DEPTH ESTIMATION FROM SINGLE IMAGE FOR 2D-TO-3D CONVERSION

Submitted towards the partial fulfillment of requirement for the award of degree of

Master of Engineering

In

Wireless Communication

Submitted by:

NIDHI JAMWAL

(801463015)

Under the guidance of:

Dr. Neeru Jindal
Assistant Professor, ECED

Dr. Kulbir Singh
Professor, ECED

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

(Established under the section 3 of UGC Act, 1956)

PATIALA – 147004, PUNJAB, INDIA

JULY 2016

DECLARATION

I hereby declare that the work which is being presented in this dissertation entitled, **"Depth Estimation from Single Image for 2D-to-3D Conversion"**, for the award of the degree of Master of Engineering in Wireless Communication at Thapar University, Patiala, is an authentic record of my own work carried out under the guidance of **Dr. Kulbir Singh**, Professor and **Dr. Neeru Jindal**, Assistant Professor, ECED, Thapar University, Patiala, and refer other's research work which are listed in the reference section. The results presented in this dissertation have not been submitted in part or in full to any other University or Institute for the award of any other degree.

Date: 13-07-2016

Nidhi Jamwal
Roll No. 801463015

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and the contents of the thesis have reached the requisite standard.

Date: 13/7/16

Dr. Neeru Jindal
Assistant Professor, ECED
Thapar University, Patiala

Dr. Kulbir Singh
Professor, ECED
Thapar University, Patiala

Countersigned by:

Dr. Sanjay Sharma
Professor and Head, ECED
Thapar University, Patiala

Dr. S. S. Bhatia
Dean of Academic Affairs
Thapar University, Patiala

ACKNOWLEDGEMENT

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my work to complete the research successfully.

Apart from my effort, the success of this thesis depends largely on the encouragement and guidelines of many others.

I take this opportunity to express my sincere gratitude to my guide **Dr. Kulbir Singh**, Professor and **Dr. Neeru Jindal**, Assistant Professor, ECED, Thapar University, Patiala, for their thorough supervision and the stream of ideas that kept me occupied during the course of the past two years. Their knowledge, discipline, availability, patience, constructive criticism of my work and support are unique.

I am also thankful to **Dr. Sanjay Sharma**, Professor and Head, ECED, **Dr. Amit Kumar Kohli**, Associate Professor and P.G. Coordinator, and **Dr. Hemdutt Joshi**, Assistant Professor, Program Coordinator of Wireless Communication, for providing us with adequate infrastructure in carrying out the work.

I would like to acknowledge all the faculty members of ECED for their full support throughout this work. I am also thankful to my friend **Amneet Singh** who was very helpful and accessible whenever I had a request.

A special thanks go to my parents and my brother for the support, the encouragement, and for all the hope they have on me.

Nidhi Jamwal

Roll No. 801463015

ABSTRACT

Images compress three-dimensional visual data to two-dimensional which exclude the third dimension of depth. Many different approaches to estimate depth information from single images have been taken towards solving this problem and significant progress has been made on the field during the last decade. But, the intrinsic limitation of these approaches is the loss of depth characteristics during the projection of the scene into the image plane. An advantage of these approaches is the relatively low amount of operations needed to process one single image, instead of two or more. However, even though many depth map generation algorithms have been developed, accurate estimation of depth information from natural scenes still remains problematic.

To confront this issue, an approach is proposed in this dissertation to estimate accurate depth maps by using information from a single small-scale blurred image. The method takes the advantage of sharpness map to estimate the amount of spatially varying subtle defocus blur. By using edge information and applying a joint bilateral filter, a sparse depth map is obtained. Then, matting Laplacian interpolation algorithm is applied to attain a full depth map.

This dissertation also introduces a new technique for converting the two-dimensional image back to three-dimensional with the help of recovered depth map. The evaluated depth estimates are added back to the corresponding two-dimensional (2D) image to give a three-dimensional (3D) impression to it. The robustness of proposed depth map estimation techniques are tested by several experiments and results are compared with the other well-documented methods, visually and quantitatively. When compared with J. Shi *et al.* [44], the average decrease in relative error, log10 error and root mean square error for real images used in this dissertation is 0.472, 0.1975, and 0.1952 respectively. The obtained results validate the effectiveness of proposed depth map estimation approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

1D	One-Dimensional
2D	Two-Dimensional
3D	Three- Dimensional
3DTV	Three-Dimensional Television
CCD	Charged Coupled Devices
CID	Computed Image Depth
CMOS	Complementary Metal Oxide Semiconductor
CoC	Circle of Confusion
DFD	Depth from Defocus
DFF	Depth from Focus
DFM	Depth from Motion
JBF	Joint Bilateral Filter
LED	Light Emitting Diode
LPF	Low Pass Filter
MRF	Markov Random Fields
PSF	Point Spread Function
SFD	Shape from Defocus
SFF	Shape from Focus
SFM	Shape from Motion
SFS	Shape from Shading
SFT	Shape from Texture
SML	Sum Modified Laplacian
STM	Spatial Domain Convolution/Deconvolution Transform Method
TOF	Time of Flight

CHAPTER 1
INTRODUCTION

The brick walls are not there to keep us out.
The brick walls are there to give us a chance
to show how badly we want something.
Because the brick walls are there to stop
the people who don't want it badly enough.
Randy Pausch [1960-2008]

1.1 Preamble

Images have been the most attractive intermediate to envision the information. They include a rich visual detail of a structure and are straightforward and simple to obtain, view, share, and publish. However, when a three-dimensional (3D) scene is projected onto a two-dimensional (2D) plane, the third dimension of depth is lost. This makes severe constraints to applications like image recognition, manipulation, and depth-enhanced image editing [1]. The most challenging job in computer vision is to retrieve this lost depth details. Photographs that have a feeling of depth are dynamic and powerful. Capturing a picture with a sense of depth and perspective, though, can be tricky because photography is very much a two-dimensional art form while eyes see everything in three dimensions. One way to give a three-dimensional impression is by adding depth to photographs**.**

1.2 Three-Dimensional Images and Generation Techniques

3D images have provided a substantial improvement in the visual quality as compared to 2D images. They allow us to view a rich 3D world, not a flat one as with the standard 2D images. 3D means the object has height, width, and depth. Human beings are able to identify the spatial interconnection between objects just by seeing them because the humans have 3D perception, generally called as depth perception. The retina creates a 2D image in each eye. These two images are processed by the human brain which develops a

3D visual experience. Besides having a vision with both eyes (stereoscopic vision), people having one eye (monocular vision) can also realize the 3D world.

The following sections provide an overview of different 2D-to-3D conversion algorithms. The 3D creation is accomplished in a variety of different manner. 3D information can be directly achieved from a scene and simultaneously 3D content is produced which is appropriate for 3D television (3DTV). But it requires segmentation of video frames which is why real-time conversion is explicitly disputable. Other ways of generating 3D data reuse the existing 2D content and include depth information to it so as to convert it to 3D.

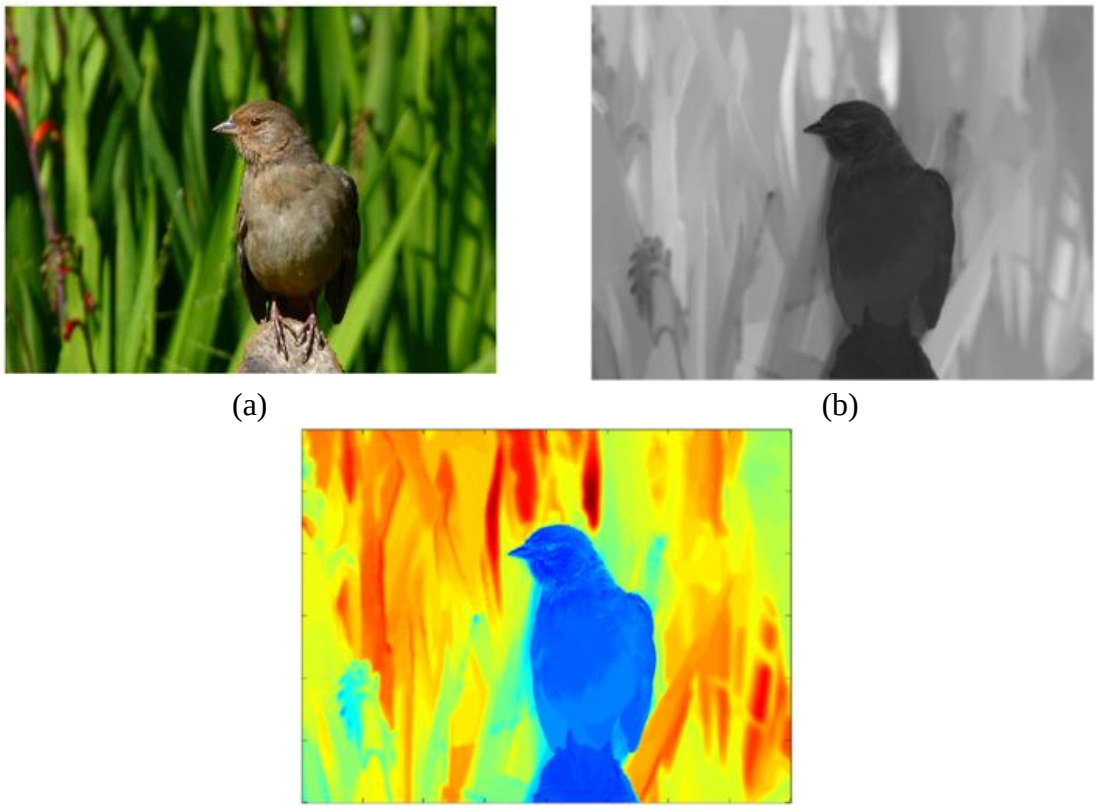
1.2.1 Cut-and-Paste Technique

A stereoscopic 3D image is actualized easily by making use of original 2D image and considering it as a left eye image. By shifting local regions of this image horizontally, the right eye view is attained. This method of generating 3D stereoscopic is named as a cut-and-paste technique [2]. Image segmentation methods are used for the segregation of local regions. The approach is well-suited for those images in which the objects are conveniently distinguished. Instead of assigning the depth to each pixel, it considers each object as independent unit and assigns depth to it. Therefore, the rendered objects might appear flat and impractical.

1.2.2 Depth Map Generation

The most frequently used methods for 2D-to-3D conversion incorporated depth information estimation. Depth estimation refers to a mechanism of restoring the third dimension from a 2D image, i.e., the distance measure of each pixel of an image. Mainly, there are two paradigms of representing depth map. First one utilizes the gray scale intensity to present the depth of every pixel. The example is shown in Figure 1.1, where Figure 1.1 (a) depicts the original image, Figure 1.1 (b) displays the corresponding depth information in which the darker pixel refers to the points that are nearby to the camera and lighter pixels are the points far away from the camera. The other paradigm is the

color representation which makes use of colors to represent depth maps. In Figure 1.1 (c), red-black colors indicate further pixels whereas blue-dark colors are nearer points.



(a) (b)

(c)

Figure 1.1: Different types of depth maps, (a) Input image, (b) Gray scale representation of depth map, and (c) Color representation of depth map [3].

1.2.2.1 Disparity Estimation Method

The basic principle of converting 2D-to-3D is built upon the binocular transform of the human optical setup of two marginally distinct images, subject to stereoscopic view. The horizontal disparities existing in the left, as well as the right eye images, are mutated into distance information. By implementing the number of steps, calibrate the images, distortion correction and image pre-processing, the disparity map is obtained. This information is then used to observe the objects at distinct depths apart from the 2D plane. The amount of horizontal shift between the pixels of both the images relies upon the

inter-lens separation and the range of the object to the camera. It is significant to note that this technique requires the usage of a stereoscopic camera for capturing the two views.

1.2.2.2 Depth Map Generation from Blur

The focal parameters of the camera have a great influence on the recorded image. There is a straightforward connection between the distance of the object to the camera (depth), and the blur quantity existing in the image. This connection between the two can be used to formulate the fundamental concept of generating Depth-from-Defocus (DFD) [2]. The approach, however, is not that easy for the reason that there are many other sources of blur such as lens peculiarity, motion blur, atmospheric obstruction and fuzzy objects [2]. This approach is greatly allied to the Shape-from-Defocus (SFD) algorithm [4] in which the information about the extent of blur is obtained by applying blur estimation technique.

1.2.2.3 Depth Map Generation from Focus

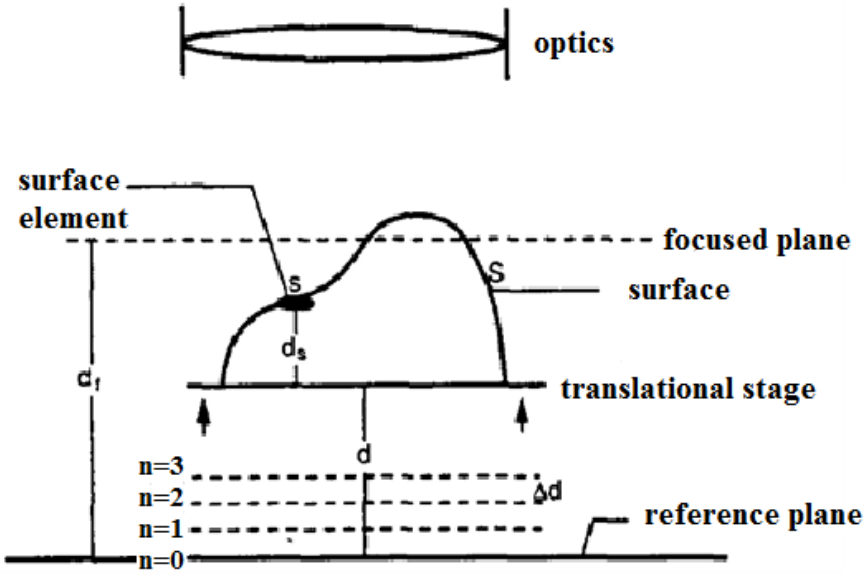


Figure 1.2: Depth map estimation using focus [5].

The Depth-from-Focus (DFF) method [5] depends on a sequence of images of the same scenery with distinct focal levels by alternating the camera range with respect to the

scene. This technique is very similar to the class of techniques utilizing blur. The only difference is that the depth map formation using defocus cues desire only one, two, or more images with the immovable camera and object positions, and utilize distinctive camera focal settings.

Figure 1.2 exemplifies the major principle of the DFF algorithm [5]. An object with erratic surface approaches towards the camera (optics) from the reference level and is located at the translational stage. The surface element on the object eventually becomes focused and achieves maximum sharpness at the time it reaches the focused plane. During the course of this mechanism, the shifts in the translational stage are recorded and the depth values relative to this stage is determined. Employing the similar process to the entire surface elements, a dense depth map is produced.

1.2.2.4 Edge Detection based Depth Estimation

These methods make use of "surrogate" depth maps which incorporate depth information at the edge locations and object boundaries of the image. Many regions of these depth maps are either absent or have imprecise depth information. However, it was figured out that the human visual structure integrates the available facts together with the depth cues to render the missing image locations. The depth maps are comparatively easy to form and are suitable for the applications in which depth accuracy is not demanding.

1.2.2.5 Depth Estimation from Geometric Perspective

The method of generating depth maps from linear and gradient perspective cues is probably more suitable for 3DTV [2]. Firstly, color segmentation is performed on a single input image. By recognizing straight lines in the image, the areas with the majority of intersection are well chosen to be vanishing points and the lines closer to these points are considered as vanishing lines which take care of linear perspective of the depth. Hence, with the vanishing point being at the furthest distant, different depth gradients are generated depending upon the slopes of the vanishing lines. Fusion of this geometric depth with the depth data of objects which is introduced by early image segmentation is then performed to produce "natural" depth map. Figure 1.3 demonstrates the technique

and the produced depth map in which a blackish gray level signifies a higher depth value. Additionally, the vanishing points are indicated by green colored points and red lines correspond to vanishing lines.

1.2.2.6 Depth Map Generation from Motion

The fundamental principle of generating depth maps is dependent on the motion of parallax, i.e., how points move relatively to one another with respect to head movements. Nearby objects move faster than the faraway objects and therefore gain a higher depth value than the background objects. In this way, relative motion can be used to estimate the depth map.

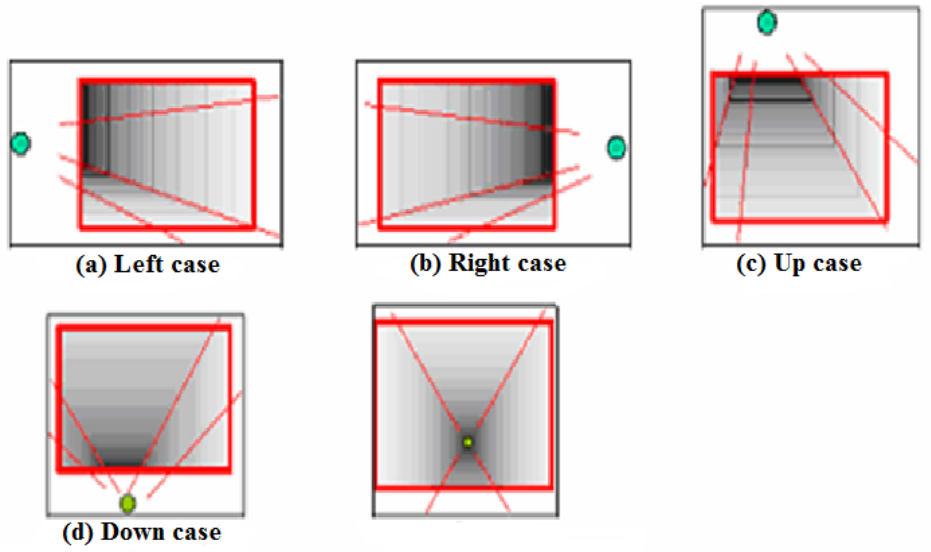


Figure 1.3: Gradient plane assignment [6].

1.2.2.7 Shading Method

The Shape-from-Shading (SFS) [7] method refers to the reconstruction of 3D shapes from intensity images by utilizing the relation between surface geometry and brightness of the image. The gradual variation of surface shading in an image reveals the shape information of the objects. Most of the SFS algorithms make use of the Lambertian reflectance model. A uniformly illuminated Lambertian surface looks identically bright in

every manner. Calculation of the depth map requires solving complex equations and therefore, the technique is considerably complicated.

1.2.3 MakeMe3D Software

The MakeMe3D software [8] uses self-developed object recognition and motion analysis to perform the 2D-to-3D conversion. This technique is mainly based on the Depth-from-Motion (DFM) concept as explained in section 1.2.2.6. The technique used by MakeMe3D is called “Morph3D” and is supposed to achieve good results for movies with image movement rotating around an object, but has difficulties creating a depth effect in scenes lacking image movement.

To enhance the 3D effect, two parameters can be adjusted: a negative and a positive percentage indicating the value of depth and the used frame offset. All the negative percentages bring the image to the back and all the positive percentages bring it to the front. MakeMe3D predicates the computation for the depth effect upon the analysis of a number of single frames. The default value is ‘1’ which means that the analysis considers frame ‘0’ as well as frame ‘1’. When adjusting the value to ‘2’, the algorithm considers frame ‘0’ and ‘2’ and skips frame ‘1’. It skips the image in between, which results in a better three-dimensionality for movies with little motion.

1.3 Image Formation Model

A brief understanding of image formation in a conventional camera is also deliberated in this chapter. When a scene is captured, whether in reality (photography) or from the computer (rendering), dissimilar images are created based on the setup of the camera. Depending on the parameters of lens and camera, the output image may distort or blur. The two utmost used camera models are described in the next sections.

1.3.1 Pinhole Camera Model

It is the simplest and straightforward model available which is used in various real-time computer graphics application. The pinhole camera has a small aperture and the lens is

also absent. Light emerging from a scene goes through the aperture and reaches the image plane forming a sharp image because only one ray per pixel is permitted through the hole. Figure 1.4 illustrates the working of this model where a single ray hits the image plane resulting in a focused image. This model can be effortlessly achieved, reflecting the main reason why it was frequently used in the past. The amount of light entering the aperture can be increased if the aperture size is large. However, this phenomenon will result in a decrease of sharpness level in the image formed. So, a new model is developed to refocus the image, which is discussed in next section.

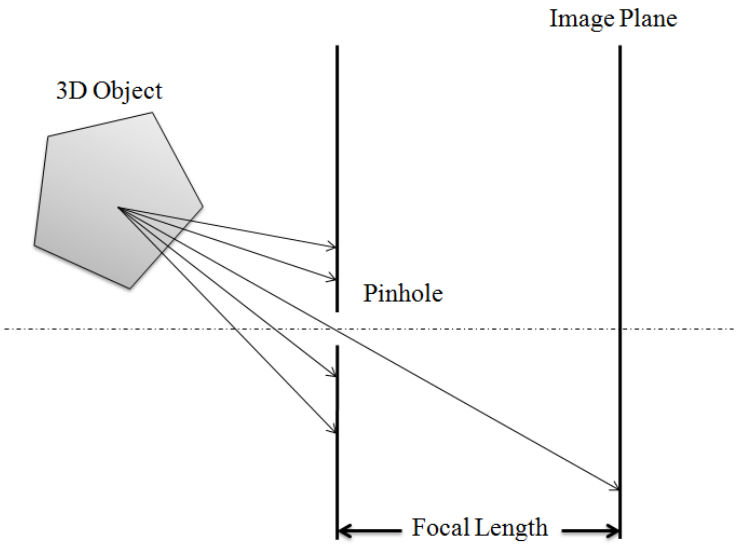


Figure 1.4: Pinhole camera model [9].

1.3.2 Thin Lens Camera Model

In this section, we will examine a widely used thin lens model [9, 10]. Digital imaging devices are preferred in various imaging areas including consumer photography, medical imaging, astronomical photography, aerial imaging, microscopy etc. But, entire imaging system goes through two frequent distortions, i.e. blur and noise. Conventional cameras having low f-number have a little depth of focus, which generally leads to defocus blurring. Small-scale defocus last in almost every image. This model converges a cluster of light rays emitted from each point in a scene by using a thin lens. The rays are refracted by the lens and thus refocused to a point either onto or beyond the image plane.

If the bundle of rays gathers onto the image plane, the image produced will be in focus. On the other hand, if the rays unite at a point beyond the image plane, they will form a Circle of Confusion (CoC) [11] as depicted in Figure 1.5. The diameter of CoC defines the measure of defocus as:

$$c = \frac{\left|d - d_f\right|}{d} \frac{f_o^2}{N\left(d_f - f_o\right)} \tag{1.1}$$

where, d_f is the distance of focus, d is the distance beyond the focal plane, f_o refers to the focal length of the camera and N is the stop number.

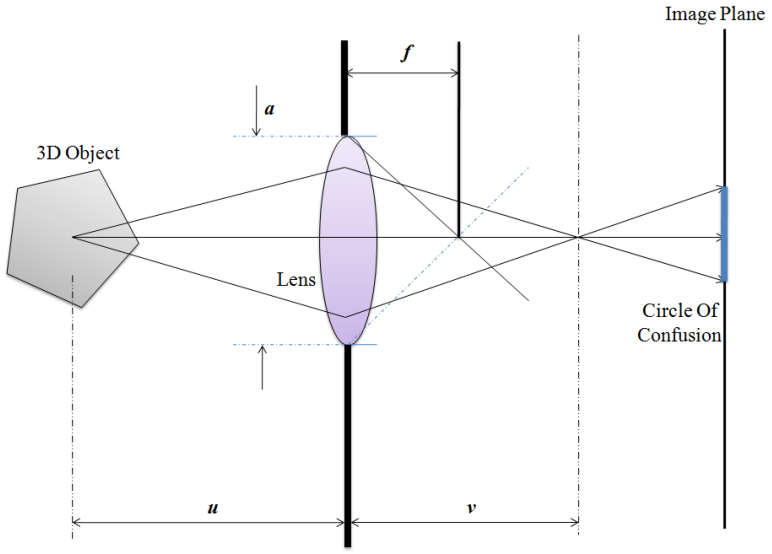


Figure 1.5: Thin lens model [9].

1.4 Three-Dimensional Image Formation

The deficiency of 3D contents has become a severe bottleneck for the complete 3D system. It includes the generation of 3D contents and its representations, transmission, visualization and coding. Various techniques have been proposed in recent years such as holographic techniques, volumetric 3D display, multiview and binocular autostereoscopic displays. Presently, the production cost is very high and therefore, there is an urgent need

of efficient 2D-to-3D conversion techniques. This section is about the generation of 3D image techniques which utilize either multiple images or a single image.

1.4.1 Using Depth Cues in Still Picture

For static images, disparity extraction method is applied which is dependent on the contrast, the sharpness, and the chrominance (croma). Sharpness is related to high spatial frequencies, whereas the contrast is affiliated to intermediate frequencies. Chroma belongs to the tint and the hue of the color. This technique which is established onto these parameters is known as the computed image depth (CID) algorithm [12, 13, 14] which converts still 2D images to 3D images. The near and far positional relationship between the objects is recognized by the sharpness and contrast existing in the input image.

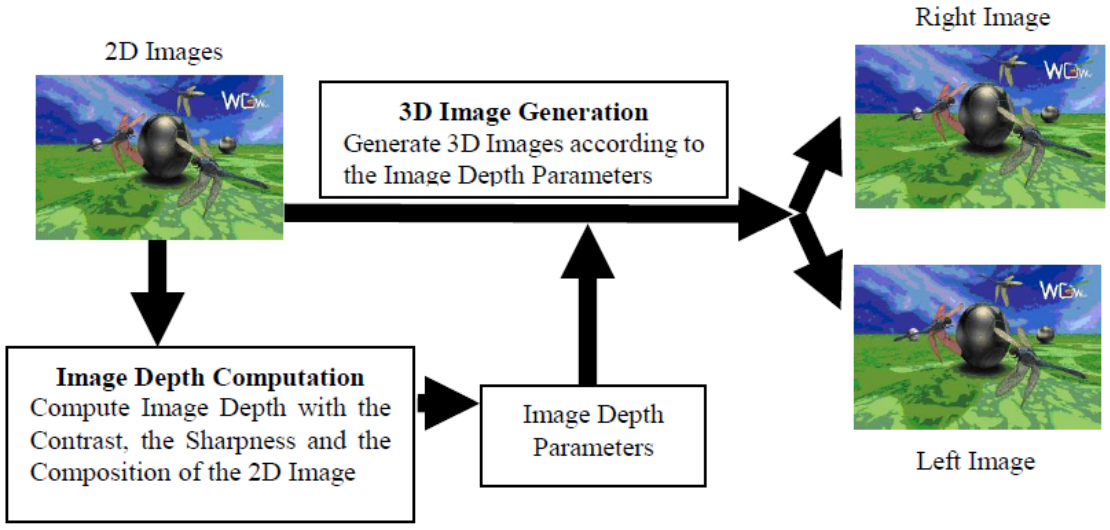


Figure 1.6: Workflow of computed image depth method [15].

CID mainly constitutes the following two steps. The computation of depth is the initial step that estimates the parameters of depth from the contrast, sharpness, and chroma of the image. The next step is the generation of the 3D images which is achieved in accordance with the depth parameters. Figure 1.6 demonstrates the fundamental basis of CID.

Firstly, the amount of sharpness, contrast, and chroma of every separated area in the given image is calculated. In addition to this, the neighboring areas which have nearly similar colors are arranged together in accordance with the chrominance values. The depth estimation process utilizes the contrast and the sharpness values. The objects that are closer exhibit a larger contrast as well as sharpness in comparison to the objects positioned at a further distance. Adjacent areas have nearly same chroma values, hence signifying that they have the similar depth. The three parameters used here grant the depth classification as far–mid–near shown in Figure 1.6.

Secondly, the process of 3D generation produces the pair of both eye images in accordance with the parameters of every grouped region. If the region parameters imply near, the left images are produced by transferring the input images to the right. Also, the right images are created by transferring them to left. However, if the parameter of a region denotes a far, the two are made by transferring to every opposite direction. The horizontal transfer value of every separated region is proportional to the 3D effect. Furthermore, when the depth parameters change abruptly, the changed images become tough to watch. Therefore, every transferred value adjusts to decreasing the quick changes of the depth parameters among the adjacent regions which result in the generation of 3D pictures which are easy to visualize. The CID is specifically appropriate for transforming still pictures as it doesn't require any motion of the objects in the pictures.

1.4.2 Using Depth Map

Three attractive features in calculating the depth information of images are used, which are as follows: gray scale analysis, relative spatial setting, and multiview 3D are rendering. A colored picture is easily transformed to an intensity value I with a gray scale as [15]:

$$I = \frac{I_R + I_G + I_B}{3} \tag{1.2}$$

where, the right-hand side incorporates the magnitude of the colors. Figure 1.7 and 1.8 illustrates the gray scale conversion. The gray scale I is extended to I' ranging from 255 to 0 considering an 8-bit word by:

$$I' = \frac{\left(I - I_{\min}\right)}{\left(I_{\max} - I_{\min}\right)} 255 \tag{1.3}$$

This is known as the dynamic contrast enhancement. Figure 1.7 illustrates the appearance of the picture after the completion of all the steps. In the further step, the luminance of the image is reset by assigning a little luminance to the uppermost part that is eventually occurring bright toward the lowermost part.

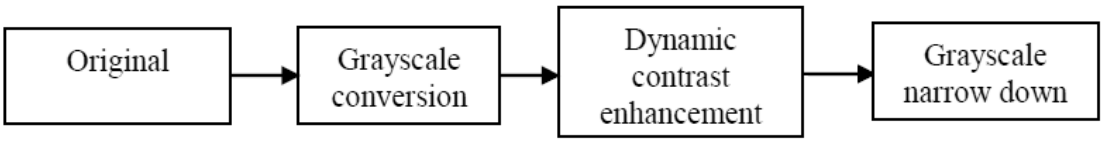


Figure 1.7: The gray scale conversions of an image [15].

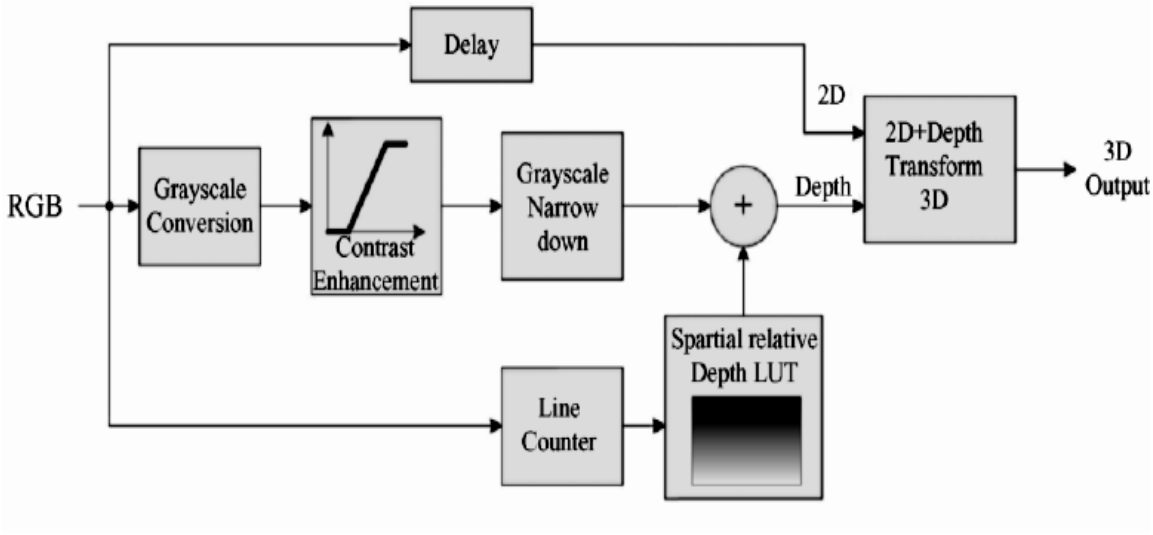


Figure 1.8: Block diagram for gray scale conversions [15].

After employment of the setting, the picture having darker gray color towards the bottom denotes a very impressive sensation of depth.

	R	G	B	R	G	B	R	G	B	R	G	B
Line 1	1	2	3	4	1	2	3	4	1	2	3	4
Line 2	4	1	2	3	4	1	2	3	4	1	2	3
Line 3	3	4	1	2	3	4	1	2	3	4	1	2
Line 4	2	3	4	1	2	3	4	1	2	3	4	1
Line 5	1	2	3	4	1	2	3	4	1	2	3	4

Figure 1.9: The pixel arrangement for four different views [15].

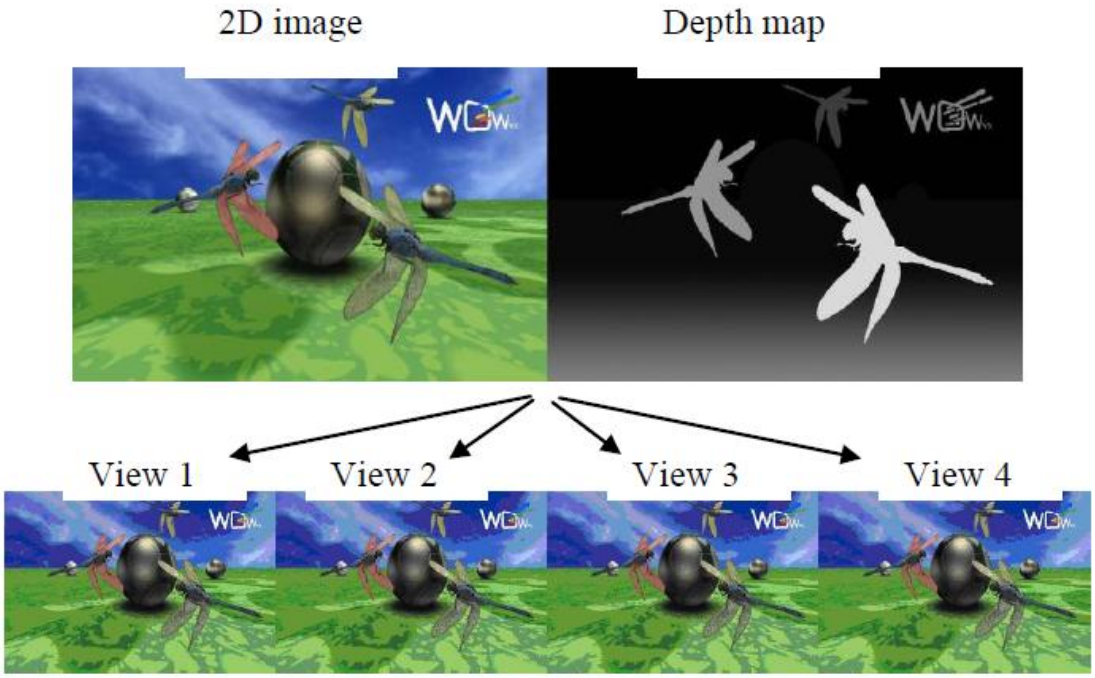


Figure 1.10: The 2D image and its depth map in the upper line. The four views are in the lower line [15].

1.5 Applications

3D models and 3D viewing are catching great pace in the area of computer vision because to its applicability in diverse fields of heath, aerospace, textile etc. 3D models

have been utilized in a wide variety of industries. The medical industry operates them to build detailed models. The movie industry employs them to manipulate characters and objects for animation and real-life moving pictures. The video game production exploits them to make resources for video games. The science section manages them to build highly detailed models of chemical fusion. The architecture area uses them for making models of advanced buildings and scenery. The engineering community makes use of them to design new tools, motor vehicles, and constructions.

1.6 Organisation of Thesis

This dissertation is divided into five chapters. The first chapter introduces different types of 3D image generation techniques which mainly focus on the approaches of depth map generation. The second chapter presents an overview of the most relevant prior work of depth information extraction. Chapter third analyses the problem of reconstructing both the depth map and 3D image from the single small-scale defocused image. Experimental results of the proposed depth estimation technique are given in the fourth chapter. It has also been shown that the estimated depth maps can be used to generate 3D images. Also, the performance comparisons are made with the previous methods. Finally, important observations are made and useful conclusions are drawn in the last chapter along with a discussion of the future scope of this work.

CHAPTER 2
LITERATURE REVIEW

Experience is what you get when you didn't get what you wanted.
And experience is often the most valuable thing you have to offer.
Randy Pausch [1960-2008]

2.1 Introduction

This chapter gives an impression of what is currently going on in the research area of 2D-to-3D conversion. The motive of this review is to throw some light on the major state-of-the-art technologies. The first part explains the concept of 3D scene capturing technologies such as single camera, multi-camera, and holographic techniques. Thereafter, the focus is on different depth estimation approaches. Finally, the conventional methods of estimating depth maps are discussed to help understand the methodology covered in this dissertation.

2.2 Scene Capturing Technologies

The capturing of the 3D content of dynamic scenes is undoubtedly vital for 3DTV implementations. In what follows, methods for generating 3D scene from single and multiple cameras data streams are outlined, as well as the holographic technique is discussed [7]. While all techniques have their difficulties, the main point remains that additional hardware or processing power is necessary to capture the depth information.

2.2.1 Single Camera Techniques

There exist several techniques for capturing 3D scenes by utilizing a single camera video sequence, described as Shape-from-Motion (SFM) methods. Currently, the SFM seems to be the best solution because of its applicability to general scenarios. All the other techniques (e.g. Shape-from-Texture (SFT), Shape-from-Focus (SFF)) are only used to achieve 3D shapes in controlled 3D environments.

E. Stoykova *et al.* discussed the SFM technique [16] which attempts to resolve for 3D geometry by making use of the relative motion of the camera and the viewed scene. This kind of relative motion produces a relevant cue to depth impression and can be perceived as a model of "disparity over time". The motion field contains the 2D velocity vectors of the image pixels, encouraged by this relative motion. Objects are assumed to be undeformable and their movements to be linear.

There are Shape-from-Texture (SFT) mechanisms which can be used to create a good 3D impression. A new technique which automatically selects scale is proposed by T. Lindeberg *et al.* [17] that depend on normalized derivatives. It is utilized for adaptive determination of the two-scale parameters in a multi-scale texture descriptor, the windowed second-moment matrix, which is defined in terms of Gaussian smoothing, first order derivatives, and non-linear pointwise combinations. The similar scale selecting technique is made to work for multi-scale blob detection not having any tuning parameters. The emerging texture description is then associated with different presumptions regarding the texture of the surface for evaluating surface orientation. Figure 2.1 shows a typical shape reconstruction based on texture features.

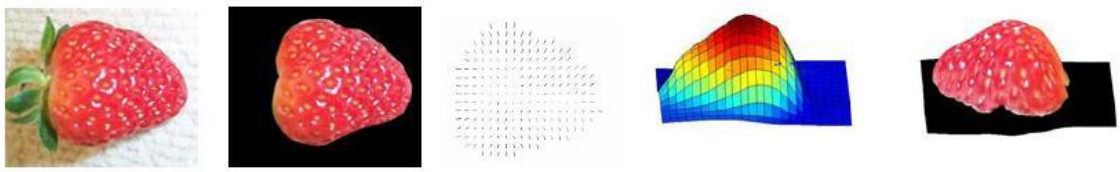


Figure 2.1: Shape-from-texture (From left to right: original image, segmented texture region, surface normals, depth map and reconstructed 3D shape) [7].

2.2.2 Multi-Camera Techniques

Multi-camera setups capture a dynamic scene from multi-view points at the similar time. Issues that relate to this kind of systems are synchronization and calibration of the cameras.

A. N. Rajagopalan *et al.* [18] proposed a method of extracting depth from pictures taken with real aperture camera by intermixing defocus and stereo cues. Markov random fields

and suitable energy function are used to model the depth maps. The technique is computationally less efficient but this method also attains simultaneous recovery of both depth map and originally focused image.

U. Mudenagudi *et al.* [19] used two defocused stereo image pairs for measuring depth information. The important advantage is that the recovery of the depth map and image restoration is achieved simultaneously. The precision and quality of depth estimation are much better in comparison to Depth-from-Defocus (DFD) methods.

A new approach named Spatial-Domain Convolution/Deconvolution Transform or S-Transform method (STM) is given by M. Subbarao *et al.* [20]. The technique estimates the distance of objects as well as fast autofocusing of the cameras. STM uses defocus information of just two images captured with particular parameters of the camera like lens position, the diameter of the aperture as well as the focal length of the camera. The images are arbitrarily blurred. Hence, STM is fast when compared to Depth-from-Focus (DFF) methods that look for the location of the lens or focal length of perfect focus. This technique includes easy operations. Furthermore, it is very easily to retrieve the depth information of the image.

2.2.3 Holographic Techniques

Holography [16] is an exclusive method which records and reconstructs the 3D information of a picture. A hologram basically records the interfering pattern acquired with the process of superimposing a reference beam with the scattering beam of the object. In traditional holography, photographic films are utilized to register holographic patterns and optical restoration is carried out. Recently, however, digital holography has replaced holographic films with charged-coupled devices (CCD) and complementary metal-oxide-semiconductor (CMOS) image sensors to record and numerically reconstruct holograms. The fundamental procedure of recording a digital hologram is illustrated in Figure 2.2.

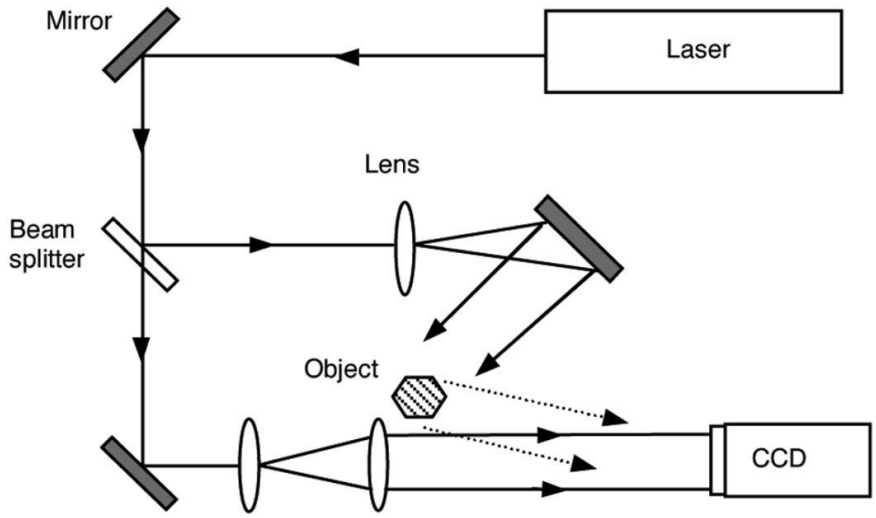


Figure 2.2: Device for recording digital holograms [16].

2.3 Depth Estimation Strategies

Depth estimation techniques are categorized into two main classes: *active* and *passive approaches*.

2.3.1 Active Approaches

In these methods, a controlled beam of energy is sent and the reflected energy is then processed. Auxiliary tools are employed for depth perception in addition to the camera. These techniques are highly accurate but the disadvantage is the need of energy, high complexity, and cost. Few of these methods are used to attain the ground truth depth maps.

B. Girod *et al.* [21] presented a different range sensing algorithm that utilizes defocus of the organized light beam to find out the depth. The technique is a continuation of the initial passive Depth-from-Defocus (DFD) concept to an actively organized light beam system. The improvement in this technique is done by utilizing anisotropic aperture or astigmatic optics as the source of light. Both of these light sources use an isotropic well-organized light arrangement and correlate blur in two perpendicular directions.

Later, B. Girod *et al.* [22] proposed another efficient method in which a range camera is equipped for the calculation of the distance to a scene by making use of a system which throws a restricted depth of field light pattern towards the scene and estimates the blurring of the pattern on that scene. This blurring is then deducted out to remove system inaccuracy from the range measurement. The range camera is very effective to yield a good range picture or depth map, with extra outputs granting localized albedo and a traditional brightness picture.

The Time-of-Flight (TOF) principle is used by S. Schuon *et al.* [23] to estimate the time taken by an emitted light pulse (e.g. IR LEDs, or Laser) to reach the camera sensor. High-quality 3D measure of a static scene is achieved with the use of TOF camera.

J. Salvi *et al.* [24] suggested an algorithm in which they throw a light pattern and see the lighted scenery from either one or several viewpoints. These patterns are coded and the comparisons in image points and projected pattern points can be easily estimated. Then, by using triangulation theory, 3D information is retrieved.

A different single view depth estimation system is proposed by F. M. Noguer *et al.* [25] which depends on defocus of a scattered bundle of dots thrown towards the image for the purpose of refocusing the image. The projected patterns of dots are depth clues for each portion of the image which can be eliminated after estimating depth. The need of utilizing a projector restricts its functional environment, for projectors are often inefficient for an outdoor scene.

C. Hua *et al.* [26] proposed a depth map optimization approach to producing a high-quality depth map. The approach is based on surface normal which is used to segment image regions. The surface normal is computed by getting a normal vector for each point after the pre-processing of the initial depth map and later, mean shift algorithm is used to get segmentation results. The linear smoothing method is adopted soon after the direction of every segmented block is figured out. Finally, the refined depth map is concluded by combining the smooth factor and the initial depth cues.

2.3.2 Passive Approaches

These are the methods which work in the presence of natural light and with the information present in the images captured by the image sensors.

A novel method was proposed by N. Asada *et al.* [27] to retrieve the reliable edge and depth information by integrating a set of multi-focus pictures, i.e., a series of images captured by systematically changing the camera focus. The blur is then measured from the difference in intensity beside the pixels in the multi-focus images. Such a blur analysis enabled not only to detect the edge points without using spatial differentiation but also to calculate the depth with high accuracy. In addition, the analysis result was stable because the proposed method involves integral computations such as summation and least-square model fitting. Two algorithms were presented: edge detection using an accumulated defocus image which represents the spatial distribution of blur, and depth estimation using a spatial-focal image which represents the intensity distribution along focus axis.

A. Levin *et al*. [28] proposed an easy alteration to the traditional cameras by which the simultaneous recovery of both high-resolution image and depth information is achieved. A patterned occlude is placed inside the aperture of the camera lens which forms a coded aperture. A criterion for depth discriminability is introduced which is utilized to draft the required pattern of the aperture. By making use of a statistical model of images, recovery of both depth map and a sharp image is accomplished. A layered depth map is thereafter produced by the user-drawn strokes to simplify layer assignments in some cases.

V. P. Namboodiri *et al*. [29] investigated the difficulty of extracting the depth layers of a single blurred image. The inhomogeneous reverse heat equation is utilized to acquire the amount of the blur which preserves the depth details distinguished by the defocus. But, due to the parabolic nature, the reverse heat equation is varying. So, this equation made to be stable with the usage of gradient degeneration considering it to be a productive stopping benchmark. The quantity of diffusion is basically an estimate of relative depth. Due to the ill-posedness, a graph-cuts algorithm is suggested to conclude the depth in the

image which uses the measure of diffusion as data likelihood and a smoothness criterion on the depth in the scene.

P. K. Chan *et al.* [30] also employed reverse heat equation for resolving the defocus cues created by the depth of the lens through which the details of depth from a single image can be extracted. Although several applications demand the depth map to be smooth, a mean shift segmentation, and graph cut based technique is suggested to deduce the depth information of an image. The important information lies in the energy function of graph cut which preserves the depth map details.

A well-organized method to measure absolute depth using a single defocused image is proposed by J. Lin *et al.* [31]. Rather than measuring defocus measure of every point, a sequence of aperture shape filters is designed for the segmentation of a defocused image according to the defocus value. A boundary-weighted belief propagation approach is applied to acquire a polished depth map. Experimental outcomes proved that the approach outperformed the previous single-image SFD algorithms both in the precision of the calculated absolute depth and time taken for the implementation.

J. Konrad *et al.* [32] proposed an algorithm which is established on the distinct method of learning the 2D-to-3D conversion using examples. Two classes of techniques are introduced. The first method is dependent on learning a point mapping from local image features like color, spatial position using a regression type idea. The other algorithm globally estimates the depth map of the picture directly from a database of 3D images (image+depth pairs) by making use of a nearest-neighbour regression type idea. The computational efficiency of both the techniques is verified on various 2D images.

Two different techniques for depth map upsampling are suggested by C. Jung *et al.* [33]. An upsampling approach for depth using image decomposition is first proposed. The colored image is segmented into its structure and texture layers, and by utilizing the structure component, the reconstruction of depth values is done. Moreover, the structure information based algorithm is expanded to a hybrid depth upsampling method. This method takes the advantage of both structure and color maps. Experimental results

demonstrated that the proposed depth map upsampling algorithms performed superior to the earlier algorithms in terms of the bad pixel rate.

2.4 Conventional Methods

The depth estimation would is altogether a complex research area, where several algorithms and setups have been presented. The previously discussed methods of extracting out the depth can be broadly classified into six main classes as discussed in Table 2.1.

2.4.1 Monocular Cues

There are a large number of monocular cues like variations in texture and gradients, defocus, color, etc. which include helpful depth information. Extracting depth from only one image by utilizing monocular cues demands a serious measure of previous understanding since there exist an elemental uncertainty among local features of the image and depth variations. Depth estimation from monocular cues is a difficult job, which incorporates the global structure of the picture.

A supervised learning method is addressed by A. Saxena *et al.* [34] for evaluating depth map by using a single monocular image. They utilize training set of monocular images (of the unconstrained environment) and their correlative depth maps. By employing Markov Random Fields (MRF) that included multiscale global and local image detail, depths and relationship between depths are captured at the distinctive point. This algorithm assesses the relative depth of scene quite well but fails in estimating absolute depths. Also, this method incurs the largest error on images which is composed of very irregular patterns, e.g. trees.

A fully automatic procedure based on image classification has been presented by S. Battiato *et al.* [6]. The technique classifies digital images as indoor, outdoor without or with geometric elements. This information is then used to calculate depth from a single input image. This technique is purely unsupervised and requires low computational devices. It is also well suited for real-time applications.

Table 2.1: Classification of depth estimation strategies.

CLASS	SUMMARY	DISADVANTAGES
Class I	This class is developed on human visual model, i.e. uses two cameras in sync placed on a horizontal plane with a specific distance between them, and focuses on the same object	The use of two cameras is the main problem because it makes it expensive and useless if one of the cameras malfunctions. It requires mechanical moving and adjustment so as to focus on the same object.
Class II	This class uses a single camera and measures the quantity of image resizing in accordance with the camera movement to create the 3D effect.	It requires prior knowledge of the size of the object along with the focal length and other parameters of the camera lens.
Class III	This class is used to develop the distance of moving objects. In this, the camera is stationary and uses the following parameters to compute the distance: maximum velocity, little velocity changes, coherent motion, and constant motion.	Due to the employment of various parameters, the computational complexity of the estimation model is very high.
Class IV	This class utilizes the sequence of images captured by the single camera for estimating the depth map, from the geometric model of the object and camera.	Depth map estimation is very difficult to estimate if the object is very near to the camera.
Class V	In this class, the methods use blurry edges to calculate the depth perception of the image. The original image is observed as a convolution of the in-focus image of the object with point spread function (PSF). The PSF depends on upon the parameters of the camera and the range of the object to the camera.	The calculations of PSF from the camera parameters and range of the object to the camera are very large and thus, difficult to compute.
Class VI	This class utilizes auxiliary devices like LEDs to develop the depth perception. The laser beam is projected on surface and camera captures the spots in the image where LEDs are situated. The triangulation of these spots with respect to camera builds the depth map of the image	The use of auxiliary devices along with the camera makes this model expensive and complex to employ on general-purpose scenarios.

B. Liu *et al.* [35] make use of semantic segmentation of a scene to guide the possible 3D reconstruction. Firstly, the semantic class of every pixel or region is predicted allowing to model the appearance and geometry constraints that were not possible in previous works (e.g. [34]), and then the depth map is estimated.

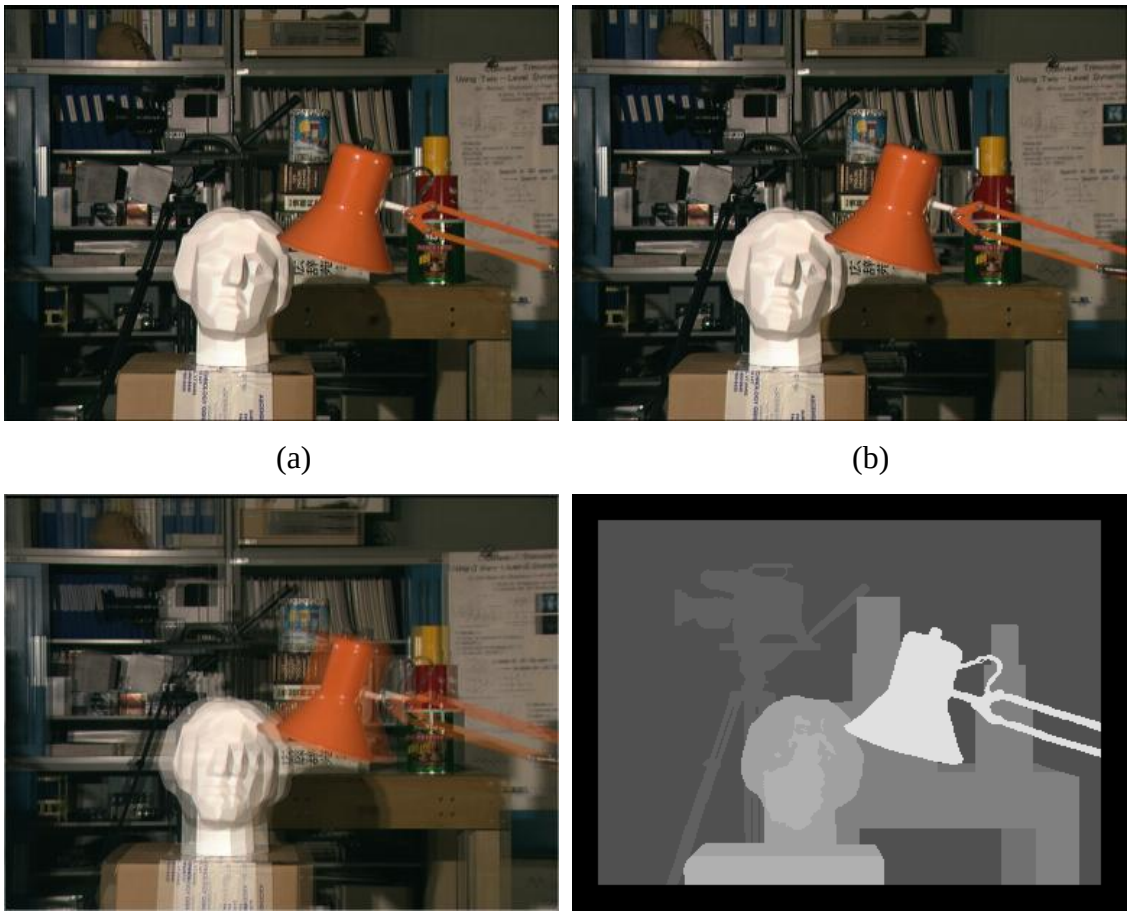
2.4.2 Stereo Cues

In computer vision and robotics, depth information can be acquired most often via *stereo vision*, also known as *stereo matching* or *stereo correspondence*. In this process, images from two different cameras (also called as stereo cameras) are utilized to triangulate and evaluate distances. When the cameras capture the scene from two distinct angles, the two images produced will be disparate. The stereo correspondence for images depends on calculating this disparity. Later, the triangulation approach is employed to regain the 3D structure.

Even though acceptable stereo vision systems are developed over the past few decades, it is essentially restricted to the baseline range between the two dissimilar cameras. Particularly, depth values are unreliable if the distance is large. Additionally, this technique fails for the regions with less texture where correspondences cannot be faithfully estimated.

Figure 2.3 shows the *Tsukuba* stereo image pair along with a disparity map [36]. The goal is to search the matching pixels of the two pictures and value of the horizontal distance between the two matching pixels which combine to form a disparity map. Then by applying triangulation technique, a depth map is finally generated. In Figure 2.3, the disparity is larger for the lamp and the head compared to the objects in the background.

While a huge count of techniques for stereo correspondence is invented, relatively less amount of work is done on characterizing their performance. D. Scharstein *et al.* [37] gave a taxonomy of dense, two-frame stereo methods. This taxonomy is modeled to assess the different components and design decisions made in individual stereo algorithms. Using this taxonomy, comparison of previous stereo methods and presented experiments is done which evaluates the achievements of several different variants.



(a) (b)

(c) (d)

Figure 2.3: The Tsukuba stereo image pair [36], (a) Left view, (b) Right view, (c) Superimposed left and right view, and (d) Disparity map.

M. Z. Brown *et al.* [38] reviewed latest advances in computational stereo, focusing importantly on three basic topics: correspondence methods, occlusion methods, and real-time implementations. The tables are presented which compiles and draw a distinction between key ideas and techniques.

Estimating the stereo images from the image series is a challenge for several years. F. Liu *et al.* [39] proposed a method for estimating the disparity maps from stereo images. They designed the temporal consistency of disparity maps by making use of scene flow. It is built into the stereo model for the next frame as a soft limitation that helps in resolving stereo vagueness from the temporal domain while in the intervening time reduces the propagating error in the disparity maps.

2.4.3 Camera Focus

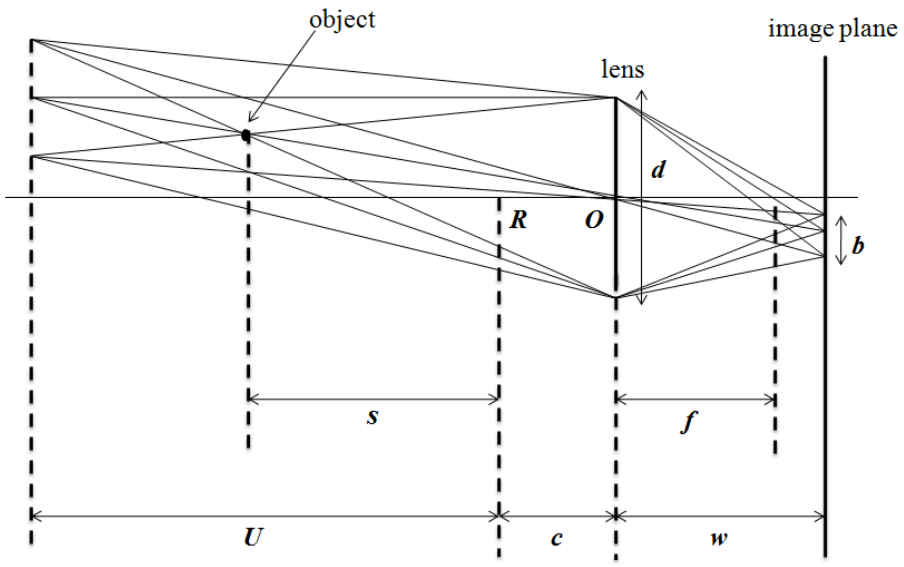


Figure 2.4: Width of blur of the point object [40].

When a point object is photographed from a camera, the image produced is also a point object provided that the camera is correctly focused by the photographer. In case the camera is not properly focused, the image produced of the point object is a circular disc constituting the blurred region with some thickness. Figure 2.4 demonstrates that the width b of defocused image has a point object P which is a function of model parameters d, w, c, U and an object distance s. Using the geometric relationship of these parameters, b is produced as [40]:

$$b = wd\left|\frac{1}{s+c} - \frac{1}{U+c}\right| = \frac{w^2}{B}\left|\frac{1}{s+c} - \frac{1}{U+c}\right| \tag{2.1}$$

where, $B = w/d$.

From the above equation, the depth s of an object is obtained from the width of blur b [40],

$$s = \begin{cases} \dfrac{wd\,(U+c)}{wd + b(U+c)} - c, & (s < U) \\ \dfrac{wd\,(U+c)}{wd - b(U+c)} - c, & (s > U) \end{cases} \tag{2.2}$$

where, w, c, d and U are produced by zoom and focus settings. Thus, s is calculated from the width of b at any settings of the three parameters. From a bundle of multiple pictures captured by different parameters, a strong depth data is achieved.

S. K. Nayar *et al.* [5] presents a fully automatic Shape-from-Focus (SFF) method which utilizes several divergent focus levels to retrieve a sequence of images. The sum-modified Laplacian (SML) operator is well-established to supply local estimates of the characteristics of focus. This operator is employed to the retrieved image sequence to conclude a series of focus estimates at every pixel in order to obtain accurate depth estimates.

A spatial domain convolution/deconvolution transform is described by G. Surya *et al.* [41], for estimating the distance of objects using defocused image. The method involves two defocused images captured with different aperture diameters. This method has less accuracy as it only gives a rough measure of distance. However, this measure can then be explicitly used by stereo algorithms to determine a more accurate estimate of distance. Also, this algorithm reduces the accuracy when cameras have a large depth of field as it minimizes the difference in blur between objects at varying distance. But this method is fast when compared to Depth-from-Focus (DFF) methods.

S. K. Nayar *et al.* [42] presented a range sensor which is dependent on focus investigation which creates a 512×480 depth map at 30Hz (video frame-rate). The textured and textureless areas are retrieved by utilizing an illumination arrangement which is protruded with the help of the similar light path utilized to obtain pictures. The illumination pattern is optimized to increase the accuracy and the resolution in estimated depth. The blur in the pictures is calculated by making use of narrow-band linear operator

which is modeled by taking into consideration all the elements of the depth from defocus system.

2.5 Gaps in Study

Previous depth estimation techniques concentrated on the use of binocular cues such as stereo vision [37], which is generated by capturing two images of the same scene taken from slightly disperse locations. In reality finding correspondence between the two images was a difficult task and furthermore, objects at far distance used to have zero disparity. Therefore, there was a need to find optimal depth in such scenarios.

In the literature, enough work has been reported based on stereo images motion, however, relatively little work is on estimating depth information from a single image. Saxena *et al.* [34] predict a depth from a set of image features using linear regression and an MRF, and later extend their work into the Make3D [14] system for 3D model generation. However, the system relies on horizontal alignment of images, and suffers in less controlled settings.

2.6 Objective of the Thesis

The main goals of this dissertation are:

- To develop a new approach of estimating the depth information from a single small-scale blurred image.
- To utilize the generated depth maps from the proposed approach for 2D-to-3D conversion of the images.
- To compare the performance of the proposed algorithm with the existing techniques.

2.7 Chapter Summary

An overview of some of the most common 2D-to-3D conversion techniques, highlighting advantages and limitations of the previous approaches, have been presented in this chapter. A review of the different scene capturing technologies along with active and

passive methods of estimating depth maps is presented additionally with a discussion of monocular and multiview systems for calculating depth information.

CHAPTER 3
DEPTH ESTIMATION FOR 2D-To-3D CONVERSION

Design is not just what it looks like and feels like. Design is how it works.
Steve Jobs [1955-2011]

3.1 Introduction

The main objective of this chapter is to describe every step of the proposed method more effectively. An efficient method is presented to estimate depth from the single small-scale blurred image by using sharpness map. The chapter also includes a method that creates a 3D sense in the input image to enhance the image quality.

First of all, the amount of sharpness is estimated at the edges to recover sparse depth map. Then, the matting Laplacian interpolation algorithm [43] is used to obtain full depth map. Finally, the retrieved depth estimates are used to attain a 3D image. A visual and quantitative comparison of the suggested technique with the existing approaches is also given.

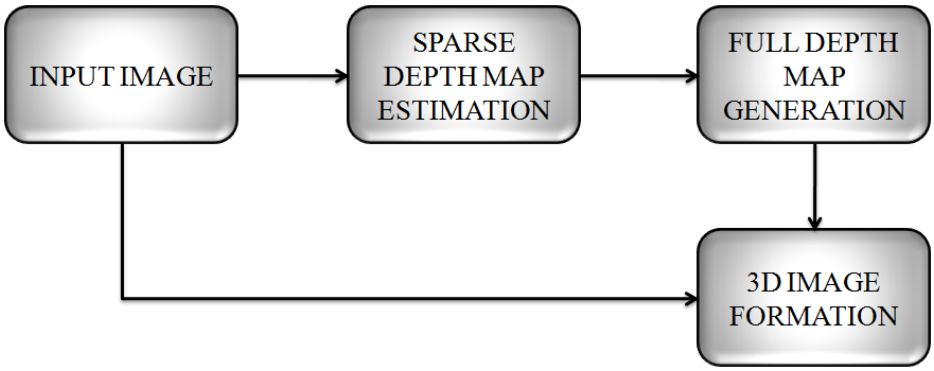


Figure 3.1: Block diagram of the proposed method.

Figure 3.1 explains the block diagram of the suggested scheme. The technique consists of three key steps: sparse depth map estimation, full depth map generation and 3D image formation In the first step of sparse depth map estimation, the original image is re-blurred and the extent of sharpness residing at the edges is estimated by utilizing image gradients. The sharpness map is used to infer the depth at edge locations, forming sparse depth map. The matting Laplacian interpolation algorithm is then used to engender the estimated depth from the edges to the entire image, forming a complete depth map. At last, this depth map along with the original image together constitute an image having 3D perception.

3.2 Dataset Used

The high-quality indoor RGB images are taken from Sony A7 camera and the corresponding ground-truth depth map is obtained by using Kinect [44]. The outdoor dataset used in this dissertation is captured by Fujifilm Real 3D camera [44]. Their ground-truth depth is extracted from stereo matching [45].

3.3 Depth Map Estimation

In this section, depth information is estimated from a small-scale defocused image with the help of spatially varying defocus blur at the edges. The subtle defocus blur present in the image restricts the measure of sharpness around edges. This characteristic feature is then used as a cue for estimating the depth of an image. The technique consists of two key steps: sparse depth map estimation and full depth map generation.

In the initial step of sparse depth map estimation, the original image is re-blurred and the extent of sharpness residing at the edges is estimated by utilizing image gradients. The extensively used PSF (Point Spread Function) approaches intend to recover considerable defocus PSFs. In the proposed method, PSF analysis has performed prosperously on the dataset. Further, the sparse depth map is modeled by utilizing sharpness amount residing at edges. The matting Laplacian interpolation algorithm is then used to propagate the

estimated depth from the edges to the entire image, forming a complete depth map. Figure 3.2 highlights the prime steps for the estimation of the depth map.

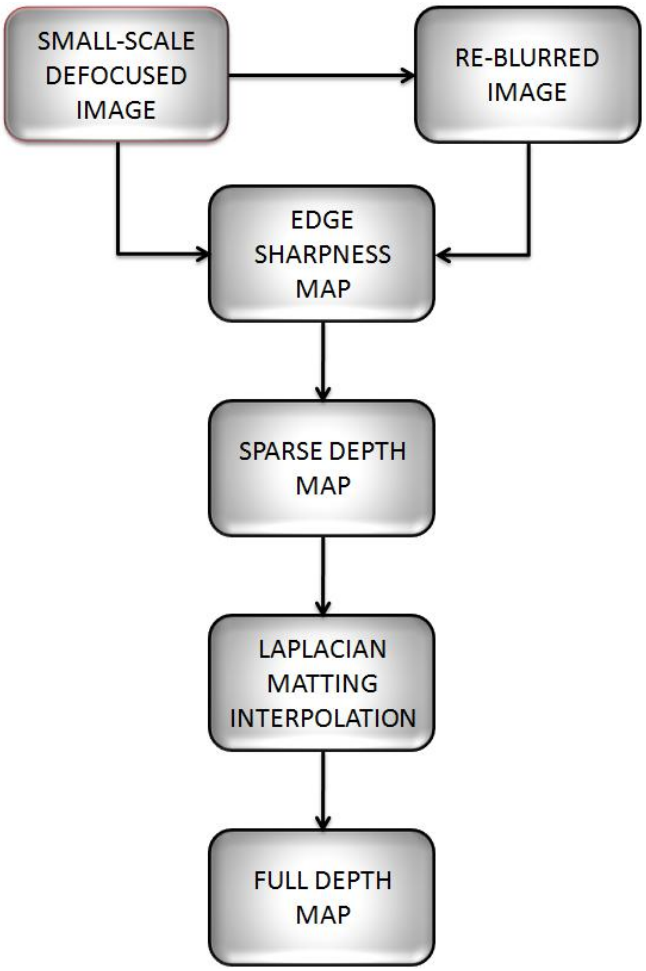


Figure 3.2: Workflow of the proposed depth estimation method.

3.3.1 Sparse Depth Map

In sparse depth map estimation, Gaussian blurring of the given image is carried out. This alteration to each pixel of the input image is exploited for measuring the sharpness amount at the edge locations. The image gradients are computed for determining the change in the image in a particular direction. The sharpness map formed by utilizing the magnitude of image gradients is used to infer the depth values at edge locations which

lead to the formation of the sparse depth map. Joint bilateral filter (JBF) is then employed onto this map which preserves the edges and reduces the noise present in it. At last, a rectified sparse depth map is obtained.

3.3.1.1 Gaussian Blurring

The quantity of blur change is appreciable at high-frequency image locations when the input image is re-blurred. Formally, the edge sharpness and contrast are reduced by slight defocus. Therefore, the sharpness measure is figured out at edges only. Gaussian filter is adopted which modifies the edges present in the image resulting in an overall effect named as Gaussian blur. This filter makes use of the Gaussian function to calculate the transformation. Figure 3.3 below shows ideal step edge function convolving with the gaussian filter resulting in a blurred edge.

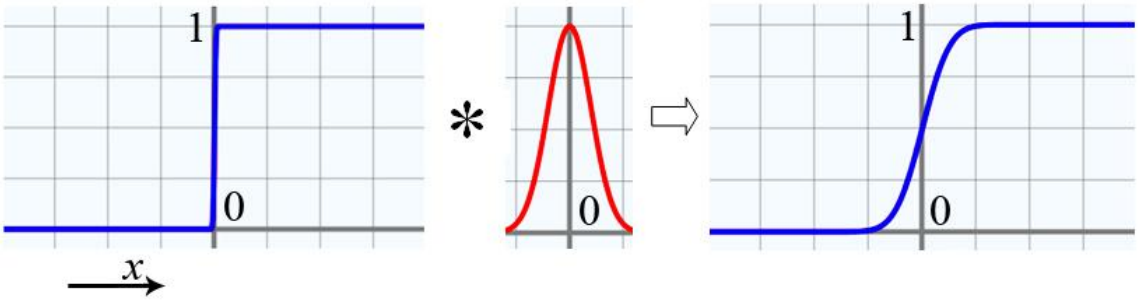


Figure 3.3: Edge model [46].

The one-dimensional (1D) equation of Gaussian function is defined as [10, 50]:

$$k(x,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{3.1}$$

And Gaussian function in 2D is the product of two Gaussians in each dimension:

$$k(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.2}$$

where, x and y is the distance in the horizontal and vertical directions respectively, σ is the standard deviation of the Gaussian distribution which evaluates the blur extent and therefore, is correlative to the diameter of CoC.

Now, the defocused image is the convolution of the sharp image with the PSF [3], i.e.,

$$i(x,y) = i_0(x,y) \otimes k(x,y,\sigma) + n \tag{3.3}$$

where, $i(x,y)$ and $i_0(x,y)$ correspond to blurred image and sharp image respectively. n is the sensor noise which is unknown and presumed to be Gaussian noise having zero mean. $k(x,y,\sigma)$ is the PSF which is approximated as 2D Gaussian function as shown in equation (3.2). In the proposed method, ideal step edges are examined which are given as [3]:

$$i_0(x,y) = Au(x,y) + B \tag{3.4}$$

where, $u(x,y)$ denote the step function as:

$$u(x,y) = \begin{cases} 0, & x,y < 0 \\ 1, & x,y \geq 0 \end{cases} \tag{3.5}$$

A represents the amplitude and B indicate the offset of the edge. The location of the edge is at $(x=0, y=0)$. With the known parameter σ_k of the Gaussian function, the original image is re-blurred and expressed as:

$$i_r(x,y) = i(x,y) \otimes k(x,y,\sigma_k) + n \tag{3.6}$$

where,

$$k(x,y,\sigma_k) = \frac{1}{2\pi{\sigma_k}^2} e^{-\frac{x^2+y^2}{2{\sigma_k}^2}} \tag{3.7}$$

The next step is to find the magnitude of gradients of the original and re-blurred images which are derived in the next section.

3.3.1.2 Gradient of Image

The magnitude of gradient expresses the change in the image as we move in either x or y directions. In the proposed scheme, the gradient magnitude of the re-blurred and input image is calculated to measure the sharpness of edges.

To calculate the gradient magnitude of the original image, the gradients along both x and y directions are independently computed first. The gradient along x axis is derived as:

$$\nabla_x i(x,y) = \frac{\partial}{\partial x}\left(i_0(x,y) \otimes k(x,y,\sigma)\right) \tag{3.8}$$

$$\nabla_x i(x,y) = \frac{\partial}{\partial x}\left(\left(Au(x,y)+B\right) \otimes \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}\right) \tag{3.9}$$

By using the derivative property of convolution:

$$\nabla_x i(x,y) = \left(\frac{\partial}{\partial x}\left(Au(x,y)+B\right)\right) \otimes \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.10}$$

Since, the derivative of step function is impulse function, therefore, the above equation reduces to:

$$\nabla_x i(x,y) = A\delta(x) \otimes \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.11}$$

$$= A\int_{-\infty}^{\infty} \delta(x-\tau) \frac{1}{2\pi\sigma^2} e^{-\frac{\tau^2+y^2}{2\sigma^2}} d\tau \tag{3.12}$$

By using sifting property of convolution,

$$\nabla_x i(x,y) = \frac{A}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.13}$$

Similarly, the gradient of the input image along y axis can be computed and the resulting equation comes out to be:

$$\nabla_y i(x,y) = \frac{A}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.14}$$

Now, the gradient magnitude of the original image is as follows:

$$\left|\nabla i(x,y)\right| = \sqrt{\nabla_x i(x,y)^2 + \nabla_y i(x,y)^2} \tag{3.15}$$

By putting values from equation (3.13) and (3.14) to the above equation, we get

$$\left|\nabla i(x,y)\right| = \sqrt{\left(\frac{A}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}\right)^2 + \left(\frac{A}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}\right)^2} \tag{3.16}$$

$$\left|\nabla i(x,y)\right| = \frac{C}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.17}$$

Here, C is a constant. In the same manner, the gradient of the re-blurred image is derived first along the x direction and then along the y direction. From equation (3.6),

$$\nabla_x i_r(x,y) = \frac{\partial}{\partial x}\left(i(x,y) \otimes k(x,y,\sigma_k)\right) \tag{3.18}$$

$$\nabla_x i_r(x,y) = \frac{\partial}{\partial x}\left(i_0(x,y) \otimes k(x,y,\sigma) \otimes k(x,y,\sigma_k)\right) \tag{3.19}$$

$$= \frac{\partial}{\partial x}\Big(\big(Au(x,y)+B\big)\otimes k(x,y,\sigma)\otimes k(x,y,\sigma_k)\Big) \tag{3.20}$$

Note that the convolution of two Gaussian is always a Gaussian function. Therefore, equation (3.20) consummate to:

$$\nabla_x i_r(x,y) = \frac{\partial}{\partial x}\Big(\big(Au(x,y)+B\big)\otimes k(x,y,\sigma+\sigma_k)\Big) \tag{3.21}$$

$$= \frac{\partial}{\partial x}\left(\big(Au(x,y)+B\big)\otimes \frac{1}{2\pi\left(\sigma^2+{\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}}\right) \tag{3.22}$$

$$= \left(\frac{\partial}{\partial x}\big(Au(x,y)+B\big)\right)\otimes \frac{1}{2\pi\left(\sigma^2+{\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}} \tag{3.23}$$

$$= A\delta(x)\otimes \frac{1}{2\pi\left(\sigma^2+{\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}} \tag{3.24}$$

$$= A\int_{-\infty}^{\infty}\delta(x-\tau)\frac{1}{2\pi\left(\sigma^2+{\sigma_k}^2\right)} e^{-\frac{\tau^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}} d\tau \tag{3.25}$$

Applying sifting property of convolution,

$$\nabla_x i_r(x,y) = \frac{A}{2\pi\left(\sigma^2+{\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}} \tag{3.26}$$

The gradient along the y direction can also be computed similarly and the resulting equation is:

$$\nabla_y i_r(x,y) = \frac{A}{2\pi\left(\sigma^2 + {\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}} \tag{3.27}$$

Finally, the magnitude of the gradient of re-blurred image is determined :

$$\left|\nabla i_r(x,y)\right| = \sqrt{\nabla_x i_r(x,y)^2 + \nabla_y i_r(x,y)^2} \tag{3.28}$$

$$= \sqrt{\left(\frac{A}{2\pi\left(\sigma^2 + {\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}}\right)^2 + \left(\frac{A}{2\pi\left(\sigma^2 + {\sigma_k}^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+{\sigma_k}^2\right)}}\right)^2} \tag{3.29}$$

$$\left|\nabla i_r(x,y)\right| = \frac{C}{2\pi\left(\sigma^2 + \sigma_k^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+\sigma_k^2\right)}} \tag{3.30}$$

Equation (3.17) and equation (3.30) represents the magnitude gradients of the original image and the re-blurred image respectively. These equations are used for calculating the edge sharpness which is broadly described in the next section.

3.3.1.3 Sharpness Map

The value of edge sharpness defines how acute the edge is. In the proposed approach, the Canny's edge operator [48] has been used for the detection of edges. This technique draws out effective structural information from the image thereby lowering the amount of information to be processed.

The edge sharpness is calculated from the equation given below:

$$E_s = 1 - \sqrt{\frac{\left|\nabla i_r(x,y)\right|}{\left|\nabla i(x,y)\right| + \rho}} \tag{3.31}$$

where, ρ is taken to be 0.0001 for avoiding the division by zero. Substituting the value of equation (3.17) and (3.30) to the equation (3.31), we get sharpness map as:

$$E_s = 1 - \sqrt{\frac{\frac{C}{2\pi\left(\sigma^2+\sigma_k^2\right)} e^{-\frac{x^2+y^2}{2\left(\sigma^2+\sigma_k^2\right)}}}{\frac{C}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} + \rho}} \tag{3.32}$$

Since, the value of ρ is very small, so it can be neglected. Therefore the above equation reduces to:

$$E_s = 1 - \sqrt{\frac{\sigma^2}{\left(\sigma^2+\sigma_k^2\right)} e^{\frac{x^2+y^2}{2\sigma^2} - \frac{x^2+y^2}{2\left(\sigma^2+\sigma_k^2\right)}}} \tag{3.33}$$

The edges are located at $(x, y) = (0, 0)$ and because the sharpness map is estimated at edges only, so the equation (3.33) truncate to:

$$E_s = 1 - \sqrt{\frac{\sigma^2}{\sigma^2+\sigma_k^2}} \tag{3.34}$$

This equation characterizes the sharpness map of the input image and is, therefore, useful for estimating the depth amount present at the edges. Finally, the amount of blur (σ) which in turn indicates the depth, can be calculated by using the equation (3.34) as:

$$\frac{\sigma^2}{\sigma^2+\sigma_k^2} = \left(1 - E_s\right)^2 \tag{3.35}$$

⇨ $$\sigma^2 = \sigma^2\left(1 - E_s\right)^2 + \sigma_k^2\left(1 - E_s\right)^2 \tag{3.36}$$

⇨ $$\sigma^2\left(1 - \left(1 - E_s\right)^2\right) = \sigma_k^2\left(1 - E_s\right)^2 \tag{3.37}$$

$$\Rightarrow \quad \sigma = \frac{(1-E_s)\sigma_k}{\sqrt{2E_s - {E_s}^2}} \tag{3.38}$$

The values of σ estimated at every location of edge form a sparse depth map. But, weak edges may result in incorrect blur value at some edge locations and hence, it will lead to the propagation of errors which will corrupt the full depth map. Therefore, to resolve this issue, joint bilateral filter (JBF) [49] is used which is reported in the next section.

3.3.1.4 Joint Bilateral Filtering

JBF is a non-linear filter and is used to conserve the edges and reduce the noise present in the images. The intensity value of every image pixel is reinstated by the weighted average of the intensity values from the neighboring pixels. It is basically a combination of a low-pass filter (LPF) with an edge-ceasing function. In the method proposed in this thesis, JBF is exercised onto the sparse depth map to obtain a refined map. The output of the JBF is a filtered depth map which can be expressed as [49]:

$$F_b\left(d_S(x)\right) = \frac{1}{W(x)} \sum_{y \in \mathcal{N}(x)} F_{g_{\sigma_s}}\left(\|x-y\|\right) F_{g_{\sigma_r}}\left(\|i(x)-i(j)\|\right) d_S(y) \tag{3.39}$$

Here, $d_s(x)$ is the sparse depth map. Also,

$$F_{g_{\sigma_s}}\left(\|x-y\|\right) = \frac{1}{2\pi{\sigma_s}^2} e^{-\frac{\left(\|x-y\|\right)^2}{2{\sigma_s}^2}} \tag{3.40}$$

and

$$F_{g_{\sigma_r}}\left(\|i(x)-i(j)\|\right) = \frac{1}{2\pi{\sigma_r}^2} e^{-\frac{\left(\|i(x)-i(j)\|\right)^2}{2{\sigma_r}^2}} \tag{3.41}$$

Equation (3.40) and (3.41) represents Gaussian functions whose spread is controlled by the parameters σ_s and σ_r respectively.

The Gaussian function (equation 3.40) is a spatial domain filter which specifies the size of neighborhood pixels $(\mathcal{N}(x))$ of x, while the edge-ceasing function (equation 3.41) smoothens the difference in intensities of the pixels. As the standard deviation of the Gaussian function varies, smoothing factor increases or decreases accordingly as demonstrated in Figure 3.4. The challenging job is to set the parameters σ_s and σ_r so as to average away the noise and yet preserving the details of the image. Only those pixels are considered which are close in space and are in range. Assume pixel size to be 1.

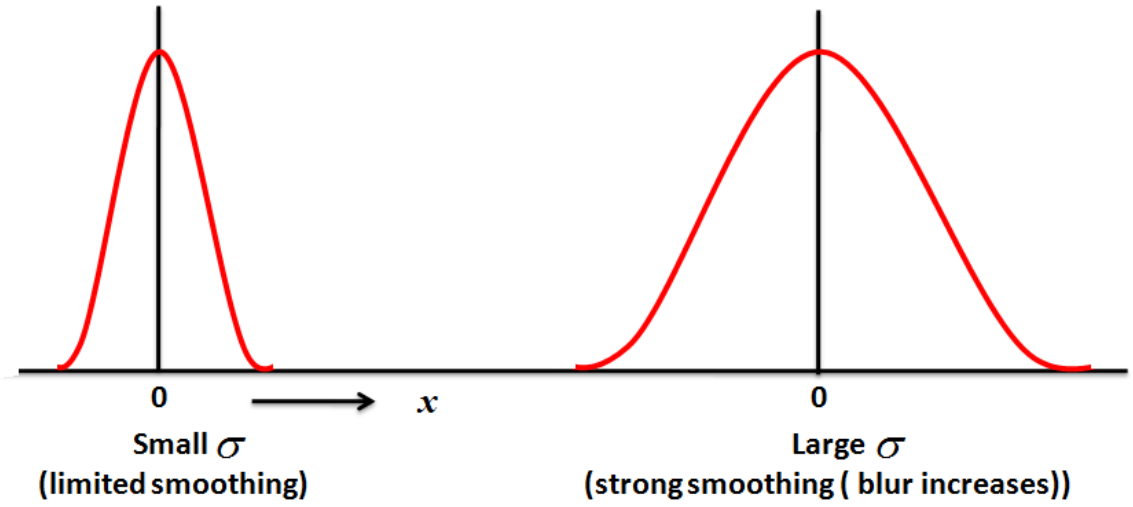


Figure 3.4: Effect of sigma on Gaussian functions.

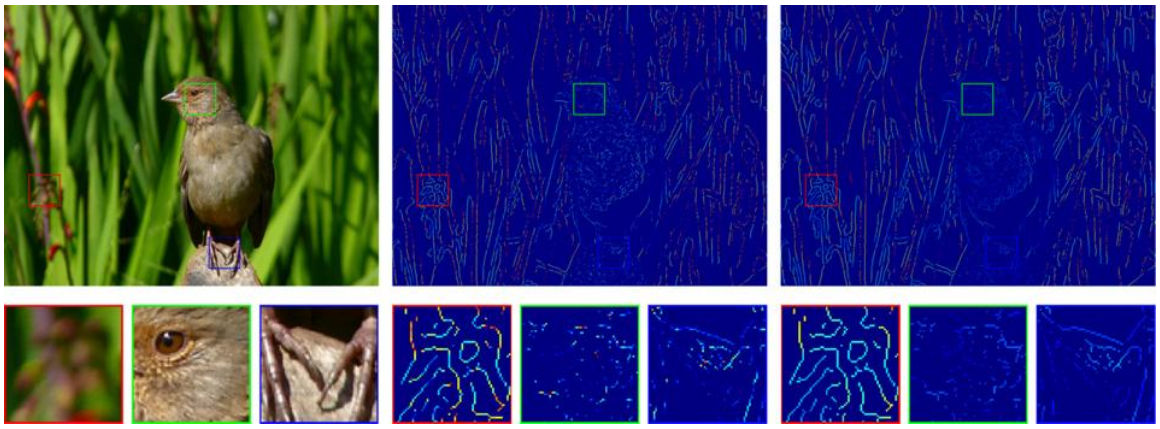
$W(x)$ is a normalization parameter that makes sure that the filter maintains image energy. The value of $W(x)$ is given as [49]:

$$W(x) = \sum_{y \in \mathcal{N}(x)} F_{g_{\sigma_s}}\left(\|x-y\|\right) F_{g_{\sigma_r}}\left(\|i(x)-i(j)\|\right) \tag{3.42}$$

The refined sparse depth map thus obtained is ready to be processed further. Figure 3.5 shows the rectified sparse depth map obtained when the JBF is applied. The final step is to retrieve a complete depth map which is explained in the later section.

3.3.2 Full Depth Map

In this section, the amount of depth present in the original image is interpolated at unknown locations. Eventually, the depth values from the edges are propagated to the whole image producing a complete depth map. To accomplish this, the sparse depth estimates should be aligned with the original image edges which can be done by the edge-aware interpolation techniques [53, 54]. Interpolation approach is applied thereupon to generate a full depth map.



(a) (b) (c)

Figure 3.5: Depth refinement using joint bilateral filtering, (a) Input image, (b) Sparse depth map, and (c) Refined depth map [3].

3.3.2.1 Interpolation Approach

For interpolating depth values at unexplored areas in the image, matting Laplacian technique [43] is adopted which propagates the sparse depth map based on color similarities in order to develop a complete depth map. The goal is to minimize the cost function which is formulated as:

$$J(D) = D^T L D + \lambda (D - D_S)^T \Sigma (D - D_S) \tag{3.43}$$

where, D_S and D signify the vectors belonging to the sparse depth map $d_s(x)$ and the full depth map $d(x)$ respectively. λ is a regularized parameter and Σ is a matrix with elements Σ_{ii} as 1 whenever the pixel i is positioned at the edge areas, else 0. Also, the matrix L is known as the matting Laplacian whose $(i,j)^{th}$ entry is expressed as:

$$\sum_{k|(i,j)\in\omega_k} (\delta_{ij} - \frac{1}{|\omega_k|}(1 + \frac{1}{\frac{\varepsilon}{|\omega_k|} I_3 + \sigma_k}(i_i - \mu_k)(i_j - \mu_k) \tag{3.44}$$

where, δ_{ij} is the Kronecker delta, μ_k is a mean vector of 1 × 3 dimensions and σ_k is a covariance matrix of 3 × 3 dimensions whose values denote the intensities in the window ω_k and $|\omega_k|$ indicates the count of pixels in this window. I_3 is a 3 × 3 identity matrix. i_i and i_j are the colors at the pixel i and j of the input image $i(x,y)$. ε is a fitting parameter whose value is taken to be 10^{-6} in the experiment. The optimal D can be achieved by elucidating the following equation:

$$D(L + \lambda\Sigma) = \lambda\Sigma D_S \tag{3.45}$$

The appropriate value of λ is taken so that a smooth constraint is put on D to purify small errors.The full depth map is illustrated by a gray scale image where each pixel represents the intensity of depth. Darker pixel in the depth map presents the area that is close to the camera and lighter pixel presents the area far from the camera as shown in Figure 3.6. Hence, the depth maps estimated from this technique is effective enough to generate good quality 3D images.

3.4. Generating Three-Dimensional Images

In addition to brightness and color, a 3D pixel adds a property of depth which signifies the position of the point on the imaginary z-axis. When each pixel is combined with its very own depth value, the effect is a 3D image.

(a) (b)

Figure 3.6: The depth recovery result of Zhuo's method [3], (a) Input image, and (b) Corresponding depth map.

3.4.1 Three-Dimensional Image Conversion

The conversion method presenting here is based on adding back the missing third dimension to the given image. The pixel transform is based on the fact that pixel values of the output depend on the pixel values of the input i.e.

$$i_{3D}(x,y) = \left(\frac{i'(x,y) - v}{u - v} \right) \alpha + \beta \tag{3.46}$$

where, $\alpha > 0$ is a constant which is often known as a gain parameter and it controls the contrast of the image and β is known as bias parameters which control the brightness. The increasing value of v emphasizes shadows while u emphasizes light. $i'(x,y)$ is the image obtained after adding the depth map to the original image which is formulated as:

$$i'(x,y) = i(x,y) + \frac{3}{2} D \tag{3.47}$$

where, $i(x,y)$ is a 2D input image and D is the calculated depth map of the picture which is multiplied by a factor so as to get optimal 3D image.

3.5 Chapter Summary

This chapter explains the proposed depth estimation approach that utilizes only single image for the assessment of depth information. The idea involves the utilization of sharpness amount at edge areas to retrieve the complete depth map. Canny's operator is used to detect the images and matting Laplacian interpolation algorithm is applied to attain the final depth map. The chapter also discusses an approach to generate 3D images by adding back the estimated depth maps to the original image.

CHAPTER 4
RESULTS AND DISCUSSIONS

If you can't make it good,
at least make it look good.
Bill Gates [1955 - present]

4.1 Introduction

The main objective of this chapter is to explore the newly proposed depth estimation approach and 2D-to-3D conversion algorithm. Experiments are conducted on indoor and outdoor scenes in which different existing depth estimation techniques, all aiming to create high-quality depth maps, are applied and evaluated. Effective visual and quantitative comparison analysis are explicitly provided in this chapter. At last, this depth map along with the original image together constitutes an image having 3D perception.

4.2 Experimental Results of Depth Estimation

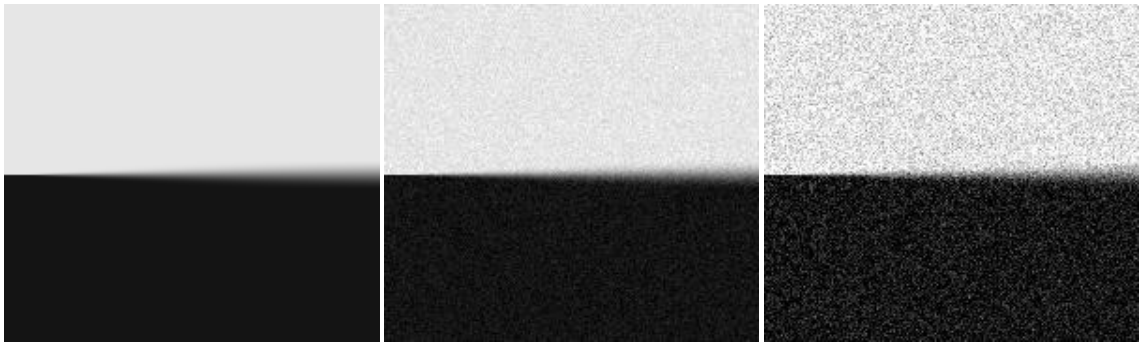
Simulation results are taken on both synthetic and real images. For this work, MATLAB is used as the development language. Not only does the MATLAB provides a lot of useful built-in functions, scripting, and testing are a lot more simplified than with other languages such as C or C++.

In the proposed experiment, the original image is re-blurred with $\sigma_k = 3$ of the Gaussian function. The edges are detected with the help of Canny method having upper threshold value to be 0.02. Also, $\sigma_s = 5$ and $\sigma_r = 0.3$ is taken in the joint bilateral filtering of sparse depth map. For the optimal depth map the value of $\lambda = 0.8$ is used in the proposed method.

4.2.1 Simulation Results on Synthetic Data

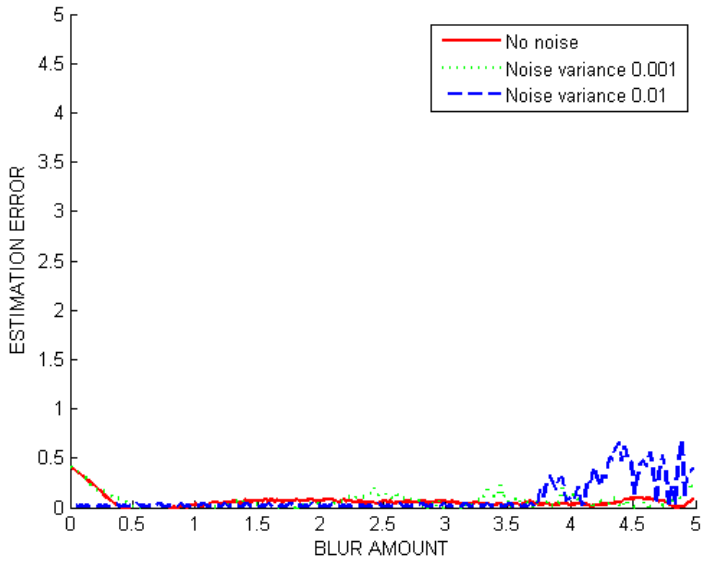
A set of bar images is created with different noise conditions as shown in Figure 4.1, in which a horizontal edge is blurred by a space-varying, one-dimensional vertical Gaussian

blur kernel. The blur amount linearly increases from 0 to 5 along the edge. Figure 4.2 demonstrates that the edges are reliably detected and the proposed methodology is robust to noise as compared to Zhuo's method [3]. As the noise variance becomes higher (in Figure 4.1(b) and 4.1(c)), the proposed method is less affected by noise.

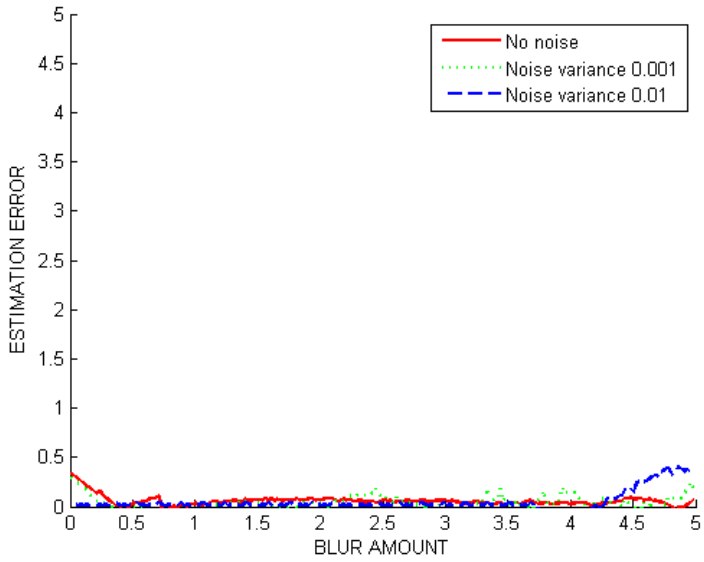


(a) (b) (c)

Figure 4.1: Synthetic bar images, (a) Without noise, (b) With noise variance 0.001, and (c) With noise variance 0.01.

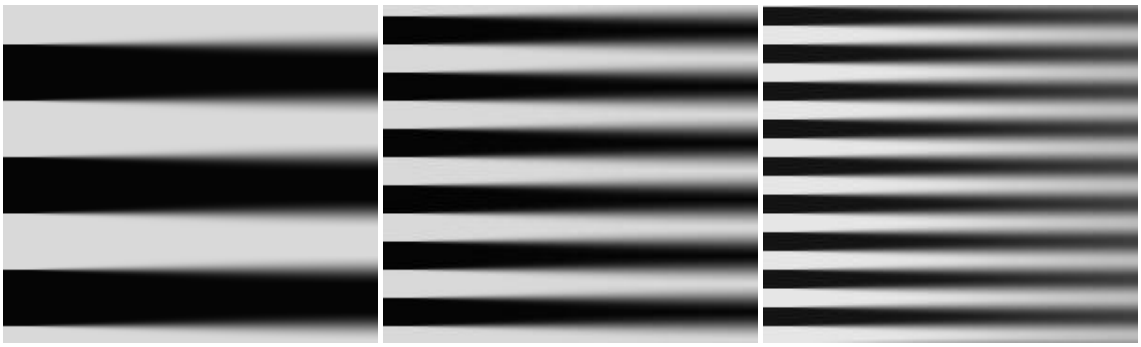


(a)



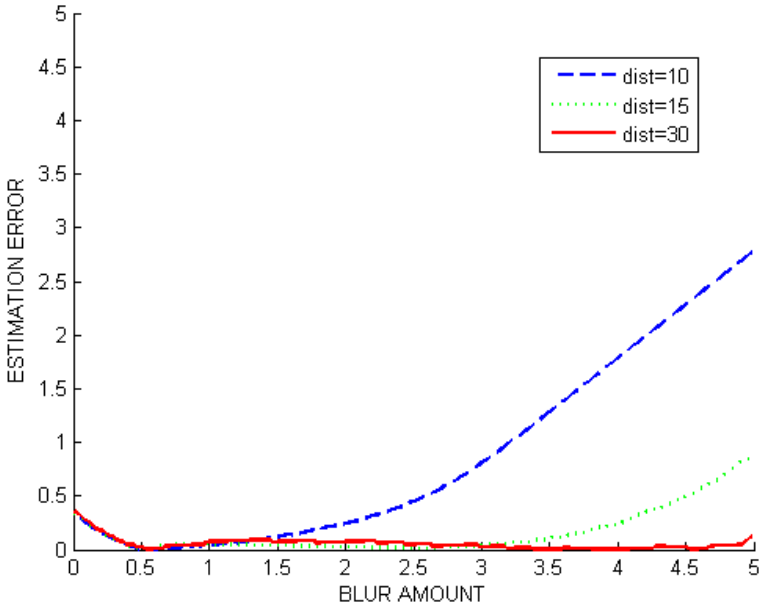
(b)

Figure 4.2: Estimation errors under noise condition, (a) The performance of Zhuo's method [3], and (b) The performance of the proposed method.

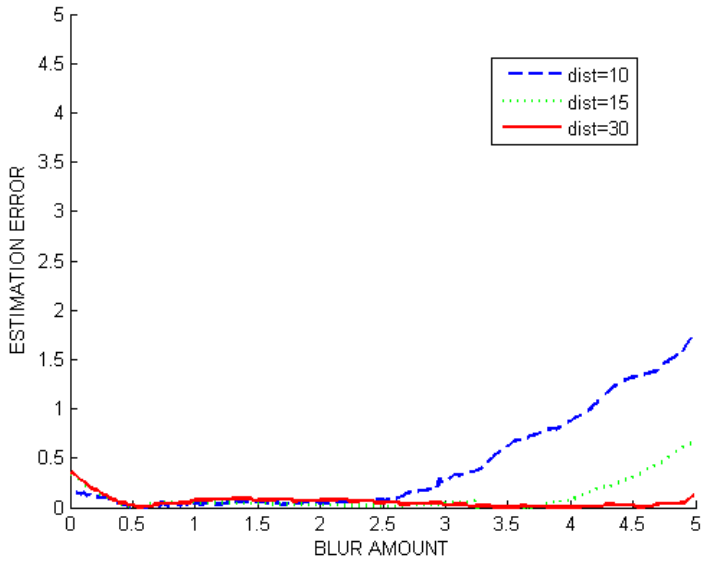


(a) (b) (c)

Figure 4.3 : Synthetic bar images, (a) With edge distance of 30 pixels, (b) With edge distance of 15 pixels, and (c) With edge distance of 10 pixels.



(a)

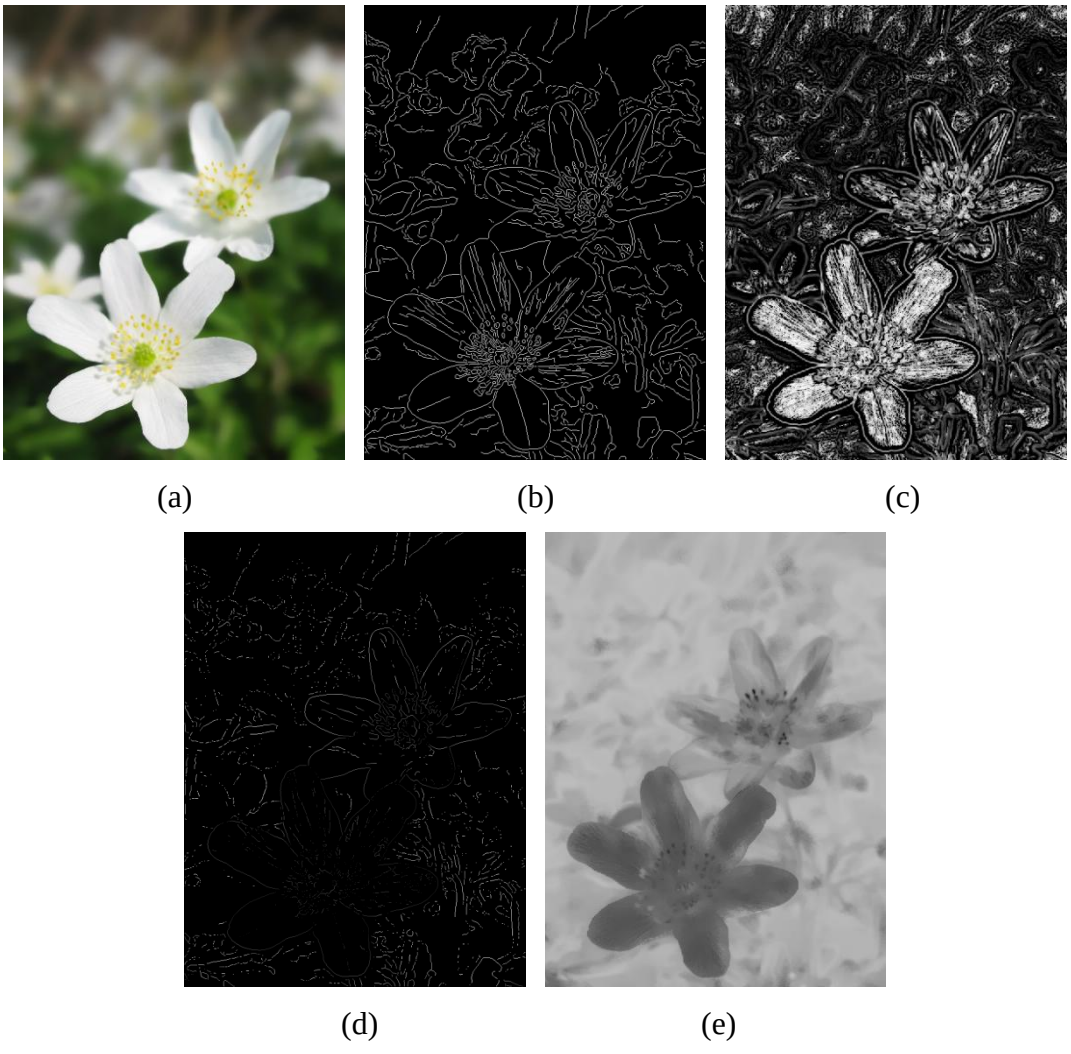


(b)

Figure 4.4: Estimation errors with different edge distance, (a) The performance of Zhuo's method [3], and (b) The performance of the proposed method.

4.2.2 Simulation Results on Real Data

The suggested approach is also tested on real images. The image of a flower is taken as input (Figure 4.5(a)) in which the scene depth changes gently from the top to the bottom. Canny's edge detection is applied to the input image as shown in Figure 4.5(b). Sharpness map is then estimated (Figure 4.5(c)) and by using the sharpness information, the sparse depth map is (Figure 4.5 (d)). The estimated depth map is shown in Figure 4.5(e) which captures all the continuous change of the depth. Another example is shown in Figure 4.6. Higher intensity depicts more depth in all the estimated depth maps.



(a) (b) (c)

(d) (e)

Figure 4.5: Depth map estimation on real images, (a) Input image, (b) Edge detection, (c) Sharpness Map, (d) Sparse depth map, and (e) Full depth map.

(a) (b)

Figure 4.6: Depth recovery result of the proposed method, (a) Input image, and (b) Depth map.

4.3 Performance Comparison

The visual comparison of the proposed and existing methods is shown in Figure 4.7, Figure 4.9, and Figure 4.11 along with a quantitative evaluation of the corresponding depth images as illustrated in Table 4.1, Table 4.2 and Table 4.3 respectively. It is clear from the performance comparisons that the proposed approach is quite reasonable under all evaluation matrices including relative error (R_{error}), log10 error ($Log10$) and root mean square error ($RMSE$). These parameters are defined as under:

$$R_{error} = \frac{1}{N}\sum_{i}\left(d_i - d_i^*\right)/d \tag{4.1}$$

$$Log10 = \sum_{i}|\log_{10}\left(d_i\right) - \log_{10}(d_i^*)|/N \tag{4.2}$$

$$RMSE = \sqrt{\sum_{i}\left(d_i - d_i^*\right)^2/N} \tag{4.3}$$

where, N is the count of pixels containing the image and d_i indicates the depth estimate of i^{th} pixel in comparison to the ground truth d^*.

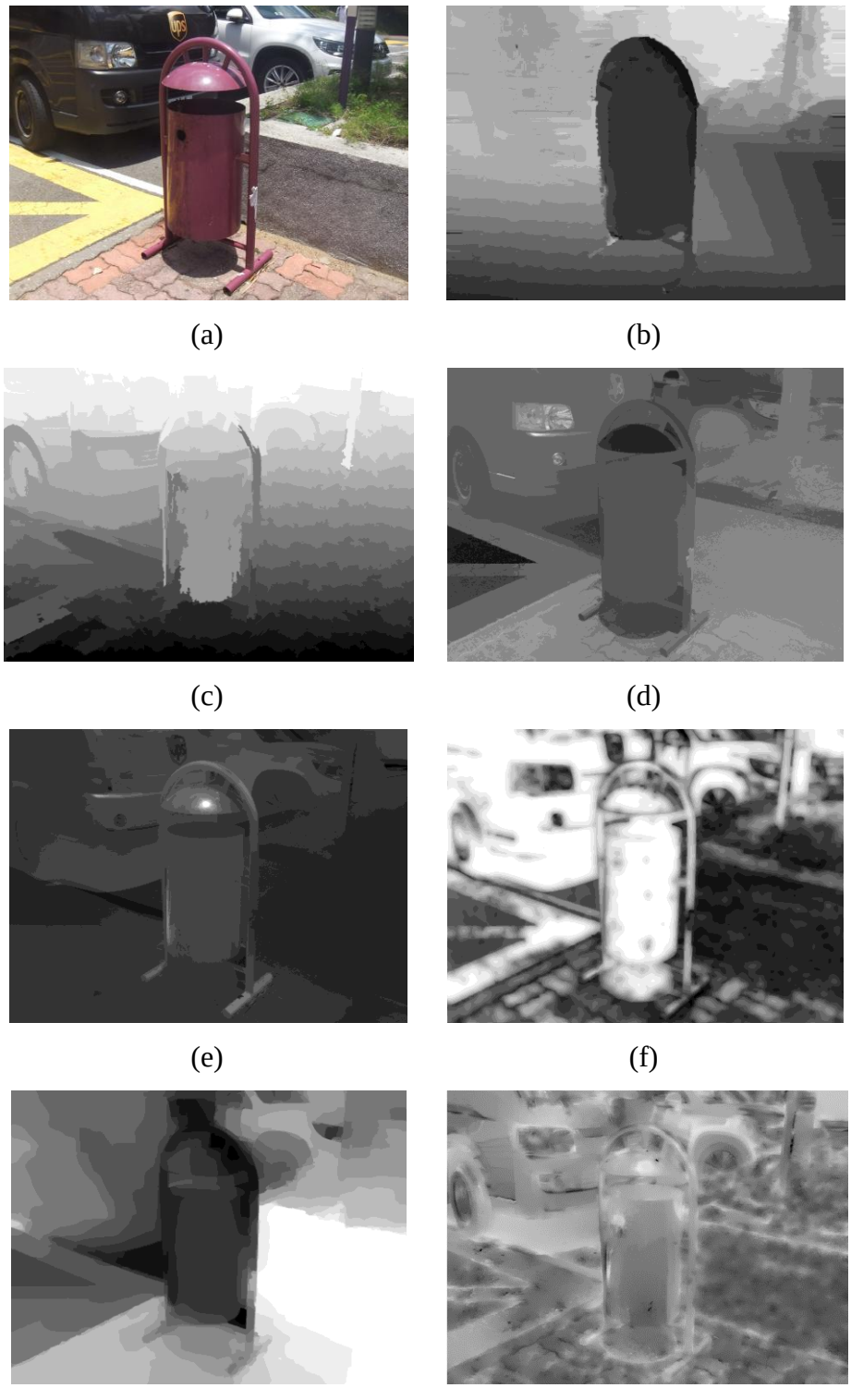


Figure 4.7: Visual comparison of depth maps of Image 1, (a) Input, (b) Ground truth, (c) A. Saxena *et al.* [34], (d) S. Bae *et al.* [52], (e) S. Zhuo *et al.* [3], (f) J. Shi *et al.* [53], (g) J. Shi *et al.* [44], and (h) Proposed.

Table 4.1: Quantitative evaluation of depth maps of Image 1.

METHODS	RELATIVE ERROR	LOG10 ERROR	RMSE
A. Saxena *et al.* [34]	2.1267	0.2709	0.2277
S. Bae *et al.* [52]	0.5988	0.2239	0.2558
S. Zhuo *et al.* [3]	1.3105	0.3336	0.3303
J. Shi *et al.* [53]	0.6798	0.2674	0.3351
J. Shi *et al.* [44]	0.3735	0.2725	0.6670
Proposed	**0.3295**	**0.1719**	**0.2184**

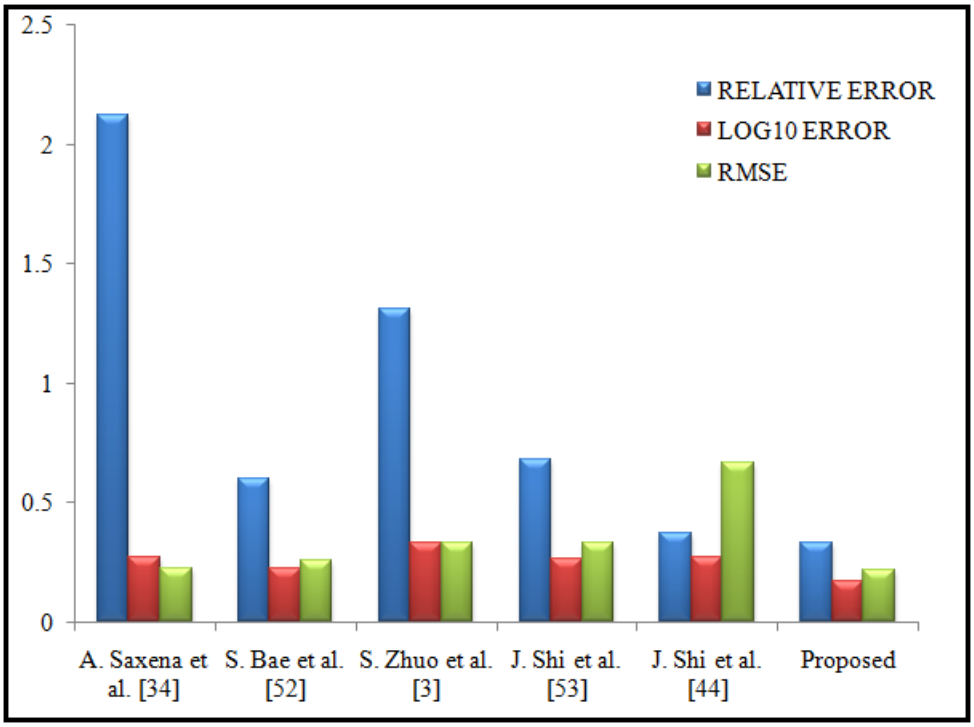


Figure 4.8: Bar chart comparison of the proposed method with other methods of Image 1.

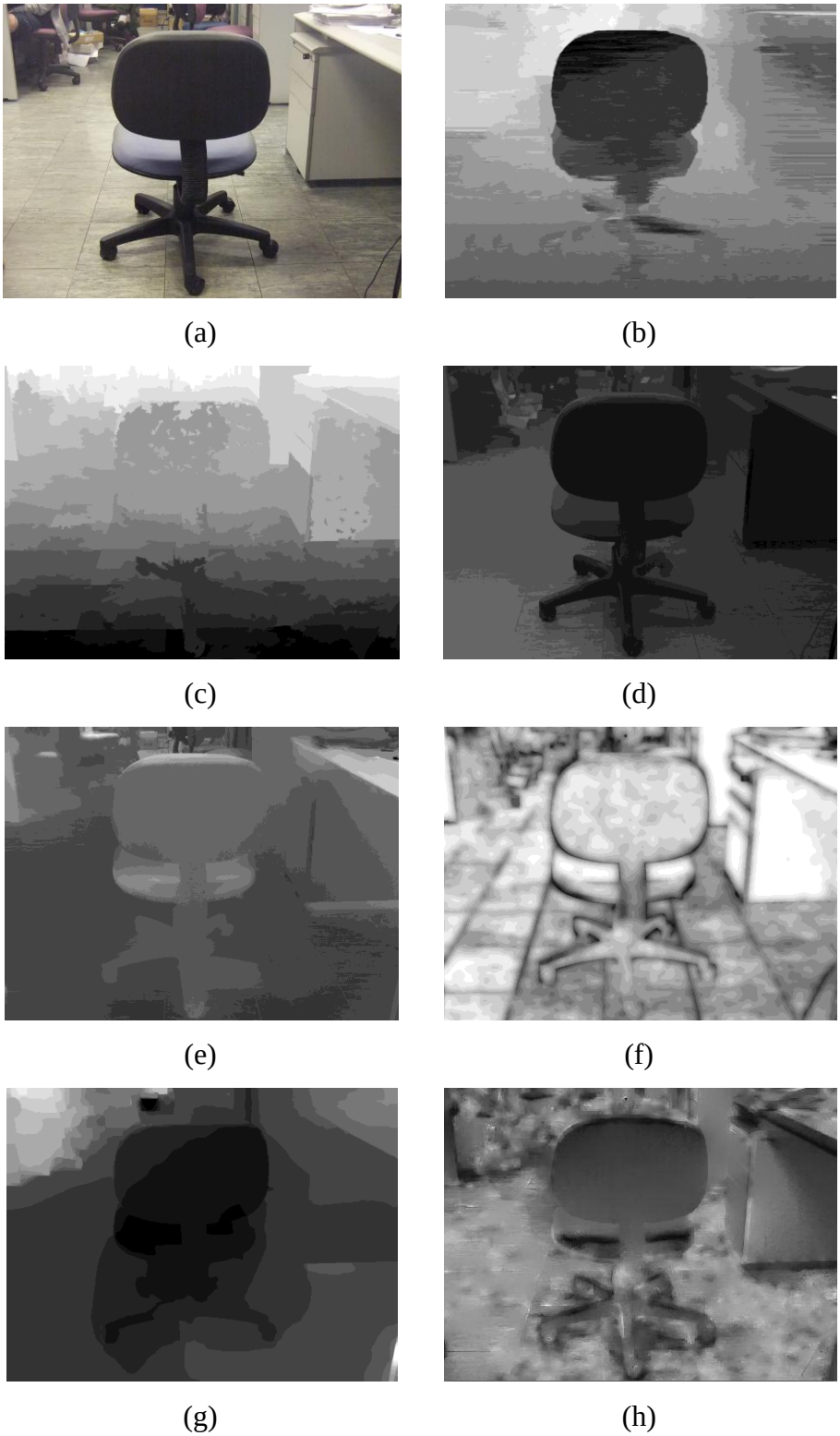


Figure 4.9: Visual comparison of depth maps of Image 2, (a) Input, (b) Ground truth, (c) A. Saxena *et al.* [34], (d) S. Bae *et al.* [52], (e) S. Zhuo *et al.* [3], (f) J. Shi *et al.* [53], (g) J. Shi *et al.* [44], and (h) Proposed.

Table 4.2: Quantitative evaluation of depth maps of Image 2.

METHODS	RELATIVE ERROR	LOG10 ERROR	RMSE
A. Saxena *et al.* [34]	3.3282	0.2907	0.2531
S. Bae *et al.* [52]	1.6878	0.3828	0.3322
S. Zhuo *et al.* [3]	0.6860	0.2404	0.2595
J. Shi *et al.* [53]	0.4086	0.2300	0.3384
J. Shi *et al.* [44]	1.4141	0.3462	0.2928
Proposed	**0.3783**	**0.1573**	**0.2059**

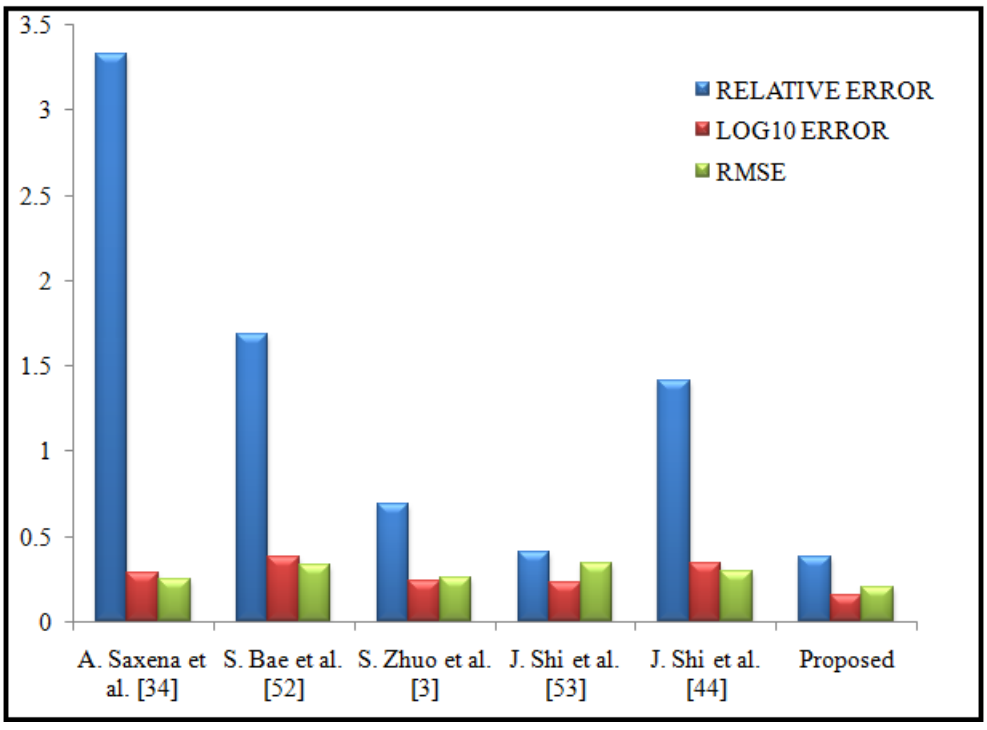
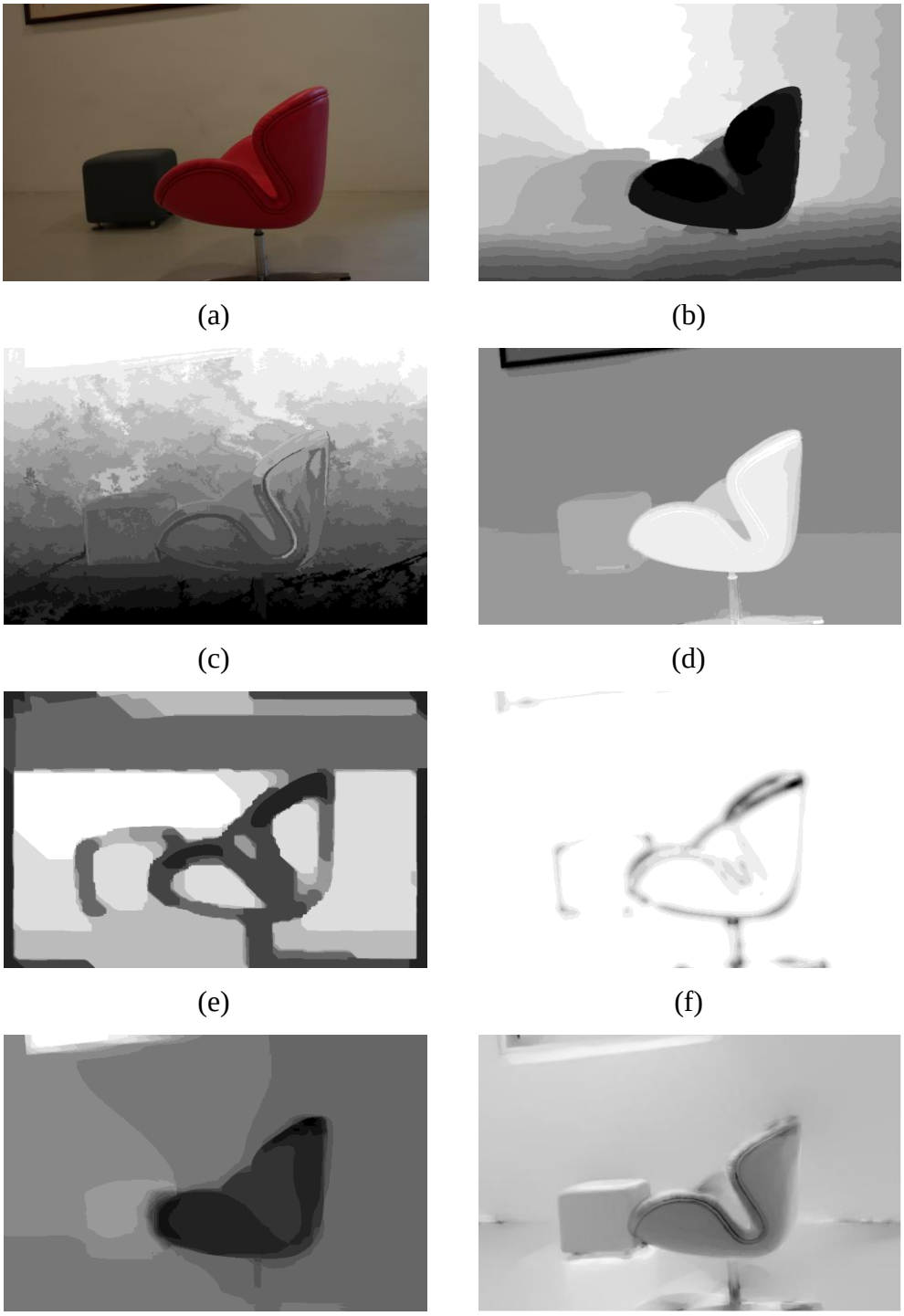


Figure 4.10: Bar chart comparison of the proposed method with other methods of Image 2.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.11: Visual comparison of depth maps of Image 3, (a) Input, (b) Ground truth, (c) A. Saxena *et al.* [34], (d) S. Bae *et al.* [52], (e) S. Zhuo *et al.* [3], (f) J. Shi *et al.* [53], (g) J. Shi *et al.* [44], and (h) Proposed.

Table 4.3: Quantitative evaluation of depth maps of Image 3.

METHODS	RELATIVE ERROR	LOG10 ERROR	RMSE
A. Saxena *et al.* [34]	1.6366	0.3256	0.2682
S. Bae *et al.* [52]	0.9134	0.3005	0.4022
S. Zhuo *et al.* [3]	0.3382	0.2561	0.3003
J. Shi *et al.* [53]	0.3551	0.2812	0.4376
J. Shi *et al.* [44]	0.6203	0.2310	0.3061
Proposed	**0.2840**	**0.2007**	**0.2558**

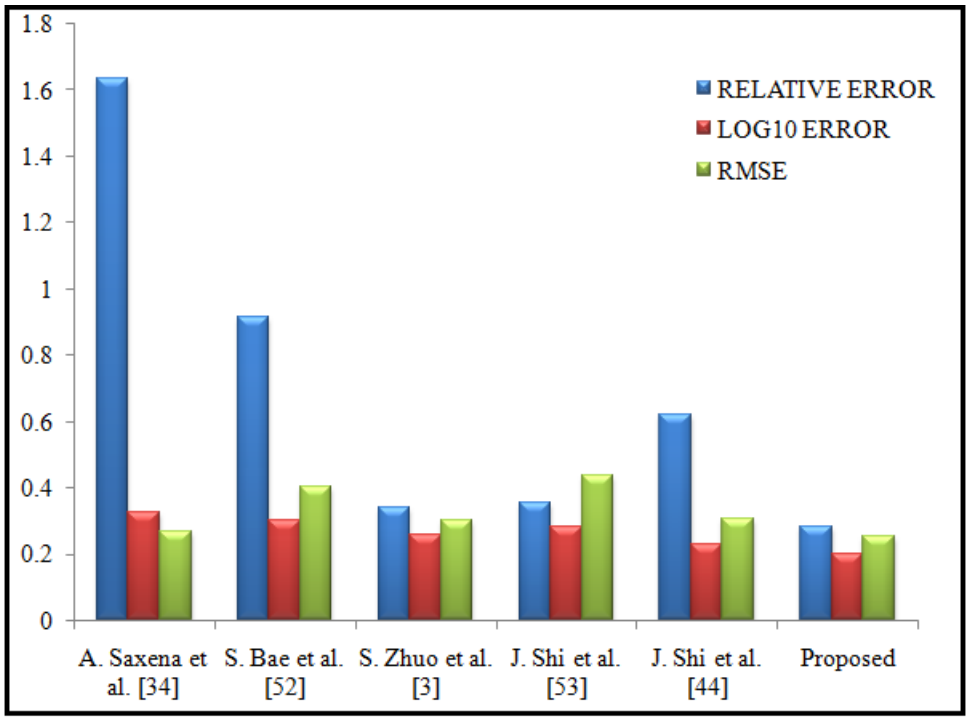
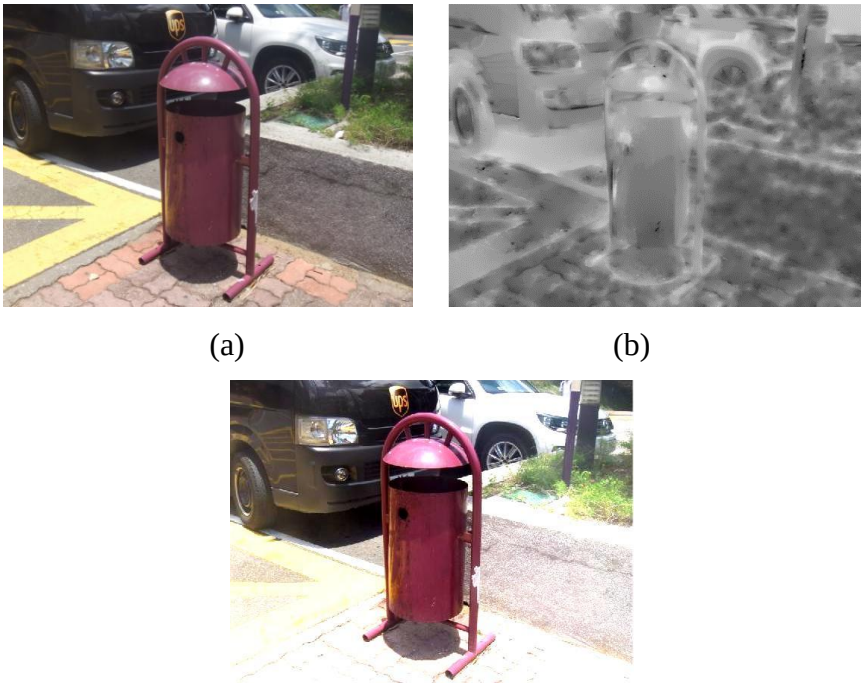


Figure 4.12: Bar chart comparison of the proposed method with other methods of Image 3.

The suggested technique is able to reproduce more accurate depth maps because it contains acceptable depth layers. But, there is one limitation that it forms incorrect depth maps when the image contains shadows. This is illustrated in Figure 4.12(b).

4.4 Three-Dimensional Images

Additionally, the 3D images are generated with the help of the estimated depth maps. An example is illustrated in Figure 4.13. Given an input image, Figure 4.13(a), and its depth map, Figure 4.13(b), 3D image is formed, Figure 4.13(c). Other examples of generated 3D images from the proposed scheme are given in Figure 4.14, Figure 4.15, Figure 4.16 and Figure 4.17.



(a) (b)

(c)

Figure 4.13: 3D image generation for Image 1, (a) Input image, (b) Depth map, and (c) 3D image.

(a) (b)

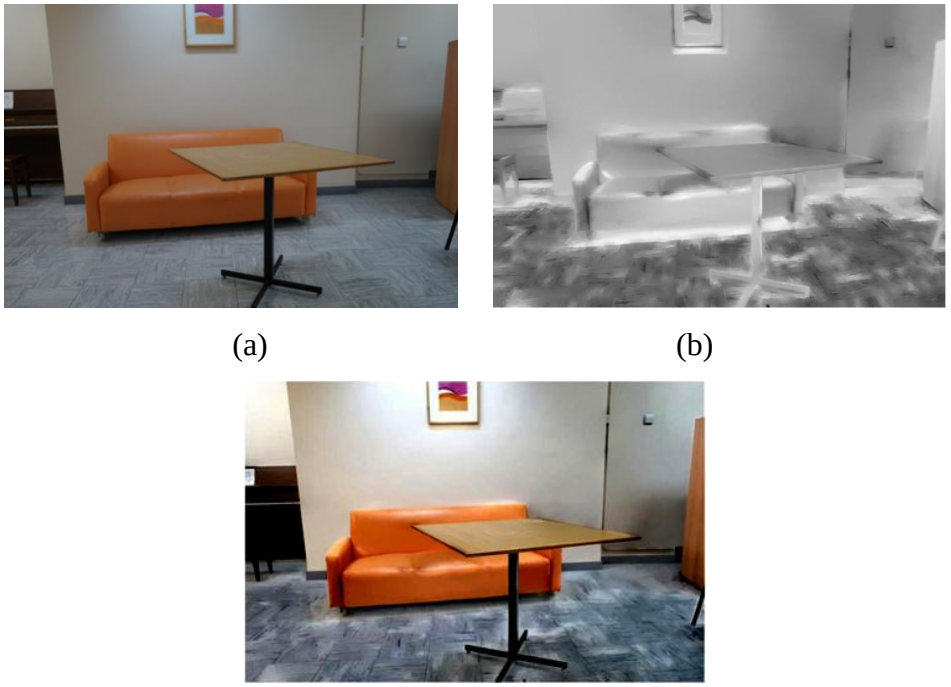
(c)

Figure 4.14: 3D image generation for Image 3, (a) Input image, (b) Depth map, and (c) 3D image.

(a) (b)

(c)

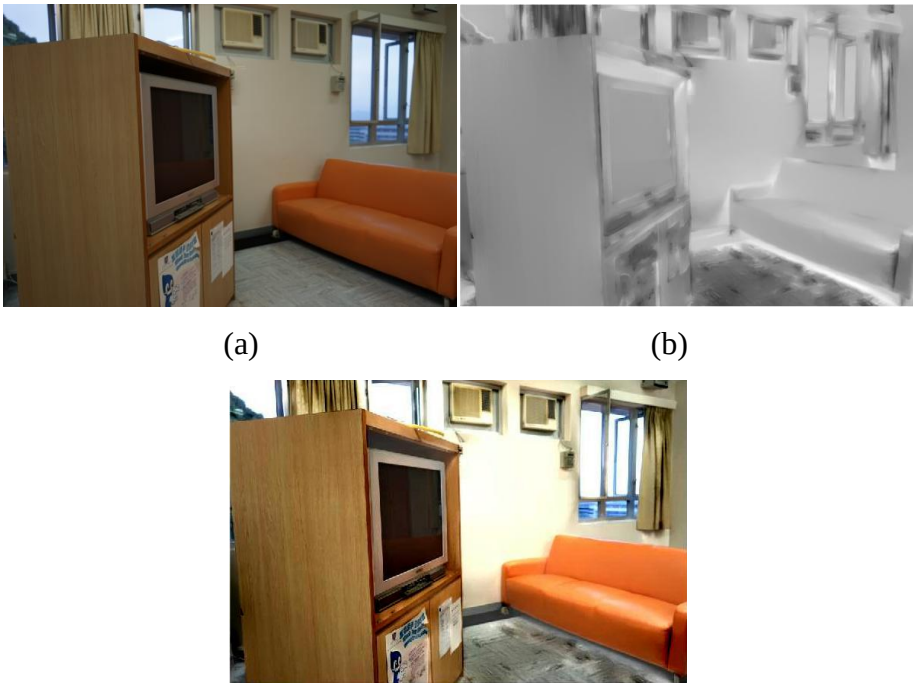
Figure 4.15: 3D image generation for Image 4, (a) Input image, (b) Depth map, and (c) 3D image.



(a) (b)

(c)

Figure 4.16: 3D image generation for Image 5, (a) Input image, (b) Depth map, and (c) 3D image.



(a) (b)

(c)

Figure 4.17: 3D image generation for Image 6, (a) Input image, (b) Depth map, and (c) 3D image.

4.5 Chapter Summary

This chapter presents an analysis of the proposed depth estimation and 3D generation technique.

The first experiment is conducted the synthetic images with varying noise variance and different edge distances. The proposed technique is very less afflicted by the noise, but the estimation errors increase when the distance between the adjacent edges become small and the blur amount becomes large.

The second experiment on real images derived that the novel depth information method proposed in this dissertation has higher performance than the previous techniques based on the single image as input.

The last experiment showed the 3D images generated by the suggested method. The images come alive when a sense of depth is added to the input image and therefore, are more comfortable to view.

CHAPTER 5
CONCLUSIONS AND FUTURE SCOPE

The questions are always more important than the answers.
Randy Pausch [1960-2008]

5.1 Conclusions

In this dissertation, a new algorithm is proposed for evaluating the depth information from a single small-scale blurred image by utilizing sharpness amount at edge locations. By using 2D Gaussian kernel, the input image is re-blurred and the gradients of the input and the re-blurred image are calculated to measure the sharpness map. Then, the sparse depth map is retrieved by evaluating the subtle defocus blur at edges and joint bilateral filter (JBF) is employed to reduce noise around the weak edges. The edges are detected by Canny's operator which gives effective structural information of the image. By using matting Laplacian interpolation method, the depth estimates are propagated from the edge locations to the whole image producing a full depth map.

Extensive experiments on the synthetic and real images have shown that the method presented is accurate and much reliable than the existing state-of-art methods. The suggested method is compared with J. Shi *et al.* [44] and it has been verified that there is an average decrease of 0.472, 0.1975, and 0.1952 in relative error, log10 error, and root mean square error respectively, for real images used in this dissertation.

The estimated depth maps from the proposed method are thereafter utilized for converting the two-dimensional image back to three-dimensional. This is achieved by adding depth maps back to the corresponding two-dimensional image along with some additional parameters. The perception of the 3D images is clearly much more interactive for the viewers and the 3D images produced are of better quality than the 2D images.

5.2 Future Work

The research work proposed here is concentrated only on the step edges. However, this approach can be extended to different types of edges in the near future. Also, different types of matting interpolation algorithms can be used to get optimal depth maps.

REFERENCES

[1] H. Su, Q. Huang, N. J. Mitra, Y. Li, and L. Guibas, "Estimating image depth using shape collections," *ACM Transactions on Graphics*, vol. 33, no. 4, July 2014.

[2] W. J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1869-1872, July 2006.

[3] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852-1858, September 2011.

[4] P. Favaro and S. Sotto, "A geometric approach to shape from defocus," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 406-417, March 2005.

[5] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824-831, August 1994.

[6] S. Battiato, S. Curti, M. L. Cascia, M. Tortora, and E. Scordato, "Depth map generation by image classification," *Proceedings of SPIE, Three-Dimensional Image Capture and Applications VI*, vol. 5304, pp. 95-104, April 2004.

[7] Q. Wei, "Converting 2D to 3D: a survey," *International Conference*, vol. 7, no. 14, pp. 1-37, December 2005.

[8] E. M. Gmbh. (2010). *MakeMe3D software* [Online]. Available: http://www.makeme3d.net/convert_2d_to _3d.php.

[9] M. Born and E. Wolf, *Principles of Optics*, 4th edition, UK: Cambridge University Press, 1999.

[10] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, 1st edition, Upper Saddle River, NJ, USA: Prentice Hall, 1998.

[11] E. Hammon, *Practical post-process depth of field, GPU Gems 3*, 2nd edition, USA: Addison-Wesley, pp. 583-605, 2007.

[12] H. Murata, Y. Mori, S. Yamashita, A. Maenaka, S. Okada, K. Oyamada, and S. Kishimoto, "A real time 2D to 3D image conversion technique using computed image depth," *SID Symposium Digest of Technical Papers*, vol. 29, no. 1, pp. 919-922, May 1998.

[13] T. Iinuma, H. Murata, S, Yamaashita, and K. Oyamada, "Natural stereo depth creation methodology for a real-time 2D-to-3D image conversion," *SID Symposium Digest of Technical Papers*, vol. 31, no. 1, pp. 1212-1215, May 2000.

[14] A. Saxena, M. Sun, and A. Y. Ng, "Make3D: Learning 3D scene structure from a single still image," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 824-840, May 2009.

[15] M. N. Galabov, "A real time 2D to 3D image conversion techniques," *International Journal of Engineering Science and Innovative Technology*, vol. 4, no. 1, pp. 297-304, 2015.

[16] E. Stoykova, A. Ayd, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis, "3-D time-varying scene capture technologies—a survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1568-1586, November 2007.

[17] T. Lindeberg and J. Garding, "Shape from texture from a multi-scale perspective," *Proceedings of Fourth International Conference on Computer Vision*, pp. 683-691, May 1993.

[18] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, "Depth estimation and image restoration using defocused stereo pairs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1521-1525, November 2004.

[19] U. Mudenagudi and S. Ghaudhuri, "Depth estimation using defocused stereo image pairs," *Proceedings of the 7th IEEE International Conference on Computer Vision*, vol. 1, pp. 483-488, 1999.

[20] M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271-294, December 1994.

[21] B. Girod and S. Scherock, "Depth from defocus of structured light," *Proceedings of SPIE, Optics, Illumination, and Image Sensing for Machine Vision IV*, vol. 1194, pp. 209-215, April 1989.

[22] B. Girod and E. Adelson, "System for ascertaining direction of blur in a range-from-defocus camera," *U.S. Patent No. 4,965,422*, October 1990.

[23] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1-7, June 2008

[24] J. Salvi, J. Pages, and J. Battle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827-849, April 2004.

[25] F. M. Noguer, P. N. Belhumeur, and S. K. Nayar, "Active refocusing of images and videos," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 671-679, July 2007.

[26] C. Hua, L. Tao, G. Mengshu, and Z. Bojin, "A depth optimization method for 2D-to-3D conversion based on RGB-D images," *Proceedings of 4th IEEE International Conference on Network Infrastructure and Digital Content*, pp. 223-227, September 2014.

[27] N. Asada, H. Fujiwara, and T. Matsuyama, "Edge and depth from focus," *International Journal of Computer Vision*, vol. 26, no. 2, pp. 153-163, February

1998.

[28] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and depth from a conventional camera with a coded aperture,” *ACM Transactions on Graphics,* vol. 26, no. 3, pp. 701-709, July 2007.

[29] V. P. Namboodiri and S. Chaudhuri, “Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera,” *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-6, June 2008.

[30] P. K. Chan, B. Z. Jing, W. Y. Ng, and D. S. Yeung, “Depth estimation from a single image using defocus cues,” *International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 1732-1738, July 2011.

[31] J. Lin, X. Ji, W. Xu, and Q. Dai, “Absolute depth estimation from a single defocused image,” *IEEE Transactions on Image Processing,* vol. 22, no. 11, pp. 4545-4550, November 2013.

[32] J. Konrad, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee, “Learning-based, automatic 2D-to-3D image and video conversion,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3485-3496, September 2013.

[33] C. Jung , S. Joo, and S. M. Ji, “Depth map upsampling with image decomposition,” *Electronics Letters*, vol. 56, no. 22, pp. 1782-1784, October 2015.

[34] A. Saxena, S. H. Chung and A. Y. Ng, “Learning depth from single monocular images,” *Advances in Neural Information Processing Systems*, pp. 1161-1168, 2005.

[35] B. Liu, S. Gould, and D. Koller, “Single image depth estimation from predicted semantic labels,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1253-1260, 2010.

[36] D Scharstein and R Szeliski. (2002). *The Middlebury Stereo Vision Page* [Online]. Available: http://vision.middlebury.edu/stereo/.

[37] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7-42, April 2002.

[38] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, August 2003.

[39] F. Liu and V. Philomin, "Disparity estimation in stereo sequences using scene flow," *British Machine Vision Conference*, vol. 1, no. 2, pp. 1-11, September 2009.

[40] M. Baba, N. Asada, A. Oda, and T. Migita, "A thin lens based camera model for depth estimation from defocus and translation by zooming," *Proceedings of the 15th International Conference on Vision Interface, Canada*, pp. 247-281, May 2002.

[41] G. Surya and S. Murali, "Depth from defocus by changing camera aperture: a spatial domain approach," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 61-67, June 1993.

[42] S. K. Nayar, M. Watanabeand, and M. Noguchi, "Real-time focus range sensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1186-1198, December 1996.

[43] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, February 2008.

[44] J. Shi, X. Tao, L. Xu, and J. Jia "Break ames room illusion: depth from general single images," *ACM Transactions on Graphics*, vol. 34, no. 6, pp. 225-236,

November 2015.

[45] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3017-3024, June 2011.

[46] A. Nasonov, A Nasonova, and A. Krylov, "Edge width estimation for defocus map from a single image," *International Conference on Advanced Concepts for Intelligent Vision Systems*, Springer International Publishing, pp. 15-22, October 2015.

[47] H. Y. Lin and C. H. Chang, "Depth recovery from motion and defocus blur," *Optical Engineering*, vol. 45, no. 12, September 2006.

[48] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, November 1986.

[49] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 664-672, August 2004.

[50] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 689-694, August 2004.

[51] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 646-653, July 2006.

[52] S. Bae and F. Durand, "Defocus magnification," *Computer Graphics Forum, Blackwell Publishing Ltd*, vol. 26, no. 3, pp. 571-579, September 2007.

[53] J. Shi, L. Xu, and J. Jia, "Just noticeable defocus blur detection and estimation," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 657-665, June 2015.

LIST OF PUBLICATIONS

1. "Depth map estimation from single image and 3D conversion," communicated to *Journal of Engineering Research*, June 2016. (*SCI Indexed*)

2. "A survey on depth map estimation strategies," communicated to *International Conference on Signal Processing*, Samrat Ashok Technological Institute (SATI), Vidisha (M.P.), May 2016.

Nidhi_JAMWAL_Thesis

by Nidhi Jamwal

FILE	NIDHI_JAMWAL_801463015.PDF (2.89M)		
TIME SUBMITTED	12-JUL-2016 11:39AM	WORD COUNT	15735
SUBMISSION ID	689198876	CHARACTER COUNT	78445

Nidhi_JAMWAL_Thesis

ORIGINALITY REPORT

19%	12%	14%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Thapar University, Patiala Student Paper	1%
2	Zhang, Xinxin, Ronggang Wang, Xiubao Jiang, Wenmin Wang, and Wen Gao. "Spatially variant defocus blur map estimation and deblurring from a single image", Journal of Visual Communication and Image Representation, 2016. Publication	1%
3	Lueder, . "Assessment of Quality of 3D Displays", 3D Displays Lueder/3D Displays, 2011. Publication	1%
4	Mohaghegh, H., S. Samavi, N. Karimi, S. M. R. Soroushmehr, and K. Najarian. "Depth estimation from single images using modified stacked generalization", 2016 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2016. Publication	1%