APPEARANCE-BASED STEREO MATCHING

Güney F., Geiger A. - Max Planck Institute for Intelligent Systems, Tübingen, Germany
{fatma.guney, andreas.geiger}@tue.mpg.de

Abstract

Stereo matching methods tend to fail on planar or specular surfaces where little or ambiguous texture information is available. With the increasing availability of large annotated datasets like KITTI[1] or SINTEL[2], we ask a natural question: Can we condition the potentials in traditional stereo CRFs on contextual appearance information to improve binocular depth estimation? While current methods focus either on depth estimation from single images or on classical stereo matching, the combination of these two tasks has been little explored so far. In our work, we focus on combining binocular and monocular cues in an efficient framework and investigate the utility of various appearance features by regressing depth and surface normals using random forests.

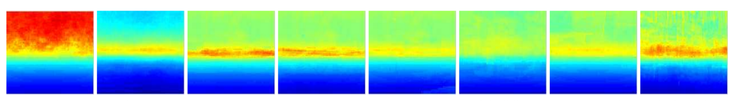
Motivation

Stereo images from the KITTI benchmark and
errors (white: >5px) made by the state-of-the-art methods.

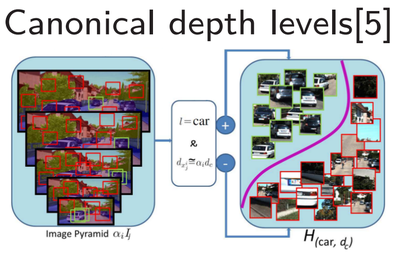
Related Work

Surface layout estimation[3]

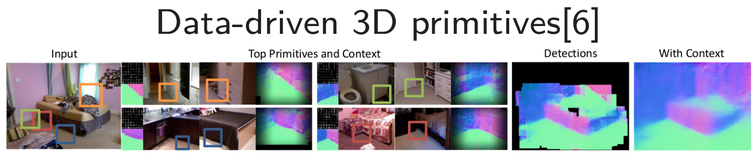
Class specific depth estimators[4]



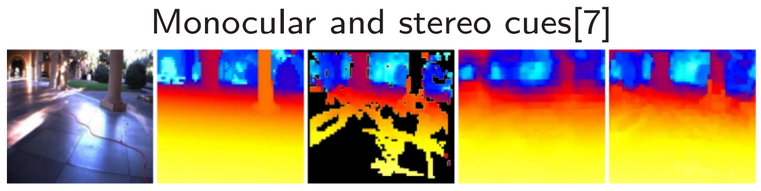
Canonical depth levels[5]



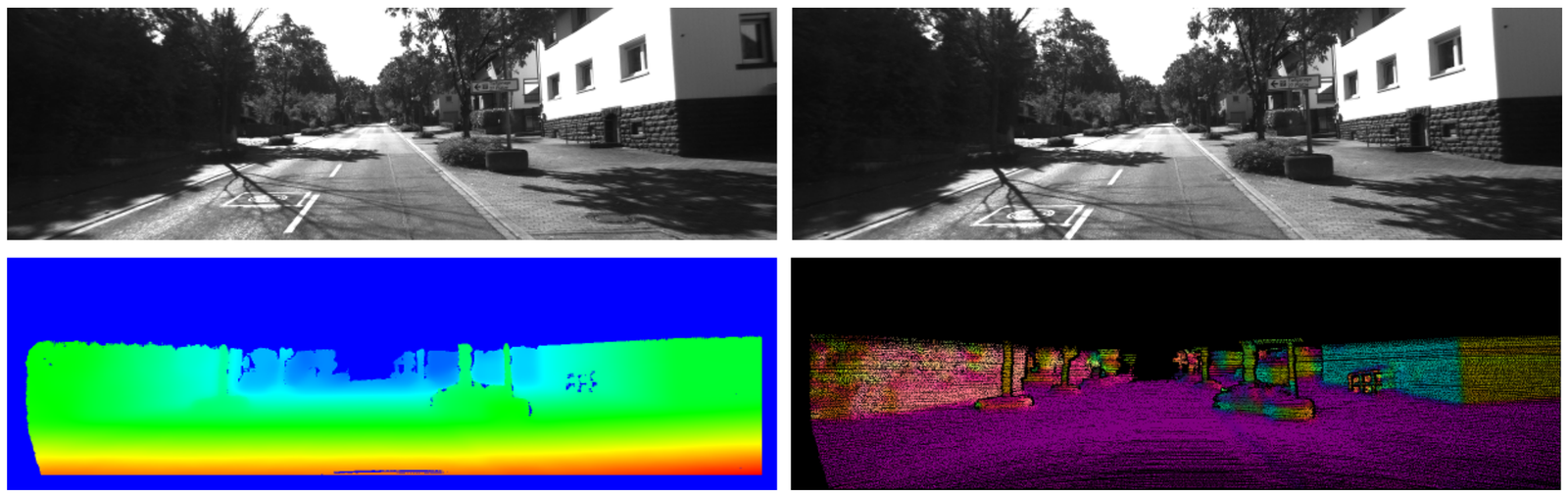
Data-driven 3D primitives[6]



Monocular and stereo cues[7]



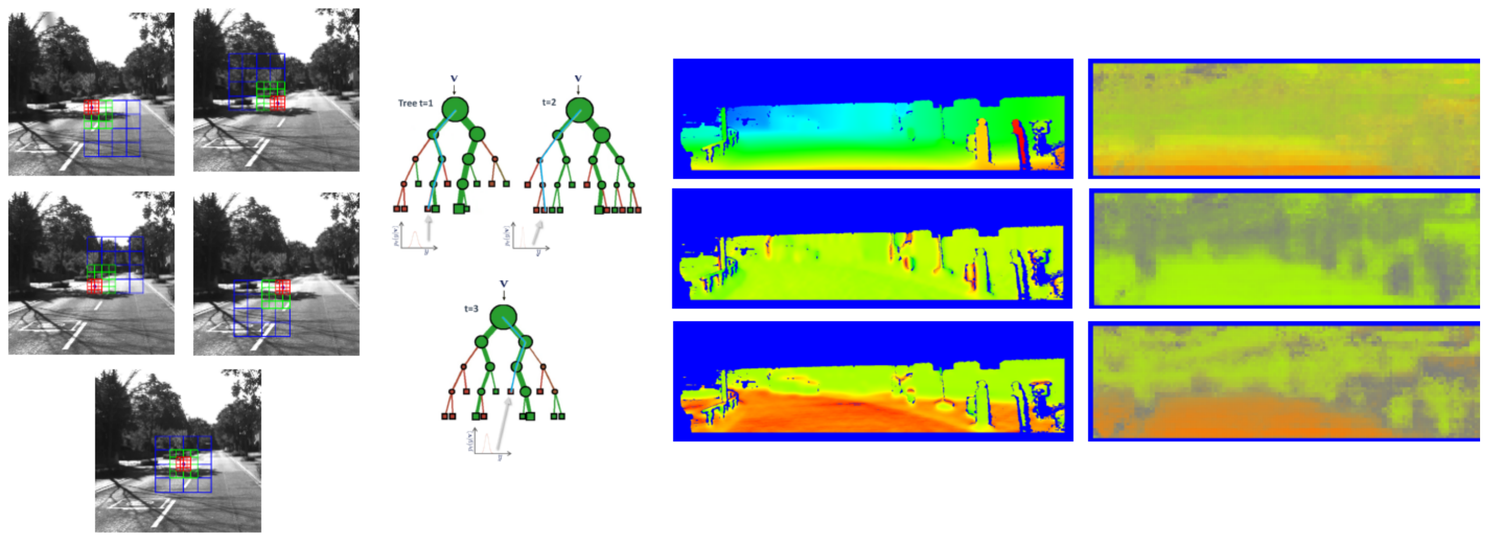
Problem Statement



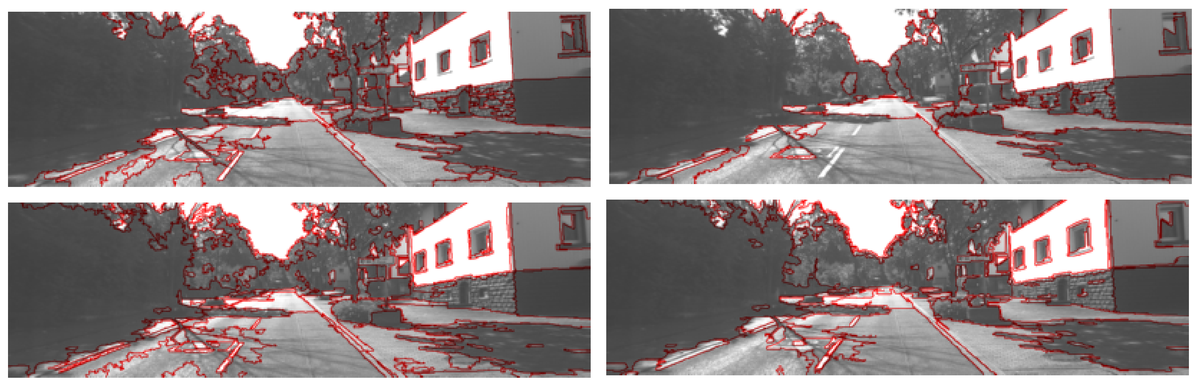
Top row: input images from left and right camera, bottom row: depth map and surface normals[1]

Approach

1. Random forest to regress gradients and disparity from HoG features



2. Random forest to regress surface normals from segment features



Features
Texton Boost
SIFT
ColourSIFT
LBP
Location

- Multiple segmentations of images
- Pixel-level features to segment features
  - Bag of Words from densely extracted features
  - Segment features as normalized histograms of words
- Additional segment features such as segment shape, vanishing points

Estimates can be added as an additional data term to the global energy:

$$E(x,y) = \sum_i E_{stereo}(x,y_i) + \sum_i E_{rf}(x,y_i) + \sum_{i,j} E_{smooth}(x,y_i,y_j)$$

x: image features, y: disparity or surface normals

References

[1] A., Geiger, P., Lenz, R., Urtasun, Are we ready for autonomous driving? The KITTI Vision Benchmark Suite, in *CVPR*, 2012
[2] D. J., Butler, J., Wulff, G. B., Stanley, M. J., Black, A naturalistic open source movie for optical flow evaluation, in *ECCV*, 2012
[3] D., Hoiem, A., Efros, M., Hebert, Recovering Surface Layout from an Image, in *IJCV*, 2007
[4] B., Liu, S. Gould, D., Koller, Single Image Depth Estimation from Semantic Labels, in *CVPR*, 2010
[5] L., Ladicky, J., Shi, M., Pollefeys, Pulling Things out of Perspective, in *CVPR*, 2014
[6] D. F., Fouhey, A., Gupta, M., Hebert, Data-Driven 3D Primitives for Single Image Understanding, in *ICCV*, 2013
[7] A., Saxena, J., Schulte, Y. Ng., Andrew. Depth Estimation using Monocular and Stereo Cues. in *IJCAI*, 2007