

DEPTH ESTIMATION: A MACHINE LEARNING BASED APPROACH

NIDHI CHAHAL



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
SEPTEMBER 2017

©Indian Institute of Technology Delhi (IITD), New Delhi, 2017

DEPTH ESTIMATION: A MACHINE LEARNING BASED APPROACH

by

NIDHI CHAHAL

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2017

Certificate

This is to certify that the thesis titled **Depth Estimation: A Machine Learning Based Approach** being submitted by **Ms. Nidhi Chahal** to the **Department of Electrical Engineering**, Indian Institute of Technology Delhi, for the award of **Doctor of Philosophy** is a record of bona-fide research work carried out by her under our guidance and supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Dr. Mona Mathur
Adjunct Faculty
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

Professor Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

To my family

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisors Professor Santanu Chaudhury and Dr. Mona Mathur for their continuous support and the stream of ideas that kept me occupied. Their expertise, vast knowledge, skills and patience added considerably to my doctoral experience. My regards to Prof. Santanu for his constant supervision which encouraged me in accomplishing my research work through all these years. I doubt that I will ever be able to convey my appreciation fully, but I owe him my eternal gratitude. Dr. Mona provided me with direction, technical support and became more of a friend and mentor, than a faculty.

I would like to thank rest of my thesis committee: Prof. S.D. Joshi, Dr. Sumantra Dutta Roy and Prof. Subhashis Banerjee for their appreciation, insightful comments and guidance. I would also like to thank my colleagues and friends for their support and encouragement through out my research work. All the discussions with them, exchange of knowledge, ideas and technical skills have been very helpful and motivational.

Finally, I must express profound gratitude to my parents who taught me good values that really matter in life and for their continuous mental support.

A very special thanks goes to my son for being such a sincere boy and always cheering me up. His love motivated me to work hard for providing him better future. Last, but not least, I am thankful to my husband for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of my entire research and writing this thesis. This accomplishment would not have been possible without support of my family. Thank you very much, everyone.

Abstract

The machine learning approach is explored in this thesis for depth estimation of two dimensional images involving various features and concepts. The depths are generated for stereo pairs using fixed point model which is obtained by learning ground truth depths of training data. For the combination of multiple cues, the depths from stereo and monocular cues are used as input features for the generation of prediction function which is the specialty of our approach for depth estimation. The accurate and reliable depths are predicted from proposed learning framework for new stereo pairs.

A novel approach of manifold learning is proposed for reliable depth estimation of single images. The deep learning process is applied which extracts CNN features from Caffe model which is available online. The features generated from this process are more authentic as compared to hand crafted features such as direct intensities of the image pixels. The complete model is trained using initial depths obtained from manifold and ground truth depths of the training data in a fixed point learning framework which gives refined and dense depth maps. The results are evaluated using various quantitative and visual measures and compared with ground truth depths of test images, if available.

For the test images of different feature distributions than training images, transfer learning is used in our work which is novel approach in the field of depth estimation. For this, manifold alignment is applied in which manifolds of source and target domains are projected to new space so that the difference in structures of both domains is reduced. After manifold mapping, the depths are generated from manifold learning and finally fixed point supervised learning process is used for prediction of refined depth maps.

सार

द्वि-आयामी के गहराई अनुमान के लिए इस शोध में मशीन सीखने की प्रक्रिया का पता लगाया गया है विभिन्न विशेषताओं और अवधारणाओं को शामिल करने वाली छवियां। गहराई स्टीरियो के लिए उत्पन्न होती है नियत बिन्दु मॉडल का उपयोग करके जोड़े जो प्रशिक्षण के आधार सच्चे गहराई सीखकर प्राप्त की जाती हैं जानकारी। कई संकेतों के संयोजन के लिए, स्टीरियो और मोनोकुलर संकेतों की गहराई होती है भविष्यवाणी समारोह की पीढ़ी के लिए इनपुट विशेषताओं के रूप में उपयोग किया जाता है जो कि हमारी विशेषता है। नए स्टीरियो जोड़े के गहराई अनुमान के लिए दृष्टिकोण प्रस्तावित से सटीक और विश्वसनीय गहराई का अनुमान लगाया गया है।

एकल के विश्वसनीय गहराई अनुमान के लिए कई गुना सीखने का एक नया दृष्टिकोण प्रस्तावित है इमेजिस। गहरी सीखने की प्रक्रिया को लागू किया जाता है जो कैफ मॉडल से सीएनएन सुविधाओं को निकालता है जो ऑनलाइन उपलब्ध है। इस प्रक्रिया से उत्पन्न सुविधाओं की तुलना में अधिक प्रामाणिक हैं जैसे कि छवि पिक्सेल की सीधा तीव्रता। तैयार की गई सुविधाओं को हाथ में लेना पूरा मॉडल प्रशिक्षण के कई गुना और जमीन सत्य गहराई से प्राप्त प्रारंभिक गहराई का उपयोग करके प्रशिक्षित किया गया है एक निश्चित बिन्दु सीखने के ढांचे में डेटा जो परिष्कृत और घने गहराई नक्शे देता है। परिणाम विभिन्न मात्रात्मक और दृश्य उपायों का उपयोग करके मूल्यांकन किया जाता है और जमीन की सच्चाई के साथ तुलना की जाती है परीक्षण छवियों की गहराई, यदि उपलब्ध हो।

प्रशिक्षण छवियों, हस्तांतरण सीखने की तुलना में विभिन्न फीचर वितरण की परीक्षण छवियों के लिए हमारे काम में प्रयोग किया जाता है जो गहराई अनुमान के क्षेत्र में उपन्यास है। इसके लिए, कई गुना संरेखण लागू होता है जिसमें स्रोत और लक्ष्य डोमेन के कई गुना नए प्रोजेक्ट किए जाते हैं अंतरिक्ष ताकि दोनों डोमेन के ढांचे में अंतर कम हो। मैनिफोल्ड मैपिंग के बाद, गहराई कई गुना सीखने से उत्पन्न होती है और आखिरकार निश्चित बिंदु पर्यवेक्षण शिक्षण होता है प्रक्रिया परिष्कृत गहराई नक्शे के पूर्वानुमान के लिए प्रयोग किया जाता है।

Contents

Certificate	i
Acknowledgments	v
Abstract	vii
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Scope and Objective	2
1.2 Major Contributions of the Thesis	3
1.3 Thesis Outline	5
2 3D Depth Estimation: A Review	7
2.1 Analysis of Depth Cues	7
2.1.1 Depth from Binocular Cues	8
2.1.2 Depth from Monocular Cues	8
2.1.3 Depth from 3D Sensors	10
2.2 Combination of Depth Cues	11
2.3 Motivation	12

3	Kinect and Depth Cues	15
3.1	Introduction	15
3.2	Kinect Depth Noise Removal	16
3.2.1	Image Registration	16
3.2.2	Algorithm	19
3.2.3	Results	21
3.3	Monocular and Motion Depth Cues	23
3.3.1	Depth From Defocus	23
3.3.2	Bokeh Effect	25
3.3.3	Depth from Motion	27
3.4	Fusion of Kinect and Depth Cues	28
3.4.1	Stereo Pair Generation	28
3.4.2	Final 3D Output	29
3.5	Conclusion	32
4	Depth Estimation of Stereo Pairs	33
4.1	Introduction	33
4.2	Proposed Approach	36
4.3	Depth Cues	38
4.3.1	Depth from Stereo Matching	38
4.4	Construction of Feature Matrix	40
4.5	Fixed Point Model	43
4.5.1	Contraction Mapping Condition	43
4.5.2	Fixed Point Learning for Depth Estimation	43
4.5.3	Multi-class SVM Learning	45
4.6	Experiments and Discussion	45
4.6.1	Cross Validation and Prediction Accuracy	47
4.6.2	Performance Evaluation	49

4.7	Conclusion	49
5	Combination of Stereo and Monocular Cues	51
5.1	Introduction	51
5.2	Depth and Feature Cues	55
5.2.1	Depth from Stereo Matching	57
5.2.2	Depth from Monocular Cue- Defocus	57
5.2.3	Feature Cues- Color, Edge	57
5.3	Feature Matrix	59
5.4	Depth Estimation using Fixed Point Learning	60
5.5	Experimental Results	61
5.5.1	Predicted Depths	61
5.5.2	Objective Measures	64
5.6	Conclusion	65
6	Depth Extraction from Single Image	67
6.1	Introduction	67
6.2	Manifold Learning	68
6.2.1	Linear Algorithms	69
6.2.2	Non Linear Algorithms	70
6.2.3	Laplacian Eigenmap	71
6.2.4	Depth Estimation using LLE	72
6.2.5	Experimental Results	75
6.3	Deep Learning	79
6.3.1	Algorithm	80
6.3.2	Reliable Features	82
6.3.3	Experimental Results	83
6.4	Fixed Point Supervised Learning	85
6.4.1	Feature Extraction	85

6.4.2	Construction of Feature Matrix	86
6.4.3	Overall Framework	87
6.4.4	Depth Estimation	88
6.4.5	Multi-Classification	88
6.4.6	Experimental Results	89
6.5	Conclusion	92
7	Transfer Learning for Depth Estimation	95
7.1	Introduction	95
7.2	Proposed Approach	97
7.3	Depth from Manifold and Fixed Point	99
7.3.1	CNN Features	99
7.3.2	Initial Depth Maps	101
7.3.3	Fixed Point Learning	101
7.4	Requirement for Transfer Learning	104
7.4.1	Solution: Manifold Alignment	109
7.5	Transfer Learning	109
7.5.1	Manifold Alignment	110
7.5.2	Algorithm	111
7.5.3	Feature Construction	112
7.5.4	Aligning Manifolds	112
7.6	Implementation Results	114
7.6.1	Final Depth Maps	114
7.6.2	Experimental Validation	118
7.6.3	Conclusion	118
8	Conclusions	121
8.1	Summary of the Contributions	121
8.2	Scope of Future Work	122

CONTENTS

xiii**Bibliography****125****Publications****137****Biography****139**

List of Figures

3.1	Image alignment of Kinect RGB and depth using intensity based registration . . .	19
3.2	Final depth after proposed hole fill algorithm for large gaps	22
3.3	Final depth after proposed hole fill algorithm for small gaps	22
3.4	An example of bokeh in a photograph	26
3.5	Depth from defocus using original image and bokeh image	27
3.6	Examples of 2D video frames from input videos	30
3.7	Initial depths from defocus and motion. Two depth images from motion cue for corresponding three input frames.	30
3.8	Conversion of 2D video frames to 3D using fusion of depth maps	31
3.9	Final depth map using fusion of Kinect, monocular and motion cues	31
3.10	Final 3D output for few video frames	32
4.1	Comparison of depths generated from MRF, SVM and fixed point learning . . .	36
4.2	Flowchart: Depth estimation using stereo cues in proposed fixed point learning algorithm	37
4.3	Feature extraction is shown by taking an example of training image from Mid- dlebury 2014 data set.	39
4.4	Testing error decreases with more contextual information	41
4.5	Training and testing error for different number of iterations	46
4.6	Depth Estimation 'Middlebury data set'	48

5.1	Flowchart: Depth estimation using combination of various features in proposed fixed point learning algorithm	52
5.2	Comparison of predicted depth maps using different features	53
5.3	Predicted depth maps using different features- The regions are marked to show difference in depth values for different features.	54
5.4	Comparison of different features using RMSE with respect to ground truth depth	56
5.5	Testing error for different features	56
5.6	Depth Estimation for data set 'Scenes'	62
5.7	An example of Sintel stereo data set- Ambush	62
5.8	Depth Estimation 'Temple data set'	63
5.9	Depth Estimation 'Ambush data set'	64
6.1	Left- An example of 3D data. Right- 2D representation, the data is embedded in two dimensions using LLE	71
6.2	Image divided into patches	74
6.3	Predicted Depth map for test image using Laplacian Eigen map and LLE algorithms	76
6.4	Predicted depth map using overlapping patches (OP)	77
6.5	Depth generated from manifold learning process- table	77
6.6	Depth generated from manifold learning process- Ballet	78
6.7	Depths from Make3D code [24] and Manifold Learning	78
6.8	Flowchart: The process of depth estimation using manifold learning, deep learning and fixed point learning	83
6.9	Comparison of predicted depths from manifold learning using different features	84
6.10	Comparison of predicted depth using different features for Make3D data set . .	84
6.11	Context span versus testing error	87
6.12	Comparison of predicted depth maps from ML (Manifold Learning) and FP (Fixed Point) with respect to GT (Ground Truth) depth	90

6.13	Comparison of predicted depth maps from ML (Manifold Learning) and FP (Fixed Point) with respect to GT (Ground Truth) depth. The depths are of different resolutions.	91
6.14	Comparison of predicted depth maps from Make3D code and FP (Fixed Point) learning with respect to GT (Ground Truth) depth.	91
6.15	Comparison of predicted depth maps from ML (Manifold Learning) and FP (Fixed Point) with respect to GT (Ground Truth) depth.	92
7.1	Block diagram: Depth estimation using proposed machine learning algorithms.	98
7.2	Results with manifold and fixed point learning. GT: Ground Truth, ML: Manifold Learning, FP: Fixed Point.	103
7.3	Training and testing images for Bayesian classification using Gaussian mixture model	105
7.4	Results for Bayesian classification using Gaussian mixture model. Mis-classified points are circled.	106
7.5	CNN visualization of train and test images. Histograms of three images in same plot.	107
7.6	Prediction results from GMM-EM algorithm for two test images	108
7.7	Depth prediction for test image 1 of similar feature distribution and test image 2 of different feature distribution than training images	109
7.8	An example of transfer learning approach. Comparison between graphs before and after manifold alignment.	111
7.9	Pictorial representation of joint manifold obtained using manifold alignment [80].	114
7.10	Comparison between graphs before and after manifold alignment	114
7.11	Depth prediction for Ambush test images	115
7.12	Predicted depths for test image from Make3D [24] data set	116
7.13	Depth prediction for test image from Desk data set.	117
7.14	Comparison of proposed approach with Eigen et al. NIPS 2014 [1]	119

List of Tables

3.1	Subjective quality measure	31
4.1	Features used for depth estimation	39
4.2	Cross validation accuracy	47
4.3	Sintel: Prediction accuracy	48
4.4	Comparison between depths from proposed learning and Middlebury code [19]	49
5.1	Features used for depth estimation	59
5.2	Quantitative Measure: PSNR (dB) of various depths with respect to ground truth depths	64
5.3	Quantitative Measure: RMSE of various depths with respect to ground truth depths	65
5.4	Quantitative Measure: SSIM of various depths with respect to ground truth depths	65
6.1	Performance evaluation using RMSE and PSNR. OP: Overlapping Patches . . .	77
6.2	Model Specification	82
6.3	Performance evaluation using RMSE and PSNR. HCF- Hand Crafted Features .	85
6.4	Performance evaluation of depths using RMSE, PSNR and SSIM	90
7.1	Performance measures of predicted depth maps	117
7.2	Performance measures of predicted depth maps	119

