Data-driven Depth Inference from a Single Still Image

Kyunghee Kim
Computer Science Department
Stanford University
kyunghee.kim@stanford.edu

Abstract

Given an indoor image, how to recover its depth information from one single image? This problem has been studied before for many years. But previous research mainly focused on using manually designed features, heuristics, or structure information. Lacking enough training data limits the methods that can be used to deal with this problem. However, with Kinect, it is now much cheaper to get ground truth depth information for indoor images. The purpose of this project is to use a lot of training data to obtain a more data-driven approach for recovering depth information given a single image.

Future Distribution Permission

The author(s) of this report give permission for this document to be distributed to Stanford-affiliated students taking future courses.

1. Introduction

Depth estimation from images and reconstruction of 3D structure of the images has been of interest to computer vision researchers for many years. Saxena et al. [1][2] used Markov Random Field (MRF) to model the depths and relation between depths at different parts of the image. Scharstein and Szeliski [3] produced a dense disparity map using two-frame stereovision. Torralba and Oliva [4] proposed a way to obtain the properties of the structure in the image from Fourier spectrum and infer the depth from this information. Saxena, Chung, and Ng [5] inferred depth from monocular image features. This project will use a MAP-MRF approach similar to [1], [2] and [6] and use massive amount of indoor images collected with Kinect [7] to infer the depth from a single image.

References

[1] Ashutosh Saxena, Sung Chung, and Andrew Ng. 3-D Depth Reconstruction from a Single Still Image. IJCV 2008.

[2] Ashutosh Saxena, Min Sun, Andrew Ng. Make3D: Learning 3D Scene Structure from a Single Still Image. PAMI 2008.

[3] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. IJCV 2002.

[4] Antonio Torralba and Aude Oliva. Depth Estimation from Image Structure. IEEE Trans Pattern Analysis and Machine Intelligence (PAMI), vol. 24, no. 9, pp.1-13, 2002.

[5] Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng. Learning Depth from Single Monocular Images. Neural Information Processing Systems (NIPS) 2005.

[6] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Joint Optimization of Segmentation and Appearance Models. ICCV 2009.

[7] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-Time Human Pose Recognition in Parts from a Single Depth Image. CVPR 2011

[8] James Hays and Alexei Efros. Scene Completion using Millions of Photographs. SIGGRAPH 2007.