers to the continuous space of all possible observations obtainable by varying the lens parameters in between those two lens settings. This concept is further elucidated in section 3.2.3. In this method, the defocus blur is never explicitly calculated as it is done in depth from defocus techniques. Instead, by using diffusion, for each pixel we can obtain the corresponding fully focused observation in the defocus space. using that observation and the corresponding virtual lens parameters we can recover the depth information from the lens equation. As a by-product, we also obtain the fully focused pin-hole image from these two defocused observations. The diffusion process simulates the depth from focus technique by generating images in the defocus space of the observation. Many separate observations as required for the depth from focus technique are no longer required. In fact, using the diffusion technique, the two modalities of estimation of depth can be considered to be equivalent. This is discussed further in section 3.2.4.

An extensive review of the literature for estimating the depth from focus and the depth from defocus was done in chapter 2. In the next section we give a brief overview of the related work done. In section 3.2 we outline the theoretical basis for the formation of the defocus image space of an observation based on the diffusion process. In section 3.3 we present the basic algorithm for depth estimation using diffusion. In section 3.4 we analyze the procedure and consider the practical issues involved in the implementation of this method. In section 3.5

we present the experimental results obtained. We conclude the chapter in section 3.6.

3.1 Related Work

In this chapter we use the technique of diffusion for synthesizing new and virtual observations in the defocus space. The idea of diffusion can be traced to that of scale space filtering by Witkin [129]. Koenderink [62] showed that this is equivalent to solving the heat equation. Since the solution of the heat equation is a temporally evolving Gaussian function, filtering a signal with which defines the scale space. This scale space approach was extended by Perona and Malik in their landmark paper [90] where they proposed a nonlinear scale space model, aimed at preserving important features such as edges. The model changes its behavior based on the conduction coefficient associated in a region of an image and achieves forward diffusion in the interior region and at the boundaries it acts in the opposite direction. In general the inverse diffusion approach can be thought of as reversing the heat equation in time. This reverse heat equation is however ill-posed and there has been a substantial amount of work done for stabilizing the reverse heat equation. Rudin, Osher and Fatemi in [109] introduced the "shock filter" where they proposed a pseudo-inverse, where the inverse diffusion propagation term is tuned by the sign of the Laplacian. There has been a lot of research done along similar lines where various nonlinear inverse diffusion models have been proposed. In linear scale space theory, recently an interesting work has been done by Lindeberg [69], where he provides a theoretical analysis of the linear scale space theory and also observes that Gaussian and higher orders of the Gaussian kernel are the only admissible kernels based on the admissibility conditions for linear scale space.

Recently, in [32] Favaro *et al.*, have used the idea of diffusion for estimating depth from defocus which is very similar to our work. However they approach the problem in the traditional manner of casting it into a variational framework. We have introduced a more general idea of generating the entire defocus space of a scene.

3.2 Defocus as a Diffusion Process

3.2.1 Diffusion Process

The diffusion equation is given by

$$\frac{\partial u(x,y,t)}{\partial t} = \nabla \cdot (\mathbf{c}(x,y)\nabla u(x,y,t)) \tag{3.2}$$

where $\nabla u(x,y,t)$ is the gradient and $\nabla\cdot$ is the divergence operator, and $\mathbf{c}(x,y)$ is a vector diffusion coefficient. This gives the general case for the anisotropic diffusion equation as given by Fick's law that is derived from the equilibrium property that restores concentration differences [127]. For the specific case where the diffusion coefficient is a scalar and homogeneous then it is given by

$$\frac{\partial u(x,y,t)}{\partial t} = c\triangle u(x,y,t)), \tag{3.3}$$

where $\triangle u(x,y,t)$ is the Laplacian of $u(x,y,t)$. The equation above is the classical equation for the isotropic diffusion of heat and can be given in extended form by the following partial differential equation:

$$\frac{\partial u(x,y,t)}{\partial t} = c\left(\frac{\partial^2 u(x,y,t)}{\partial x^2} + \frac{\partial^2 u(x,y,t)}{\partial y^2}\right) \tag{3.4}$$

Here the constant c is the thermometric conductivity or diffusivity [128]. The equation above describes how heat diffuses over a surface, given an initial temperature distribution with time. It is assumed here that the diffusion of heat is uniform in all directions. Consider that $u(x,y,t=0)$ is an image $I_1(x,y)$. The solution of the heat equation can be obtained in terms of convolution of the image with a temporally evolving Gaussian kernel [114]. This is known as the source solution for the heat equation [128] and is given by $\sigma^2 = 2ct$ where σ denotes the spread of the Gaussian kernel and is used through out in this thesis. The observations in our case corresponds to $u(x,y,t)$ at two distinct time instants t_1 and t_2 where $0 \leq t_1 < t_2 < \infty$. As the image is progressively convolved with a Gaussian kernel, it gets increasingly more blurred thereby representing the image information at a different scale. Note that as $t \to \infty$ this corresponds to a fully diffused image. This is the basic idea underlying scale space analysis. Also note that the process is not defined for $t < 0$, a fact that will be utilized later to define the *extended defocus space*. Also note that we use an infinitely extended image domain while solving eqn.(3.4) and hence this would correspond to the Neumann boundary condition.

3.2.2 Basic Model of Defocus

Consider the image formation process in a real aperture camera employing a thin lens [19]. 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. Following geometric optics 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 [19]

$$\sigma = \rho r v \left(\frac{1}{F} - \frac{1}{v} - \frac{1}{Z} \right) \tag{3.5}$$

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 ρ is a camera constant that depends on the sampling resolution on the image plane. Let $I_0(x, y)$ be the pin-hole image of the scene. From the eqn.(3.5) 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.

The depth related defocus process is linear but not space invariant. Assuming a diffraction-limited lens system (i.e. using wave optics) and a constant depth in the scene (this assumption will be relaxed at a later stage), the point spread function (PSF) of the camera system at a point (x, y) may be approximately modeled as a circularly symmetric 2-D Gaussian function [19]:

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2} \right) \tag{3.6}$$

where the blur parameter σ is obtained from eqn.(3.5).

Note that some researchers have also used circular pillbox blur and each model of blur function has its own advantages and disadvantages. However, both models assume a perfectly circular aperture and no self-occlusion in the scene [8]. For a Gaussian model of blur σ is related to the spread of the PSF rather than the radius as defined in eqn.(3.5).

Assuming the depth to be constant everywhere, the observed defocused image $I(x, y)$ is given by

$$I(x, y) = I_0(x, y) * h(x, y). \tag{3.7}$$

This equation can be directly related to the solution of the diffusion equation in terms of the Gaussian kernel as discussed in section 3.2.1. The real aperture imaging can thus be thought of as providing a real world example of scale space theory. The eqn(3.7) can be represented by

taking its Fourier transform. Denoting the Fourier transform of a function $f(x, y)$ by $\hat{f}(\omega_x, \omega_y)$ we obtain

$$\hat{I}(\omega_x, \omega_y) = \hat{I}_0(\omega_x, \omega_y)\hat{h}(\omega_x, \omega_y) = \hat{I}_0(\omega_x, \omega_y)\exp(-\frac{\sigma^2(\omega_x^2 + \omega_y^2)}{2}) \tag{3.8}$$

3.2.3 Defocus Space

For a given scene, one can have two defocused observations E_1 and E_2 corresponding to two different camera parameter settings C_1 and C_2, such that the resulting blur parameters are σ_1 and σ_2, assuming $\sigma_1 > \sigma_2$ without loss of generality. For the two observations E_1 and E_2, a defocus space can be defined.

Definition 1 *(Defocus space) :*
The defocus space is defined to be the set of all possible observations E for a given scene generated by varying the blur σ as a combination of the associated blur parameter σ_1 and σ_2 in the two observations E_1 and E_2 respectively, by the following relation

$$\sigma^2 = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 \tag{3.9}$$

for all values of $0 \leq \alpha \leq 1$.

This is equivalent to generating $I(x, y, t)$ for $t_1 \leq t \leq t_2$ given the states $I(x, y, t_1)$ and $I(x, y, t_2)$ at two specified time instants t_1 and t_2 in the heat diffusion eqn(3.4). Substituting eqn(3.9) in eqn(3.8) we obtain:

$$\begin{aligned}
\hat{I}(\omega_x, \omega_y) &= \hat{I}_0(\omega_x, \omega_y)\text{exp}\left[-\frac{1}{2}(\alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2)(\omega_x^2 + \omega_y^2)\right] \\
&= \left\{\hat{I}_0(\omega_x, \omega_y)\text{exp}\left[-\frac{\sigma_1^2(\omega_x^2 + \omega_y^2)}{2}\right]\right\}^{\alpha}\left\{\hat{I}_0(\omega_x, \omega_y)\text{exp}\left[-\frac{\sigma_2^2(\omega_x^2 + \omega_y^2)}{2}\right]\right\}^{1-\alpha}
\end{aligned}$$

or

$$\hat{I}(\omega_x, \omega_y) = \hat{I}_1^{\alpha}(\omega_x, \omega_y)\hat{I}_2^{1-\alpha}(\omega_x, \omega_y). \tag{3.10}$$

The relation given in eqn(3.10) is equivalent to the notion of scale space as formed by the diffusion equation. This can be noticed as eqn(3.10) can be thought of as convolving the image $I(x, y)$ with a time varying Gaussian kernel. This is because convolving a Gaussian function with another Gaussian function always results in a Gaussian function. The eqn(3.10) effectively reduces to convolving the original image $I_0(x, y)$ with a Gaussian kernel which varies with time (in this case α) according to the relation given in eqn(3.9) and eqn.(3.10).

The defocus blur σ could be present physically due to any of the following camera parameters : aperture, the lens to image plane distance, the focal length or even a combination of these, as shown in eqn(3.5). A monotonic variation in any of the lens parameters can generally result in a non-monotonic variation in the blur (for instance as v is changed from an initial value, σ reduces, becomes zero and then subsequently increases), signifying both sides of the defocus cone (see Fig.3.1 for illustration). The diffusion based defocus space generation process however generates the blur in a monotonic manner, i.e we are restricted to one side of the defocus cone. By continuously varying the parameter α, we can generate any virtual observation for defocus setting lying between the lines AB and CD in Fig.3.1 using the eqn(3.10). The utility of such a variation of blur in defocus morphing has been demonstrated in [18]. The defocus space thus consists of all possible observations of the defocus blur $\sigma_1^2 \leq \sigma^2 \leq \sigma_2^2$.

Corresponding to the notion of continuous defocus space as introduced in the previous section, a practical counterpart of this defocus space would be a sampled defocus space. This corresponds to generating the defocus space for discrete values of α between 0 and 1. In Fig.3.1, the lines corresponding to A_1B_1,A_2B_2, ...A_nB_n may represent one such possible set of sampled defocus space. The sampled defocus space generated for an image is similar to the physically obtained focused image space described in [119].

3.2.4 Equivalence of DFF and DFD

So far we have considered a restricted range of α between $[0, 1]$. Now we relax this condition and something interesting happens. If the values of α beyond the range $[0, 1]$ are considered then the defocus space generated is the extended defocus space.

Definition 2 *(Extended defocus space):*
The extended defocus space is defined to be the set of all possible observations I for a given scene generated by varying the blur σ as a combination of the associated blur parameter σ_1 and σ_2 (assumed $\sigma_1 > \sigma_2$ in the two observations I_1 and I_2 respectively, by the following relation

$$\sigma^2 = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 \tag{3.11}$$

for all values of $\beta \leq \alpha \leq \infty$.

where β is

$$\beta = \frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \tag{3.12}$$

Here the value of β is the value of α such that $\sigma^2 = 0$ in eqn(3.11), resulting in a fully focused observation. The observations in the extended defocus space can be obtained from the diffusion equation since corresponding to the image $I_0(x, y, t)$ with $t \to \infty$ we can obtain an observation $I(x, y)$ with $\alpha \to \infty$. This represents the fully diffused image. Similarly for each point there exists a value $\alpha = \beta < 0$ corresponding to $t = 0$. This corresponds to a fully focused observation, i.e. $\sigma^2 = 0$. Thus the extended defocus space is defined for the range $\alpha \in [\beta, \infty)$. In the range $\alpha = [\beta, 0]$ the process, instead of being a diffusion becomes an inverse diffusion. Beyond this range, the defocus space is undefined since one cannot have the blur $\sigma^2 < 0$. Thus $\sigma^2 = 0$ corresponds to the convergence of all rays at the imaging plane. This is illustrated in Fig.3.1.

Depth from defocus(DFD) methodology estimates the space variant blur whereas depth from focus (DFF) methodology estimates the focused image point. It is possible to use the techniques in DFD methodology to estimate the space variant blur using just two observations, whereas DFF requires many samples to estimate the fully focused point. Here as we have shown, it is possible to generate the extended defocus space for the image using just two observations. Thus both the techniques can be considered fundamentally equivalent, rendering the need for multiple samples to be redundant. The diffusion based process thus provides an equivalent means for estimating the depth from the known lens parameters using either depth from defocus or depth from focus.

3.3 Algorithm for Depth Estimation

The derivation of eqn(3.10) is based on the assumption of constant depth. When there is depth variation in the scene, eqn(3.10) is no longer valid as the blurring process becomes shift variant, implying an inhomogeneous diffusion process. This corresponds to the following diffusion equation

$$\frac{\partial u(x,y)}{\partial t} = c(x,y)\left(\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2}\right) \tag{3.13}$$

Here c is no longer a constant but is now a function $c(x, y)$. We assume that the directional change in c i.e. $\nabla c(x, y)$ is assumed to be negligible. Here $c(x, y)$ is still a scalar quantity and the diffusion is isotropic but inhomogeneous and this equation varies spatially from eqn.(3.2). The inhomogeneity is handled by forming a small MxM window about a point over which the depth can be assumed to be constant as is done commonly in all literature. M is related

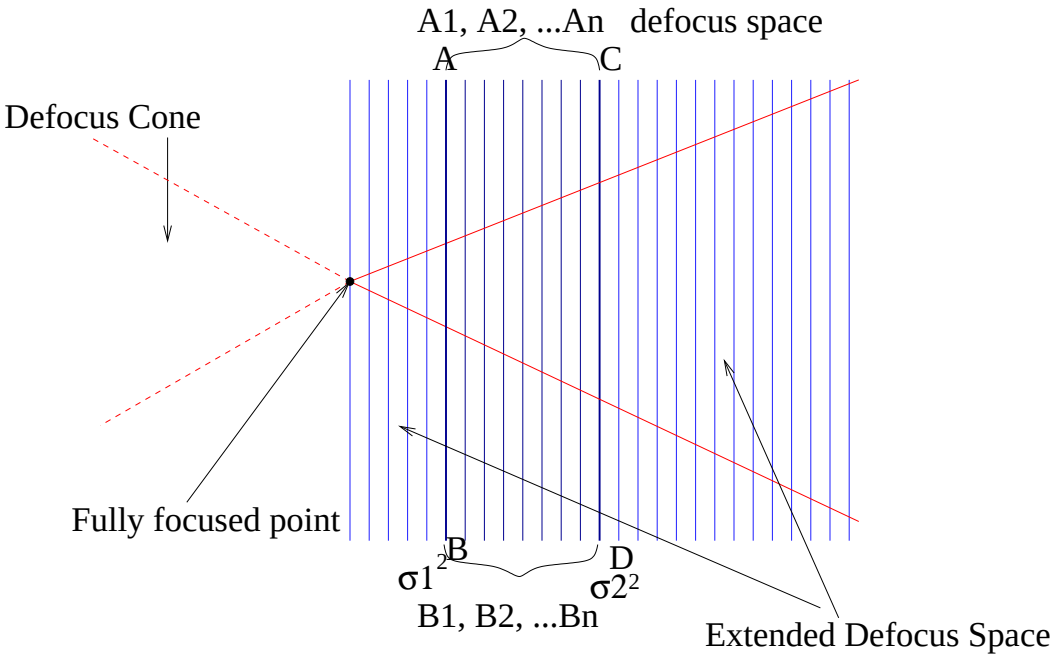


Figure 3.1: Illustration of the concept of defocus space for a particular scene

to the amount of blur in the observation and we select a value such that $M > 6\sigma$. Using this modification the defocus space for a scene can be created locally even in the depth varying case. The depth estimation is done by obtaining the fully focused point for each image. The process of creating the defocus space is a monotonic process. As α reduces, the characteristics of the process changes from diffusion to inverse diffusion and the deblurring of the defocused observations takes place. In obtaining the fully focused image the value of α is not restricted to lie between 0 and 1, rather we go for values of $\alpha < 0$. The characteristic of the convolution changes from a low pass filter to a high pass filter for $0 \leq \beta$. The defocus process has to be stopped when the fully focused point is reached. This stopping point is estimated empirically from the virtually synthesized observations using a band pass filter, similar to the way it is done in DFF methods [66]

The various steps of the algorithm for depth estimation are as follows:

Step 1: Divide the observed images E_1 and E_2 into overlapping MxM windowed representations.

Step 2: Obtain the FFT of the corresponding windows in E_1 and E_2.

Step 3: Synthesize a sample of the defocus space corresponding to a particular value of $\alpha \in [\beta, 0)$ for each window using eqn(3.10). Note that β is unknown as the value of β would

give us the depth. Hence α should be changed incrementally.

Step 4: Estimate the amount of focus using a sharpness criterion function [66] which is essentially a band pass filter and decide whether a fully focused point is reached. Else reduce the value of α and go to Step 3.

Step 5: Using the corresponding values of the virtual lens parameters, calculate the value of depth at the point. Save the pixel value as the restored one.

This algorithm is sequentially executed for all pixels in the image till the corresponding pin-hole observation of the scene is obtained and a dense depth map is generated.

3.4 Computational Difficulties

The algorithm uses the windowed Fourier transform. In some cases, especially, where the gray level variance in the window is very low, signifying a texture-less scene, there might be a problem as the spectral components are nearly zero. When the value of α goes beyond the 0 to 1 range, potentially a division by zero can occur in eqn(3.10). This can be avoided by marking such windows out of computation. Mathematically it signifies that the depth cannot be estimated for homogeneous regions.

Another factor which adversely affects the accuracy is its sensitivity to quantization error. Generally, an 8 bit quantization of the scene results in a very noisy virtual observations. This is because the defocus space generating process acts as a high pass filter when we take $0 \leq \beta$, which greatly enhances the quantization error. Further the inverse diffusion process is inherently unstable and the quantization error aggravates the instability. Practical implementation suggests the use of a 16-bit representation of the intensity function.

Generation of virtual observations using eqn(3.10) locally may demand quite a bit of computation. This is more so due to the fact that a finer sampling of the extended defocus space would lead to a better accuracy in the depth estimate. To obtain better estimates of the fully focused points efficiently, a hierarchical virtual sampling technique over nested intervals is used wherein, using the algorithm defined earlier a value of α is quickly estimated using coarser discrete steps in the range $[\beta, 0]$. Then a further dense sampling is performed in a small neighborhood ϵ around the best current estimate of α, i.e. $\alpha \in [\hat{\alpha} - \epsilon, \hat{\alpha} + \epsilon]$ and the estimate of α is refined in a hierarchical manner.

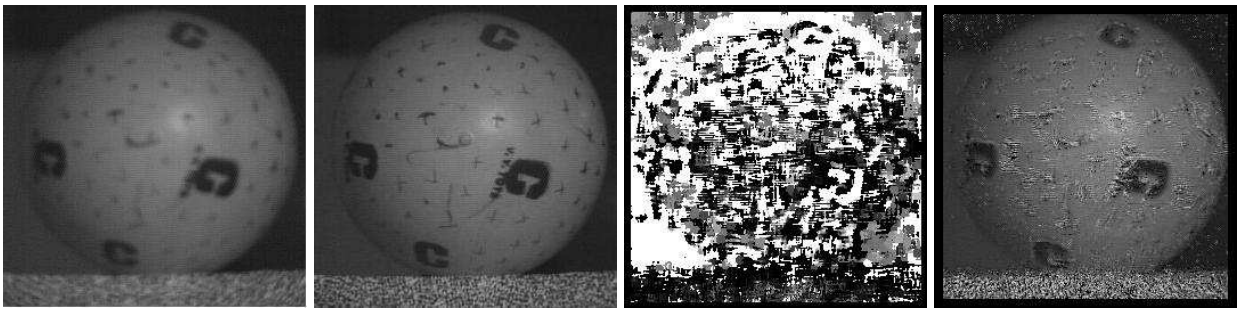
3.5 Experimental Results

The algorithm has been tested with real as well as simulated data. In the case of real data, there is a substantial amount of noise in the recovered structure. This is mainly because the real world data is in eight bit form and the resultant quantization error is quite significant. However, the overall structure recovered still resembles the true structure in the scene. In a similar way, the corresponding deblurred observation, in general, does not resemble the actual pin-hole image, but the result is definitely better focused and less blurred than the observations given as input to the algorithm. The results obtained with synthetic data can be observed to be of better quality due to the use of 16 bit representation.

In Fig.3.2, two images of a ball are taken with varying lens-to-image plane distances. In the experimental setup the base was at a distance of 117 cm. from the camera. The point on the ball nearest to the camera was at 121.8 cm. while the points lying on the occluding boundary of the ball were at a distance of 132.3 cm. from the camera. The change in the lens-to-image plane distance introduces a small amount of change in magnification. This was taken into account and corrected using a simple resizing operation. Fig.3.2.(c) shows the recovered dense depth map with the darker shading corresponding to a nearer distance. The darkest points (gray level 0) refer to the homogeneous regions for which the depth cannot be estimated as explained in section 3.4. Fig.3.2.(d) shows the corresponding deblurred image obtained.

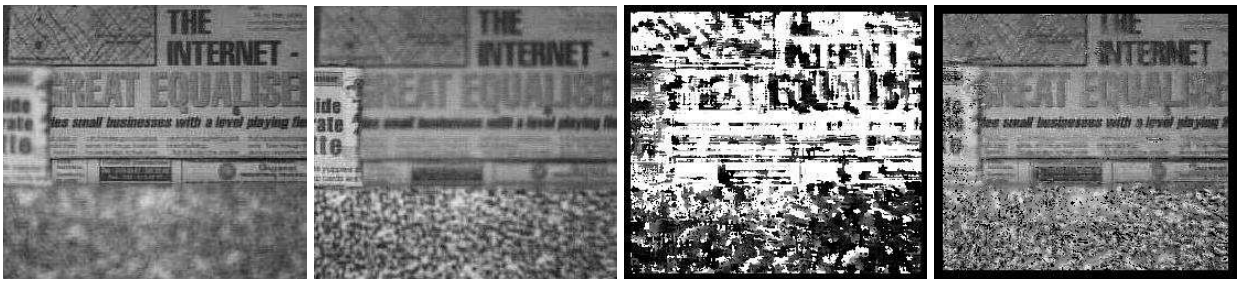
The second experimental setup was the "blocks world" where three blocks were arranged at different depths (see Fig. 3.3(a,b)), the nearest one at a distance of 73 cm., another at 82.7 cm. and the farthest block at 96.6 cm. Again images were taken with varying lens-to-image plane distances to obtain different amount of defocus in different observations. Fig.3.3.(c) shows the dense depth map estimated in this case and Fig.3.3.(d) shows the deblurred image obtained.

The Fig.3.4 shows a test data where a textured image is synthetically blurred with a continuously varying Gaussian kernel. The variance of the Gaussian kernel was increased in a ramp like manner from left to right. The second observation was simulated using different values of the blur kernel. Fig.3.4.(c) shows the corresponding dense depth map and Fig.3.4.(d) shows the deblurred image. The left to right variation in depth is clearly visible. Similarly the restored image is much sharper, although it contains dark spots where depth could not be estimated due to reasons mentioned in section 3.4. Although we could have copied the intensities at these pixels from one of the observations as they correspond to fairly homogeneous regions, we refrain from



(a) (b) (c) (d)

Figure 3.2: Ball Image: (a,b) Two observations with the right one being less blurred, (c,d) recovered structure and the deblurred image, respectively.



(a) (b) (c) (d)

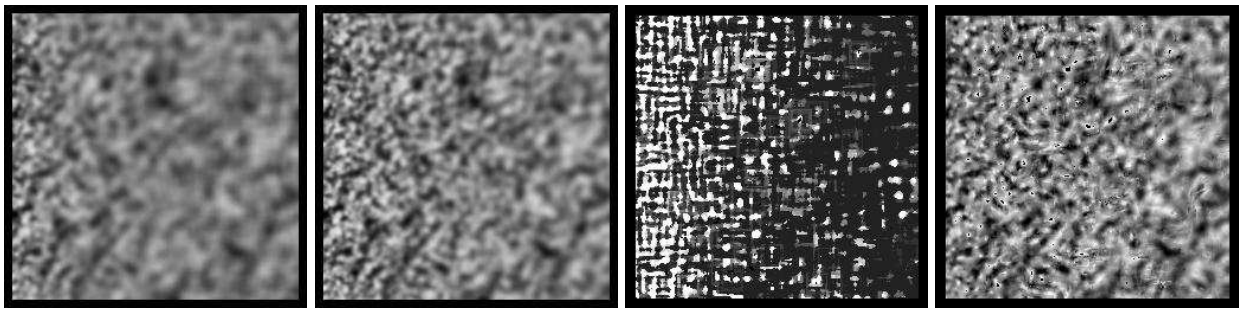
Figure 3.3: Two observations of the Block World. (a) The furthest block is in focus, (b) the nearest block is in focus, (c,d) recovered structure and the deblurred image, respectively.

doing it so that the effect can be highlighted. Fig.3.5 shows another synthetically generated test data where a textured image is blurred with a continuously varying Gaussian kernel. However, here the variance of the blur was increased in a radially outward manner. Fig.3.5.(c) shows the corresponding dense depth map and Fig.3.5.(d) shows the corresponding deblurred image. Once again the depth variation is quite clear from the plot.

The results appear to be noisy as the linear diffusion process suffers from instability in the extended defocus space as the process corresponds to inverse diffusion. One does require a suitable regularizing functional to make the problem better posed. This method, however, presents a theoretical basis for understanding depth from focus/defocus in the light of the diffusion equation and is thus important in its own merit.

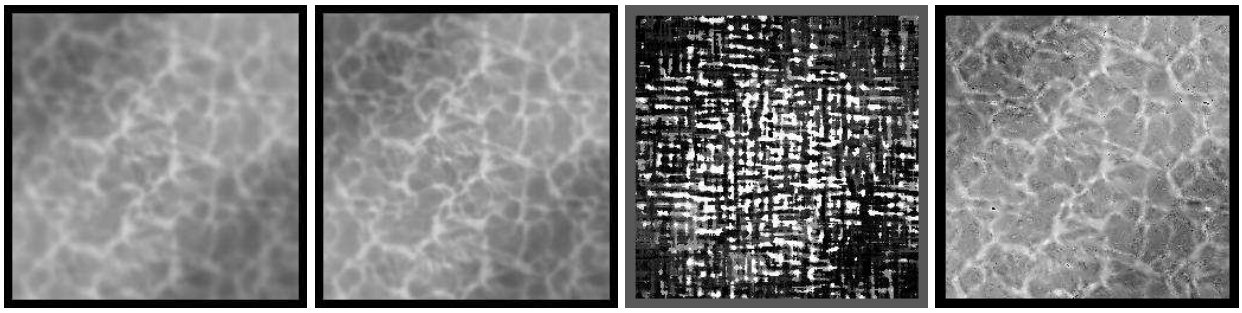
3.6 Conclusion

For a given scene in the real world, we have defined a defocus space which is a virtual space of all observations based on the properties of a real aperture imaging system. A method for



(a) (b) (c) (d)

Figure 3.4: (a,b) Two synthetically generated blurred observations. Here the blur increases progressively from left to right. (c) Recovered structure, and (d) the deblurred image.



(a) (b) (c) (d)

Figure 3.5: (a,b) Two simulated observations. Here the blur increases radially outward. (c,d) Recovered structure and the deblurred image, respectively.

generating the defocus space based on the diffusion equation has been presented. We have also presented an algorithm for recovering the scene structure based on the defocus space. An interesting outcome of this work is that it brings out the equivalence of the depth from focus and depth from defocus modalities for depth estimation. This algorithm has been tested with real as well as synthetic images.

A problem that is faced in the method described in this chapter is the drawback faced while using windowed Fourier transform. This problem is not faced when one does a spatial diffusion based approach for depth estimation from defocus as done by Favaro *et al.* [32]. However there are problems associated with this approach where occlusion effects are not handled appropriately. We therefore consider a stochastically perturbed diffusion model which is considered in the following chapter.

Chapter 4

Stochastically Perturbed Diffusion

In the previous chapter the linear diffusion process was used to model the defocus blur. The defocus morphing relationship presented there allowed us to generate the defocus space, with a frequency domain based morphing relation. Similarly the linear diffusion process has been used by Favaro *et al.* [32] to also estimate depth from defocus. There they use linear diffusion to estimate the relative blur between two defocused images. They have also provided the analysis required for estimating depth from defocus using linear diffusion. However there are several shortcomings to their approach which are addressed in this chapter. A particular observation made by Hasinoff and Kutulakos [50] is that this method of shape recovery is particularly relevant for complex scenes which have a large amount of geometric detail and complex self occlusion relationships which make it difficult to estimate the shape using stereo based methods. However, the method proposed by Favaro *et al.* could not handle depth estimation in such complex scenes. This was because, their method could not handle departure from Gaussian assumption in case of self-occlusions. Second, the method proposed obtains a local minima for the depth in the scene as it is based on a conjugate gradient based method and may not result in the true depth for complex scenes. In this chapter we address both these shortcomings. Here, we propose a model wherein the heat equation is perturbed stochastically. In this approach the departure from the Gaussian blur model is implicitly accounted for in the stochastic perturbation of diffusion. The mathematical existence for the stochastically perturbed heat equation, which is used here, has been analyzed by Yip[131] and he has used it to model the dendritic growth of crystal structures. Here we adapt the model for solving the depth from defocus problem by correlating the stochastic heat equation to the defocus blurring process.

We have considered various methods for estimating depth from defocus in chapter 2. How-

ever, most of the works done assume that the observations do not suffer from self-occlusion. The handling of occlusion effects in depth computation has been addressed in [3],[8],[33]. The extent of departure from the Gaussian shape depends on the nature of depth discontinuity in the scene, which is unknown. Any imperfection in the lens aperture would also change the shape of the blur kernel. Unlike earlier methods, the proposed method can handle such an effect under a unified framework without having to estimate the departure from the assumed model. An interesting recent work has been by Hasinoff and Kutulakos [50], where the authors consider depth from focus as a pixel matching operation. However, this method requires many high resolution observations.

4.1 Defocus as a Stochastically Perturbed Diffusion

In this section we discuss the mathematical basis of stochastic perturbation of the heat equation as a tool to analyze defocused images.

4.1.1 Stochastic Perturbation

The stochastic perturbation of the heat equation is achieved by perturbing the deterministic heat equation with a stochastic process. An archetype stochastic process that concerns us is the Brownian Motion W(t). A Brownian motion has independent normal increments. The interesting aspect of the Brownian motion is that the transition probability for a Brownian motion is the conditional Gaussian probability density function and over time it results in the heat equation. A general diffusion of the stochastic process can be given by [86]

$$dX = b(t)dt + a(t)dW(t) \tag{4.1}$$

where $W(t)$ is the standard Brownian motion, $b(t)$ is the drift and $a(t)$ is the diffusion coefficient. Here $dW(t)$ is the Ito derivative of the Brownian motion. The heat equation is essentially an Ito-diffusion with only a diffusion coefficient and no drift coefficient associated to the diffusion process, i.e., $b(t) = 0$.

The addition of stochastic perturbation to the deterministic diffusion equation can be physically thought of as adding thermal fluctuations to the heat diffusion equation. The issues like existence and regularity of the evolution arise by such an addition. These were rigorously studied and proved by Yip[131]. They were studied in the context of crystal growth. However, the

same formulation is valid for the defocus problem. The form of the stochastically perturbed diffusion or the stochastic heat equation is given by

$$du = (c\triangle u)dt + a(t)dW(t) \tag{4.2}$$

where $W(t)$ is a spatially correlated infinite dimensional Brownian motion, $dW(t)$ is the Ito's differential and $\triangle u$ corresponds to the Laplacian of u in space. The spatial correlation of W is essential for proving the Gibbs-Thomson condition [131]. This implies that the movement of each particle is not stochastic in space but in time. The Gibbs-Thomson condition is related to the regularity and existence of the solution of eqn(4.2). Gibbs-Thomson relation is a function which relates the temperature and curvature values in equilibrium for the interface of evolution. Loosely speaking the Gibbs-Thomson condition essentially prescribes an equality between the variation of the energy of the interface and the total divergence of the Gibbs-Thomson relation. These are discussed in detail by Yip in his work[131] where he gives a proof of the Gibbs-Thomson condition for eqn(4.2).

4.1.2 Defocusing as a Stochastically Perturbed Diffusion

The defocusing phenomenon has a specific space varying characteristic at surface edges and occluding edges. Consider the particular case as shown in fig. 4.1. Here we consider the specific case of a surface edge discontinuity which results in self-occlusion. In depth from defocus, self occlusion results when a continuum of rays is partially occluded and results in the blur kernel being modified [112]. This is illustrated in fig.4.1. Here, the rays emanating from the point P are partially blocked due to the surface discontinuity. The image plane is at a distance from the focus point and so the observation of point P results in a blur with radius R_b. However, due to the partial occlusion due to the near edge, the resultant blur instead of being circular is deformed (being R_{eff}). This artifact is present for all points in the observation from the surface edge to the point A. From point A onwards, the blur kernel is unaffected. A similar effect can be observed in the case of an occluding edge as well [3].

There have been a few approaches [3], [8], [33] where the authors have tried to address this problem by explicit modeling of this phenomenon or by adding a post-processing step. However, in our model, due to the stochastically perturbed curvature driven motion along the level sets, it is possible to incorporate this variation implicitly. This is particularly important in correctly estimating the blur kernels along discontinuities like surface edges and occluding

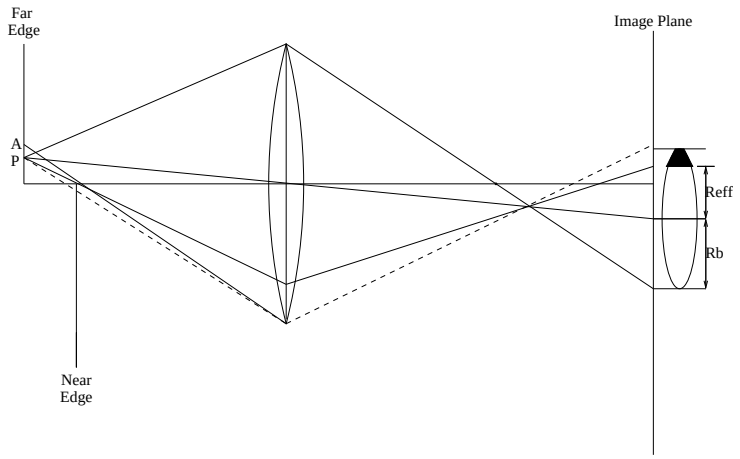


Figure 4.1: Illustration of the self-occlusion on account of surface discontinuity. For the point P, the point spread function (PSF) is the circular region devoid of the darkened region. For the point A in the scene, the PSF is circular as there is no self-occlusion.

edges. This is depicted in fig.4.1. As shown in the figure, along the surface edge, the contributions from the near and far surface are inhomogeneously mixed and this results in an anisotropic nature to the resultant blur kernel. So, when one does a stochastic curvature driven motion along the level sets, the blur contribution along the surface edge can be appropriately estimated. The non-uniformity of the kernel is implicitly handled in this model. There exists a similar effect when one has an occluding edge as well[3] .

4.1.3 Evolution Equation

We now proceed to obtain an explicit evolution equation. In order to do this we first obtain an expression for the stochastic perturbation part of eqn(4.2). Here we consider the recent work done in stochastic level sets [56] and stochastic curvature driven motion [71]. In [56], the authors discuss the evolution for the stochastic perturbation where $a(t)$ in the stochastic heat equation is the normal at the surface **n**, where we consider the surface as being the image $u(x, y, t)$, i.e. the image being evolved at time t. The stochastic perturbation of the eqn(4.2) can be therefore given by

$$du = \mathbf{n}\, dW(t). \tag{4.3}$$

This equation can be given in the differential form as

$$\frac{\partial u(x, y, t)}{\partial t} = \mathbf{n}(x, y, t) dW(t), \tag{4.4}$$

where $\mathbf{n}(x, y, t)$ is the normal to the interface $u(t)$ (i.e. the interface $u(x, y, t) \forall x, y$). The equivalent deterministic evolution using the level set framework for the geometric heat equation

is given by the following equation

$$\frac{\partial u(x,y,t)}{\partial t} = \kappa(x,y,t)\mathbf{n}(x,y,t) \tag{4.5}$$

where $\kappa(x,y,t)$ is the mean curvature of the level set and $\mathbf{n}(x,y,t)$ is the normal to the level set. Here κ is given by

$$\kappa = \frac{u_x^2 u_{yy} - 2u_x u_y u_{xy} + u_{xx} u_y^2}{(u_x^2 + u_y^2)^{3/2}}, \tag{4.6}$$

where u_x refers to $\frac{\partial u}{\partial x}$. The normal $\mathbf{n}$ is given by

$$\mathbf{n} = \frac{\nabla u}{\|\nabla u\|},$$

where $\nabla u = [u_x u_y]$. The geometric heat equation is similar to the linear heat equation except that it diffuses orthogonal to its gradient and does not diffuse along the direction of the gradient. As a result the stochastic perturbation mainly affects the level set curves and does not affect the homogeneous regions. This is appropriate since any kernel variation for instance due to self occlusion would mainly occur along edges and would be reflected in the stochastic perturbation. The effect of the perturbation is further spread on the homogeneous regions through the deterministic diffusion component.

Now, the stochastic formulation of the above deterministic formulation according to eqn(4.4) could be written as

$$du(x,y,t) = \mathbf{n}\,(x,y,t)dW(t), \tag{4.7}$$

The differential in eqn(4.7) is the *Ito differential*. This suffers from problems like it is not invariant to the parametrization of the curve, i.e., the evolution depends on the implicit representation of the initial curve and ill pastiness, i.e., under certain conditions it approaches the inverse heat equation which is unstable[56],[71]. These difficulties are overcome by introducing the *Stratonovich differential*[86] given by

$$du(x,y,t) = \mathbf{n} \circ dW(t). \tag{4.8}$$

The Stratonovich form is in an implicit form and converting it to the explicit Ito form results in an added second order term. This is because of the difference in estimating Ito and Stratonovich differentials. In Ito diffusion the integration happens at the left end point whereas in the Stratonovich case the integration happens at the mid-point while evaluating the integration of the differential[86]. With a single Gaussian perturbation in space, the eqn(4.8) is written

as

$$du(x,y,t) = \mathbf{n}dW(t) + \frac{1}{2}\triangle u(x,y,t)\left[\frac{\nabla u(x,y,t)}{|\nabla u(x,y,t)|}\right]. \tag{4.9}$$

The numerical implementation of the scheme for evolution is done by considering a step $\triangle t$ in time and $\triangle x$ in space and is given by[56]

$$u(x,y,t+\triangle t) = u(x,y,t) + \mathbf{n}\sqrt{\triangle t}\,\mathcal{N}_{(0,1)}(t) + \frac{1}{2}\triangle u(x,y,t)\left[\frac{\nabla u(x,y,t)}{|\nabla u(x,y,t)|}\right]. \tag{4.10}$$

where $\mathcal{N}$ is the noise term and it denotes a standard Gaussian random variable, and the second order term is introduced because of the Stratonovich differential component. This term is a kind of smoothing term and is nothing but the degenerate diffusion component along the edges with the stochastic term corresponding to the diffusion component across the edges. Hence the complete stochastic heat equation would then be

$$u(x,y,t+\triangle t) = u(x,y,t)+\mathbf{n}\sqrt{\triangle t}\,\mathcal{N}_{(0,1)}(t)+\frac{1}{2}\triangle u(x,y,t)\left[\frac{\nabla u(x,y,t)}{|\nabla u(x,y,t)|}\right]+c(x)\triangle u(x,y,t).$$

Since the stochastic perturbation appropriately handles the deformation of the kernel, the diffusion coefficient c is taken to be only a single inhomogeneous coefficient value and not a diffusion tensor.

4.2 Depth Estimation

We consider the case when we are given two images $I_1(x,y)$, $I_2(x,y)$ with different defocus blurs. Then the resultant formulation is

$$\begin{aligned} u(x,y,t) &= I_1(x,y) \\ u(x,y,t+m\delta t) &= I_2(x,y), \end{aligned} \tag{4.11}$$

and where the term $u(x,y,t+m\triangle t)$ is obtained from $u(x,y,t)$ by the evolution in eqn(4.11) and m is the number of iterations in going from image I_1 to I_2. The evolution equation in eqn(4.11) blurs the image I_1 with a space-variant blur till it approximates the image I_2 closely enough which is tracked by a discrepancy measure ϕ. The blur parameter σ is related to the diffusion coefficient by the eqn(4.12)

$$\sigma^2 = \frac{tc}{\gamma} \tag{4.12}$$

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. The blur parameter σ is directly proportional to the depth in the scene[19] . In order to estimate the depth in the scene one therefore has to estimate the diffusion coefficient for the evolution equation. In a deterministic case one would obtain the following minimization problem:

$$\hat{c}(x,y) = \arg\min_{c(x,y)\geq 0} \int\int\int \phi(u(x,y,t+dt), I_2(x,y))dxdydt. \tag{4.13}$$

where $\phi(.)$ is a discrepancy measure and $\hat{c}(x,y)$ is the diffusion coefficient for the deterministic diffusion equation. However in the stochastically perturbed case, the resultant diffusion coefficient is a combination of deterministic and stochastic diffusion. The deterministic diffusion coefficient is obtained from the contribution from the following part of the evolution equation:

$$du_{\text{det}} = (c_{\text{det}}(x,y)\triangle u)dt \tag{4.14}$$

which is the deterministic part of eqn(4.2). The stochastic diffusion coefficient contribution is obtained by normalizing the stochastic perturbation component in the evolution equation. We recall that the stochastic perturbation component is given by

$$\begin{aligned} du_{\text{st}} &= \mathbf{n}\sqrt{\triangle t}\,\mathcal{N}_{(0,1)}(t) + \frac{1}{2}\triangle u(x,y,t)\left[\frac{\nabla u(x,y,t)}{|\nabla u(x,y,t)|}\right] \\ &= \mathbf{n}\circ dW(t) \end{aligned} \tag{4.15}$$

The stochastic diffusion coefficient is then given by normalizing the stochastic contribution by the corresponding deterministic evolution:

$$c_{\text{st}}(x,y) = \frac{\mathbf{n}\circ dW(t)}{\kappa\mathbf{n}} \tag{4.16}$$

where κ is the curvature and $\mathbf{n}$ is the normal.

Thus the combined diffusion coefficient is given by

$$d(x,y) = c_{\text{det}}(x,y) + \eta c_{\text{st}}(x,y) \tag{4.17}$$

where η is the weight factor which determines the relative weight of the stochastic perturbation. The depth in the scene is obtained by solving for $d(x,y)$ in a minimization problem of the form

$$\hat{d}(x,y) = \arg\min_{d(x,y)\geq 0} \int\int\int \phi(u(x,y,t+dt), I_2(x,y))dxdydt. \tag{4.18}$$

We adopt a Euclidean distance measure for ϕ. Here the image $I_2(x,y)$ is assumed to be the more defocused image. However, that may not always be the case, and one can have sections

in an image which are more in focus and other sections which are more defocused compared to the corresponding sections in the second image. In that case as an initial step the images are preprocessed and the regions which are more in focus are identified. The diffusion always happens in a forward direction to avoid instabilities that may arise due to backward diffusion. The method used to ensure this is similar the one suggested in [32]. The minimization in eqn.(4.18) cannot be done using conjugate gradient descent algorithm due to the stochastic perturbation. We adopt a simple simulated annealing scheme to perform the stochastic optimization. The various steps for the algorithm for depth estimation are as follows:

STEP 1: Given the initial images $I_1(x, y)$, and $I_2(x, y)$ divide them into sections such that the diffusion is always in the forward direction using the preprocessing step discussed earlier.

STEP 2: Compute u_{n+1} from u_n using the formula for du given in eqn(4.11).

STEP 3: Compute the discrepancy measure ϕ_n

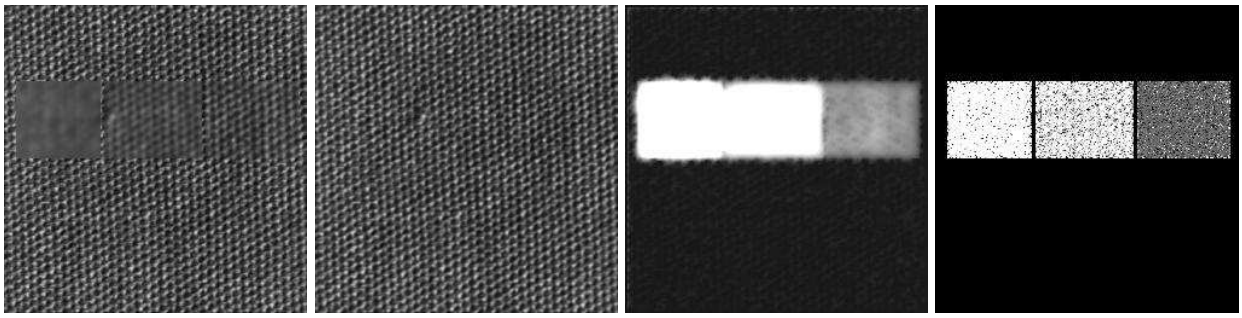
STEP 4: Accept u_{n+1}

- if $\phi_{n+1} < \phi_n$
- otherwise, accept u_{n+1} with probability $\exp\left(\frac{-(\phi_{n+1}-\phi_n)}{T(n)}\right)$.

STEP 5: Loop back to STEP 2 till the stopping criterion is satisfied.

Here $T(n)$ is a time-dependent function that plays the same role as a decreasing temperature. Its choice is crucial. If the temperature decreases too fast the process may get stuck in a local minimum, else if it decreases slowly the convergence is delayed. Here we adopt $T(n) = T_0/\sqrt{n}$ as suggested by Juan *et al.*[56]. The stopping criterion is based on the Euclidean distance measure approaching zero.

The depth estimate is then obtained by considering the deterministic and the stochastic parts separately. For the deterministic part, we assume a constant diffusion coefficient and relate the blur to the time of evolution. The blur cannot be related directly in the stochastic part due to the non-uniform nature of evolution. Hence, in each iteration we normalize the stochastic perturbation with the corresponding orthogonal diffusion component. We then integrate the corresponding contributions over time to obtain the contribution of the stochastic perturbation to the blurring process. The final depth estimate is obtained as the joint contribution of the deterministic and stochastic components.



(a) (b) (c) (d)

Figure 4.2: 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.*[32]. (d) shows the corresponding result obtained by the proposed method.

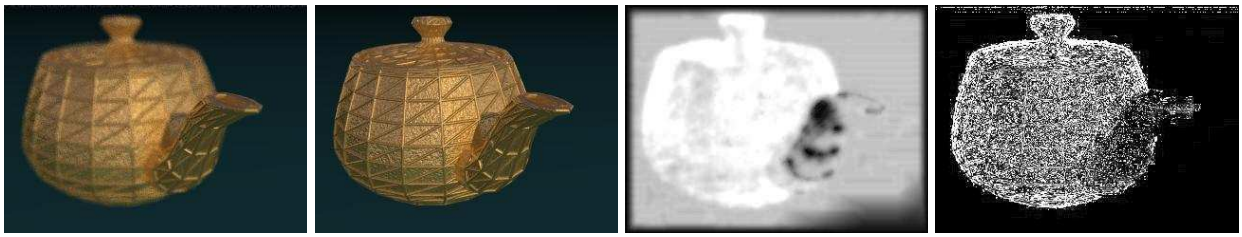
The depth obtained in this method has a space-varying characteristic, i.e., the problem solved is equivalent to space varying point spread function (PSF) estimation. Further due to the stochastic nature, the self occlusion effects and other imperfections are implicitly handled by the method when it does a stochastic perturbation of the blur model.

4.3 Experiments

The algorithm has been tested with real and simulated observations. The method works quite well on all these test data sets.

4.3.1 Simulated Data

Fig.4.2 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 4.2(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 [32] results in the regions being connected as can be seen in Fig. 4.2(c). Fig. 4.2(d) shows the corresponding estimated depth map obtained from the proposed technique where the depth in the different regions is seen separately. The brighter areas correspond to regions that are more defocused. The accuracy is confirmed against the expected depth map.



(a) (b) (c) (d)

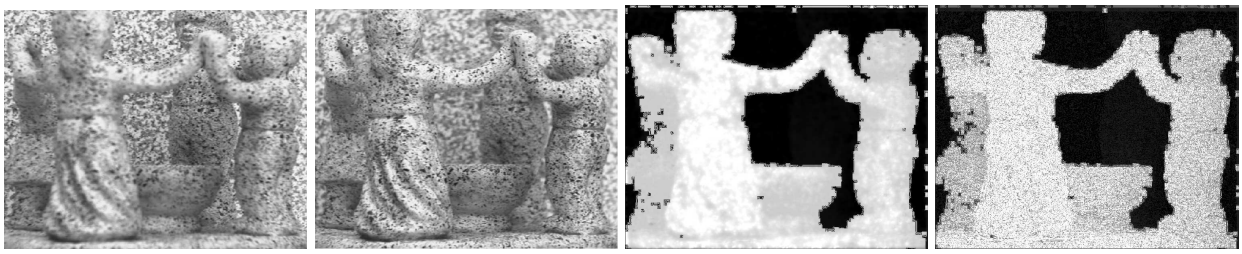
Figure 4.3: Here (a,b) are two different blurred observations generated using Povray toolkit. The defocus blur is proportional to actual depth in the scene demonstrated. (c) shows the resulting structure recovered from the deterministic method[32] . (d) shows the corresponding depth map from the proposed method.

The fig.4.3 shows a test data generated using the Povray toolkit. This is a standard ray tracing toolkit. It allows one to generate data sets which are blurred using physically correct depth blur model (pill box blur) based on geometric optics. The proposed method is still able to obtain a reliable estimate for the depth which is shown in fig. 4.3(d). Fig. 4.3(c) shows the depth map obtained by the deterministic method [32] and the proposed technique compares very favorably. This is especially indicative of the adaptability of the method for blur models different from the assumed Gaussian blur model.

4.3.2 Real Data

The experimental setup shown in fig.4.4 is the “dolls” data set[32]. The images were taken with varying lens to image plane distances to obtain different amount of defocus in different observations. The Fig.4.4(c) shows the depth map estimated by the deterministic method[32] and Fig.4.4(d) shows the depth map obtained by the proposed method. Once again we can clearly identify the depth boundaries from the recovered depth map, justifying the usefulness of the proposed algorithm. The different dolls are clearly visible to be at different depths.

A challenging data set is the “hair” data set used in [50]. The data set is of a wig with a messy hairstyle surrounded by several artificial plants. This data poses challenging self occlusion and complex structure issues. Fig. 4.5(a,b) shows the 2 input images used. Fig. 4.5(c) shows the depth map obtained for the deterministic method [32]. As can be seen, the method does not handle the self occlusion and non-convex diffusion coefficient issues efficiently. Fig. 4.5(d) shows the depth map obtained from the confocal stereo method [50]. However, they have images from 13 aperture settings each with 61 focal settings. Fig. 4.5(e) shows the depth map



(a) (b) (c) (d)

Figure 4.4: Here (a,b) are two real data sets showing the dolls placed at different depths (Images courtesy Favaro [32]). (c) shows the resultant depth map for the deterministic method[32]. (d) shows the corresponding result from the proposed method.

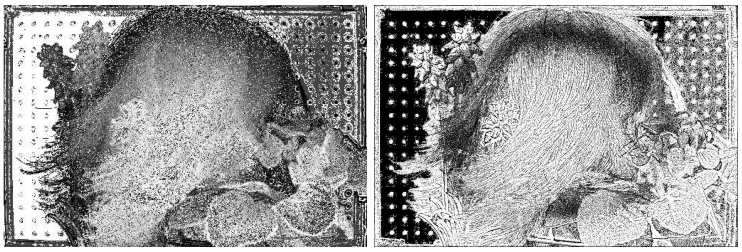
obtained from the proposed method using just two input images which is comparable to the depth map in [50] obtained from many images.

4.4 Conclusion

In this chapter we have proposed a method based on stochastic perturbation of diffusion for solving the depth from defocus problem. The main contribution here has been in incorporating a stochastic formulation of the blur model which can effectively handle variations in the blur from the standard Gaussian blur model. The variations arise in the real world due to aberrations in the lenses and aperture and are experimentally too elaborate to measure. Further the problem of deformation of the Gaussian kernel due to self occlusion is also implicitly handled. We demonstrate that improved results can be obtained using the proposed technique. The proposed method also takes into account the non-convex nature of depth in the scene and the depth map is obtained reliably for complex scenes.

It may be noted that most researchers in the area of structure recovery have pointed out the need for regularization of the recovered surface. The proposed method does not impose any such constraint while recovering the depth. In the next chapter we use a Markov random field representation for regularizing the estimated diffusion coefficient and obtain a regularized estimate of depth from defocus using deterministic diffusion. This representation improves the accuracy of the depth estimated.

(a) (b) (c)



(d) (e)

Figure 4.5: Here (a,b) are two real data sets showing a wig and flowers (Images courtesy [50]). (c) shows the resultant depth map for the deterministic method[32]. (d) shows the corresponding result from [50] (which uses images from 13 aperture, each with 61 focus settings) and (e) depicts the result from the proposed method (using only 2 images).

Chapter 5

Regularized Depth from Defocus

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 by adding a regularization term 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 [55]. However this approach results in overly smooth solutions. An approach made towards solving this problem is by using total variation based regularization [122]. A more principled approach is by using energy minimization using the discrete optimization framework of graph-cuts as proposed by Boykov *et al.* [12]. This approach can be mathematically formulated as an approach towards exact maximum a posteriori (MAP) estimation of a Markov random field (MRF) [45].

In the subsequent section we discuss the related work. In section 5.2 we consider the MAP-MRF representation for representing the diffusion coefficient which is correlated to the defocus cue. In section 5.3 the solution of the resulting problem is discussed. In section 5.4 we present the experimental results which demonstrate the efficacy of the method.

5.1 Related Work

The earliest works towards the use of graph-cuts in image processing has been towards denoising of images where Greig *et al.* [45] used the Ford-Fulkerson idea of Graph-Cuts towards solving the problem of denoising of images by solving in a MAP-MRF framework proposed earlier by Besag [7]. The use of MAP-MRF towards solving the problem of stereo was proposed by Roy and Cox [107]. An important contribution was by Boykov *et al.* [12] who demonstrated a fast approximate energy minimization technique for solving computer vision problems by using the idea of alpha expansion and alpha swap. A theoretical understanding of the energy functions that can be minimized using graph cuts was done by Kolmogorov and Zabih [64]. Further work done by Kolmogorov and Zabih showed effective use of graph cuts for computation of depth from stereo in the presence of occlusion [63]. 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 [19]. 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 [32], 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 graph-cut allows use of robust regularization like the Huber function and total variation seamlessly.

5.2 MAP-MRF representation of defocus cue

We now consider the representation of the diffusion coefficient to be estimated.

We recollect that, the defocusing of a scene can be formulated in terms of the isotropic heat equation [62] given by

$$\begin{aligned} \frac{\partial u(x,y,t)}{\partial t} &= c\left(\frac{\partial^2 u(x,y,t)}{\partial x^2} + \frac{\partial^2 u(x,y,t)}{\partial y^2}\right) \\ u(x,y,0) &= I_0(x,y) \end{aligned} \tag{5.1}$$

Here the solution $u(x, y, t)$ taken at a specific time $t = \tau$ plays the role of an image $I(x, y) = u(x, y, \tau)$ and $I_0(x, y)$ corresponds to the initial condition, i.e. the pin-hole equivalent observation of the scene. Note that we have used $u(x, y, t)$ to represent the evolution of heat everywhere in the thesis. 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. The equivalent Gaussian point spread function is given by

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \tag{5.2}$$

where σ is a blurring parameter that is a function of depth at a given point and σ is related to the diffusion coefficient

$$\sigma^2 = \frac{tc}{\gamma} \tag{5.3}$$

where t is the time variable in the diffusion equation, c is the diffusion coefficient.

Here we directly estimate the σ. Let w_i denote the label or σ 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 that image $\hat{w}$ 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 the following energy function $E(w)$.

$$E(w) = \sum_i \left(\phi(z|w_i) + \sum_{j \in \mathcal{N}} \psi(w_i, w_j)\right) \tag{5.4}$$

Here the first term is the data likelihood and the second term 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.

$$\phi(z|w_i) = ||z - I_0 * w_i||_2 \tag{5.5}$$

The interaction potential is given by

$$\psi(w_i, w_j) = M(i, j)|w_i -, w_j|. \tag{5.6}$$

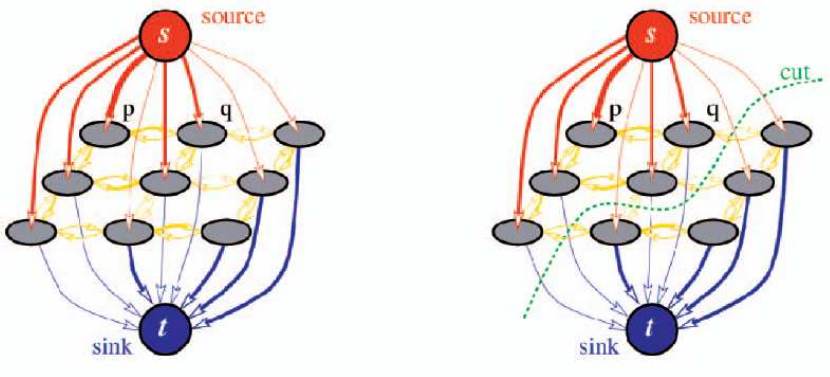


Figure 5.1: Illustration of graph-cut in a 3×3 Markov random field, courtesy [11].

where $M(i,j)$ is a truncation factor used to obtain a robust error term between the two labels w_i, w_j. In the experiments the truncated linear term was used after experimental comparison.

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 the previous chapter and in the approach by Favaro *et al.* [32], 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 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.

5.3 Graph-Cuts for solving MAP-MRF framework

We minimize eqn.(5.4), thereby maximizing the posterior probability using graph cuts ([11],[12]). 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 [12] which allows us to consider this method of using binary labels to minimize the cost over the entire set w. The graph cut procedure is depicted in the following figure from [11].

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

$$E(w_1, \cdots, w_n) = \sum_{i<j} E^{i,j}(w_i, w_j). \tag{5.7}$$

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 [64], 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:

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \tag{5.8}$$

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 method defined.

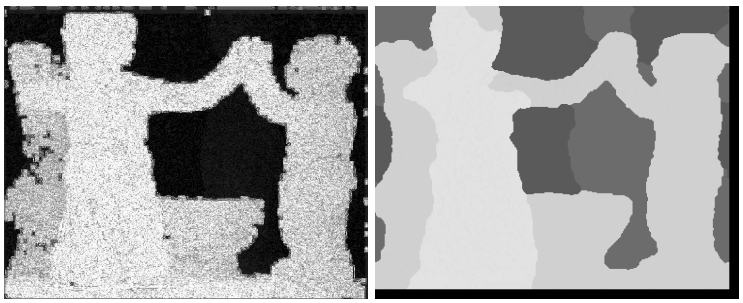
5.4 Experimental Results

Since the usefulness of this method was for regularizing the depth from defocus estimate obtained by linear diffusion, we have directly evaluated this method with real image data sets and have tried it on some of the challenging image data sets.

The first data set used for evaluation is the "dolls" data set [32]. 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 [32] can be seen in 5.2(c). Here the authors have used $\mathcal{L}_2$ regularization. The result obtained by stochastically perturbed depth from defocus method is shown in 5.2(d). Here no regularization has been used. The result obtained by using regularized depth from defocus using graph-cuts is shown in 5.2(e). Here, the regularization used is truncated $\mathcal{L}_2$. It can be seen that the result obtained by the technique proposed in this chapter is definitely much 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 few vegetables. Fig. 5.3(a) shows the image where the near vegetables are in focus and fig. 5.3(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. 5.3(c). The resultant depth map

(a) (b) (c)



(d) (e)

Figure 5.2: Here (a,b) are two real data sets showing a few dolls at different depths(Images courtesy [32]). (c) shows the resultant depth map for the method by Favaro *et al* [32]. (d) shows the resultant depth map obtained by the stochastic depth from defocus method explained in the previous chapter and (e) shows the resultant obtained by the regularized depth from defocus method explained in this chapter.

(a) (b) (c)

Figure 5.3: Here (a,b) are two real data sets showing a few vegetables at different depths [picture: Courtesy Dr. Sunil Hadap]. (c) shows the resultant obtained by the regularized depth from defocus method explained in this chapter.

in this challenging data set clearly shows the different vegetables and we are able to correctly estimate the depth.

5.5 Conclusion

We have seen the need for regularization and have provided a principled method for regularizing the deterministic 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.

While the method described in this chapter enables accurate depth estimation from two defocused images, a more challenging problem is estimation of depth from a single defocused image. Towards solving this problem we consider an approach based on restoring a defocused image and thereby obtaining the multiple observations required for estimating depth in the scene. We therefore now consider the problem of deblurring a blurred observation in the next chapter.

Chapter 6

Non-Parametric Image Restoration

Image restoration has been one of the classical problems in image processing. The degradation of an image is due to noise and blur and the problem is formulated as

$$Y(x,y) = \int\int U(t,\tau)h(x,y;t,\tau)dtd\tau + N(x,y). \tag{6.1}$$

Here $Y(x,y)$ is the observation, $U(x,y)$ is the original undegraded image, $h(x,y;t,\tau)$ is the space varying blurring kernel and $N(x,y)$ is the noise. The noise $N(x,y)$ is often assumed to be additive white Gaussian noise. No prior knowledge of the point spread function (PSF) of the blurring kernel is assumed in this study and hence the problem of recovering the original image $U(x,y)$ given an observation $Y(x,y)$ is known as space varying blind image restoration.

The specific case of eqn(6.1) where no blurring is assumed is given by

$$Y(x,y) = U(x,y) + N(x,y). \tag{6.2}$$

This has been addressed using the mean shift filter by Comaniciu and Meer [23].

A detailed review of the traditional methods for approaching the problem of image restoration has been done in chapter 2. We recall that these methods are based on regularized least squares technique [5], harmonic analysis based techniques ([16], [116], [6]), statistical methods [68] and partial differential equation (PDE) based methods ([4], [87], [90], [15]). These approaches have had considerable success in dealing with the problem of image restoration, more so when the blur kernel is known. However, here we approach the problem of blind image restoration using the framework of cluster analysis. Some of the very recent work on restoration include the work by Foi *et al.* [59], Figueiredo and Nowak [36]. Foi *et al.* solve the denoising problem, but they do not address the deblurring problem. The work by Figueiredo and Nowak

[36] addresses the problem of image deconvolution. However, they do not consider a spatially varying point spread function. Our interest in this chapter lies in solving the blind restoration problem when the PSF is spatially varying. The literature is very sparse when one considers the problem of spatially varying PSF. Rajagopalan and Chaudhuri [101] have solved the corresponding restoration problem using a Markov random field (MRF) prior. However, their method is strictly speaking, non-blind as they assume availability of multiple observations. In this chapter we use a single observation and make no assumption about the image intensity function such as being an MRF. Our approach yields a simple mean shift based technique for solving the problem of blind image restoration which performs very well as compared with the above mentioned methods. We now discuss and motivate the non-parametric mean shift based method.

A basic methodology in solving problems in image processing has been that based on analysis of the feature space of an image where the features are either intensity based or other specific features (for e.g. texture). This approach is centered around a mapping of the image(s) to a multi-dimensional feature space. Lately this particular methodology has become popular for solving interesting problems in low-level vision. A central task for feature space analysis is that of cluster analysis. There have been various parametric and non-parametric approaches in literature for cluster analysis [105]. These methods usually assume knowledge of the number of clusters or that the clusters have same shape. There have also been methods based on Gaussian mixture models, however, the mixture models also require the number of clusters as a parameter and, in general, the unstructured feature space cannot always be characterized in terms of a mixture model. In the analysis of arbitrarily structured feature space the non-parametric approaches have been more suitable. The non-parametric approaches have been either hierarchical or based on density estimation. In this chapter we use a method introduced in 1975 by Fukunaga and Hostetler [39] where the clustering is done by recursively shifting each data point to the average of data points in its neighborhood. Cheng in [20] developed a more general formulation and pointed out applications for the same in clustering and Hough transform, as well as its role in global optimization. Recently Fashing and Tomasi [28] have improved the understanding of mean shift as an optimization procedure by considering its equivalence to Newton's method and they have also proved that for all kernels the mean shift procedure is a quadratic bound maximization. There has lately been considerable interest in applying mean shift to problems in computer vision, based mainly on the work of Comaniciu and Meer in areas

like discontinuity-preserving smoothing and segmentation [23] and tracking [23]. The idea of mean shift based tracking has also been extended by accounting for a variable-bandwidth mean shift by Comaniciu [22]. Related to this, Collins [21] has looked into the problem of tracking blobs across scale space which accounts for variation in the scale of the kernel. The idea of mean shift has also become popular for problems like appearance based clustering [103].

In low-level vision, non-parametric clustering has been used mainly in the areas of denoising and segmentation [23]. The non-parametric clustering technique has been particularly apt as the problems of low-level vision cannot be parametrized explicitly. Recently there has been work done by Wang *et al.* in extending this approach for image and video segmentation using anisotropic kernel mean shift [58] and in interesting applications like video tooning [57].

In the next section we discuss the method of gradient based cluster separation. In section 6.2 we consider the mixed diffusion process which involves mean shift in both forward and reverse directions. We then use these techniques to solve the problem of image restoration in section 6.3. The experimental results for these methods are presented in section 6.4. We conclude in section 6.5.

6.1 Gradient Based Cluster Separation

A technique for clustering a set of points is to explicitly move the points in the direction of the gradient of the kernel density estimates [23], [39]. Since the true probability density function or even its form is not known, non-parametric techniques are used to obtain estimates of the density gradient [26]. The approach is to obtain a differentiable, nonparametric estimate of the probability density function and then its gradient is computed.

Let $\mathbf{e_1}, \mathbf{e_2}, ...\mathbf{e_n}$ be a set of n independent and identically distributed random vectors in the d-dimensional feature space $\mathcal{R}^d$ and $\mathbf{G}$ be a symmetric positive definite $d \times d$ bandwidth matrix [23], [26]. Here, in the case of image restoration, e represents the gray/color intensity. However, e in general could denote other vectors like texture. A fully parametrized $\mathbf{G}$ increases the complexity of the estimation and, in practice, the bandwidth matrix $\mathbf{G}$ is chosen to be the identity matrix $\mathbf{G} = g^2\mathbf{I}$. The bandwidth matrix determines the window width at each data point. Therefore taking the bandwidth matrix as identity implies that the window width is constant at each data point and is determined by the scalar value g. Then the kernel density

estimator takes the form

$$\hat{f}(\mathbf{e}) = \frac{1}{ng^d} \sum_{j=1}^{n} k\left(\left\| \frac{\mathbf{e} - \mathbf{e}_j}{g} \right\|^2 \right). \tag{6.3}$$

where $k(\mathbf{e})$ is a bounded kernel function with compact support satisfying

$$\lim_{\|\mathbf{e}\| \to \infty} \|\mathbf{e}\|^d k(\mathbf{e}) = 0 \text{ and } \int_{\mathcal{R}^\Gamma} k(\mathbf{e}) d\mathbf{x} = 1.$$

A differentiable kernel function is used and then the density gradient is estimated as gradient of eqn.(6.3). This gives the estimate of density gradient as

$$\nabla \hat{f}(\mathbf{e}) = \frac{1}{ng^d} \sum_{j=1}^{n} \nabla_e k\left(\left\| \frac{\mathbf{e} - \mathbf{e}_j}{g} \right\|^2 \right) \tag{6.4}$$

$$\tag{6.5}$$

Eqn.(6.4) is the general form of the estimate of the density gradient. If one uses the Gaussian kernel the resulting estimate of the density gradient is

$$\nabla \hat{f}(\mathbf{e}) = \frac{1}{n(2\pi)^{n/2} g^{(n+2)}} \sum_{j=1}^{n} (\mathbf{e} - \mathbf{e}_j). \exp\left[-(\mathbf{e} - \mathbf{e}_j)^T \left(\frac{\mathbf{e} - \mathbf{e}_j}{2g^2} \right) \right]. \tag{6.6}$$

In [39], the authors point out how eqn.(6.6) is essentially a weighted measure of the mean shift of the observations about the point $\mathbf{e}$. In order to move the values, the estimate of mean shift of the normalized gradient is used. The mean shift of the normalized gradient is

$$\frac{\nabla \hat{f}(\mathbf{e})}{\hat{f}(\mathbf{e})} = \nabla \ln \hat{f}(\mathbf{e}). \tag{6.7}$$

This method is termed the mean shift algorithm by Comaniciu and Meer in [23]. In this chapter we refer to this method as the forward mean shift algorithm, the reason for which will be clear in a short while. When we relate to the diffusion process, it is termed as forward diffusion. The method for gradient based clustering is a recursive algorithm to transform each observation according to the clustering algorithm

$$\mathbf{e}_j^{i+1} = \mathbf{e}_j^i + a \nabla \ln \hat{f}(\mathbf{e}_j^i). \tag{6.8}$$

Here a is a constant which determines the rate of convergence of the clusters.

Figure 6.1 illustrates the process of clustering that occurs as a result of eqn.(6.8). The top part of the figure shows the way the function $\nabla \ln \hat{f}$ helps to move the features to cluster centers. As can be seen, the features are moved along the gradient, bringing them gradually closer and finally they all move towards the mean of all clusters.

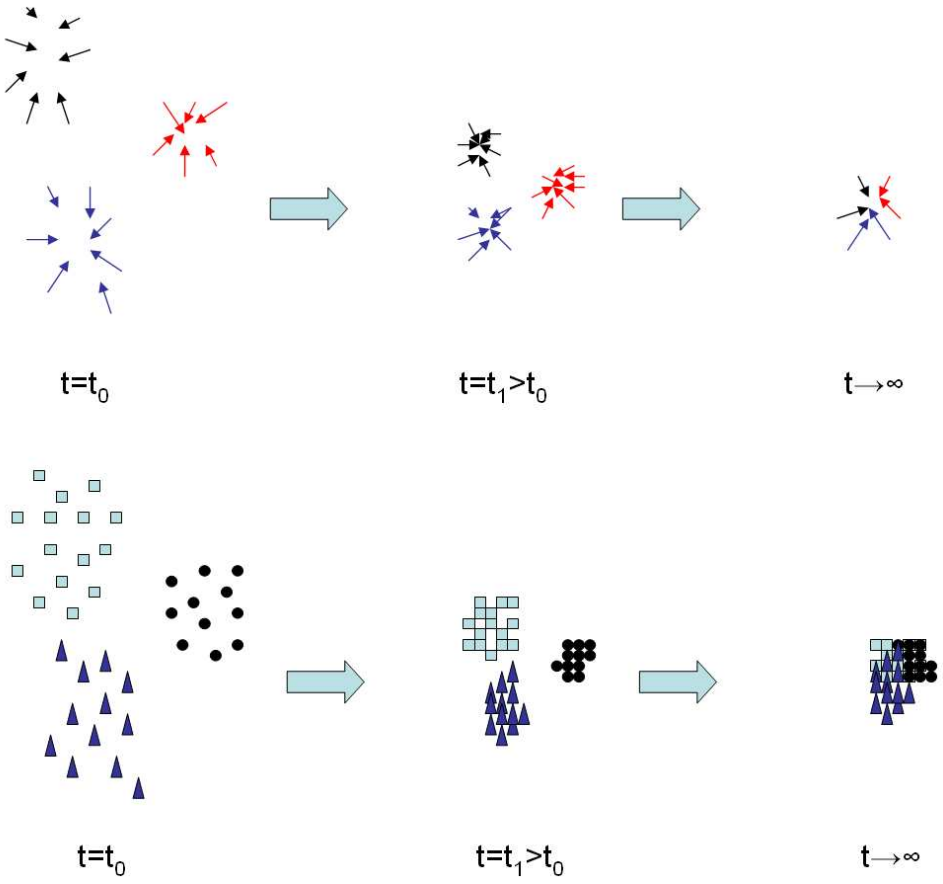


Figure 6.1: Illustration of the forward diffusion process in terms of image features. (a) At $t = t_0$, (when there is no blurring), the feature clusters are well separated. (b) At $t = t_1 > t_0$, due to forward diffusion which introduces blurring, the features move closer to each other and (c) at sufficient blurring, the feature clusters merge when they are indistinguishable. In the top row we show $\nabla \ln \hat{f}(e)$ representing the gradient flow and below we show the time evolution in feature space.

This process of clustering is valid when one has a cluster containing erroneous measurements and it is required to remove them, thus effecting denoising of the signal. An interesting alternate case could be when the clusters are somehow mixed along the cluster boundaries and it is required to separate the clusters, effecting opposite of what was achieved in the previous case. In other words, the signal would then get sharpened. This can be done by moving the points away along the direction of the gradient estimate. Fig. 6.2 illustrates the resultant difference in the movement of the points. As previously, the upper half indicates the motion of the estimate of the gradient of kernel and the lower half the actual evolution of feature values. This is a type of inverse diffusion as the cluster separation process when done for an image results in sharpening of the line fields. We term this as the reverse mean shift in line with the terminology used in [23] and [20]. As can be seen from the gradient based part of the image, when the clusters are mixed, the points along the cluster boundary are influenced by the neighboring mode instead of their own cluster mode. Hence, when we move the points away (as can be seen in the shaded region in Fig. 6.2), the values are moved away from the incorrect mode towards the actual cluster center and then they get associated with the correct mode. Quite naturally, when the mixing in the feature space is nearly complete, i.e., the resulting image is nearly homogeneous after having lost most of the spectral information, the reverse mean shift process would fail to move the feature points, offering no deblurring. But this is quite expected as the gradient would then be nearly zero. It has already been proved in the restoration literature [95] that the quality of restoration goes down with increase in the blurring which implies that the feature space becomes more and more inseparable and the reverse mean shift process would not be of much help either. The equation for the gradient based cluster separation algorithm now takes the form

$$\mathbf{e}_j^{i+1} = \mathbf{e}_j^i - a\nabla \ln \hat{f}(\mathbf{e}_j^i). \tag{6.9}$$

This is however, a divergent process. Hence, it is required to have a proper stopping criterion for terminating this.

6.2 Mixed Diffusion

The generalized mean shift filter is defined to be a process of incorporating both forward and reverse mean shift procedures in a unified manner. This is because the process of denoising is

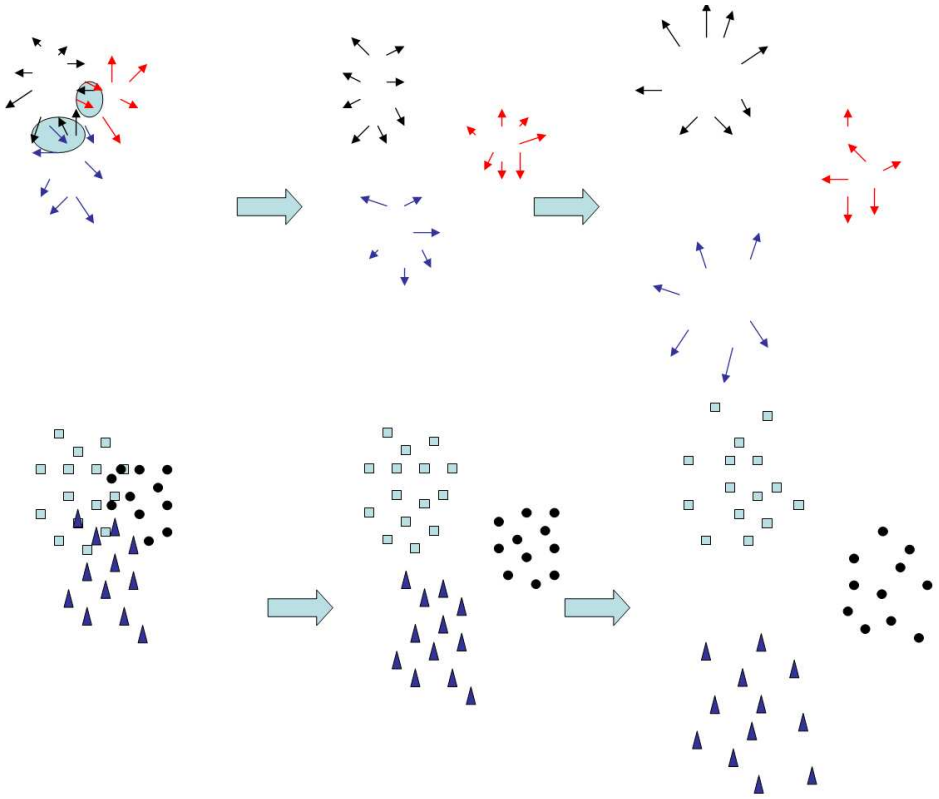


Figure 6.2: Illustration of inverse diffusion in the feature space (which has some blurring). As we do the inverse diffusion, the feature space (bottom row) separate out by moving farther away from each other. The gradient flow is shown in the evolution of top row that explains why the feature space gradually separate out. This is a divergent process.

achieved through forward mean shift and the process of deblurring is achieved using reverse mean shift. Both these operations being of contrasting nature, they should be combined very judiciously.

The forward and reverse mean shift procedures are applied in this chapter in an iterative though inhomogeneous manner. The basic idea behind such a mixed diffusion is to cluster the points around the "true" modes. This is due to the fact that the clusters with closely spaced modes may be mixed together and they would have to be separated and the noisy data present in a cluster may have to be removed. The method then is to first perform cluster separation so as to separate the modes that are mixed. Being an inverse diffusion, it results in deblurring, but this accentuates the noise as the intra-cluster distance is also increased. The forward diffusion is subsequently carried out to reduce the intra-cluster distance and remove the noisy data. The process is repeated with either forward or backward mean shift being done till a stopping criterion is satisfied. This is illustrated in Fig. 6.3. As can be seen from the gradient flow field, initially the mixing of clusters along boundaries get resolved as was described in the reverse mean shift procedure earlier. The points in the cluster are then moved along the gradient to the cluster modes by the forward mean shift. This results in the individual clusters moving to their correct or properly accentuated modes.

A pertinent point in forward diffusion for denoising has been a proper stopping criterion. This is necessitated in order to control the diffusion when the noise is removed. This has been addressed by Sporring and Weickert [115], Weickert [125] and Mrazek and Navara [77]. Sporring and Weickert [115] suggest the use of generalized entropies with intervals of minimal entropy change indicating stable scales with respect to evolution time. However the entropy may be stable over the entire interval and hence may not be a good criterion for stopping. Weickert in [125] suggests that since the relative variance decreases monotonically from 1 to 0, it can be used to measure the distance of $U(t)$ from the initial state $U(0)$ and therefore prescribing a certain value for the variance may serve as a criterion for stopping. A better criterion has been proposed by Mrazek and Navara [77] based on signal-noise decorrelation. They propose a stopping time $t = T$ such that the correlation given by

$$corr(U(0) - U(t), U(t)) = \frac{cov(U(0) - U(t), U(t))}{\sqrt{var(U(0) - U(t)).var(U(t))}} \tag{6.10}$$

is minimum. Here $U(t)$ is the diffused image $U(x, y)$ at time t and $U(0)$ is the initial image, cov is the covariance and var the variance. The idea here is that the noise N and the signal $U(t)$

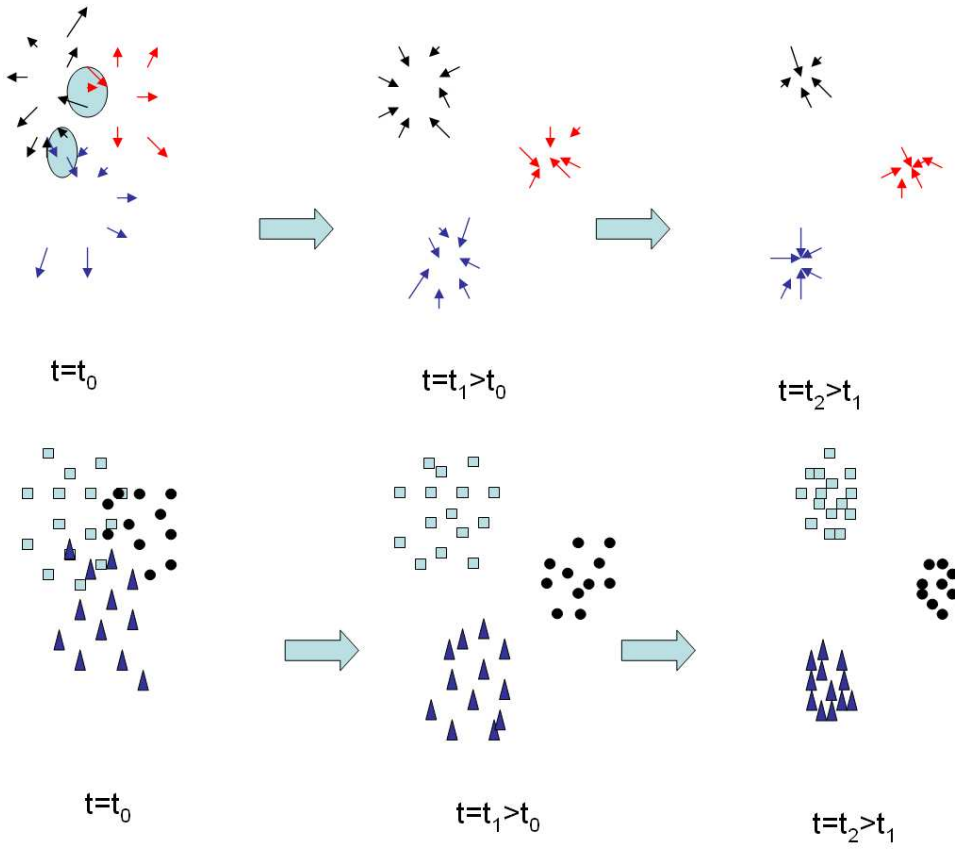


Figure 6.3: Illustration of mixed diffusion in the feature space (which has both blurring and noise). The inverse diffusion till time t_1 results in the the mixed clusters being pulled apart and the individual elements of clusters coming closer together at time t_2 due to forward diffusion. The top row shows the negative of the gradient flow till time $t < t_1$ effecting reverse mean shift, and subsequently the positive gradient field to effect a forward mean shift.

are de-correlated and the residual $U(0) - U(t)$ would initially be dominated by the noise N, and hence as the image is denoised, the correlation criterion, which is basically the normalized covariance between the residual and the image would be minimized. We use the same criterion for obtaining a stopping criterion for the forward diffusion. Here we must mention that in the initial formulation Comaniciu and Meer [23] had argued that there is no need for proposing a stopping criterion during forward diffusion as the kernel bandwidth decides the convergence to the mode and due to its convergent nature an explicit stopping criteria is not required. However, the use of the above criterion helps to achieve a tighter bound on the stopping criterion irrespective of the kernel bandwidth.

For the reverse mean shift filter, the stopping criterion is even more important as this being an inverse diffusion process, the procedure could very soon diverge unless it is stopped properly. We formulate a stopping criterion similar to the one for forward mean shift, but instead of minimization, one now has to maximize the expression. We consider the noise N and signal $U(t)$ to be decorrelated. The residual $U(0) - U(t)$ during the reverse mean shift process would contain elements of both the noise and the signal. In the reverse mean shift process, the residual would initially be very much correlated to the signal $U(t)$. Beyond a certain point the residual would mostly be dominated by noise N. Hence, the stopping criterion used for the reverse mean shift procedure is to do reverse mean shift till the decorrelation criterion achieves a maximum, indicating that the residual at this point is maximally correlated to the signal $U(t)$.

The actual procedure that we follow is to first start with the reverse mean shift procedure (see illustration in Fig. 6.3. If the covariance of the residual with the signal $U(t)$ increases, this indicates that the residual is dominated more by the actual signal than the noise, signifying sharpness of the signal and the reverse mean shift procedure is continued. Else, if the covariance actually reduces, this indicates that the noise dominates the residual and no further sharpening of the signal is possible and hence the direction is switched and the forward mean shift procedure is carried out until the stopping criterion is met.

6.3 Image Restoration

In this section we first consider the case of deblurring in the absence of noise using reverse mean shift and then use the mixed diffusion procedure to perform image deblurring in the presence of noise. The noiseless case is initially used to bring out the usefulness of the reverse mean shift

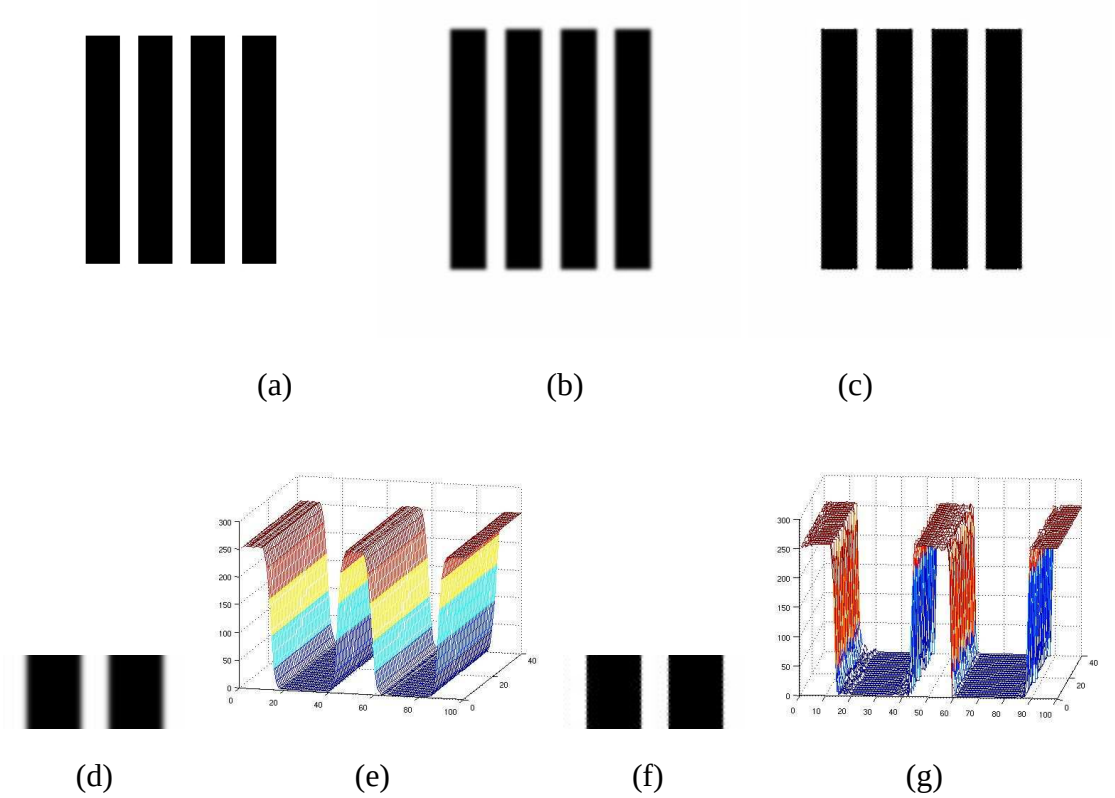


Figure 6.4: Result of inverse diffusion of the original test image (a) which has been blurred (b) using the gradient based cluster separation process is shown in (c). (d) depicts a zoomed portion of the blurred input and (e) depicts the mesh plot for the same, (f) depicts a the corresponding deblurred image and (g) shows a mesh plot for the same.

procedure.

6.3.1 Deblurring using Reverse Mean Shift

Consider a cluster in the case of image deblurring based on the intensity feature. The neighboring pixels having the same intensity belong to the same cluster. The blurring process results in bringing the clusters closer. Therefore, one can use the reverse mean shift procedure to perform deblurring or cluster separation. Since, the point spread function (PSF) is not known, the kernel density function $k(x)$ in eqn(6.3) is not known. Hence, the non-parametric approach of modeling the kernel density estimate is apt for this problem. The process of deblurring using the reverse mean shift procedure is illustrated in the Fig. 6.4.

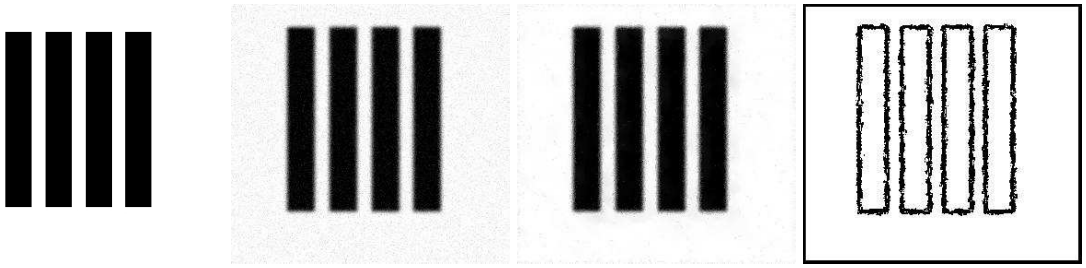
Fig. 6.4 illustrates how the process of gradient based cluster separation effectively deblurs a blurred test image. Here Fig. 6.4(a) is the original image, and Fig. 6.4(b) is the blurred observation. The process of blurring converts the step edge into a ramp edge. Fig. 6.4(c) shows

the deblurred image which has been restored using the reverse mean shift method. This is further illustrated in Fig. 6.4(d) by considering a zoomed-in portion of the blurred image. Fig. 6.4(e) displays the corresponding region as a mesh plot showing the step edge being replaced by a ramp edge. Fig. 6.4(f,g) shows the corresponding restored image and the mesh plot. The process of restoration does not make any assumption on the form and variance of the blur kernel. This substantiates our claim that the process of gradient based cluster separation is able to reverse the blurring process without any apriori knowledge of the PSF.

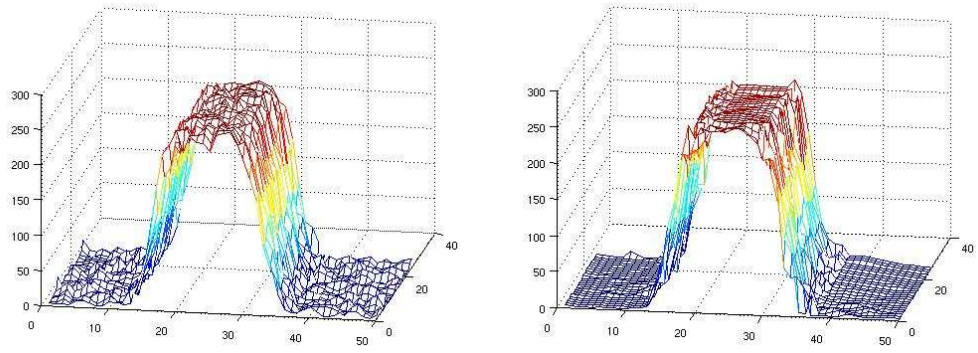
Mathematically, a Gaussian kernel used for kernel density estimation is able to exactly restore the blur caused by any Gaussian PSF. Since any Gaussian kernel can be generated from another Gaussian kernel by appropriately evolving it through a linear diffusion process either in forward or backward direction. However, the kernel density estimation using the Gaussian function can approximate other types of PSF quite well also. Strictly speaking, a Gaussian kernel cannot be used to deblur an averaging kernel. Yet, as we have found empirically and is shown using experimental results, the deblurred approximation is quite close to the original image. Another alternate function which can be used for kernel density estimation is one based on polynomial functions [113]. However, the Gaussian kernel being a very well behaved one, it is used for the kernel density estimation in this chapter.

6.3.2 Restoration using Generalized Mean Shift

We now consider solving the complete problem as specified in eqn(6.1). The proposed solution involves an iterative inhomogeneous application of the reverse and forward mean shift filter with an appropriate switching criterion as described in the previous section. The cluster, as before, is based on intensity of pixels. The resultant procedure when observed from the view-point of cluster analysis works by first separating the mixed clusters by increasing the inter-cluster separation. This however increases the intra-cluster distance as well. This is reduced by the forward mean-shift which smoothens the noise present in the data. An iterative application of forward mean shift along the cluster interiors and reverse mean shift filter along the cluster boundaries result in isolation of the relevant clusters and thereby one achieves deblurring as well as denoising. However, a key factor which needs to be noted that in the region where the reverse mean shift filter is applied (mainly along the edges in the image), the denoising is de emphasized. During the process of reverse mean shift, the intensities are moved opposite to the gradient of kernel density estimate, and hence the noise values are also enhanced. This is



(a) (b) (c) (d)



(e) (f)

Figure 6.5: Result of generalized mean shift of (a) the original test image which has been degenerated with noise and blur as shown in (b) is obtained in (c) with (d) depicting the switching criterion showing where the reverse and where forward mean shift happens. (e) depicts a mesh plot of a cross-section of the noisy blurred input and (f) shows a mesh plot of the corresponding restored result.

illustrated in Fig. 6.5.

Fig. 6.5(a) shows the initial test image. This image is degenerated by blurring it and adding noise to it and this is shown in Fig. 6.5(b). The blur kernel used is Gaussian shaped and Gaussian noise has a standard deviation of 20. The degenerated image is restored using the generalized mean shift filter and the resultant image is shown in Fig. 6.5(c). The Fig. 6.5(d) shows how the reverse mean shift occurs along the edges(shown in dark) and forward mean shift occurs along the interior regions. This shows how the switching criterion is able to correctly identify the cluster boundaries and the cluster interiors and deblur and denoise appropriately. Fig. 6.5(e) shows an enlarged portion of the input image, showing its noisy and ramp nature as a mesh plot. Fig. 6.5(f) shows the corresponding result of restoration as a mesh plot. It can be clearly seen that the result is deblurred and denoised. However, due to the difficult nature of the problem, one cannot do as much deblurring and denoising as was possible in the previous

case. Hence it is possible to deblur the edges only to a limited extent since the noise gets boosted. Thus, using the generalized mean shift procedure, it is possible to do both denoising and deblurring.

6.3.3 Implementation Issues

The formulation presented in the eqn.(6.9) is quite general enough as various forms of kernel density functions can be incorporated. One of the common reasons for the blurring in an observation is the capturing of real images using a finite aperture lens where the object is not in focus. In such a case the point spread blur function can be approximately modeled as Gaussian [19]. Even if the original estimate of the kernel density function is not explicitly known, the non-parametric kernel function generally provides a good enough approximation to the unknown underlying model. Another criterion to be considered is the bandwidth parameter g in eqn.(6.4). Here the bandwidth parameter is assumed to be small and practically we have used the bandwidth parameter to be 0.8 in this study. Moreover, the problem of deblurring also makes an implicit assumption of being spatially correlated. Hence, the eqn(6.4) is modified as

$$\nabla \hat{f}(\mathbf{e}) = \frac{1}{ng^d} \sum_{j \epsilon \mathcal{N}_\mathbf{e}} \nabla_e k \left(\left\| \frac{\mathbf{e} - \mathbf{e}_j}{g} \right\|^2 \right) \tag{6.11}$$

$$\tag{6.12}$$

where $\mathcal{N}_\mathbf{e}$ is a small neighborhood over which the computation is restricted to. In our study we choose $\mathcal{N}_\mathbf{e}$ to be a 7x7 window making the iterative updates very quick. The diffusion coefficient used in the reverse mean shift procedure is set as a small value of 0.1 for stable reverse mean shift. This is necessary to prevent any fast inverse diffusion due to spurious noise that may be present in the data.

A factor which has to be taken into account during the implementation is the threshold value to be used for saturating the inverse diffusion. The saturation threshold is a function of the amount of blur in an image. If the saturation threshold is low, then the reverse mean shift is stopped before the deblurring is complete. If the saturation threshold is high, then the reverse mean shift process starts degenerating. This is because, though the clusters have been separated, further reverse mean shift diverges the values within the cluster as well. This is illustrated in the fig. 6.6, which is a plot of saturation threshold versus PSNR achieved for a constant Gaussian blur kernel with variances 2.3 and 4.6 respectively applied on the Lena image. As can be seen, for a Gaussian blur with PSF 2.3, the optimal cluster separation threshold is 0.4 and for a PSF

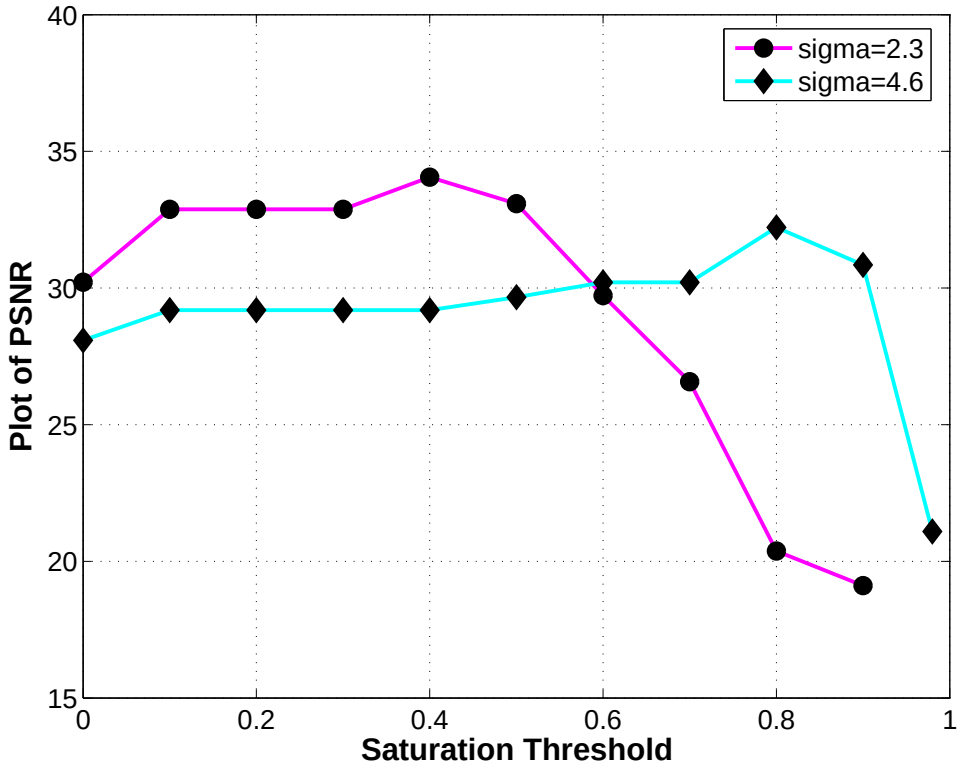


Figure 6.6: Plot of the cluster separation threshold (X Axis) versus the PSNR achieved for Gaussian PSF with variances 2.3 and 4.6 applied on the Lena image.

of 4.6, the optimal saturation threshold is 0.8. Since, the amount of blur is not known apriori, we set the threshold to be 0.5 times the current gradient value, i.e. we set $a = 0$ in eqn(6.9) whenever

$$|e_j^{i+1} - e_j^i| \geq 0.5|e_j^i| \quad \forall j. \tag{6.13}$$

Another aspect that is of concern in the implementation of the generalized mean shift filter is the kernels that have to be used for doing forward and reverse mean shift. For the forward mean shift, we use the Epanechnikov kernel [20] and for reverse mean shift we use the Gaussian kernel. The forward mean shift procedure being stable, for faster implementation the accelerated convergence provided by the Epanechnikov kernel can be used. The reverse mean shift, being a divergent process, it is required to use a more well behaved kernel like a Gaussian one.

6.4 Experimental Results

We now present the experimental results for both cases of deblurring alone and the simultaneous deblurring and denoising problems using the proposed method. Since there are a large

number of image restoration techniques, we compare the performance of the proposed method with only those handling a similar problem, i.e. blind and space varying deconvolution. Initially we demonstrate the performance of the proposed reverse mean shift alone. This is tantamount to deblurring an observation in absence of any noise. First, Lena image has been blurred with a spatially invariant Gaussian blur with variance 3.0. This is shown in fig.6.7(a). The fig.6.7(b) shows the result of the Osher-Rudin filter. The result appears to be an "impressionistic" output of the original. This has been pointed out by the authors themselves in their paper [87]. Figures 6.7(c) and 6.7(d) show the results obtained from the Alvarez-Mazorra method [2] and the Gilboa *et al.* complex diffusion shock filter [41], respectively. The results obtained using these shock filters are not impressive. This is particularly because they are primarily designed to handle both noisy and blurred images. In case there is no noise and only de-blurring needs to be done, then they do not perform well. Fig.6.7(e) shows the result of using a blind deconvolution algorithm ([10]). One can clearly observe certain ringing effects in the result which is very disturbing. Fig.6.7(f) shows the result of the proposed method. Our method is able to successfully restore most of the blurred edges to their original form. This is primarily because the density gradient determines a better estimate of the gradient direction in which the image has to be restored compared to the original gradient being computed. The quantitative evaluation of the various methods based on the peak signal to noise ratio (PSNR) metric is given in table 6.1. The quantitative metric evidently proves that the proposed method performs significantly better than the other methods. There is approximately a 6.5dB improvement over the classical shock filter. It can also be observed that the method performs around 3.5 dB better than existing blind deconvolution algorithm [10].

Next we consider the case where a space varying Gaussian blur is applied to the Lena image. We have applied a radially varying Gaussian blur with the variance ranging from 1.0 in the center to 2.0 at the boundaries in a radially symmetric manner. The results are shown in fig. 6.8. Fig. 6.8(a) shows the input image that has been blurred with a space varying blur. Fig. 6.8(b) shows the result of deblurring using Osher-Rudin shock filter, fig. 6.8(c) shows the result of applying Alvarez-Mazorra shock filter, fig. 6.8(d) shows the result as obtained by applying the complex shock filter proposed by Gilboa *et al.*. Fig 6.8(e) shows the result obtained by applying the blind deconvolution algorithm and fig. 6.8(f) shows the result obtained by the proposed method. As can be seen from the results, the proposed method performs well as

(a) (b) (c)

(d) (e) (f)

Figure 6.7: Result of inverse diffusion of the Lena image which has been blurred as shown in (a) using (b) Osher-Rudin shock filter, (c) Alvarez-Mazorra shock filter, (d) Gilboa *et al.* complex shock filter, (e) blind deconvolution, and (f) the proposed method.

Table 6.1: Quantitative Evaluation of deblurring results based on the PSNR metric. The values are in decibels.

Image Data Set	Input	Osher Rudin	Alvarez Mazorra	Gilboa *et al.*	Blind Deconvolution	Reverse Mean Shift
Lena (Constant Blur)	28.09	21.51	20.95	21.50	28.96	32.22
Lena (Space Varying Blur)	30.03	21.65	20.89	21.73	26.98	33.31
Satellite (Averaging Blur)	29.64	22.18	22.38	22.29	28.99	32.62

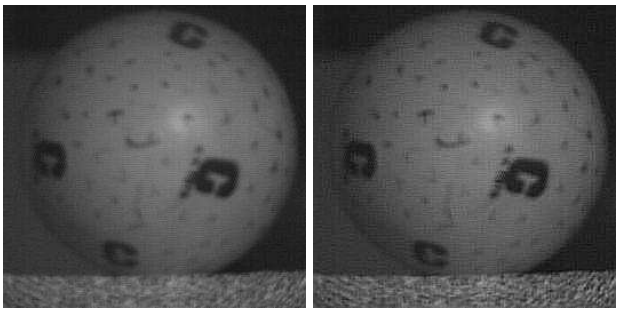
(a) (b) (c)

(d) (e) (f)

Figure 6.8: Result of space varying blind deconvolution with (a) input image using (b) Osher-Rudin shock filter, (c) Alvarez-Mazorra shock filter (d) Gilboa *et. al* complex shock filter, (e) blind deconvolution and (f) using the proposed method.

compared to all other existing methods. This shows that the method can also handle spatially inhomogeneous blur very well. The quantitative comparison based on PSNR as illustrated in table 6.1 establishes that the proposed method successfully achieves much higher performance compared to the shock filter or the blind deconvolution technique. It may be noted that the PSF being spatially varying, the blind deconvolution results in a poorer restoration compared to what was achieved in fig. 6.7(f).

The next experiment was done using real data set where a defocused image of a ball was captured. The result of de-blurring is shown in fig.6.9(b). In this case a few disturbances can be noticed. This is primarily a result of the quantization inherent when the data set is stored using 8 bits/pixel. As a result the de-blurring process generates a few anomalies due to spurious shocks being generated. We suggest the use of a higher accuracy of 16 bits while storing the real data set to avoid the anomalies. Notwithstanding the above, it can still be observed that the



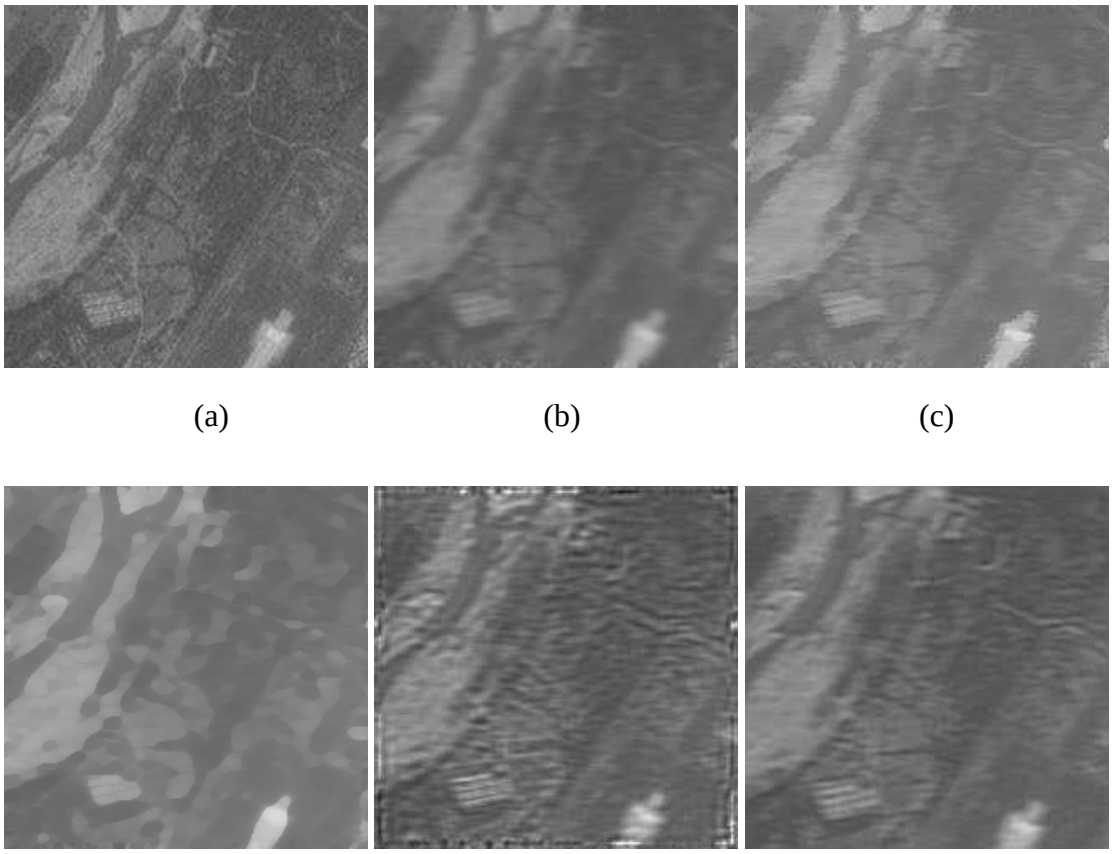
(a) (b)

Figure 6.9: Performance of space varying blind deconvolution for a real aperture image where (a) is the input image and (b) the result using the proposed method.

deblurring by the proposed method does successfully result in a better and sharper image with the letters on the ball being more clearly visible and the mat on which the ball rests also being more sharpened.

Now we consider a case where the PSF is not Gaussian. A spatial averaging blur model (rectangular function) is considered for a satellite image. This is a practical aspect particular to a kind of satellite imagery. Since a finite duration rectangular function can never be obtained by time evolving a Gaussian kernel either in the forward or in the reverse direction, one can never undo the effect of blurring completely in this case. Notwithstanding above, we would like to show that the proposed method yields a very good approximation. We can observe in the fig.6.10 that the method is able to successfully resolve the underlying details in the picture. Fig. 6.10(a) shows the original satellite image and fig. 6.10(b) shows the input observation which is blurred with a rectangular PSF of width 8 in horizontal direction. Fig. 6.10(c) shows the result of applying Osher-Rudin shock filter, fig. 6.10(d) shows the result of applying Alvarez-Mazorra shock filter. The results demonstrate that the existing shock filters do not have much success when the actual PSF is very different from a Gaussian one. Fig. 6.10(e) shows result of deblurring using the blind deconvolution algorithm. Fig. 6.10(f) shows the result of deblurring using the proposed method which is able to resolve the fine details very well. The PSNR metric in table 6.1 clearly shows that the performance of the proposed method for this case is indeed significantly much better (around 10 dB improvement over classical shock filter and 7dB over blind deconvolution).

After having demonstrated the usefulness of the reverse mean shift procedure, we now present the results for simultaneous deblurring and denoising using the proposed generalized mean shift filter. We first test our method on the Peppers image. This image is blurred with a



(a) (b) (c)

(d) (e) (f)

Figure 6.10: Illustration with a 1-D blur kernel where (a) is the original image which has been blurred as shown in (b) in the horizontal direction, and the image is de-blurred using (c) Osher-Rudin shock filter, (d) Alvarez-Mazorra shock filter, (e) blind deconvolution, and (f) using the proposed method.

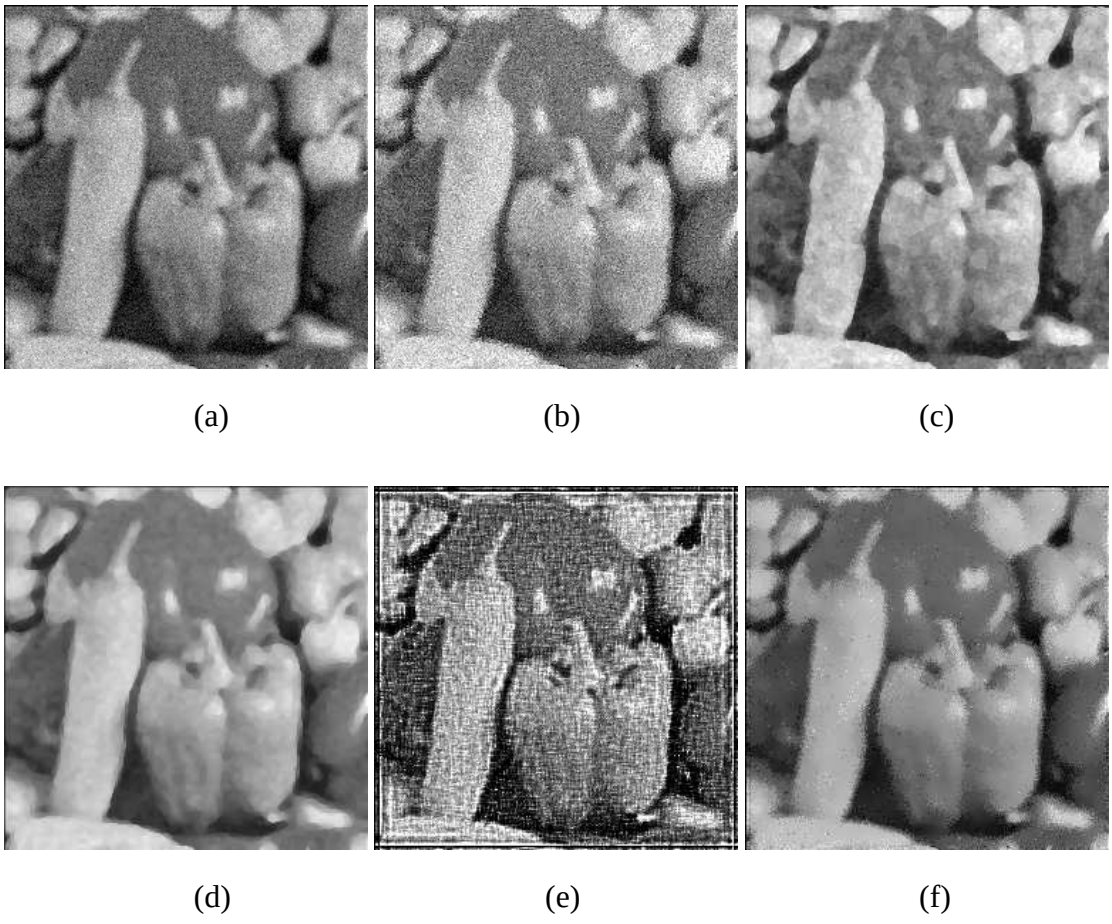
Gaussian PSF with a variance of 2.0 and corrupted with Gaussian noise with a standard deviation of 20. The results are shown in Fig. 6.11. Fig. 6.11(a) is thus, a very noisy input image. Figs. 6.11(b), (c) and (d) are the results of image restoration using Osher-Rudin, Alvarez Mazorra and Gilboa *et al.* complex shock filters respectively. Of the three methods, the result from the complex shock filter appears to be better although the texture is mostly gone. The Osher-Rudin method just does deblurring but no denoising; hence the noise gets boosted. Fig. 6.11(e) shows the result of restoration using the blind deconvolution method [10] which is sharp but very noisy. The result of restoration by the proposed method is given in fig. 6.11(f). As can be seen the result by the proposed method is better than the results of other methods. The noise has been smoothened, the texture has been mostly retained and yet the edges are quite sharp. This is also verified by a quantitative comparison based on PSNR values which is given in table 6.2. We obtain an improvement of about 3.5 dB PSNR over its nearest competitor.

We now consider the case where the standard Mandrill image is blurred with an averaging (rectangular) blur kernel of width 8 pixels and height 4 pixels and is perturbed with an additive Gaussian noise having a standard deviation of 10. Purpose of the experiment, as previously explained in fig.6.10, is to study the performance when the PSF is very different from being a Gaussian one. The noisy input image is shown in Fig. 6.12(a). The result using Osher-Rudin's shock filter is shown in Fig. 6.12(b). The edges are considerably sharpened in this result, however, the noise is not removed as expected (see the nose region). Fig. 6.12(c) shows the result of Alvarez-Mazorra method. The noise is removed, however the resultant image mostly has a piecewise constant appearance. Some of the whiskers are lost here and the facial hairs pick up a different texture. Fig. 6.12(d) shows the result of Gilboa's complex shock filter which does a good job of deblurring and denoising, but the facial texture is lost. Fig. 6.12(e) shows the result of deblurring using the blind deconvolution method. Here, the noise gets boosted very significantly during the deconvolution process. Fig. 6.12(f) shows the result using the proposed method which shows a denoising performance equivalent to that of Gilboa's method. But the edges are much better preserved in this method compared to Gilboa's method as can be seen from the eyes and the whiskers. The PSNR based comparison in table 6.2 shows that the proposed method indeed does a much better restoration as compared to the other methods with a significant 4dB improvement over the next best method. Hence we infer that the proposed method can also efficiently handle PSF that is not specific to the chosen kernel.

Table 6.2: Quantitative Evaluation of simultaneous deblurring and denoising results based on the PSNR metric. The values are in decibels.

Image Data Set	Input	Osher Rudin	Alvarez Mazorra	Gilboa	Blind Deconvolution	Generalized Mean Shift
Peppers Constant blur	21.19	17.75	19.66	20.42	16.10	22.98
Mandrill Averaging blur	19.02	16.65	17.95	18.54	15.77	20.19
Barbara Space varying blur	21.69	18.48	19.42	19.86	18.61	22.76

The algorithm is then tested on the Barbara test image which is blurred with a space varying Gaussian blur kernel where the variance is varied from 1.0 in the center to 3.5 at the boundaries. The image is also perturbed by an additive Gaussian noise with a standard deviation of 15. Fig. 6.13(a) shows the noisy input image. Fig. 6.13(b) shows the result of restoration using Osher-Rudin shock filter, fig. 6.13(c) shows the result of applying Alvarez-Mazorra shock filter, fig. 6.13(d) shows the result of applying the complex shock filter by Gilboa *et al*. Of the three methods the result from the complex shock filter is better in terms of denoising and deblurring, although the scene texture is badly affected. Fig 6.13(e) shows the result of blind deconvolution. The noise gets badly boosted in this method, though the image is much sharper. Fig. 6.13(f) shows the result of deblurring and denoising using the proposed method and the result is much better as compared to the result obtained by the complex shock filter as the noise is substantially eliminated and the strong edges are better deblurred. However, we do seem to have lost some texture in the scene in this case. A quantitative comparison in terms of PSNR improvement given in table 6.2. The efficacy of our method is evident from the PSNR measures given in table 6.2. The proposed method clearly performs much better (at least 3dB) than the other methods. All these results substantiate our claim that the forward and reverse mean shift algorithm can be effectively combined to simultaneously deblur and denoise images.



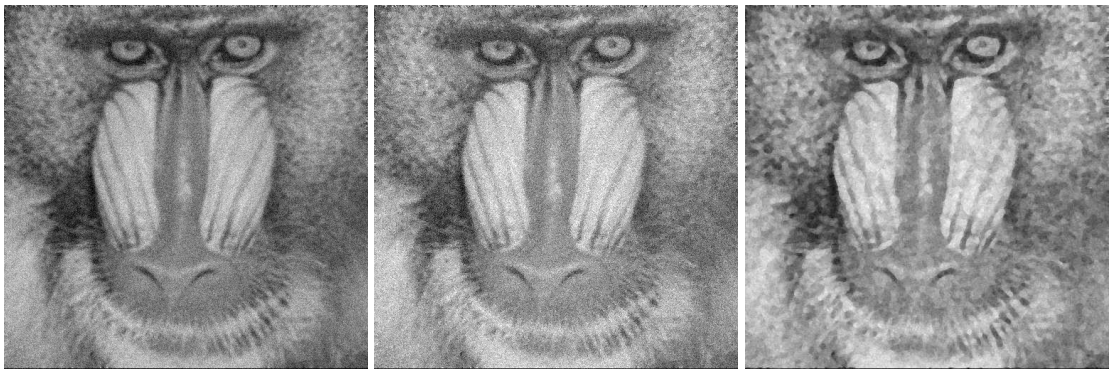
(a) (b) (c)

(d) (e) (f)

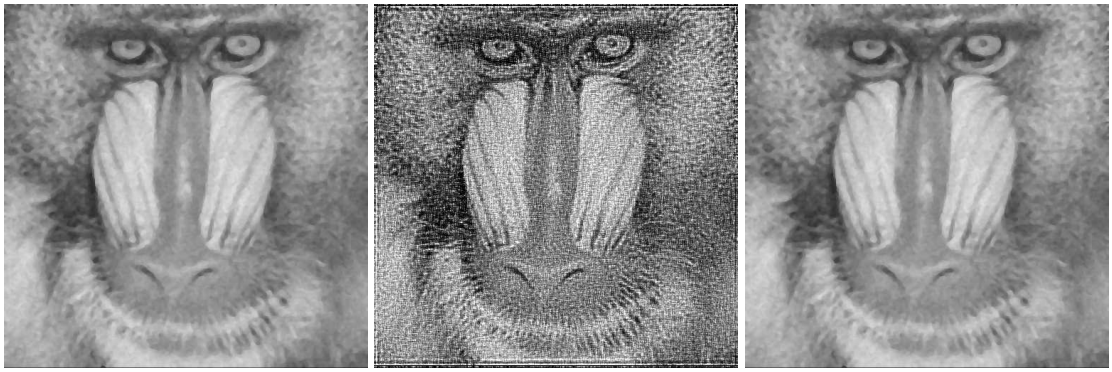
Figure 6.11: Denoising and deblurring results of the Pepper image shown in (a) using Osher-Rudin shock filter in (b), using the Alvarez-Mazorra filter in (c), using Gilboa's complex shock filter in (d), using blind deconvolution method in (e) and the proposed method in (f).

6.5 Conclusion

In this chapter, we have addressed the problem of blind restoration of images and have shown that it can be convincingly solved using the proposed idea of generalized mean shift filter which combines both forward and reverse mean shift filters using a switching criterion. The results presented justify the suitability of the non-parametric approach towards addressing the blind and space varying image restoration problem. The proposed technique does not require the PSF to be of any specific form. It can also handle a PSF that is spatially varying. Since the PSF is never estimated explicitly, no modeling of the PSF or the image field is required. It is a very simple yet powerful technique for image restoration, which is also computationally very efficient. We rarely required more than 10 iterations of eqn. to converge to the quality solution as per the suggested stopping criterion.



(a) (b) (c)



(d) (e) (f)

Figure 6.12: Blind restoration of the Mandrill image shown in (a) using Osher-Rudin shock filter in (b), using the Alvarez-Mazorra filter in (c), using Gilboa's complex shock filter in (d), using blind deconvolution method in (e) and the proposed method in (f).

While the current method handles the problem of deblurring quite well, and also handles deblurring in the presence of noise, the results obtained by deblurring do not preserve the correlation of the restored image with depth in the scene. This is because, the depth in the scene is correlated by the diffusion equation. Hence, in the next chapter we explore a technique for restoring blurred images by using the reverse heat equation so as to also preserve the correlation of the restored result with depth in the scene.

(a) (b) (c)

(d) (e) (f)

Figure 6.13: Space varying restoration results of the Barbara image shown in (a) using Osher-Rudin shock filter in (b), using the Alvarez-Mazorra filter in (c), using Gilboa's complex shock filter in (d), using blind deconvolution method in (e) and the proposed method in (f).

Chapter 7

Restoration using Stabilized Reverse Heat Equation

The problem that is addressed in this chapter is one of de-blurring an image $Y(x, y)$ that has been blurred by a blurring kernel $h(x, y)$ representing some physical process. We recollect that as seen in eqn.(6.1)this problem is modeled by the following convolution relation:

$$Y(x, y) = \int \int U(t, \tau)h(x, y; t, \tau)dtd\tau \tag{7.1}$$

As is normally assumed the function $h(x, y)$ has the properties that it is non-negative, and the integral of the function $h(x, y)$ is unity.

As shown by Guichard and Morel [48], the convolution of an image with a kernel is proportional to its Laplacian and can be modeled by the heat equation as follows:

$$\frac{\partial u(x, y, t)}{\partial t} = c\triangle u(x, y, t),\ u(x, y, 0) = I_0(x, y) \tag{7.2}$$

Here u represents the image being diffused using the heat equation, c is the diffusion coefficient, $\triangle u$ is the Laplacian of u and $I_0(x, y)$ is the initial deblurred image. The use of the heat equation has also been used by Witkin [129] and Koenderink [62] in the formation of the notion of scale space. An important work along these lines has been use of anisotropic diffusion for edge preserving denoising by Perona and Malik [90]. While there has been much work done on the forward aspect of heat diffusion [106], relatively less work has been done on the reverse aspect of the heat equation. The reverse heat equation is ill-posed and so its use has been limited. Osher and Rudin [109] in their work proposed the use of "shock" filters which are hyperbolic partial differential equations. These are stable and have good convergent properties. However,

they provide piecewise constant results and do not achieve true deblurring. Another work has been the use of stabilized inverse diffusion equations [91] by Pollak *et al.*. They also use an approximation to the inverse diffusion which has a physical motivation. However, they also do not approach the true reverse heat equation. A very recent work [14], has explored the use of reverse heat equation with a non-local means based additional criterion. They perform alternating steps of reverse heat and non-local regularization. The alternate formulation that we provide is simpler. Here, we solve the problem of deblurring by using the reverse heat equation. Since the reverse heat equation is ill-posed we stabilize it by controlling the disruption of edges. This is achieved by adding a normal component of the heat equation in the forward direction. We also formulate a stopping criterion for terminating the reverse heat equation process when the deblurring of the image is completed. In the next section we discuss the reverse heat equation and its stabilization.

7.1 Stabilized Backward Heat Equation

The reverse heat equation is given as

$$\begin{aligned} \frac{\partial u}{\partial t} &= c\triangle u \\ u(x,y,\tau) &= I(x,y) \end{aligned} \tag{7.3}$$

where $\triangle u$ denotes the Laplacian of u, $I(x,y)$ is the blurred observation and c is the diffusion coefficient. We have to find the solution

$$u(x,y,0) = I_0(x,y). \tag{7.4}$$

This is achieved by reversing time in the heat equation

$$\frac{\partial u}{\partial t} = -c\triangle u,\ u(x,y,0) = I(x,y). \tag{7.5}$$

However implementing eqn(8.3) can be done only for a few time steps and then the resulting image blows up due to the high pass nature of the resulting operation. It boosts the noise, especially along the edges where the Laplacian has high values. Explicit edge information can be considered in the heat equation by considering the geometric form of the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial \eta^2} + \frac{\partial^2 u}{\partial \zeta^2} = u_{\eta\eta} + u_{\zeta\zeta}. \tag{7.6}$$

Here η refers to the normal and ζ to the tangential direction. The diffusion along the normal is given by

$$u_{\eta\eta} = \frac{u_{xx}u_x^2 + 2 * u_{xy}u_x u_y + u_{yy}u_y^2}{u_x^2 + u_y^2} \tag{7.7}$$

and the diffusion along the tangent is given by

$$u_{\zeta\zeta} = \frac{u_{xx}p_x^2 - 2 * u_{xy}u_x u_y + u_{yy}u_y^2}{u_x^2 + u_y^2} \tag{7.8}$$

Since the diffusion along the normal diffuses across the edges and diffusion along the tangent continues along the edges, the blurring in an image is caused more due to diffusion along the normal. Therefore in order to stabilize the reverse diffusion, the reverse diffusion across the edges has to be done at a slower rate as compared to reverse diffusion along the tangent. The diffusion along the normal is a more divergent process and has to be done at a slower rate. Thus in order to stabilize the reverse diffusion we add a forward component of diffusion along the normal. The resultant stabilized form of the heat equation is given by

$$\frac{\partial u}{\partial t} = -c\triangle u + \beta u_{\eta\eta} \tag{7.9}$$

Here we use $c > \beta$ in order to ensure the overall reverse nature of the diffusion. The diffusion is carried out until a stopping criterion is reached which corresponds to the initial required solution $I(x, y, 0)$.

7.2 Relation to Other Techniques

We now consider the analysis of shock filters and Kramer's algorithm as explored by Guichard and Morel [48]. Osher and Rudin in their "shock filter" formulation, proposed the following equation

$$\frac{\partial u}{\partial t} = -sign(\triangle u)|\nabla u| \tag{7.10}$$

where ∇u is the gradient of u. This equation enhances the Marr-Hildreth edges. Kramer defined a filter that sharpens blurred images by replacing the gray level value at a point x by either the minimum or the maximum of the gray level values in a circular neighborhood. Guichard and Morel [48] proved that the PDE underlying the Kramer filter is

$$\frac{\partial u}{\partial t} = -sign\left(\nabla^2 u(\nabla u, \nabla . u\right) \tag{7.11}$$

where ∇u is the gradient of u and instead of the Laplacian, the directional second derivative is used. This filter enhances the Canny edges. While, both these filters perform edge enhancement,

they are not equivalent to the actual reverse heat equation, as compared to the proposed approach which is based on the reverse heat equation itself.

The non-local reverse heat equation proposed recently [14] is closely comparable to the proposed technique. The non-local reverse heat equation is given as

$$\frac{\partial u}{\partial t} = -\triangle u + \lambda N L_0 u \tag{7.12}$$

where

$$N L_0 u(x) = \frac{1}{C(x)} \int \exp \frac{G_\sigma * |u(x) - u(y)|^2(0)}{h^2} u(y) dy, \tag{7.13}$$

where $C(x)$ is the normalizing factor, h acts as a filtering parameter and G_σ is the Gaussian kernel with standard deviation σ. Here NL_0 is the non-local means filter [13] and it means that $u(x)$ is replaced by a weighted average of $u(y)$. The weights are significant only if a Gaussian window around y looks like the corresponding Gaussian window around x. This approach is certainly interesting. The main difference, as is evident by comparing eqns(7.9) and (7.12), is that in our approach we rely more on the local normal component of the heat equation to stabilize the equation as compared to the non-local component used by Buades *et al.*. Since the objective has been to closely approximate the reverse heat equation, the damping by using a normal component of the heat equation itself satisfies this criterion in a better way.

7.3 Stopping Criterion

Consider the eqn(7.3) using which we have to estimate the initial condition given in eqn(7.4), i.e. we have to estimate the value of $u(x, y, 0) = I_0(x, y)$. The eqn(7.5) has to be stopped when $u(x, y, t) = u(x, y, 0)$. However, here we do not know the value of $u(x, y, 0)$. An observation that can be used is that the eqn(7.3) is valid only till time $t = 0$ and it breaks down if we go beyond this time. The modified reverse heat equation given in eqn(7.9) will not be valid for the value of time $t < 0$. Hence, beyond this point the solution will degenerate rapidly. This observation can be used for stopping the reverse heat equation. If we consider the image as a manifold with at least C_2 continuity, the degeneration of the solution can be detected by the divergence of the curvature. In eqn (7.9), since the normal component is added, the tangential term is diffused in reverse direction more rapidly. The tangential term corresponds to curvature driven motion. Since, the curvature driven term is reversed at a faster rate, the degeneration in this term happens before degeneration in the normal component. Hence, the divergence of

curvature can be used as an indicator that the image approximates the desired initial image. Further, when the divergence of the curvature happens, the degeneration in the normal component would happen in a few more time steps, based on the difference of the weightage given to them. And hence, the degeneration of the curvature is a good indicator for stopping. The curvature is given by

$$\kappa = \frac{u_{xx}u_x^2 - 2u_{xy}u_x u_y + u_{yy}u_y^2}{(u_x^2 + u_y^2)^{3/2}} \tag{7.14}$$

The eqn(7.9) is stopped when the change in curvature exceeds a threshold, i.e. $\kappa_t > \theta$. Comparatively, the shock filter formulation [87] is a convergent procedure and does not require a stopping criterion. In the non-local means based reverse heat equation [14], the authors suggest stopping the reverse heat equation when the value of the Laplacian exceeds twice the value of the initial Laplacian. But using this criterion results in certain artifacts being generated due to the degeneration of the solution.

7.4 Implementation Details

In the implementation of the reverse heat equation, the boundary conditions were assumed to be Neumann boundary conditions, i.e. the gradient is zero along the boundary. We now consider the values of the various constants. In eqn(7.9) the values of c and β are chosen to be small and $c > \beta$. Additionally, they must be small enough to maintain Courant-Friedrichs-Lewy (CFL) conditions. Here we have chosen values of c as 0.2 and β as 0.02. These values have been empirically chosen. The value of θ used for setting the threshold for change in curvature was 0.3

7.5 Results

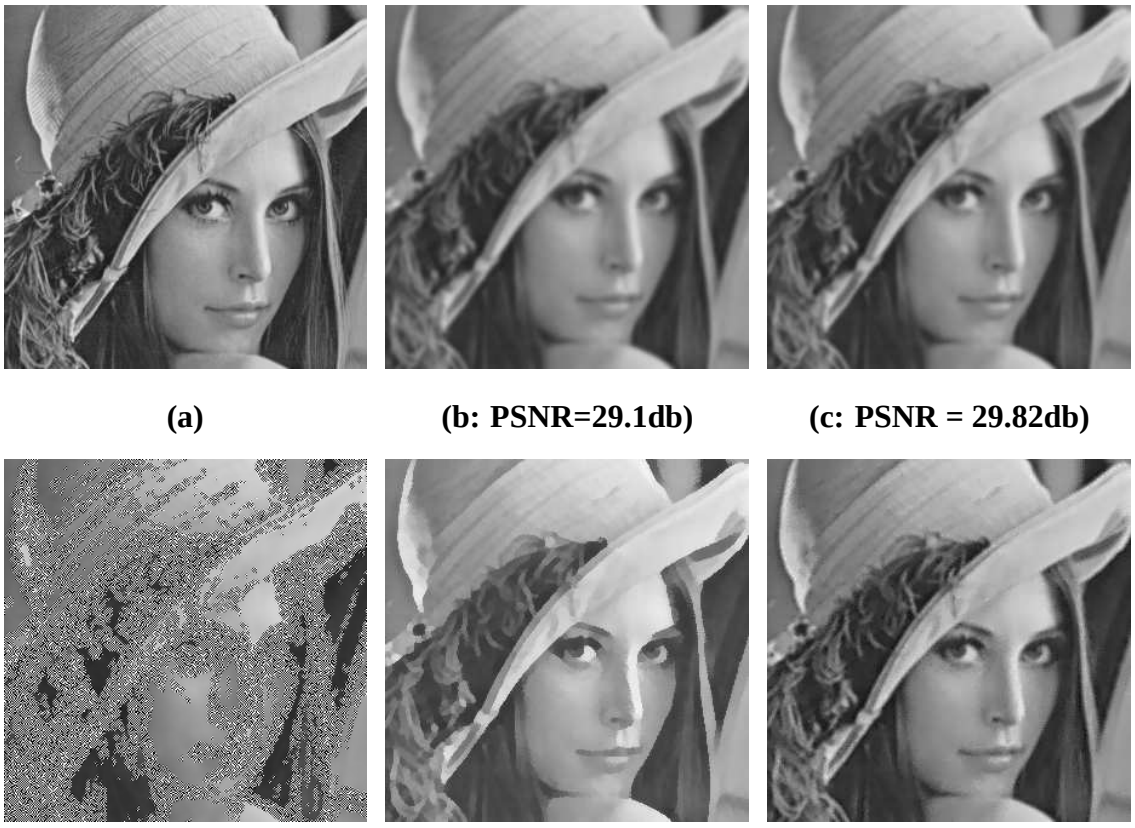
We first justify the use of the proposed technique by considering the performance of the reverse heat equation when used for deblurring without any modification. The results using the reverse heat equation are shown in fig. 7.1(c)&(d). Fig. 7.1(a) shows the original Lena image that is blurred with a constant Gaussian blur with standard deviation 3.0. Fig. 7.1(b) shows the blurred input image. Fig. 7.1(c) shows the result of using the reverse heat equation for 2 iterations. As seen in the figure use of reverse heat effectively starts deblurring the input. But as can be seen in fig. 7.1(d), which shows the resultant image after applying the reverse heat equation

for 10 iterations, this equation is unstable and the values blow up quickly. We next evaluate the proposed technique by experimentally comparing the method with the shock filter method [87]. Note that the proposed technique as well as the shock filter method do not use any information about the nature of the blurring function and both perform blind deconvolution. Fig. 7.1(e) shows the result of applying the shock filter. As can be seen while the shock filter preserves the strong edges, the weak texture edges are strongly affected in this method. This is because the shock filter does not approximate the reverse heat equation appropriately. Fig. 7.1(f) shows the result using the proposed technique. As can be seen, the result achieves true deblurring as can be seen from the texture on the hat and hair. Here the method took 19 iterations (as compared to original reverse heat blowing up in 10 iterations) before the stopping criterion was satisfied over the entire image. The result is closely comparable to the original image. Quantitative comparison in terms of PSNR values establish around 1.5 db improvement over the input image and around 10db improvement over the shock filter method.

We next consider an experiment of performing blind *space varying* deblurring. We blur a sand texture image obtained from the Brodatz texture database with a Gaussian blur function with the standard deviation being increased from 1.0 to 1.5 from left to right. Fig. 7.2(a) shows the input image and Fig. 7.2(b) shows the blurred input image. The input image is restored using the shock filter and this result is shown in fig. 7.2(c). As can be seen, the shock filter method results in a piecewise constant resultant image where much of the texture information is lost. This particularly emphasizes the need for the proposed technique. Fig. 7.2(d) shows the result of the proposed technique. As can be seen, using the proposed technique one can obtain deblurring of the input image with very little artifacts being present. The results quantitatively show around 3db improvement over the input image and around 3.7db improvement over the shock filter method.

7.6 Conclusion

Here we present a technique which addresses the challenging problem of blind space varying deblurring. The problem is modeled using the heat equation and deblurring is framed as a problem of solving the reverse heat equation. The unstable nature of the reverse heat equation is addressed by adding the normal component of the reverse heat equation in the forward direction. A curvature based stopping criterion appropriately stops the reverse heat equation



(a) **(b: PSNR=29.1db)** **(c: PSNR = 29.82db)**

(d: PSNR=9.33db) **(e: PSNR=21.56db)** **(f: PSNR=30.48db)**

Figure 7.1: Deblurring results for a constant Gaussian blurred Lena image (a) is the original Lena image which is blurred as seen in (b) using constant Gaussian Blur. (c) shows the result of using the original reverse heat equation for 2 iterations and (d) shows the result on using 10 iterations of the reverse heat equation. This shows the instability of the reverse heat equation. (e) shows the result of applying the shock filter [87] and (f) shows the result by using the modified reverse heat equation.

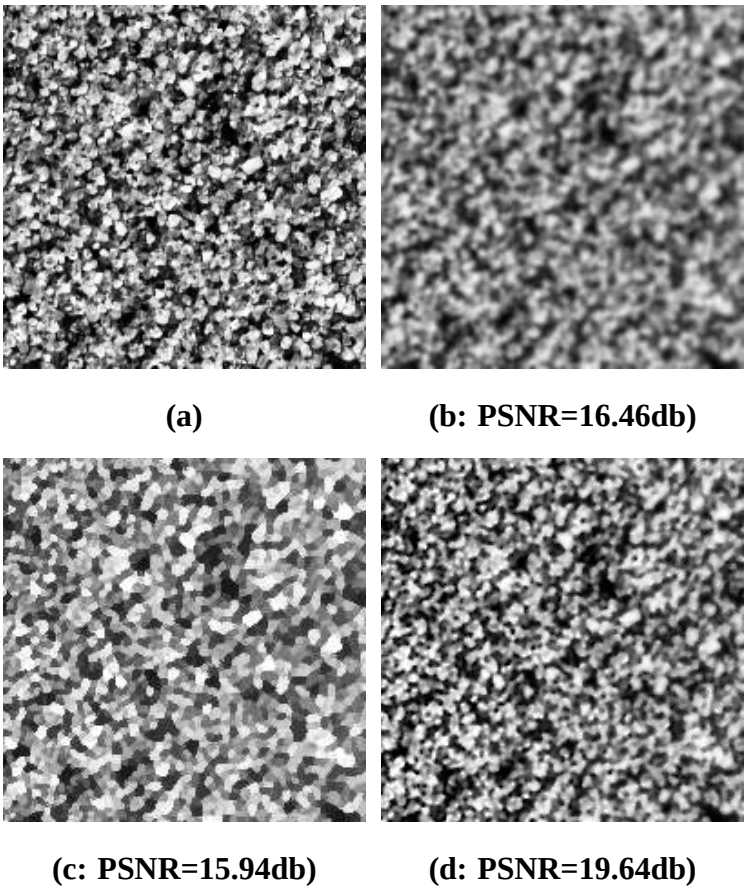


Figure 7.2: Deblurring of a space varying blurred input image where (a) is the input sand texture image that is blurred with a space varying (ramp) Gaussian blur. (c) is the result of applying the shock filter [87] and (d) is the result by using the proposed method.

without artifacts being introduced in the solution. The results obtained justify the feasibility of the proposed theory.

We next consider the use of the reverse heat equation to obtain depth from a single defocused image in the next chapter.

Chapter 8

Depth from a Single Defocused Image

Shape from a single image [1] 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, inspite of the ease with which the human system is able to achieve the same.

(a) **(b)**

Figure 8.1: A sample image of a scene captured with a low depth of field and the rendering of the same scene in 3D based on the depth estimated using the proposed method.

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

[1] Shape from shading problem [55] which recovers depth from a single observation is not considered here as it is a model based approach, the model being a known reflectance map.

lenses based on the range of depth variation in a scene. This artifact has been used in computer vision 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 ([19],[35]) 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.8.1(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.8.1(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 rendering of the 3D scene shown in fig.8.1(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 sketch was proposed by Marr [75]. This forms a philosophical basis for our work. A successful approach for single image based structure recovery has been that by Criminisi *et al.* [24]. 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 ([53],[54]). Another approach [110] based on similar lines uses images of scenes and their associated ground-truth depth maps. It discriminatively trains an MRF using multi-scale 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 [123], 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.

8.1 Is depth from a single defocused image possible?

In shape from a single image using geometry, as done by Criminisi *et al.*, one can use the vanishing points and obtain distances from the vanishing point. This would establish a relative scaling. Hence, it is feasible to obtain shape from a single image using geometric techniques. One could also learn the geometric classes using data sets and classify the various geometric classes as done by Hoiem *et al.*. The methods mentioned would not work if one is provided with a single defocused image. This is because there would be ambiguity in the precise location of the vanishing point. Moreover, the blur would also affect the classification of the geometric classes. In such a case depth from a single defocused image would seem impossible.

However, here one could make use of the depth from defocus techniques which we have considered so far. The techniques for depth from defocus are based on estimating the blur. All the methods proposed so far (as reviewed in chapter 2) use the relative blurring between two differently defocused images. Given two differently defocused images one could use the techniques discussed in chapters 3, 4 and 5 to estimate the depth accurately. However, given a single defocused image one cannot use the relative blurring between the two images. Here, the challenge is to estimate the blur from a single defocused image. One approach would be to restore the image and estimate the relative blur difference between the restored and the original defocused image. This can however not be used generally as the process of deblurring or blind deconvolution destroys the depth relationship between objects. The resultant depth estimate would not be coherent. One can estimate the depth from a restored image only if the process of deblurring carefully preserves the depth relationship.

In order to understand the depth preserving deblurring process we must take into account the characteristics of the defocus blur. We have seen in chapter 3 that the defocus blur can be modeled using the linear diffusion process. Hence, the depth based blur characteristic can be preserved if we use the linear diffusion process in the reverse direction. The reverse heat equation however is unstable. We have seen in the previous chapter how we could stabilize the reverse heat equation for deblurring. We build up on the ideas of the previous chapter to obtain a method for estimating the relative depth.

8.2 Reverse Heat Equation

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

$$\frac{\partial u}{\partial t} = c\left(\frac{\partial^2 u(x,t)}{\partial x^2}\right)$$
$$u(x,\tau) = I(x), \tag{8.1}$$

where $I(x)$ is the blurred observation and c is the diffusion coefficient. 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

$$u(x,0) = f(x). \tag{8.2}$$

This is achieved by reversing time in the heat equation and the resultant is

$$\frac{\partial u}{\partial t} = -c\left(\frac{\partial^2 u(x,t)}{\partial x^2}\right),\ u(x,0) = I(x). \tag{8.3}$$

As discussed in the previous chapter, the above equation degenerates rapidly. However, this equation has been used for restoring images. The idea of using the reverse heat equation for restoring images was first proposed by Gabor in 1965 [70]. Recently the use of the reverse heat equation has been advocated by Buades *et al.* [14]. They propose the use of the reverse heat equation regularized by using the "non-local means" constraint. In this chapter we use the reverse heat equation as given in eqn.(8.3). 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 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. Note that here we do not use the geometrically stabilized form of the reverse heat equation as proposed in the previous chapter so as to preserve the depth based correlation of the deblurring as closely as possible. Based on the experience of deblurring in the previous chapter we propose a way of using the reverse heat equation to get an estimate of depth.

Koenderink [62] 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.(8.1) 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.(8.1) 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

$$\frac{\partial u}{\partial t} = -\beta(x)c\left(\frac{\partial^2 u(x,t)}{\partial x^2}\right), \tag{8.4}$$

where $\beta(x)$ is given by

$$\beta(x) = \begin{cases} 1 \text{ if } |\nabla u - \overline{\nabla u}| < \theta \\ 0 \text{ else} \end{cases} \tag{8.5}$$

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.(8.4) 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 d is then given by

$$\hat{d}(x) = \int_0^{t(x)} c(x,t')dt'. \tag{8.6}$$

Here, $\hat{d}(x)$ is the approximate estimate of the depth at the location x. An estimate of $\hat{d}(x)$ for the scene in fig. 1 is shown in fig. 3. Using the stopping condition proposed in the previous chapter results in better deblurring performance, but the depth estimate obtained is quite weak since the stopping condition proposed in the previous chapter stops the reverse heat equation relatively early so as to prevent any degeneration in the restored image. Here, we need a better depth estimate and so we use the stopping criterion as given in eqn.8.5.

It was demonstrated in [19] 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.

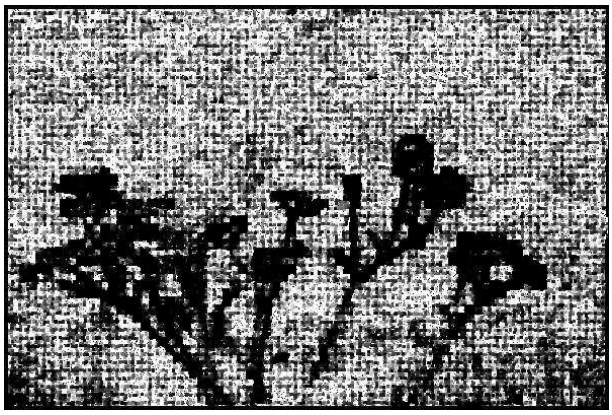


Figure 8.2: The integrated diffusion coefficient of the reverse heat equation for the scene shown in fig.8.1. The convention darker is the focused region (closer) is used throughout this paper.

8.3 DFD Using Graph Cuts

Using the reverse heat equation coefficient we can obtain an approximate estimate of the depth in the scene using eqn.(8.6). 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 [45]. We have discussed the use of this MAP-MRF framework for estimating the depth using two images in chapter 5. Here, however, since we are using only a single image, the current usage of the MAP-MRF framework is different, since here the likelihood used is different.

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.(8.6) 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. $\mathcal{L} = \{l_1, \cdots, l_n\}$. These labels are assigned over the image for the relative depth in the scene given by $\mathbf{w}$, one label w_x at each pixel x, that maximizes the posterior probability given by the Gibbs distribution

$$p(\mathbf{w}|\mathbf{D}) = \frac{\mathbf{p}(\mathbf{D}|\mathbf{w})\mathbf{p}(\mathbf{w})}{\mathbf{p}(\mathbf{D})} = \frac{\mathbf{1}}{\mathbf{Z_w}} \exp(-\mathbf{E}(\mathbf{w})), \tag{8.7}$$

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

$$E(\mathbf{w}) = \sum_{\mathbf{x}} \left(\phi(\mathbf{D}|w_x) + \sum_{\mathbf{y} \in \mathcal{N}} \psi(w_x, w_y). \right) \tag{8.8}$$

Figure 8.3: The binary valued edge map obtained using Canny edge detector that is included in the data likelihood for the scene shown in fig.8.1(a).

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

The prior in the scene $\psi(w_x, w_y)$ chosen to have the form

$$\psi(w_x, w_y) = ||(w_x - w_y)||_2 \tag{8.9}$$

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

$$\phi(\mathbf{D}|\mathbf{w_x}) = \begin{cases} \left((w_x - \hat{d}(x)\right)^2 \text{ if } M(x) = 1 \\ \eta \text{ if } M(x) = 0 \end{cases} \tag{8.10}$$

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 $M(x)$ and fig.8.3 shows a scene with its edge map M. Since the nearly homogeneous regions do not offer any information about the depth or the associate diffusion, we give more weight to the edge pixels. Accordingly the value of η is appropriately chosen.

We minimize eqn.(8.8), thereby maximizing the posterior probability using graph cuts ([11],[12]). 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 [12] 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

$$E(w_1, \cdots, w_n) = \sum_{i<j} E^{i,j}(w_i, w_j). \tag{8.11}$$

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. These labels provide a discrete approximation of w and the corresponding minimization is same as minimization of $E(\mathbf{w})$ in eqn. 8.8. For an energy function of this form it has been proved by Kolmogorov and Zabih [64] 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:

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \tag{8.12}$$

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 $\hat{\mathbf{w}}$ by minimizing eqn.(8.8).

8.4 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 a defocused input image fig. 8.4(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. 8.4(b) and the ground truth for the data is shown in fig. 8.4(c). The quantitative comparison of the estimated depth map with the ground truth showed that 95% of the pixels are labeled accurately.

We then tested our method on a general outdoor image of a beautiful spring scene fig.8.5(a). This image is captured with a moderated 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.8.5(b). The result is obtained using 16

(a) **(b)** **(c)**

Figure 8.4: 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.

labels of depth with a gradient degeneration threshold θ in eqn.(8.5) of 0.2. The corresponding data likelihood term from the reverse heat equation is shown in fig.8.2 and the edge map for the scene is shown in fig.8.3. 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.8.5(a). Fig.8.1(b) shows a rendering of the 3D scene.

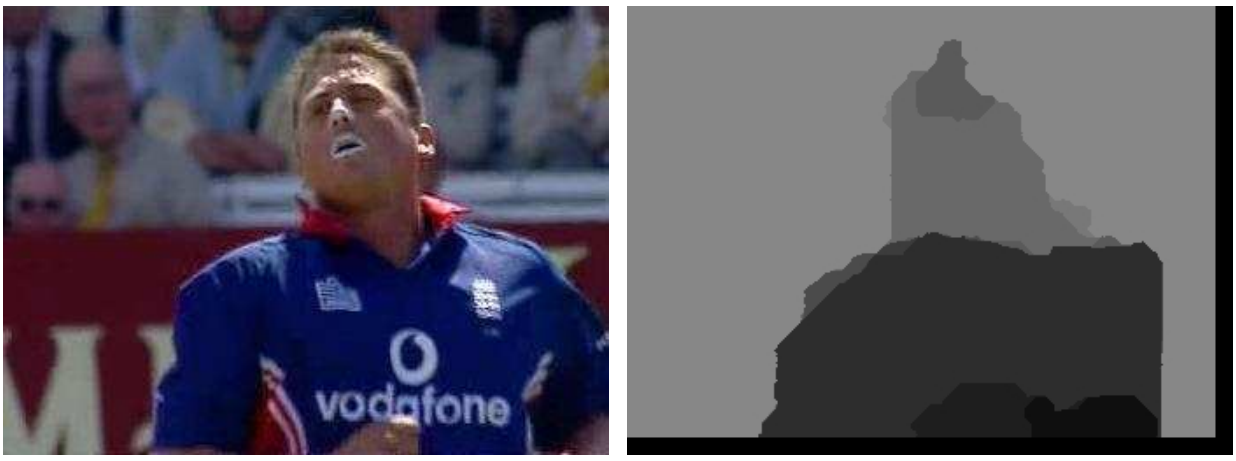
We next consider an image taken from a sports scene shown in fig.8.6(a). As can be seen, here the player [2] is in focus and the spectators are out of focus. This image is interesting due to the very low amount of texture present in the scene. The scene structure is estimated using the proposed technique and the result is shown in fig.8.6(d). 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.8.7(a) shows a room with various artifacts. 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.8.7(b). This shows that the algorithm is able to estimate the relative layers of depth even in such challenging situations.

[2] Darren Gough of England

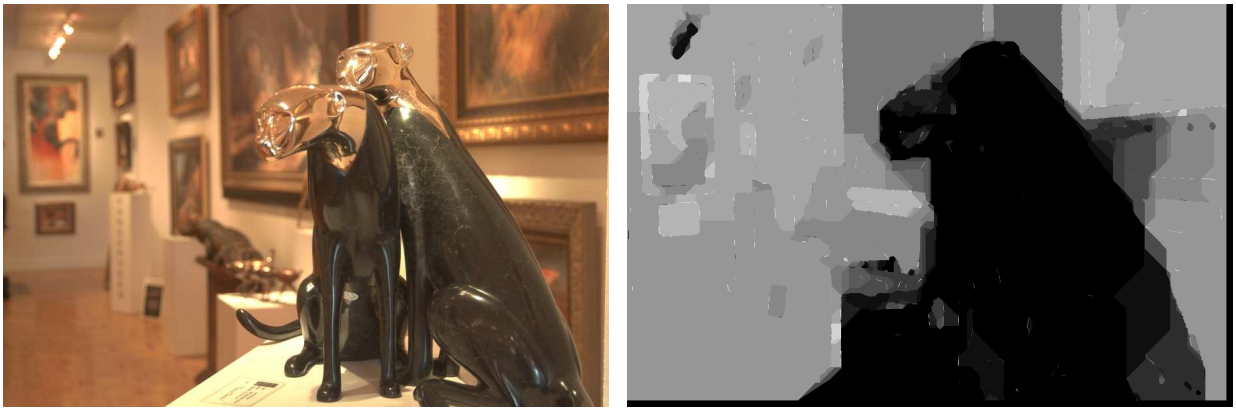
(a) **(b)**

Figure 8.5: An outdoor scene shown in (a) and the resultant depth map estimated using the proposed method. Note that darker regions correspond to focused regions and lighter regions correspond to defocused regions.



(a) **(b)**

Figure 8.6: 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).



(a) **(b)**

Figure 8.7: A scene with complex lighting conditions is shown in (a) and the resultant depth map is shown in (b).

8.4.1 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. 8.8. A similar amount of defocus blur is generated on both sides of the defocus cone as is illustrated in fig.8.8. 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 [32]) where the focus is interchanged between background and foreground objects. In the first case as seen in fig.8.9(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.8.9(b). Note, that the depth map shown also captures the details like the rightmost doll has its front portion more in focus than the back part and the depth variation around the hands is also reflected properly.

When we consider the other image of the dolls where the foreground is in focus and the background is out of focus as seen in fig.8.10(a), then the resulting depth map shown in fig.8.10(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.8.9(b) and 8.10(b) shows that the recovered depth maps are mutually very consistent.

Favaro *et al.* [32] have used the two images given in figures 8.9(a) and 8.10(a) to compute

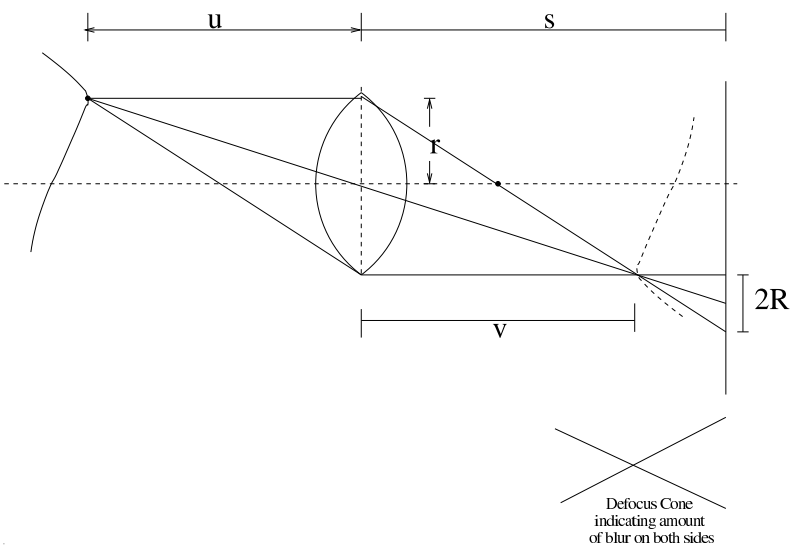
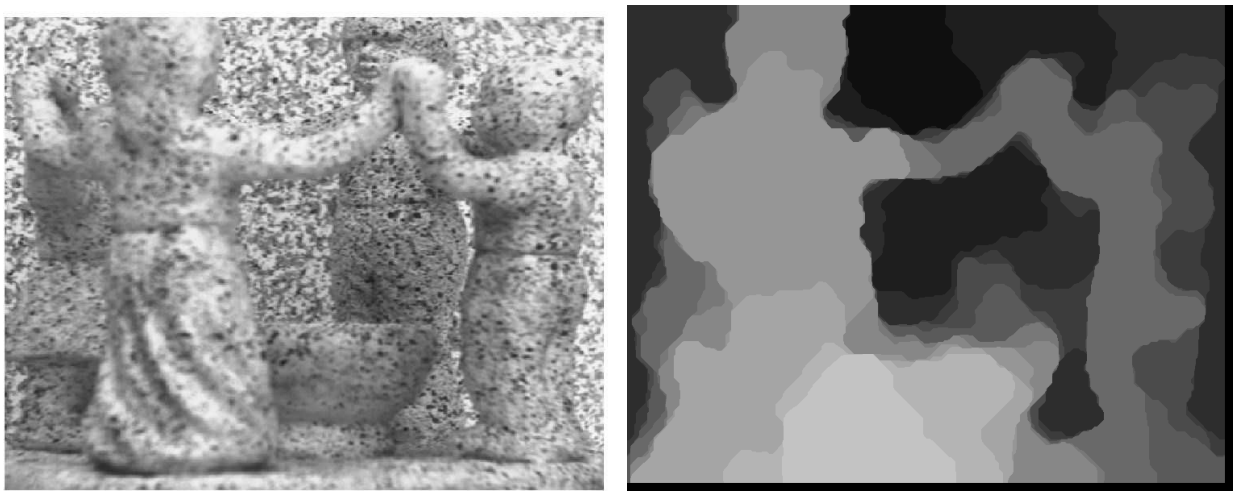
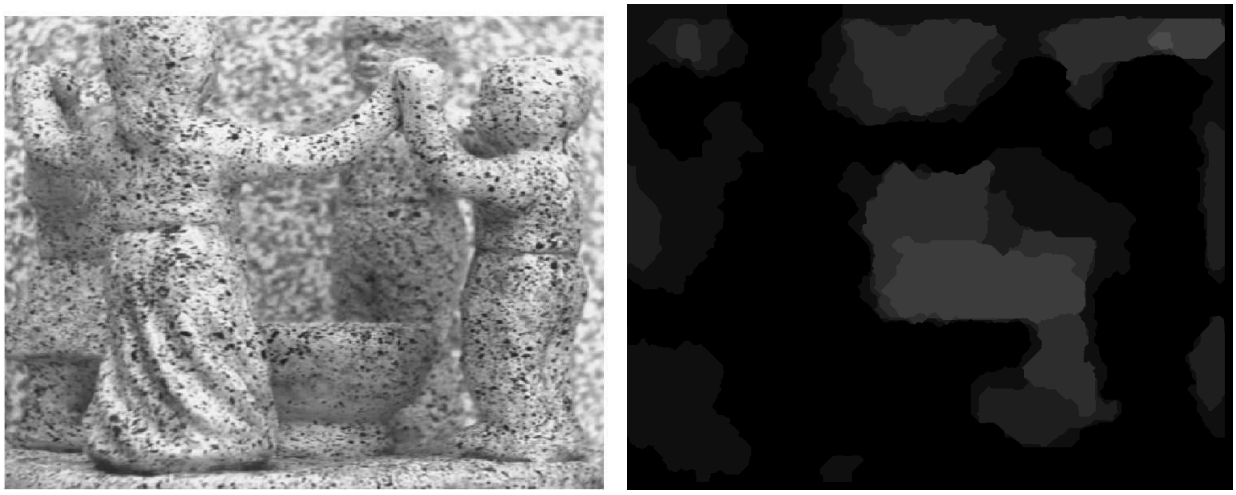


Figure 8.8: Illustration of image formation in a real aperture lens.



(a) **(b)**

Figure 8.9: The dolls scene (courtesy [32]) with the foreground defocused as shown in (a). The depth map from the proposed method is shown in (b).

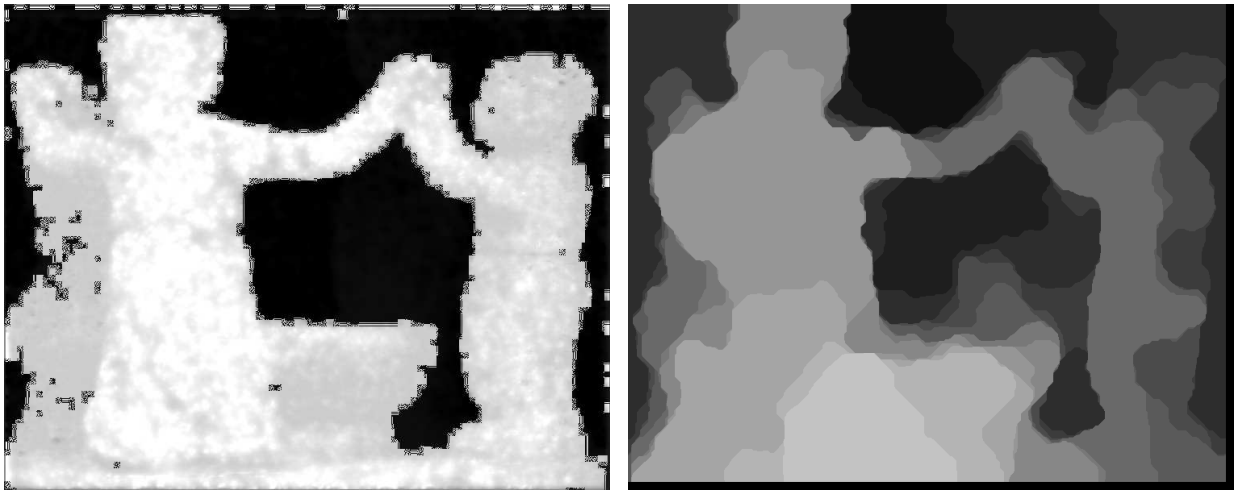


(a) **(b)**

Figure 8.10: The dolls scene (courtesy [32]) with the background defocused as shown in (a).The depth map from the proposed method is shown in (b).

the depth. In fig.8.11, we compare their results, with the results obtained from the proposed method. The methods cannot ideally be compared since the authors in their method [32] have used two images while we have used a single image. The result obtained by the method proposed by Favaro *et al.* is shown in fig.8.11(a) and those obtained by the proposed method (for the foreground defocused case for uniformity in comparison) is shown in fig.8.11(b). As can be seen in fig.8.11, the results obtained by the proposed method are definitely comparable and in some cases, as in around the rightmost doll, the depth map from the proposed method shows more detail.

In a recent work, Hasinoff and Kutulakos [50] have proposed a method based on confocal constancy property in which the method of depth from focus can be seen to be a pixel matching operation. However, as is common in depth from focus techniques, their technique requires a large number of observations (hundreds of images, 13 aperture and 61 focal settings) of high resolution. We have used just one of their images as shown in fig.8.12(a). Using the proposed technique we obtain the depth map as seen in fig.8.12(b). As can be seen, we are able to obtain consistent results with the most focused region being away from the center towards the left side of the box. However, the left most end of the box is also slightly defocused. This is clearly seen in our result. The right most end of the box is most defocused. Even this fact is seen in the result. An important point here is the background which is defocused is correctly identified as being at the same level. Fig.8.12(c) shows the result obtained by using their confocal stereo method



(a) **(b)**

Figure 8.11: Depth map obtained from the dolls scene for the method proposed by Favaro *et al.* [32] using two images is shown in (a) and the result obtained by the proposed method is shown in (b).

which uses hundreds of images. As can be seen, the results obtained are very comparable. In certain regions like towards the center (near the edge), the depth map shown in the confocal stereo method appear to be at variance from the adjoining areas, whereas, the depth map from the proposed technique is consistent. A very important fact is that the proposed method clearly identifies the background as being at a same depth which is definitely not the case in the result using confocal stereo.

(a) **(b)** **(c)**

Figure 8.12: A box image (courtesy [50]) with a slope is shown in (a). The depth map obtained using the proposed method with a single image is shown in (b). The result obtained from the method in [50] using around 100 images

8.5 Conclusion

In this chapter 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.

Chapter 9

Conclusion

In this thesis we consider in a unified manner the problem of depth estimation and image restoration based on the defocus cue. The heat equation effectively models the defocus cue. In this thesis we consider various ways in which this model can be used for depth estimation and image restoration. The use of a common underlying diffusion model helps us in solving both these problems by considering various aspects of the diffusion process.

The first basic approach adopted by us was using linear diffusion. The linear diffusion was done using defocus morphing of two defocused observations to generate the entire defocus space of the observations. This method also demonstrated the equivalence between depth from focus and depth from defocus methods as this method solved the problem of depth from defocus using two observations in a manner similar to the depth from focus technique where many observations are used to estimate the depth in the scene. This method used the spectral interpolation property of defocus morphing to generate the defocus space. Hence, the characteristics of local spectral operations of windowing and inverse filtering in textureless regions characterized this method. However, this first effort provided the base for exploring the use of diffusion for both depth estimation and image restoration.

The limitations of the spectral method provided the necessary impetus for exploring the problem in the spatial domain. In the spatial domain the problem of depth from defocus was considered lately by Favaro *et al.*. However, unfortunately their method had a few shortcomings. One of the important facts was that the linear diffusion based method suggested did not adequately handle departures from the Gaussian assumption of the point spread function. These departures occur specifically around self-occlusion edges and aperture imperfections. We therefore considered the use of stochastically perturbed diffusion (SPD). We have shown that use of

SPD implicitly handles departure from the Gaussian assumptions. Moreover the stochastic level sets based evolution also handles the non-convex nature of the diffusion.

We then used an MRF framework for representing the diffusion coefficient. This framework was necessary in order to incorporate the neighborhood information in the estimation of diffusion coefficient. The MAP estimate of the MRF framework was done using the graph cuts formulation, i.e using the alpha expansion algorithm. This method has advantages like symmetric handling of diffusion coefficient and generalized framework for incorporating robust regularization functions. These advantages helped in providing accurate results.

While the forward diffusion methods are quite stable, the reverse diffusion method is unstable due to the ill-posed nature of the reverse heat equation. We therefore proposed an alternative non-parametric reverse mean shift technique which improved deblurring results as compared to other partial differential equations based techniques like the provably convergent "shock filters" of Osher and Rudin. We have used a stopping criterion based on cluster separation in order to terminate the reverse mean shift procedure. The reverse mean shift was further extended in terms of a generalized mean shift technique that could be used for deblurring in the presence of noise. An optimal switching criterion was used to determine the switching between forward and reverse mean shift.

The non-parametric technique is a non-linear method for deblurring that approximates the reverse heat equation. A better approximation is obtained by stabilizing the linear reverse heat equation itself. The reverse heat equation is stabilized using geometric form of the heat equation. Since the blur is predominant in the normal form rather than the tangential form, the reverse heat equation is stabilized by damping the reverse heat using the normal form of the heat equation. The stopping criterion is selected based on the degeneration of the tangential component as compared to the normal component. Using this technique appropriate blind space varying deconvolution is achieved.

The use of reverse heat equation enables us to consider the challenging problem of depth estimation from a single defocused image. While, depth estimation from multiple defocused images has been explored, depth estimation from a single defocused image was considered to be impossible. However, using the reverse heat equation we are able to estimate an initial likelihood estimate of depth in the scene. The likelihood coupled with the edge information is used in an MRF framework to estimate the dense depth in the scene. Thus we are able to demonstrate depth estimation using a single defocused image.

We have presented all the methods developed in the thesis with extensive experimentation. The results are also compared in each case with state of the art related techniques. The ideas thus proposed in the thesis provide an extensive exploration of the defocus cue in depth estimation and image restoration based on the common underlying idea of the use of the diffusion equation.

9.1 Future Work

The focus of the thesis has been on analysis of defocused images for depth estimation and defocused images. While we have built a coherent body of work in the thesis, there are several interesting aspects which can be considered for future work.

- The defocus space concept allows one to generate a continuum of defocused images and one can synthesize defocused observations. However, the windowed Fourier transform used for generating the defocus space resulted in windowing artifacts being generated. One would like to explore a corresponding spatial or spatio-temporal representation of the defocus space which would result in accurate defocus space being generated. This would have important applications in refocusing of images.

- The methods proposed for depth estimation from two defocused images did not assume an explicit parametric form on the scene. There have been methods for structure recovery that have leveraged the parametric form of structure in the scene. Hence, we would like to explore in the future whether a corresponding parametric structure technique can be used for depth estimation.

- We would also like to further explore the possibility for joint depth estimation and image restoration in a discrete optimization framework. Current work in discrete optimization using graph cuts suggests that it should be possible to solve the problem in such a framework.

- There has been interesting work done where use of depth from defocus has been used as a cue for super-resolution and inpainting. We would like to explore in the future whether it is possible to use the diffusion framework to simultaneously perform depth estimation and solve problems like super-resolution and inpainting.

- While exploring the problems of image restoration we have seen that the problem of joint deblurring and denoising of images is especially hard. We would like to consider reaction-diffusion schemes or stochastic diffusion schemes which could adapt better locally to the image information and appropriately carry out deblurring and denoising.

- We would like to further explore the use of priors in diffusion frameworks so as to be able to perform better depth estimation and image restoration. There has been recent work that have used prior information effectively for object segmentation and structure recovery.

- There has been considerable interest in machine learning techniques for depth estimation from a single image using geometric class information [54],[110]. We would like to consider using similar techniques for identifying the blur in images. This would help us in devising better stopping criterion.

- We would lastly like to explore use of the techniques proposed for depth estimation from a single defocused image, image restoration to improve the performance of other computer vision tasks like content based retrieval, object segmentation and computational photography.

Bibliography

[1] N. Ahuja and A. L. Abbott. Active stereo: Integrating disparity, vergence, focus, aperture and calibration for surface estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(10):1007–1029, 1993.

[2] L. Alvarez and L. Mazorra. Signal and image restoration using shock filters and anisotropic diffusion. *SIAM Journal of Numerical Analysis*, 31(2):590–605, April 1994.

[3] N. Asada, H. Fujiwara, and T. Matsuyama. Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2):155–167, February 1998.

[4] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*. Springer Verlag, New York, USA, 2002.

[5] M. Banham and A. Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 14:24–41, March 1997.

[6] M. Belge, M. E. Kilmer, and E. L. Miller. Wavelet domain image restoration with adaptive edge-preserving regularity. *IEEE Trans. Image Process.*, 9:597–608, 2000.

[7] 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.

[8] S.S. Bhasin and S. Chaudhuri. Depth from defocus in presence of partial self occlusion. In *Proc. Eighth IEEE Int'l Conf. on Computer Vision*, volume 1, pages 488–493, July 2001. Vancouver, Canada.

[9] J. Biemond, R.L. Lagendijik, and R.M. Mersereau. Iterative methods for image deblurring. *Proceedings of IEEE*, 78(5):856–883, 1990.

[10] D.S.C. Biggs and M. Andrews. Acceleration of iterative image restoration algorithms. *Applied Optics*, 36(8):1766–1775, 1997.

[11] Yuri Boykov and Vladimir 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.

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

[13] A. Buades, B. Coll, and J-M. Morel. A review of image denoising methods, with a new one. *Multicscale Modeling and Simulation*, 4(2):490–530, 2005.

[14] 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.

[15] F. Catte, T. Coll, P. L. Lions, and J. M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal*, 29:182–193, 1992.

[16] R.H. Chan, T.F. Chan, L. Shen, and Z. Shen. Wavelet deblurring algorithms for spatially varying blur from high-resolution image reconstruction. *Linear Algebra Appls.*, 366:139–155, 2003.

[17] T.F. Chan and C.K. Wong. Total variation blind deconvolution. *IEEE Trans. on Image Processing*, 7:370–375, 1998.

[18] S. Chaudhuri. Defocus morphing in real aperture images. *J. Optical Soc. America, Ser-A*, 22(11):2357–2365, November 2005.

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

[20] Y. Cheng. Mean shift, mode seeking and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, August 1995.

[21] Robert T. Collins. Mean shift blob tracking through scale space. In *CVPR 2003 Conference Proceedings*, pages 234–240, June 2003. held in Madison, Wisconsin, June.

[22] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):1–8, February 2003.

[23] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[24] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.

[25] S. Das and N. Ahuja. Performance analysis of stereo, vergence and focus as depth cues for active vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(12):1213–1219, 1995.

[26] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, 2nd ed.* Wiley Interscience, New York, NY, 2001.

[27] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:97–108, 1993.

[28] Mark Fashing and Carlo Tomasi. Mean shift is a bound optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):471–474, 2005.

[29] Olivier Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, 1993.

[30] 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.

[31] P. Favaro, A. Mennucci, and S. Soatto. Observing shape from defocused images. *International Journal of Computer Vision*, 52(1):25–43, April 2003.

[32] 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.

[33] P. Favaro and S. Soatto. Seeing beyond occlusions (and other marvels of a finite lens aperture)learning shape from defocus. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 03)*, volume 2, pages 579–586, 2003. Madison, Wisconsin, USA.

[34] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):406–417, March 2005.

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

[36] M. Figueiredo and R. Nowak. A bound optimization approach to wavelet-based image deconvolution. In *IEEE International Conference on Image Processing - ICIP'2005, Genoa*, 2005.

[37] D. A. Forsyth. Shape from texture without boundaries. In *Proc. European Conf. Computer Vision*, volume 3, pages 225–239, 2002.

[38] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, USA, 2002.

[39] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, January 1975.

[40] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 11(7):689–703, July 2002.

[41] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Image enhancement and denoising by complex diffusion processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1020–1036, 2004.

[42] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Image sharpening by flows based on triple well potentials. *Journal of Mathematical Imaging and Vision*, 20:121–131, 2004.

[43] B. Girod and S. Scherock. Depth from defocus of structured light. In *Proc. Optics, Illumination, and Image Sensing for Machine Vision IV, SPIE vol. 1194*, pages 209–215, 1989. held in Philadelphia, Penn., November 1989.

[44] M. Gokstorp. Computing depth from out-of-focus blur using a local frequency representation. In *Proc. Intl Conf. Pattern Recognition*, pages 153–158, 1994. Jerusalem, Israel.

[45] 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.

[46] P. Grossman. Depth from focus. *Pattern Recognition Letters*, 5(1):63–69, January 1987.

[47] F. Guichard and J.-M. Morel. A note on two classical shock filters and their asymptotics. In M. Kerckhove, editor, *Scale-Space and Morphology in Computer Vision*, *volume 2106 of Lecture Notes in Computer Science*, pages 75–84, 2001.

[48] 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.

[49] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd Edition, 2004.

[50] S.W. Hasinoff and K.N. Kutulakos. Confocal stereo. In *Proc. Ninth European Conference on Computer Vision*, pages 620–634, May 2006. Graz, Austria.

[51] S.W. Hasinoff and K.N. Kutulakos. A layer-based restoration framework for variable-aperture photography. In *Proc. IEEE International Conference of Computer Vision (ICCV)*, October 2007. Rio de Janeiro, Brazil.

[52] A.D. Hillery and R.T. Chin. Iterative wiener filters for image restoration. *IEEE Trans. on Signal Processing*, 39(8):1892–1899, 1991.

[53] D. Hoiem, A.A. Efros, and M. Hebert. Geometric context from a single image. In *Proc. IEEE International Conference of Computer Vision (ICCV)*, pages 654–661, October 2005. Beijing, China.

[54] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Automatic photo pop-up. *ACM Trans. Graph.*, 24(3):577–584, 2005. SIGGRAPH 2005 Conference Proceedings.

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

[56] O. Juan, R. Keriven, and G. Postelnicu. Stochastic motion and the level set method in computer vision: Stochastic active contours. *Internation Journal of Computer Vision*, 69(1):7–25, August 2006.

[57] Heung-Yeung Shum Jue Wang, Yingqing Xu and Michael Cohen. Video tooning. *ACM Trans. on Graphics (Proc. of SIGGAPH2004)*, 23(3):574–583, 2004.

[58] Yingqing Xu Jue Wang, Bo Thiesson and Michael Cohen. Image and video segmentation by anisotropic kernel mean shift. In Toms Pajdla and Jir Matas, editors, *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part II*, pages 238–249, May 2004. held in Prague, Czech Republic, June.

[59] V. Katkovnik K. Dabov, A. Foi and K. Egiazarian. Image denoising with block-matching and 3d filtering. In *Electronic Imaging 06, Proceedings SPIE 6064, no. 6064A-30*, 2006. held in San Jose, California, USA.

[60] A.K. Katsaggelos, J. Biemond, R.W. Schafer, and R.M. Merseraeau. A regularized iterative image restoration algorithm. *IEEE Trans. on Accoustics, Speech and Signal Processing*, 39:914–929, 1991.

[61] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: Movies, color, texture and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, 2000.

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

[63] 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.

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

[65] H. P. Kramer and J. B. Bruckner. Iterations of a non-linear transformation for enhancement of digital images. *Pattern Recognition*, 7(1-2):53–58, 1975.

[66] E. Krotkov. Focusing. *International Journal of Computer Vision*, 1:223–237, 1987.

[67] D. Kundur and H. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13:43–64, May 1996.

[68] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Verlag, 2nd. Edition, New York, USA, 2001.

[69] T. Lindeberg and B. M. ter Haar Romeny. Linear-Scale -Space. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, pages 1–41. Kluwer Academic Publisher, Boston, 1994.

[70] M. Lindenbaum, M. Fischer, and A. Bruckstein. On gabor's contribution to image enhancement. *Pattern Recognition*, 27(1):1–8, 1994.

[71] P.-L. Lions and Panagiotis E. Souganidis. Fully nonlinear stochastic partial differential equations. *C. R. Acad. Sci. Paris, t. 331, Srie I*, pages 1085–1092, 1998.

[72] Y. Lou, P. Favaro, A.L. Bertozzi, and S. Soatto. Autocalibration and uncalibrated reconstruction of shape from defocus. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007. Minnesota, US.

[73] M. Vergauwen F. Verbiest Cornelis K. J. Tops R. Koch M. Pollefeys, L. Van Gool. Visual modeling with a hand-held camera. *Int. Journal of Computer Vision*, 59(3):207–232, 2004.

[74] Yi Ma, Stefano Soatto, Jana Kosecka, and Shankar S. Shastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, New York, U.S.A., 2004.

[75] David Marr. *Vision: A computational representation into the human representation and processing of Visual Information*. W.H. Freeman and Company, New York, 1982.

[76] F. Moreno-Noguer, P.N. Belhumeur, and S.K. Nayar. Active Refocusing of Images and Videos. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, 26(3):67–1–67–9, July 2007.

[77] P. Mrazek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *International Journal of Computer Vision*, 52(2/3):189–203, 2003.

[78] M. Namba and Y. Ishida. Wavelet transform domain blind deconvolution. *Signal Processing*, 68:119–124, July 1998.

[79] V.P. Namboodiri and S. Chaudhuri. Use of linear diffusion in depth estimation based on defocus cue. In *Proc. Fourth Indian Conference on Computer Vision, Graphics & Image Processing*, pages 133–138, December 2004. Kolkata, India.

[80] Takashi Matsuyama Naoki Asada, Hisanaga Fujiwara. Edge and depth from focus. *International Journal of Computer Vision*, 28(2):153–163, 1998.

[81] Takashi Matsuyama Naoki Asada, Hisanaga Fujiwara. Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2):155–167, 1998.

[82] S. K. Nayar and Y. Nakagawa. Shape From Focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.

[83] S.K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(12):1186–1198, Dec. 1996.

[84] R. Neelamani, H. Choi, and R. Baraniuk. Wavelet domain regularized deconvolution for ill-conditioned systems. In *IEEE International Conference on Image Processing*, pages 58–72, 1999.

[85] M. Noguchi and S.K. Nayar. Microscopic shape from focus using active illumination. In *Proc. of 12th IAPR Int. Conf. on Pattern Rec.*, volume 1, pages 147–152, 1994. Jerusalem, Israel.

[86] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Verlag, New York, sixth edition, 2004.

[87] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal of Numerical Analysis*, 27(4):919–940, August 1990.

[88] A. P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[89] A. P. Pentland, S. Scherock, T. Darrell, and B. Girod. Simple range cameras based on focal error. *J. Optical Soc. America A*, 11(11):2925–2934, november 1994.

[90] P. Perona and J. Malik. Scale Space and Edge Detection using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[91] I. Pollak, A. S. Willsky, and J. Krim. Image Segmentation and Edge Enhancement with Stabilized Inverse Diffusion Equations. *IEEE Trans. on Image Processing*, 9(2):256–266, February 2000.

[92] M. Pollefeys and L. Van Gool. From images to 3d models. *Communications of the ACM*, 45(7):51–55, 2002.

[93] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.

[94] E. Prados and O. Faugeras. Shape from shading: a well-posed problem? In *Proc. Computer Vision and Pattern Recognition, (CVPR)*, volume 2, pages 870– 877, 2005.

[95] A. N. Rajagopalan and S. Chaudhuri. Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters. *International Journal of Computer Vision*, 30:175–190, 1998.

[96] A. N. Rajagopalan and S. Chaudhuri. A recursive algorithm for maximum-likelihood based identification of blur from multiple observations. *IEEE Trans. Image Processing*, 7(7):1075–1079, July 1998.

[97] A. N. Rajagopalan and S. Chaudhuri. Mrf model based identification of shift variant point spread function for a class of imaging systems. *Signal Processing*, 76(3):285–299, 1999.

[98] A.N. Rajagopalan and S. Chaudhuri. A block shift-variant blur model for recovering depth from defocused images. In *Proc. Intl. Conf. Image Processing*, pages 636–639, 1995. Washington D.C., USA.

[99] A.N. Rajagopalan and S. Chaudhuri. Space-variant approaches to recovery of depth from defocused images. *Computer Vision and Image Understanding*, 68(3):309–329, 1997.

[100] A.N. Rajagopalan and S. Chaudhuri. A variational approach to recovery of depth from defocused images. *IEEE Trans. on PAMI*, 19(10):1158–1164, October 1997.

[101] A.N. Rajagopalan and S. Chaudhuri. Optimal recovery of depth from defocused images using an mrf model. In *Proc. Intl. Conf. Computer Vision*, pages 1047–1052, 1998. Bombay, India.

[102] 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.

[103] D. Ramanan and D. A. Forsyth. Using temporal coherence to build models of animals. In *Proceedings of the International Conference on Computer Vision*, pages 338–345, 2003. held in Nice, France.

[104] L. Remaki and M. Cheriet. Numerical schemes of shock filter models for image enhancement and restoration. *Journal of Mathematical Imaging and Vision*, 18(2):129–143, 2003.

[105] Stephen J. Roberts. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition*, 30(2):261–272, 1997.

[106] Bart M. Romeny. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

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

[108] L. I. Rudin. Images, numerical analysis of singularities and shock filters. Technical Report TR 5250:87, Computer Science Department, California Institute of Technology, December 1987. Ph.D. Thesis.

[109] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259–268, 1992.

[110] Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng. Learning depth from single monocular images. In *Proceedings of Conference on Neural Information Processing Systems, NIPS 18*, December 2005.

[111] J. G. M. Schavemaker, M. J. T. Reinders, J. J. Gerbrands, and E. Backer. Image sharpening by morphological filtering. *Pattern Recognition*, 33:997–1012, 2000.

[112] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162, September 2000.

[113] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[114] N. Sochen, R. Kimmel, and A. M. Bruckstein. Diffusions and confusions in signal and image processing. *J. Math. Imaging and Vision*, 14:195–209, 2001.

[115] Jon Sporring and Joachim Weickert. Information measures in scale spaces. *IEEE Trans. Information Theory*, 45(3):1051–1058, April 1999.

[116] J.-L. Starck, E. Cands, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):131–141, 2002.

[117] J-L. Starck, M.K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83:2279 2283, 2003.

[118] M. Subbarao. Parallel depth recovery by changing camera aperture. In *Proceedings of International Conference on Computer Vision*, pages 149–155, 1988. held in Florida, USA.

[119] M. Subbarao and T. Choi. Accurate Recovery of Three-Dimensional Shape From Image Focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):266–274, march 1995.

[120] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *Computer Vision and Pattern Recognition*, pages 498–503, 1988.

[121] M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *Intl J. Computer Vision*, 13(3):271–294, 1994.

[122] 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.

[123] James Z. Wang, Jia Li, Robert M. Gray, and Gio Wiederhold. Unsupervised multiresolution segmentation for images with low depth of field. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(1):85–90, January 2001.

[124] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, May 1998.

[125] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201–212, 1999.

[126] J. Weickert. Coherence enhancing shock filters. In *Pattern Recognition, Lecture Notes in Computer Science, Vol. 2781, Springer, Berlin, Proceedings of DAGM 2003*, pages 1–8, 2003. in Magdeburg, Germany -September 10-12 2003.

[127] Joachim Weickert. *Anisotropic Diffusion in Image Processing*. ECMI Series, Teubner, Stuttgart, 1998.

[128] D. V. Widder. *The Heat Equation*. Academic Press Inc., London, UK, 1975.

[129] A. P. Witkin. Scale-Space Filtering. In *Proc. of the 4th International Joint Conference on Artificial Intelligence*, pages 1019–1022, Karlsruhe, West Germany, 1983.

[130] Y. Xiong and S.A. Shafer. Moment filters for high precision computation of focus and stereo. In *Proc. Intl Conf. Intelligent Robots and Systems*, pages 108–113, August 1995. Pittsburgh, Pennsylvania, USA.

[131] N.K. Yip. Existence of dendritic crystal growth with stochastic perturbations. *Journal of Nonlinear Science*, 8:491–579, 1998.

[132] Li Zhang and Shree K. Nayar. Projection defocus analysis for scene capture and image display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, 25(3):907–915, July 2006.

[133] D. Ziou and F. Deschenes. Depth from defocus estimation in spatial domain. *Computer Vision and Image Understanding*, 81(2):143–165, February 2001.

Publications

1. V.P. Namboodiri, S. Chaudhuri and S. Hadap, ”Regularized depth from defocus”, in *Proc. of IEEE Intl. Conference on Image Processing (ICIP)*, San Diego, USA, 2008.

2. V.P. Namboodiri and S. Chaudhuri, ”Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera”, in *Proc. of IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, U.S, 2008

3. V. P. Namboodiri and S. Chaudhuri, “On Defocus, Diffusion and Depth Estimation”, *Pattern Recognition Letters*, Vol. 28, Issue 3, pp. 311-319, February 2007.

4. V. P. Namboodiri and S. Chaudhuri, “Shape Recovery Using Stochastic Heat Flow”, in *Proc. of British Machine Vision Conference (BMVC)*, Warwick, UK, 2007.

5. V. P. Namboodiri and S. Chaudhuri, “Image Restoration using Geometrically Stabilized Reverse Heat Equation”, in *Proc. of IEEE Intl. Conf. on Image Processing (ICIP)*, San Antonio, Texas, USA, 2007.

6. P. Chatterjee, V. P. Namboodiri and S. Chaudhuri, “Super-resolution Using Sub-band Constrained Total Variation”, in *Proc. of 1st Scale Space and Variational Methods (SSVM) Conference*, Ischia, Italy, 2007.

7. V. P. Namboodiri, A. Ghorawat and S. Chaudhuri, “Improved Kernel-Based Object Tracking Under Occluded Scenarios”, in *Proc. 5th Indian Conference on Computer Vision, Graphics and Image Processing,(ICVGIP)*, Madurai, LNCS 4338, pp. 504-515, 2006.

8. V. P. Namboodiri and S. Chaudhuri, “Shock Filters based on Implicit Cluster Separation”, in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, USA, 2005.

9. V. P. Namboodiri and S. Chaudhuri, “Use of Linear Diffusion in depth estimation based on defocus cue”, in *Proc. Indian Conf. on Vision, Graphics and Image Processing (ICVGIP)*, Kolkata, Dec. 2004.

Acknowledgments

I take this opportunity to thank everyone who has helped me in the making of this dissertation.

Foremost I would like to thank my advisor, Prof. Subhasis Chaudhuri. He has been extremely helpful and highly motivated me in my research. The discussions with him were always very insightful and would inspire me further in research. His constant encouragement and deep insights helped me keep this research on course. I am very thankful for the constant help and support that he has given me throughout the course of my doctoral research. It has been indeed a great pleasure working with him.

I am thankful to Prof. Harish Pillai, Prof. S.N. Merchant and Prof. S. Dutta Roy for agreeing to be on my research committee. Their suggestions and remarks were always valuable. I would like to thank Prof. Subhasis Banerjee and Prof. Alfred Bruckstein for kindly reviewing my thesis. Their suggestions were very useful in improving my thesis. I would also like to thank Prof. Sharat Chandran for his suggestions.

I am deeply indebted to Dr. Vladimir Kolmogorov for hosting me at UCL, Adastral Park and for the guidance, advice and support that he gave me during my visit there. The discussions with him helped me a lot in carrying out the research in this dissertation. He very kindly explained the details of discrete optimization and graph cuts during the visit. I would also like to thank the British Council, UKIERI and the members Ms. Manjula Rao and Anjali Rao who helped in making this visit happen.

I am very thankful to Prof. Vasudeva Murthy for kindly hosting me at TIFR Bangalore. It was a pleasure to visit the centre and work with him. His suggestions were quite useful and he was very helpful during the visit.

I would like to now thank all my colleagues in the lab: Dr. Manjunath Joshi, Dr. Rajashekhar, Dr. C.V. Jiji, Tanmay Pawar, V. Srikrishnan, Malay Nema, R. Shanmuganathan, Deven Patel, Ranjith Ram, Neethu Prakash, Priyam Chatterjee, Hitesh Shah, Amit Prabhudesai, Shilpa Inamdar, Trupti Thete, Smita Nair, Satya, T. Nagaraj, Mayank Tyagi, Ankit Mehra,

K. Mrityunjay Singh, Gourav Dhavale, Amit Ghorawat, Sujata Mukherjee, Amit Tilwankar. They have been very helpful and supportive all along. I have been fortunate in having worked with such wonderful people.

I would like to thank all my friends from Vision and Graphics Lab, Aniruddha Joshi, Appu Shaji, Biswarup Chowdhuri, Neela Sawant, Anish Chandak, Kashyapp Paidmarri, Behjat Siddique, Deepali Singla, Rhushabh Goradia, M. Nitya, Rocky Agrawal, Shyamsunder, K. Anil, Aman Parnami. They have been the most amazing friends and their camaraderie helped me immensely during my dissertation.

I would like to thank all my friends in the department and hostel who have been responsible for making my stay here joyful. There are just a whole lot of them who have been totally awesome.

I would like to now thank my family who have been extremely kind, supportive, helpful, caring, considerate. This thesis would have been impossible without their understanding, patience and love. This thesis is dedicated to them.

Date: _____ Vinay P. Namboodiri