Gaussian-Hermite Moment-Based Depth Estimation from Monocular Image for Stereo Vision

by

Samiul Haque

M.Sc. Engg. (EEE)

Department of
ELECTRICAL AND ELECTORINC ENGINEERING

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
July 2015

Declaration of Authorship

I, Samiul Haque, declare that this thesis titled, 'Gaussian-Hermite Moment-Based Depth Estimation from Monocular Image for Stereo Vision ' or any part of it has not been submitted elsewhere for the award of any degree or diploma and that all sources are acknowledged.

Signature: _____

Date: _____

The thesis titled 'Gaussian-Hermite Moment-Based Depth Estimation from Monocular Image for Stereo Vision ' submitted by Samiul Haque, Student-ID: P0412062209, Session: April 2012 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Electrical & Electronic Engineering on July 29, 2015.

Board of Examiners

1.

_____

Dr. S. M. Mahbubur Rahman
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000.

Chairman
(Supervisor)

2.

_____

Dr. Taifur Ahmed Chowdhury
Professor and Head
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000.

Member
(Ex-Officio)

3.

_____

Dr. Mohammed Imamul Hassan Bhuiyan
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000.

Member

4.

_____

Dr. Md. Ibrahim Khan
Professor
Department of CSE
Chittagong University of Engineering and Technology.
Chittagong-4349.

Member
(External)

Acknowledgements

I would like to express my sincerest appreciations and gratitude to Dr. S. M. Mahbubur Rahman,Professor of the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, for his all out support and keen supervision throughout the work. It is his inspiration and guidance that have enabled me to finish this work. I am truly owed to have the experience of doing research work under his supervision.

I am grateful to Professor Dr. Taifur Ahmed Chowdhury, Head of the Department of Electrical and Electronic Engineering, BUET for rendering me support and departmental assistance during my M.Sc. study in BUET.

I would also like to convey my thanks to all the supervisors, teachers, colleagues and friends for their support and encouragement during the work.Last but not least, I am indebted to my parents and my wife for thorough encouragement and psychological support to complete this work.

Samiul Haque
July 2015

Abstract

Depth estimation has turned into an emerging and challenging field of research in computer vision. The ubiquitous availability of stereo display technology has increased the demand of visual media with depth information. All major blockbuster movies and games are now being released with stereo versions. Strict parameters are required for generating stereo image in multiview set up. Thus the high-complexity of multiview imaging arrangement of capturing stereo image has motivated researchers to find a robust method of estimating depth from conventional 2D imaging set up. In addition to that 3D display technology has evolved into an advanced stage, but huge amount of media are still in 2D, in such a case depth estimation from monocular image is the only solution for generating a stereo view. Hence, research efforts are ongoing to develop low-complexity depth estimation algorithm for a scene specially from its monoscopic images captured using a CCD camera. Existing depth calculation methods from monocular images include depth from motion, depth form geometry and depth estimation using a learned database. These methods are limited by object geometry, prior knowledge of the scene as well as highly prone to noise. So, there is still a search for robust and autonomous depth calculation algorithm which does not depend on specific scene classes.

Motivated by the noise robust and invariant properties of orthogonal moments to the geometry of objects, this thesis presents a new moment based depth estimation method, which is independent of any prior knowledge about the scene. In particular, the Gaussian-Hermite moments (GHMs) which are very popular in visual signal processing are chosen to estimate the focus cue of a pixel from its neighborhood. It is known that there exists a significant correlation among the neighboring pixels in terms of depth information except for the sharp edges of an object. Hence a closed from expression of image matting is applied on the focus map of the image to generate the desired depth map. Extensive experiments are carried out in order to compare the proposed GHM-based depth estimation method with the existing methods using commonly-referred images in the literature. Performance comparisons of depth estimation in terms of visual quality, stereo generation and mean opinion score show that the proposed method performs significantly better than other methods.

Contents

List of Figures

List of Tables

List of Abbreviations

DDE	**D**epth from **D**efocus **E**stimatoin
DDK	**D**epth from **D**isparity of **K**inect
DEI	**D**epth using **E**dge **I**nformation
DGHM	**D**epth from **G**aussian-**H**ermite **M**oments
DVP	**D**epth from **V**anishing **P**oint
GHM	**G**aussian-**H**ermite **M**oments
HVS	**H**uman **V**isual **S**ystem

List of Symbols

p	Order of polynomial
$H_p(x)$	Hermite polynomial of order p
δ_{pq}	Kronecker delta
σ	Spread Factor
$I(x, y)$	Continuous square integrable image signal
$\eta_{p,q}$	Gaussian-Hermite moments of order (p, q)
γ	Parameter denoting lower order moments in focus estimation.
W	Sliding image window
$\eta_{p,q}^{W}$	Moments of a pixel from a sliding window W
$F_T^W(i, j)$	Focus cue of a pixel under the sliding window W
$\mathbb{L}(I^W(i, j, \gamma))$	Image intensity function reconstructed using lower order moments
$\mathbb{H}(I^W(i, j, \gamma))$	Image intensity function reconstructed using higher order moments
$\mathbf{F_m}$	Sparse focus map
$\mathbf{d_m}$	Depth map
$\mathbf{L}$	Matting matrix
$\mathbf{D}$	Diagonal matrix
$\mu_\mathbf{k}$	Mean matrix of the colors in window W_k
$\mathbf{\Sigma_k}$	Covariance matrix of the colors in window W_k
$\mathbf{U_3}$	3×3 identity matrix.
λ	Regularization parameter in matting process

Chapter 1

Introduction

1.1 Introduction

In the recent years, the 3D scene contents have become ubiquitous in entertainment industry, e.g., films and games, robotics, news and media. Disparity map generation by capturing the scene in multiview camera provides an option to generate 3D media content from known camera parameters, but these settings are complex and costly. Besides, there are huge amount of 2D media contents produced so far, which have neither any depth information nor any calibration parameters to convert them to 3D. In such a scenario, the only option for generating stereo version of such media is to estimate the depth map from monocular image using efficient depth cues. This section summarizes the motivation for depth estimation from monocular images. Then it briefly describes the background of depth estimation. It also focuses on the major challenges in depth estimation and the necessity of a robust method for depth estimation of monocular images. This section ends up with discussing the objective of this research work which is to use Gaussian-Hermite moments for depth estimation.

1.2 Depth of Image: A Background

There are several approaches for depth calculation of monocular images. For example, the traditional object based techniques [1–5] assign depth to segmented foreground and background objects. These methods are highly dependent on the

performance of the segmentation algorithm. Vanishing lines are also used for depth estimation [6], but their use is limited only to certain categories of scenes. Depth estimation from scene classification [6–8] requires predefined outdoor and indoor annotations. Machine learning approaches [9–13] require scene database with prior depth knowledge, the performance of which is highly dependent on the training data set, feature extraction and classification algorithm. The Kalman filtering approach [14, 15] commonly used in the depth estimation of dynamic scenes, has very slow convergence speed. Depth from defocus cue [16] method uses known low pass filter to blur the original image and estimates the depth from the blurring amount. This method shows a poor performance when the background is not defocused. In general the existing depth estimation methods can be classified into following five major categories

- Depth from disparity
- Depth from motion and interpolatoin
- Depth from geometry
- Depth from recursive filtering
- Depth from focus cue

In the following subsections we discuss these methods in brief. An excellent overview of these methods can be found in [17].

1.2.1 Depth from Disparity

Depth from disparity is one of the most common approaches for depth estimation. For a 1D parallel camera arrangement according to the triangle similar geometry, the depth of a view is inversely proportional to the disparity. For estimating the disparity of each pixel between views correspondences are estimated. The traditional practice for disparity estimation were fixed size block-matching approaches [18]. This method relies on blue screening method followed by Bayesian matte [19] extraction for separating foreground from background. In recent years researchers have preferred variable block size and segment based disparity estimation techniques [20] for their high adaptivity. Disparity plane is refined according to the steps described in [21] in order to increase accuracy.

The depth sensing can be made more accurate by arranging more than two cameras to capture image or video. The problem definition in multiview processing is commonly described as the reconstruction of 3D image models from two or more captured image views. This images are captured at a same time from cameras positioned at different location. By calibrating these cameras carefully and determining the necessary camera and geometric parameters (e.g. focal length, inter camera distance, distortion coefficients), 3D models of captured objects hence their depth can be estimated.

1.2.2 Depth from Motion and Interpolation

Choi et al.[22] utilized motion and pixel value as the feature for image segmentation. The picture of a sequence is first horizontally divided into region. The mean and standard deviation of gray level of motion pixels is then calculated for each region. Foreground is defined to be the set made up of the pixels whose pixel values are within the mean and standard deviation of each region. Although this mechanism is simple and can be implemented in very large scale integration (VLSI), the number of scenes it can cope with is limited. In [1] an alternative way of segmentation with motion was proposed. In this algorithm the scene is segmented considering the motion distribution based on Gibbs formulation [23] and E-matrix method [24]. 3D motion estimation with the help of segmentation and 2D correspondence is then used to obtain the depth map. This method is applicable only for video or consecutive image sequence. Then machine learning techniques are used with prior information to interpolate depth of frames between two key frames. The depth information of the key frames are already known. The relationship between input and output was studied and employed for depth estimation in [9].

1.2.3 Depth from Geometry

These methods usually intend to generate accurate depth maps by determining vanishing line and vanishing point of a scene. Battiato et al. propose to classify images as outdoor, outdoor with geometric appearances and indoor.For outdoor images without geometric appearance, the algorithm detects the position of the horizon, which is usually present in outdoor scenes, and assigns depth according

to the horizon. The depth monotonically increases as approaches to the horizon. As for outdoor scenes with geometric appearance and indoor scenes which usually contain objects, the algorithm relies on the vanishing lines and vanishing points to generate depth map in different cases. The straight lines, which are originally parallel to each other, intersect at a vanishing point in images. The orientation of vanishing lines and the location of vanishing points reveal the depth map of images. False regions that do not follow the depth distribution from horizon and vanishing points may exist and for such cases a segmentation is required. The performance of this method is greatly dependent on the image type as in many cases, there are no vanishing points available in an image. A similar depth estimation approach presented in [7], which classifies indoor and outdoor scenes according to their spectrum signature and a well trained probabilistic model.

1.2.4 Depth from Recursive Filtering

Hung et al. [14] developed a pixel-based model of Kalman filter which iteratively update the estimation. The optical flow model, derived from image motion equation and spherical projection, is clearly described in [15]. The kalman filter approaches, due to their nature, require longer time to converge. This characteristic makes it less attractable especially in real-time application.

1.2.5 Depth from Focus Cue

Wavelet and edge detection is used by Aguirre et al. [2] to separate foreground and background under the assumption that background is defocused and has no sharpness and spectral high frequency. Wavelet transform is performed in block-based manner to obtain a coarse spectrum measure. Pixel-wise refinement using edge Lipschitz regularity [3] folowed with edge strength and depth of each pixel as its output. Similar approach was taken in [4]. Zhuo & Sim [16] addresses the problem of recovering defocus map from a single image by re-blurring the input image using a Gaussian kernel and then taking the ratio between the gradients of input and re-blurred images. They also provides a way to propagate the defocus estimates from edge location to the entire image and obtain a full depth map using defocus map interpolation method.

1.3 Scope of the Work

For estimating depth from disparity multiview arrangement is necessary . Two or more cameras are arranged at a measured known distance, the camera parameters are also necessary for calculation of disparity. The complexity of imaging setting makes it a poor choice for depth estimation. Moreover this method can not be applied to the images and videos which have been already captured as 2D media content. Depth from motion technique is simple to be implemented in VLSI but it works with only limited number of scenes. For interpolation method prior depth information is required which also involves machine learning scheme. Kalman filter approach is time consuming and computationally expensive.

From the above discussions, we find that according to the imaging set up depth extraction methods can be classified in two main categories as *singleview* and *multiview* depth estimation techniques. In multiview techniques multiple cameras should be present during the capture. Also these methods are sensitive to camera calibration and camera positions. There are huge amount of image and videos captured with single view set up. In our day to day life, we use cameras to capture monoscopic image. This motivates us to focus on estimating depth maps for such monocular images.

In general, all the existing methods are very much sensitive to noise and geometry of the objects in the scene, as well as they require prior knowledge of the scene. Hence, there is a scope of developing a low complexity depth estimation scheme for monocular images, which does not require any prior knowledge and is also robust to noise and object geometry. Certain affine transformation of the geometric and orthogonal moments such as the Gaussian-Hermite, Krawtchouk, Tchebichef and Zernkie have shown to posses invariance properties to scaling, shifting and rotation of a pattern. In addition to the fact that the orthogonal moments can provide the sharpness or smoothness, in other words an overall focus content of an image [25]. Hence, orthogonal moments are expected to provide a focus cue, which is robust to noise and object geometry. Among all the orthogonal moments Gaussian-Hermite moments (GHMs) are chosen in the proposed study due to their popularity in visual signal processing and also for its noise robust and shape invariance property. The popularity of GHMs lies in the fact that the width of the Gaussian weight function of the Hermite polynomial expansion provides a flexibility in isolating the visual features just as the Human Visual System (HVS) does. Moreover, the

zero-crossings of the Gaussian-Hermite basis functions are more evenly distributed than that of the other orthogonal moments and thus better for noise robust feature representation [26]. Hence there exist a suitable scope of a thorough research study that considers GHMs for depth estimation for monocular images.

1.4 Objective

In this work the primary objective is to develop a novel depth estimation method using orthogonal GHMs. In particular, first the moments are chosen to determine the focus value of each pixel from information of its neighboring pixels. A focus map of the image is generated using a sliding window technique. Next the obtained sparse focus map is targeted to be used in depth map estimation by using a suitable matting algorithm. Finally the performance of the proposed method is evaluated by comparing the result with that of the other existing methods and ground truth depth data.

1.5 Outline

The thesis is organized as follows:
Chapter 2 introduces the depth estimation method proposed in this thesis. It also explains the methods of extracting focus cue of an image using GHM and finally the application of a matting algorithm to get a depth map.
In Chapter 3 the results of the method on different types of images are presented. The comparison of proposed method with existing method is also illustrated in this chapter.
Finally, Chapter 4 draws the conclusion of the work. This chapter summarizes the findings of the entire study. It also suggests the scope for further study regarding this thesis.

Chapter 2

Proposed Depth Estimation Method

2.1 Introduction

This chapter presents the mathematical formalism used in this thesis with necessary explanation. At first a brief overview of GHMs is given. The formulation of generating GHMs from an image is illustrated in the following section. In the later section the method for generating focus map is explained in detail. The process of generating the depth map of the image from the sparse focus map using matting process is described in the last section before conclusion.

2.2 GHM: A Brief Review

Since GHMs are estimated using Hermite polynomials that are orthogonal to Gaussian weight function, first we present a brief introduction to Hermite polynomials and their orthogonality with the Gaussian weighting function. Next, the method of obtaining the GHMs of the image using this polynomials are presented.

2.2.1 Gaussian-Hermite polynomials

The Hermite polynomial of order where $p \in \mathbb{Z}^1$ on the real line $x \in \mathbb{R}^1$ is defined as [27]

$$H_p(x) = (-1)^p \exp(x^2)(d^p/d\,x^p)\exp(-x^2) \tag{2.1}$$

It can also be written in the form of following series

$$H_p(x) = \sum_{p/2}^{k=0} \frac{(-1)^k p!}{k!(p-2k)!}(2x)^{p-2k} \tag{2.2}$$

The property of Hermite polynomials is such that they satisfy the following orthogonal relation with respect to the weight function $w(x) = \exp(-x^2)$

$$\int_{-\infty}^{\infty} \exp(-x^2) H_p(x) H_q(x) dx = 2^p p! \sqrt{\pi} \delta_{pq} \tag{2.3}$$

here δ_{pq} is the Kronecker delta. Hermite polynomials can be computed efficiently using the following recursive equation

$$H_{p+1}(x) = 2xH_p(x) - 2pH_{p-1}(x), \quad for \quad p \geq 1 \tag{2.4}$$

With initial conditions set to $H_0(x) = 1$ and $H_1(x) = 2x$. Equation (2.3) suggests that Hermite polynomial is orthogonal, but not orthonormal. In order to obtain an orthonormal version a weighted Hermite polynomial has to be adopted which can be defined by [28]

$$\hat{H}_p(x) = (2^p p! \sqrt{\pi})^{-1/2} \exp(-x^2/2) H_p(x) \tag{2.5}$$

Equation (2.5) is orthonormal Hermite polynomial and it is a Hermite polynomila modulated by a Gaussian function with variance 1. For general cases the Gaussian-Hermite (GH) polynomial with spread factor σ ($\sigma > 0$) on the real line $x \in \mathbb{R}^1$ has the following definition [28]

$$\hat{H}_p(x;\sigma) = (2^p p! \sqrt{\pi}\sigma)^{-1/2} \exp(-x^2/2\sigma^2) H_p(x/\sigma) \tag{2.6}$$

.

Including the spread factor σ, the orthogonality property described in equation (2.3) becomes

$$\int_{-\infty}^{\infty} \hat{H}_p(x;\sigma)\hat{H}_q(x;\sigma)dx = \delta_{pq} \tag{2.7}$$

Equation (2.7) satisfy the condition $\hat{H}_p(-x,\sigma) = (-1)^p\hat{H}_p(x,\sigma)$. So we can say Gaussian-Hermite polynomials have the symmetry property with respect to x.

2.2.2 Gaussian-Hermite moments

Let $I(x,y) \in L_2(\mathbb{R}^2)$ be a continuous square integrable 2D image signal. The set of 2D Gausian-Hermite moments of order $(p,q)(p,q \in \mathbb{Z}^1))$ is denoted as $\eta_{p,q}$ are obtaind from 2D GH basis functions expressed in termos of p^{th} and q^{th} order GH polynomials using the following relation [29]

$$\eta_{p,q} = \int\int I(x,y)\hat{H}_p(x;\sigma)\hat{H}_q(y;\sigma)d\,x d\,y \tag{2.8}$$

The orthogonality property of Gaussian-Hermite polynomials enables us to express an image intensity function in terms of its GHMs. The GHMs of the 2D image singal may be considered as the projections of the signal onto these 2D basis function. Thus, these moments charecterize the local neghboring image regions at different spatial modes that are defined by certain combinatoin of the derivatives of the Gaussian functions. Ideally from all possible moments the image $I(x,y)$ may be reconstructed without any error as

$$\hat{I}(x,y) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \eta_{pq}\hat{H}_p(x;\sigma)\hat{H}_q(y;\sigma) \tag{2.9}$$

An image can be approximately restored from its GHMs of orders $(0,0)$ up to $(N-1,N-1)$ by the following equation

$$\hat{I}(x,y) = \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} \eta_{p,q}\hat{H}_p(x;\sigma)\hat{H}_q(y;\sigma) \tag{2.10}$$

where $\hat{I}(x,y)$ is the restored approximate 2D image intensity function. The GHMs are obtained from two real lines $x \in \mathbb{R}^1$ and $y \in \mathbb{R}^1$ and hence a modification is required to obtain moments from the discrete coordinates of practically available images. Let, $I^W(i,j)((i,j) \in \mathbb{Z}^2)$ be a portion of image with window W having a size of $N \times N$. Using a similar approach considered in [30] , to obtain the 2D GHMs using (2.8), we have normalized the coordinates such that $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$ by choosing only the discrete values

$$x = \frac{2i - N + 1}{N - 1} \quad i = 0, 1, 2, \ldots, N-1 \tag{2.11}$$

$$y = \frac{2j - N + 1}{N - 1} \quad j = 0, 1, 2, \ldots, N-1 \tag{2.12}$$

In terms of discrete implementation the 2D moments of a pixel within the sliding image window W, $\eta^W_{m,n}$ may be obtained as

$$\eta^W_{p,q}(i,j) = \sum_{i,j\in\omega} I^W(i,j)\hat{H}_p(i;\sigma)\hat{H}_q(j;\sigma) \tag{2.13}$$

2.3 Methodology

2.3.1 Focus Estimation

Inspired from previously studied Chebyshev moments-based focus measure [25] we propose Gaussian-Hermite moment based overall focus measure for each pixel. The focus of an image or a pixel can be measured by taking the ratio of Higher order Chebyshev moment components to lower order moment components. In our thesis, we introduce a novel approach to generate a smooth focus map. We consider each pixel separately having some focus cue. In a 2D image signal each element is usually highly correlated to neighboring pixels. In the edge points of objects the correlation becomes weaker. For measuring focus cue of a pixel we take a $N \times N$ sliding window and slide it all over the image to calculate moment

components. In our work, focus cue of this window over a pixel can be calculated as

$$F_T^W(i,j) = \frac{\left\|\mathbb{H}(I^W(i,j),\gamma)\right\|}{\|\mathbb{L}(I^W(i,j),\gamma)\|} \tag{2.14}$$

Here γ is the limiting value by which the lower order moments are decided. γ is a function of window size N where $\gamma = \alpha(N-1)$ where $0 < \alpha \leq 1$.

$\mathbb{L}(I^W(i,j),\gamma)$ is identified as the low pass component which is actually the image intensity function reconstructed using the lower order Gaussian-Hermite moments and $\mathbb{H}(I^W(i,j),\gamma)$ is considered as the high pass components corresponding to the higher order Gaussian-Hermite moments.

For each pixel of the monocular image, we consider the neighbouring pixels, the GHMs of which is first calculated. Higher-order moments for which the width of GH polynomial is very small, capture minute details or variations around the pixel of interest. These higher-order moments can be considered as high pass filtered components. On the other hand, lower-order moments capture the significant variations around the pixel of interest and thus represent the lower-order filtered components. Ratio of the energy of reconstructed regions obtained from higher-order moments to that obtained from the lower-order moments gives the focus cue of the pixel of interest. Since the neighbouring pixels of the image have high dependency among themselves, it is found that the depth cue have similar dependency. Hence, same procedure is implemented to each pixel of the image using a sliding window with varying sizes and shapes.

From (2.10) the low pass component $\mathbb{L}(I^W(i,j),\gamma)$ can be written as

$$\mathbb{L}(I^W(i,j),\gamma) = \sum_{p=0}^{\gamma}\sum_{q=0}^{\gamma}\eta_{p,q}\hat{H}_p(i;\sigma)\hat{H}_q(j;\sigma) \tag{2.15}$$

The reconstructed image intensity function is the summation of high pass a low pass components. So, we can write

$$\left\|\mathbb{H}(\hat{I}^W(i,j),\gamma)\right\| = \left\|\hat{I}^W(i,j)\right\| - \left\|\mathbb{L}(\hat{I}^W(i,j),\gamma)\right\| \tag{2.16}$$

Here the $\hat{I}^W$ indicated that the reconstructed image is an approximation of the original image signal. By sliding the window over the image along horizontal and vertical axes we can generate the focus map for entire image from (2.14). In general, sharp objects are considered closer and on the foreground and blurred objects are considered to be in background.

The following illustrations in Figure 2.1 and Figure 2.2 show the GHM based focus map estimated for two images

(a) (b)

Figure 2.1: Visual output of result of focus map obtained using GHM. (a) original image (*Boy*), (b) Focus map estimated using GHM

(a) (b)

Figure 2.2: Visual output of result of focus map obtained using GHM. (a) original image (*Flower*), (b) Focus map estimated using GHM

2.3.2 Depth Map Generation

The focus map obtained from GHM based approach is a sparse focus map $\hat{\mathbf{F}}_\mathbf{m}$. By propagating the defocus blur estimates from edge locations to the entire image we obtain a full depth map $\mathbf{d_m}$). We have followed the similar approach taken in [16]. We try to align the focus discontinuities with image edges. Matting Laplacian is applied to perform the focus map interpolation. The interpolation problem can be formulated as minimizing the following cost function [16]

$$E(\mathbf{d_m}) = \mathbf{d_m}^T \mathbf{L} \mathbf{d_m} + \lambda(\mathbf{d_m} - \mathbf{F_m})^T \mathbf{D}(\mathbf{d_m} - \mathbf{F_m}) \tag{2.17}$$

here $\mathbf{d_m}$ is the vector form of the sparse depth map, $\mathbf{F_m}$ is the vector form of sparse focus map, $\mathbf{L}$ is the matting Laplacian matrix and $\mathbf{D}$ is a diagonal matrix. An element of $\mathbf{D}$ is defined as

$$D_{ii} = \begin{cases} 1, & \text{if } ii \text{ is at edge location} \\ 0, & \text{otherwise} \end{cases}$$

The fidelity to the sparse depth map and smoothness of interpolation are balanced by the regularization parameter λ. An element of the matting Laplacian matrix L is defined as [31]

$$L_{ij} = \sum_{k|(i,j)\in W_k}^{max} \left(\delta_{ij} - \frac{1}{|W_k|} \left(1 + (\mathbf{I_i} - \mu_\mathbf{k})^T (\mathbf{\Sigma_k} - \frac{\epsilon}{|\omega_k|}\mathbf{U_3})^{-1} (\mathbf{I_j} - \mu_\mathbf{k}) \right) \right) \quad (2.18)$$

Where δ_{ij} is the Kronecker delta, U_3 is a 3×3 identity matrix, μ_k and Σ_k are the mean and covariance matrix of the colors in window W_k centered at (i, j), I_i and I_j are the colors of the input image I at the pixel i and j respectively represented as a vector of length 3. ϵ is a regularization parameter and $|W_k|$ is the size of the window W_k.

The optimal depth map is obtained by solving the following sparse linear system [16]

$$(\mathbf{L} + \lambda\mathbf{D})\mathbf{d_m} = \lambda\mathbf{D}\mathbf{F_m} \quad (2.19)$$

thus,

$$\mathbf{d_m} = \lambda\mathbf{D}\mathbf{F_m}(\mathbf{L} + \lambda\mathbf{D})^{-1} \quad (2.20)$$

The depth map from Gaussian-Hermite moments (DGHM) $\mathbf{d_m}$ in 2D matrix form can be retrieved by unwrapping the vector $\mathbf{d_m}$ into a matrix of the original image size.

2.4 Conclusion

In this chapter we have discussed the mathematical and algorithm details of proposed GHM based depth estimation method. In the next chapter, we will present the results of depth estimation obtained using this method and compare them with several existing methods and ground truth depth images.

Chapter 3

Experimental Results

3.1 Introduction

In Chapter 2, the proposed method for depth estimation is being described in detail. This chapter highlights the result of DGHM method. The chapter also compare the results achieved from DGHM with several existing methods and also present a comparison with ground truth depth data. The robustness of DGHM is also illustrated in this chapter using noticeable visual outputs.

3.2 Experimental Images

For evaluating our method, we have used color images from different datasets. Focused images are available from [16] maintained by Zhuo & Sim. For ground truth data we have used the datasets from [32] [33] provided by Middleburry college. Images captured through Kinect with depth data are provided by University of Washington [34]. To compare our results with machine learning approaches, the image and depth data from [10] and [35] have been used. Table 3.1 gives brief description of the images used in the evaluation process.

Table 3.1: Significant Features of Experimental Images

Experimental Images						
Image Name	Focus	Edges	Occlusion	Size	Source	Ground Truth
Boy	Single Object	High Detail	Nil	667 × 500	[16]	No
Flower	Single Object	High Detail	Minor	800 × 600	[16]	No
Pumpkin	Single Object	High Detail	Nil	800 × 600	[16]	No
Bird	Single Object	Medium Detail	Nil	800× 600	[16]	No
Building	All Focused	High Detail	Major	800× 600	[10]	No
Teddy	All Focused	High Detail	Major	450 × 375	[32]	Yes
Cones	All Focused	High Detail	Major	450× 375	[32]	Yes
Books	All Focused	High Detail	Major	695 × 555	[33]	Yes
Laundry	All Focused	Medium Detail	Minor	671 × 555	[33]	Yes
Car	All Focused	Low Detail	Minor	579 × 430	[35]	No
Room	All Focused	Medium Detail	Minor	640 × 480	[34]	Yes

In our image set, we have used different types of images. Table 3.1 describes the significant features of the experimental images. The feature focus describes whether the image contain only one focused object. Ground truth data is considered available when depth is obtained at the time of image capture by special camera and imaging set up. The experimental image is being made robust by including images of different sizes having different levels of occlusion of image objects.

3.3 Performance Evaluation

In evaluating the performance of proposed method, we have taken three measures. Initially, visual comparison of estimated depth from different methods to that of proposed method is illustrated. There are two aspects of visual comparison -

- First, depth map quality in terms of object separation and smoothness of the depth map is observed using naked eye.
- Second, the depth map quality is evaluated by generating stereo anaglyph image and by visualizing them using chromatic 3D glasses. There are several commonly referred methods for generating stereo image from monocular image and depth map which can be found in [36] and [37]. We have used a web application platform provided by DX interactive [38].

Beside comparing the proposed method with other existing methods visual comparison of the results with ground truth depth map is also performed. Finally, robustness of the proposed method is evaluated. For evaluation of the robustness three results are demonstrated:

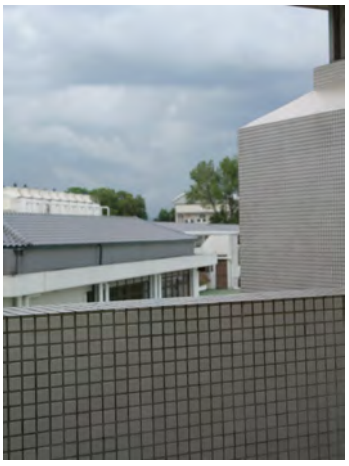
- Comparison of error percentage on mismatched focused edge is evaluated on a synthetic image.
- Consistency of the depth estimation is shown by adding white Gaussian noise to the images.
- The performance of the proposed method is evaluated using geometric distortion cause by multiview scenes.

3.4 Results

Five representative depth estimation methods, namely Depth from edge information (DEI) [39], Depth from scene classification (DSC) [35], Depth from vanishing point (DVP) [6], Depth from defocus estimation (DDE) [16], Dept from disparity of Kinect (DDK) are compared with proposed DGHM method. In order to improve the readability of this section, five comparing methods and corresponding performance comparisons in terms of visual output are illustrated in separate subsections. Finally the overall results of the methods are illustrated using a experiment that uses the scores given by unbiased viewers. The robustness of proposed DGHM method is also presented in a later subsection. In our experimental setup of DGHM method the regularization parameter λ was set to 0.001, sliding image window size was 9×9, value of γ was set to 16.

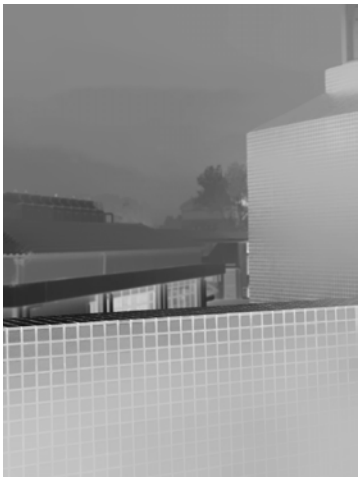
3.4.1 DGHM vs. DEI

Figure 3.1 and Figure 3.2 demonstrate a comparison of proposed method with DEI [39]. This method utilizes the edge information to segment the image into object groups. A depth map is then assigned based on a hypothesized depth gradient model. Figure 3.1 and Figure 3.2 illustrates a comparison of estimated depth map between DGHM and DEI. DEI uses predefined gradients for assigning depth to segmented regions. In our experiment the assumed predefined gradient was such that objects in the bottom of the image was considered to be closer than the objects in top portion of the image. Figure 3.1 and Figure 3.2 shows that the DGHM based method provides a smoother depth map. The stereo image generated from the depth maps are depicted in Figure 3.3 and Figure 3.4. The effect of smoother depth maps could be visualized in stereo images by viewing them with a anaglyph 3D glass. In the anaglyph image of *Building* the lines along the wall is distorted in case of DEI.



(a)

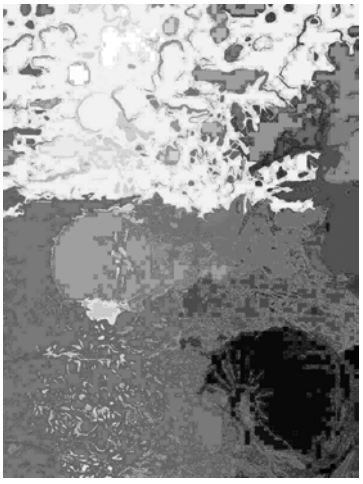
(b)



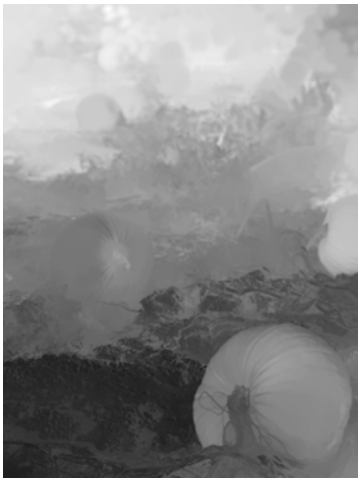
(c)

Figure 3.1: Visual output of depth images obtained using DGHM and DEI. (a)Original image (*Building*), (b) Depth Output from DEI method, (c) Depth output from DGHM method

(a)

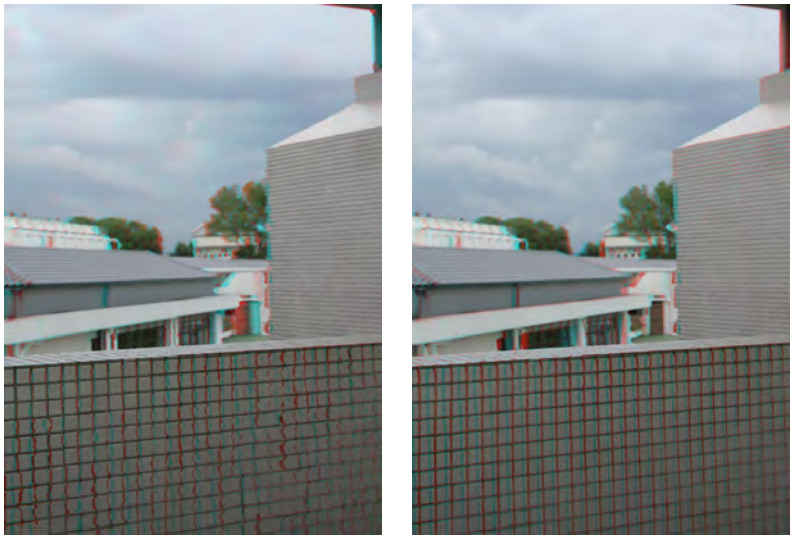


(b)



(c)

Figure 3.2: Visual output of depth images obtained using DGHM and DEI. (a)Original image (*Pumpkin*), (b) Depth Output from DEI method, (c) Depth output from DGHM method



(a) (b)

Figure 3.3: Visual Comparison of stereo images generated using depth from DEI and DGHM methods. (a)Stereo image from DEI,(b)Stereo image from DGHM

(a) (b)

Figure 3.4: Visual Comparison of stereo images generated using depth from DEI and DGHM methods. (a)Stereo image from DEI,(b)Stereo image from DGHM

3.4.2 DGHM vs. DSC

Research conducted by Karsch et al. [35] at Microsoft introduces a technique that automatically generates plausible depth maps from videos using non-parametric depth sampling. This technique is applicable to single images as well as videos. For training and evaluation, DSC uses a Kinect-based system to collect a large dataset containing stereoscopic videos with known depths. The proposed DGHM show better performance in distinguishing objects in the depth map compared to [35]. Code of DSC is provided by Microsoft research [40]. Figure 3.5 compare depth maps of the image *Car* generated from DGHM and Karsch et al. From the depth map it is evident that DGHM provides a more detailed depth map by separating the objects by depth levels. The stereo image generated from the depth maps are depicted in Figure 3.6

(a)

(b)

(c)

Figure 3.5: Visual output of depth images obtained using DGHM and DSC. (a)Original image (*Car*), (b) Depth Output from DSC method [35], (c) Depth output from DGHM method

(a) (b)

Figure 3.6: Visual Comparison of stereo images generated using depth from DSC and DGHM methods. (a)Stereo image from DSC,(b)Stereo image from DGHM

3.4.3 DGHM vs. DDE

Depth from defocus exploits the blurring of image to estimate a depth map [16]. At first spatially varying defocus blur at edge locations are detected later the The input defocused image is re-blurred using a Gaussian kernel and the defocus blur amount is obtained from the ratio between the gradients of input and re-blurred images. By propagating the blur amount at edge locations to the entire image, a full depth map is obtained. Figure 3.7 and Figure 3.8 illustrates a comparison between DGHM and DDE method proposed by Zhuo & Sim [16]. The code of DDE method is provided through [41]. In defocus estimation standard deviation of the Gaussian filter was set to 1, the maximum re-blurring parameter was set to 3. A canny edge detector was used for detecting edge locations. The stereo image generated from the depth maps are depicted in Figure 3.8

(a)

(b)

(c)

Figure 3.7: Visual output of depth images obtained using DGHM and DSC. (a)Original image (*Bird*), (b) Depth Output from DDE method [16], (c) Depth output from DGHM method

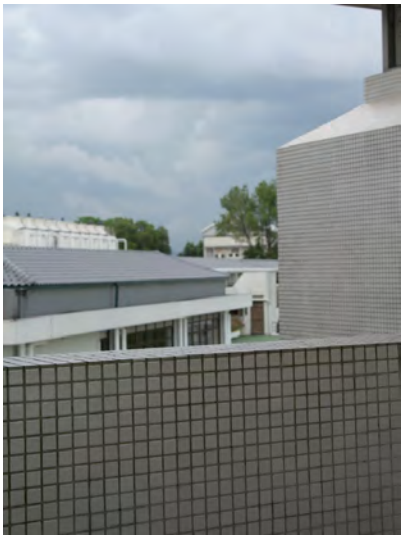
(a)

(b)

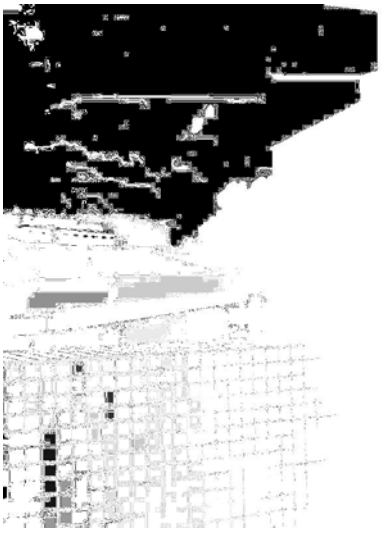
Figure 3.8: Visual Comparison of stereo images generated using depth from DDE and DGHM methods. (a)Stereo image from DDE,(b)Stereo image from DGHM

3.4.4 DGHM vs. DVP

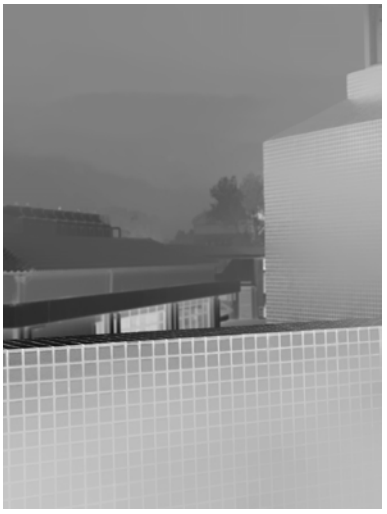
In DVP method, at first vanishing point of the image is calculated then gradient plain is assigned along the vanishing lines [6]. This method is heavily dependent on the accuracy of vanishing point detection. Numerous images have zero or multiple vanishing points so this method fails in most of the images. A comparison between visual output of DVP and DGHM is illustrated in Figure 3.9. In our experimental set up depth gradient plane was assigned according to the relative position of vanishing point in the image plane. The stereo image generated from the depth maps from two methods are depicted in Figure 3.10



(a)

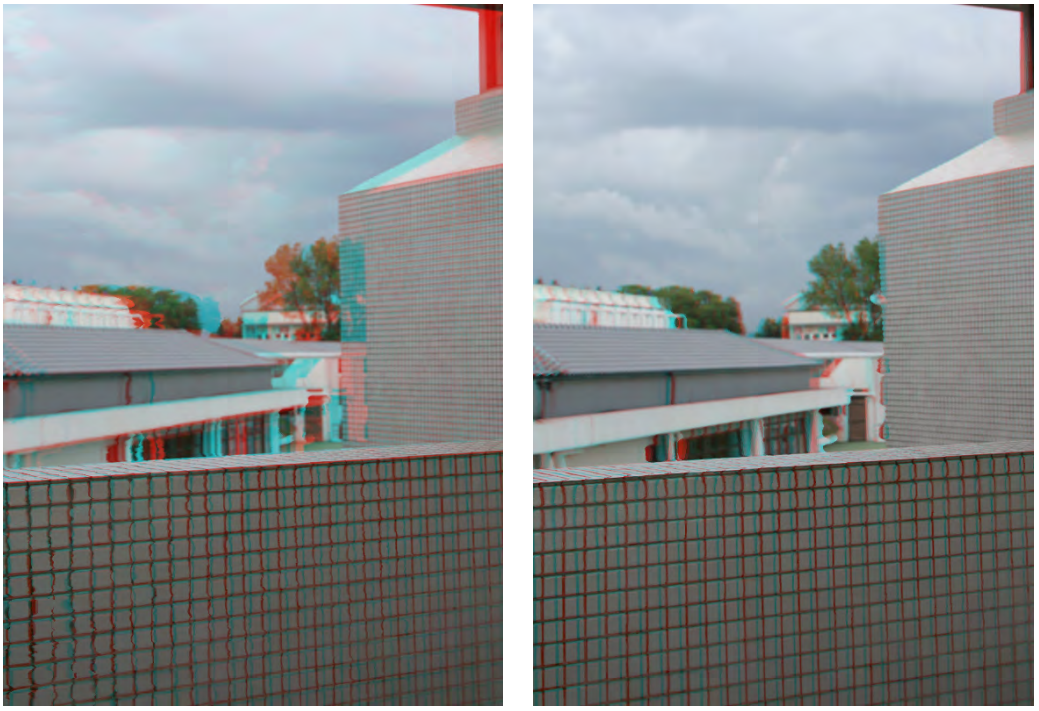


(b)



(c)

Figure 3.9: Visual output of depth images obtained using DGHM and DSC. (a)Original image (*Building*), (b) Depth Output from DVP method, (c) Depth output from DGHM method



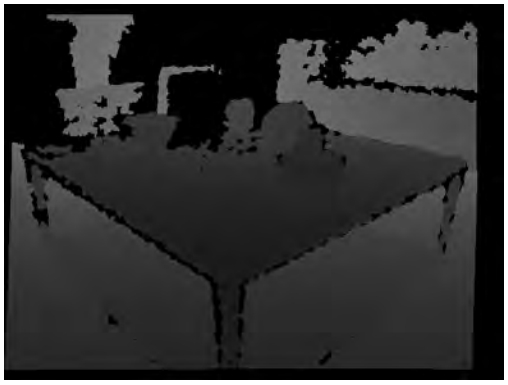
(a) (b)

Figure 3.10: Visual Comparison of stereo images generated using depth from DVP and DGHM methods. (a)Stereo image from DVP,(b)Stereo image from DGHM

3.4.5 DGHM vs. DDK

Kinect camera from mircosoft corporation provides a easy way to obtain depth map using disparity with dual camera and infrared laser light setting. The Kinect image data is taken from University of Washington [34]. Kinect camera has specialized 3D depth sensors. The depth map is constructed by analysing a speckle pattern of infrared laser light. It is speculated that the Kinect combines structured light with depth from focus and depth from stereo techniques [42]. The details of Kinect is not publicly available. It is evident that DGHM produces a depth map which is smoother than the depth map obtained by Kinect. The stereo image generated from the depth maps are depicted in Figure 3.12. From Figure 3.12, it can be observed using 3D glasses that DDK depth map from Kinect produces distortion around the legs of the table which is not present in the stereo anaglyph image generated using depth from DGHM.

(a)

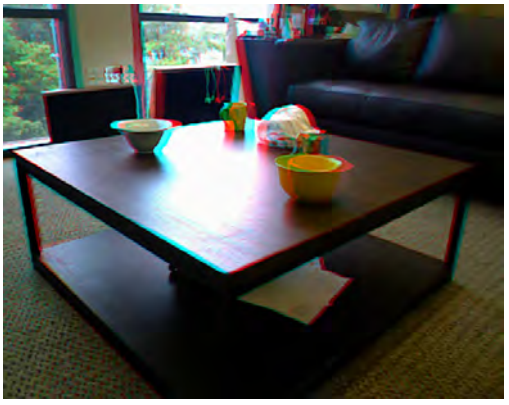


(b)

(c)

Figure 3.11: Visual output of depth images obtained using DGHM and DSC. (a)Original image (*Room*), (b) Depth Output from Kinect, (c) Depth output from DGHM method

(a)



(b)

Figure 3.12: Visual Comparison of Stereo Images generated using depth from DVP and DGHM methods. (a)Stereo image generated from depth map of Kinect,(b)Stereo image from DGHM

3.4.6 Scores from Visual Quality

Table 3.2 represents the Mean Opinion Score (MOS) of stereo images generated using different methods. For obtaining MOS, 20 unbiased viewers were asked to score stereo anaglyph images according to their quality in a scale of 10. Stereo images generated from the depth of images described in Table 3.1 were used for evaluation. The result indicates that DGHM performance is better than the other methods.

Table 3.2: Mean Opinion Score from unbiased viewers on Anaglyph stereo images

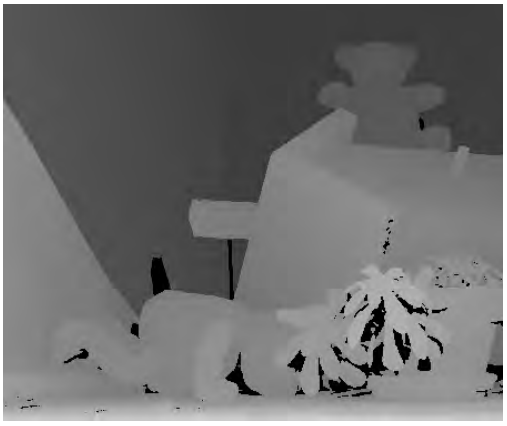
Image	DEI	DVP	DDE	DSC	DDK[1]	DGHM
Boy	6.5	4	7.5	4	–	**7**
Flower	5	4	7	4	–	**7**
Pumpkin	6.5	5	5.5	4	–	**7**
Bird	6	2	7.5	4.5	–	**7.5**
Building	7	5	7	6	–	**7**
Teddy	4	5	7	4	–	**7.5**
Cones	5	4.5	6	6	–	**7**
Books	5.5	3	7	6	–	**8**
Laundry	6	3	6	5.5	–	**7**
Car	5.5	4	6	6	–	**6.5**
Room	6	5	6.5	5.5	6	**8**
Average	5.7	4	6.6	5.4	6	**7.2**

3.4.7 Comparison with Ground Truth

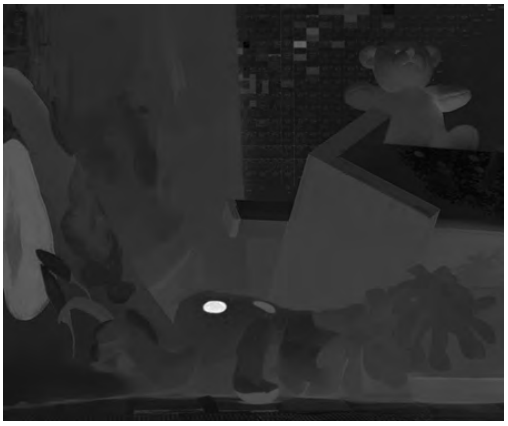
Figure 3.13 and Figure 3.14 present a comparison of our proposed method with ground truth depth image found from [32] . The stereo results demonstrate that the stereo image generated using the depth map from GHM is equivalent to the stereo image generated from ground truth depth map. Table 3.3 presents the MOS obtained by asking 20 unbiased viewers to score the quality of stereo anaglyph images generated from DGHM and ground truth depth map.

[1]Kinect depth data is not available for all images

(a)



(b)



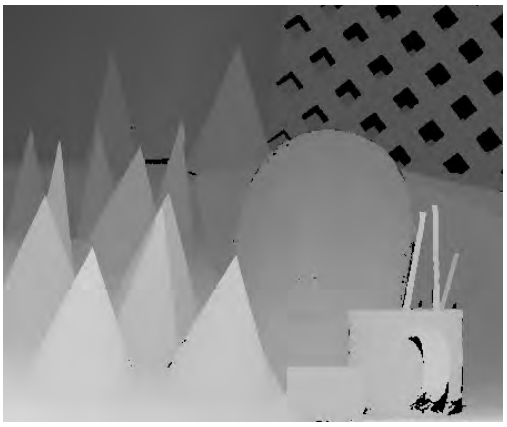
(c)

(d)

(e)

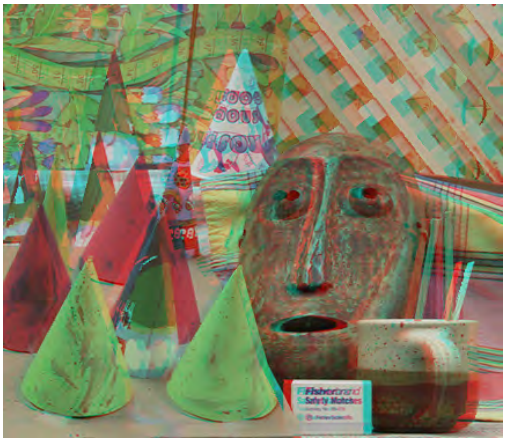
Figure 3.13: Visual output of stereo image generated using ground truth depth and DGHM. (a) Original image (*Teddy*), (b) Ground truth depth [32], (c) DGHM output, (d) Stereo image generated using (b), (e) Stereo image generated using (c)

(a)

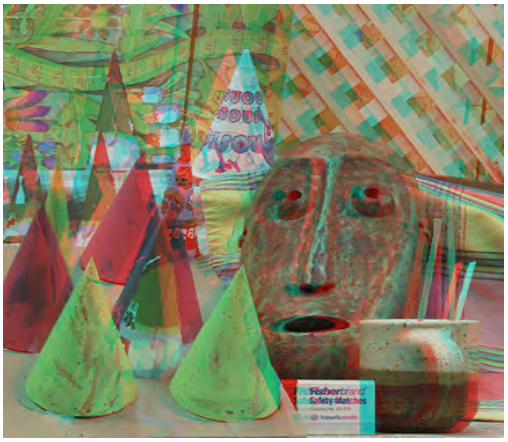


(b)

(c)



(d)



(e)

Figure 3.14: Visual output of stereo image generated using ground truth depth and DGHM. (a) Original image (*Cones*), (b) Ground truth depth [32], (c) DGHM output, (d) Stereo image generated using (b), (e) Stereo image generated using (c)

Table 3.3: Mean Opinion Score from unbiased viewers on Anaglyph stereo images obtained from DGHM and ground truth depth data

Image	Ground Truth Depth	DGHM
Teddy	7.5	**7**
Cone	7	**7**
Books	6.5	**6**
Laundry	7	**7**

3.4.8 Robustness of DGHM

Robustness of DGHM is evaluated in three ways. These evaluation results are illustrated below

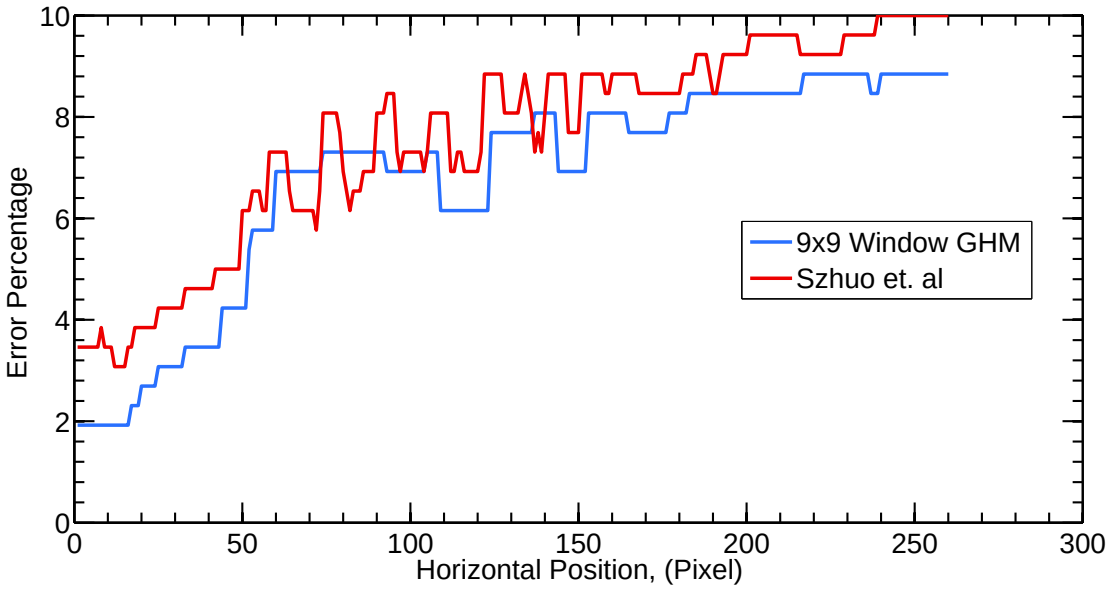
Focus Estimation on Synthetic Image

A synthetic image of size 260×260 is generated having a horizontal edge at its middle position. The edge is generated in such a way that at left most position it is extremely sharp. From left to right gradually the edge becomes blurred linearly. So, at the right side the edge is deformed. Figure 3.15 illustrates this synthetic image and also the performance of edge detection of proposed DGHM method and DDE method introduced in [16]. The synthetic image is used for focus estimation using DDE and DGHM method. As in the original image there are only two gray levels in two side of the edge, the focus estimation result is also expected to produce a binary image. The percentage of mismatched pixels on each column is considered as the error percentage. In naked eye the depth estimation result from GHM based method and depth from [16] appears to be almost same. However the proposed method performs better around the blurred edge and in presence of noise. The edge detection performance of DGHM is clearly better that DDE [16].

(a)

(b)

(c)



(d)

Figure 3.15: (a) Synthetic image with gradually increasing blur amount. (b) Focus estimation result of DDE [16], (c) Focus estimation result of DGHM (d)Error percentage in focus estimation of the edge using DGHM method and Zhuo et al.

Depth Estimation under Noise

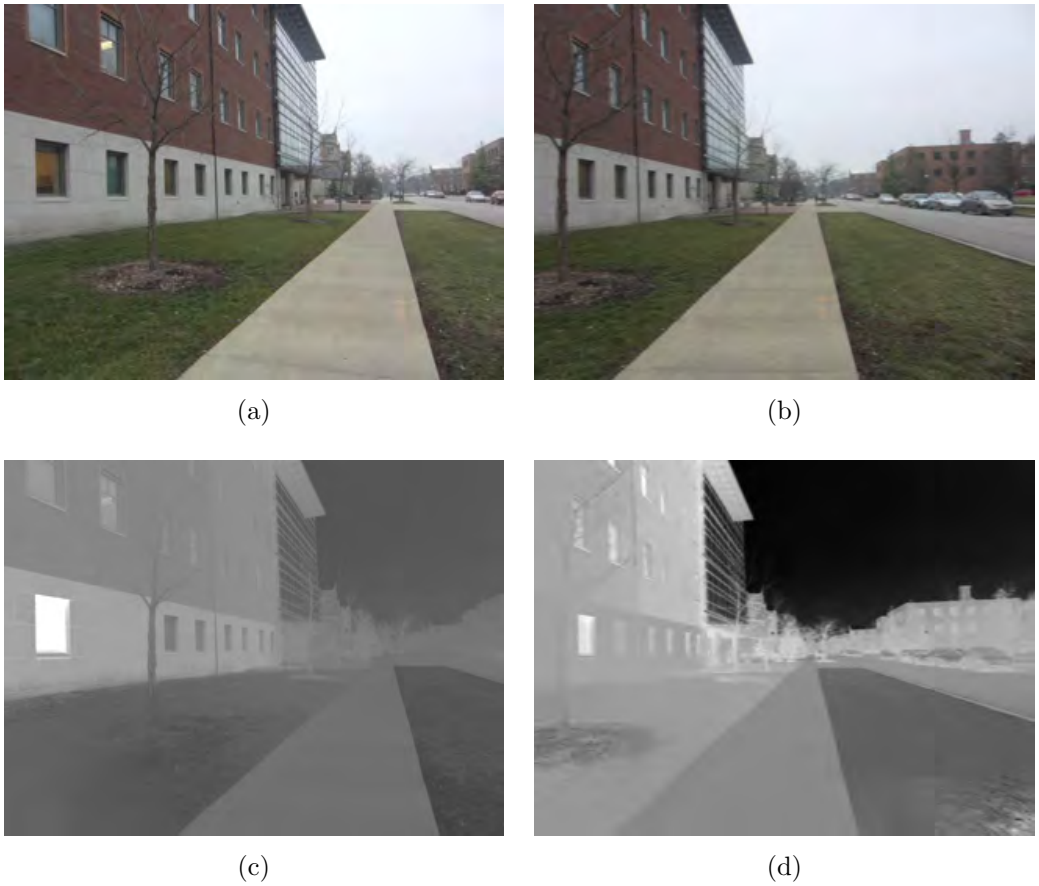
To test the noise robustness of DGHM, additive white Gaussian noise is blended with the image *Flower* the result of depth esimation in Figure 3.16 suggests that DGHM can estimate depth properly under noisy condition.

(a) (b)

Figure 3.16: Depth Estimation using DGHM in presence of additive white Gaussian noise. (a) Original Image (Flower) with noise, (b) DGHM result

DGHM under Geometric Distortion

For testing DGHM robustness under distortion resulting from multiview scenes we have selected image of same scenes from two different angle this images are collected from [35]. The DGHM results demonstrate that even after the distortion relative depth results are consistent.



(a) (b)

(c) (d)

Figure 3.17: Robustness of DGHM under distortion caused by multiview scene. (a) Initial view, (b) Final View (distortion), (c) DGHM result for (a), (d) DGHM result for (b)

3.5 Conclusion

In this chapter, the performance of proposed DGHM method is evaluated by comparing the depth map and generated stereo image with that of several existing methods. A MOS table is prepared from unbiased viewers' opinion to illustrate the performance of different methods including the proposed algorithm. The MOS table demonstrates that depth estimated using proposed GHM based method produces better stereo image compared to state of the art methods. The robustness of the DGHM method is also demonstrated with suitable examples.

Chapter 4

Conclusion

4.1 Summary of the work

Stereo vision has numerous application in entertainment information transfer and automated systems. The 3D vision is also crucial in the field of robotics. In forming 3D image or video the depth of the scene plays the most important role. In commonly-used existing systems, the depth information of a scene is extracted in a highly complex multiview imaging arrangement. In this thesis, considerable attention has been given to make a sound,robust and low complexity based depth estimation method for single image which does not require any prior knowledge. Requirement of prior knowledge and susceptibility to noise and scene class are the major limitations of the existing methods. In this work, 2D-GHM are used for depth estimation which show invariance to noise and shape geometry. GHM based focus estimation is employed to generate a focus map. By exploiting the fact that neighboring pixels maintain significant correlation among themselves we used Laplacian matting to estimate depth from the focus cue. Extensive experiments have been carried out in order to compare the proposed GHM based depth estimation method with the existing methods such as depth from disparity, depth from edge information, depth from vanishing point, depth from defocs estimation, depth from scene classification etc. using commonly-referred images having depth data. Performance comparison of depth estimation in terms of visual quality, stereo generation and MOS show that the proposed method significantly overcomes the limitations of existing methods in most of the cases.

4.2 Future Scope

There are tremendous scopes of future extension of this work by using GHMs in depth estimation. Some of the scopes are listed out below:

- We also would like to investigate the optimum highest order of moment γ to distinguish between highpass and lowpass component in focus map estimatoin using GHMs.
- There is high probability that optimum value of γ will vary according to the scene classification. We may use scene classification scheme for deciding the optimum value of γ
- In general close object appears to be sharper in an focused image, but in some cases distant objects may have sharper appearance than the close objects in the image. In that case DGHM may produce confusing results. To improve the performance of DGHM in such conditions this method can be fused with other methods like DEI, DVP or DSC.

References

[1] A Aydin Alatan and Levent Onural. Estimation of depth fields suitable for video compression based on 3-d structure and motion of objects. *Image Processing, IEEE Transactions on*, 7(6):904–908, 1998.

[2] Sergio Aguirre Valencia and Ramon M Rodriguez-Dagnino. Synthesizing stereo 3d views from focus cues in monoscopic 2d images. In *Electronic Imaging 2003*, pages 377–388. International Society for Optics and Photonics, 2003.

[3] Stephane Mallat and Sifen Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on pattern analysis and machine intelligence*, 14 (7):710–732, 1992.

[4] Cassandra Swain and Tsuhan Chen. Defocus-based image segmentation. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 4, pages 2403–2406. IEEE, 1995.

[5] Tae-Woo Kim, Jitae Shin, and Byung Tae Oh. Moving object-based depth map estimation using relabeling and hybrid matching. *Optical Engineering*, 53(3):033108–033108, 2014.

[6] Sebastiano Battiato, Alessandro Capra, Salvatore Curti, and Marco La Cascia. 3d stereoscopic image pairs by depth-map generation. In *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, pages 124–131. IEEE, 2004.

[7] Antonio Torralba and Aude Oliva. Depth estimation from image structure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(9): 1226–1238, 2002.

[8] Yannan Ren and Ju Liu. Single outdoor image depth map generation based on scene classification and object detection. In *Ubi-Media Computing and*

Workshops (UMEDIA), 2014 7th International Conference on, pages 91–94. IEEE, 2014.

[9] Philip V Harman, Julien Flack, Simon Fox, and Mark Dowley. Rapid 2d-to-3d conversion. In *Electronic Imaging 2002*, pages 78–86. International Society for Optics and Photonics, 2002.

[10] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng. 3-d depth reconstruction from a single still image. *International journal of computer vision*, 76(1):53–69, 2008.

[11] Hao Su, Qixing Huang, Niloy J Mitra, Yangyan Li, and Leonidas Guibas. Estimating image depth using shape collections. *ACM Transactions on Graphics (TOG)*, 33(4):37, 2014.

[12] Janusz Konrad, Meng Wang, and Prakash Ishwar. 2d-to-3d image conversion by learning depth from examples. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 16–22. IEEE, 2012.

[13] Yea-Shuan Huang, Fang-Hsuan Cheng, and Yun-Hui Liang. Creating depth map from 2d scene classification. In *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*, pages 69–69. IEEE, 2008.

[14] YS Hung and HT Ho. A kalman filter approach to direct depth estimation incorporating surface structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(6):570–575, 1999.

[15] V Shantaram and Madasu Hanmandlu. Recursive depth estimation from a sequence of images. In *Information Technology: Coding and Computing, International Conference on*, pages 201–201. IEEE Computer Society, 2000.

[16] Shaojie Zhuo and Terence Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852–1858, 2011.

[17] Gwo Giun Lee, He-Yuan Lin, Ming-Jiun Wang, Chun-Fu Chen, and Wang-Yang Li. 3d video system: Survey and possible future research.

[18] Peter J McNerney, Janusz Konrad, and Margrit Betke. Block-based map disparity estimation under alpha-channel constraints. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(6):785–789, 2007.

[19] Yung-Yu Chuang, Brian Curless, David H Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–264. IEEE, 2001.

[20] Andreas Klaus, Mario Sormann, and Konrad Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 15–18. IEEE, 2006.

[21] Li Hong and George Chen. Segment-based stereo matching using graph cuts. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–74. IEEE, 2004.

[22] Chul-Ho Choi, Byong-Heon Kwon, and Myung-Ryul Choi. A real-time field-sequential stereoscopic image converter. *Consumer Electronics, IEEE Transactions on*, 50(3):903–910, 2004.

[23] M Chang, M Ibrahim Sezan, and A Murat Tekalp. A bayesian framework for combined motion estimation and scene segmentation in image sequences. In *Proc. 1994 IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP*, volume 94, pages 221–224, 1994.

[24] Juyang Weng, Narendra Ahuja, and Thomas S Huang. Optimal motion and structure estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):864–884, 1993.

[25] PT Yap and P Raveendran. Image focus measure based on chebyshev moments. *IEE Proceedings-Vision, Image and Signal Processing*, 151(2):128–136, 2004.

[26] Bo Yang, Gengxiang Li, Huilong Zhang, and Mo Dai. Rotation and translation invariants of gaussian–hermite moments. *Pattern recognition letters*, 32 (9):1283–1298, 2011.

[27] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Number 55. Courier Corporation, 1964.

[28] Bo Yang, Tomas Suk, Mo Dai, and Jan Flusser. 2d and 3d image analysis by gaussian hermite moments.

[29] Jun Shen, Wei Shen, and Danfei Shen. On geometric and orthogonal moments. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(07):875–894, 2000.

[30] Bo Yang and Mo Dai. Image reconstruction from continuous gaussian–hermite moments implemented by discrete algorithm. *Pattern Recognition*, 45(4): 1602–1616, 2012.

[31] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):228–242, 2008.

[32] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–195. IEEE, 2003.

[33] Daniel Scharstein and Chris Pal. Learning conditional random fields for stereo. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[34] Koonchun Lai, Liefeng Bo, and Dieter Fox. Unsupervised feature learning for 3d scene labeling. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 3050–3057. IEEE, 2014.

[35] Kevin Karsch, Ce Liu, and Sing Bing Kang. Depth extraction from video using non-parametric sampling. In *Computer Vision–ECCV 2012*, pages 775–788. Springer, 2012.

[36] Salvatore Curti, Daniele Sirtori, and Filippo Vella. 3d effect generation from monocular view. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, pages 550–553. IEEE, 2002.

[37] Wan-Yu Chen, Yu-Lin Chang, Shyh-Feng Lin, Li-Fu Ding, and Liang-Gee Chen. Efficient depth image based rendering with edge dependent depth filter and interpolation. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1314–1317. IEEE, 2005.

[38] http://www.dxinteractive.com/.

[39] Chao-Chung Cheng, Chung-Te Li, and Liang-Gee Chen. A 2d-to-3d conversion system using edge information. In *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, pages 377–378. IEEE, 2010.

[40] http://research.microsoft.com/en-us/downloads/29d28301-1079-4435-9810-74709376bce1/.

[41] http://www.sjzhuo.net/defocusestimation/.

[42] John MacCormick. How does the kinect work? *Presented at Dickinson College*, 6, 2011.