Int J Comput Vis (2012) 99:319–337
DOI 10.1007/s11263-010-0385-z

On Learning Conditional Random Fields for Stereo

Exploring Model Structures and Approximate Inference

Christopher J. Pal · Jerod J. Weinman · Lam C. Tran · Daniel Scharstein

Received: 22 March 2010 / Accepted: 15 September 2010 / Published online: 20 October 2010

Abstract Until recently, the lack of ground truth data has hindered the application of discriminative structured prediction techniques to the stereo problem. In this paper we use ground truth data sets that we have recently constructed to explore different model structures and parameter learning techniques. To estimate parameters in Markov random fields (MRFs) via maximum likelihood one usually needs to perform approximate probabilistic inference. Conditional random fields (CRFs) are discriminative versions of traditional MRFs. We explore a number of novel CRF model structures including a CRF for stereo matching with an explicit occlusion model. CRFs require expensive inference steps for each iteration of optimization and inference is particularly slow when there are many discrete states. We explore belief propagation, variational message passing and graph cuts as inference methods during learning and compare with learning via pseudolikelihood. To accelerate approximate inference we have developed a new method called sparse variational message passing which can reduce inference time by an order of magnitude with negligible loss in quality. Learning using sparse variational message passing improves upon previous approaches using graph cuts and allows efficient learning over large data sets when energy functions violate the constraints imposed by graph cuts.

Keywords Stereo · Learning · Structured prediction · Approximate inference

This research was supported in part by: a Google Research award, Microsoft Research through awards under the eScience and Memex funding programs, a gift from Kodak Research and an NSERC discovery award to C.P. Support was also provided in part by NSF grant 0413169 to D.S.

C.J. Pal (✉)
École Polytechnique de Montréal, Montréal, QC, Canada
e-mail: christopher.pal@polymtl.ca

J.J. Weinman
Dept. of Computer Science, Grinnell College, Grinnell, IA, USA
e-mail: weinman@grinnell.edu

L.C. Tran
Dept. of Electrical and Computer Engineering, University of California San Diego, San Diego, CA, USA
e-mail: lat003@ucsd.edu

D. Scharstein
Middlebury College, Middlebury, VT, USA
e-mail: schar@middlebury.edu

1 Introduction

In recent years, machine learning methods have been successfully applied to a large number of computer vision problems, including recognition, super-resolution, inpainting, texture segmentation, denoising, and context labeling. Stereo vision has remained somewhat of an exception because of the lack of sufficient training data with ground-truth disparities. While a few data sets with known disparities are available, until recently they had been mainly been used for benchmarking of stereo methods (e.g., Scharstein and Szeliski 2002). Our earlier work in this line of research (Scharstein and Pal 2007) sought to remedy this situation by replacing the heuristic cues used in previous approaches with probabilistic models for structured prediction derived from learning using real images and ground truth stereo imagery. To obtain a sufficient amount of training data, we used the structured-lighting approach of Scharstein and Szeliski (2003) to construct a database of 30 stereo pairs