Int J Comput Vis (2013) 105:269–297
DOI 10.1007/s11263-013-0639-7

Variational Recursive Joint Estimation of Dense Scene Structure and Camera Motion from Monocular High Speed Traffic Sequences

Florian Becker · Frank Lenzen · Jörg H. Kappes · Christoph Schnörr

Received: 26 May 2012 / Accepted: 8 June 2013 / Published online: 3 July 2013

Abstract We present an approach to jointly estimating camera motion and dense structure of a static scene in terms of depth maps from monocular image sequences in driver-assistance scenarios. At each instant of time, only two consecutive frames are processed as input data of a joint estimator that fully exploits second-order information of the corresponding optimization problem and effectively copes with the non-convexity due to both the imaging geometry and the manifold of motion parameters. Additionally, carefully designed Gaussian approximations enable probabilistic inference based on locally varying confidence and globally varying sensitivity due to the epipolar geometry, with respect to the high-dimensional depth map estimation. Embedding the resulting joint estimator in an online recursive framework achieves a pronounced spatio-temporal filtering effect and robustness. We evaluate hundreds of images taken from a car moving at speed up to 100 km/h and being part of a publicly available benchmark data set. The results compare favorably with two alternative settings: stereo based scene reconstruction and camera motion estimation in batch mode using multiple frames. They, however, require a calibrated camera pair or storage for more than two frames, which is less attractive from a technical viewpoint than the proposed monocular and recursive approach. In addition to real data, a synthetic sequence is considered which provides reliable ground truth.

Keywords Structure from motion · Variational approach · Recursive formulation · Dense depth map

Electronic supplementary material The online version of this article (doi:10.1007/s11263-013-0639-7) contains supplementary material, which is available to authorized users.

F. Becker (✉) · F. Lenzen
Heidelberg Collaboratory for Image Processing,
University of Heidelberg, Speyerer Str. 6,
69115 Heidelberg, Germany
e-mail: becker@math.uni-heidelberg.de
URL: http://hci.iwr.uni-heidelberg.de

F. Lenzen
e-mail: frank.lenzen@iwr.uni-heidelberg.de
URL: http://hci.iwr.uni-heidelberg.de

J. H. Kappes · C. Schnörr
Image and Pattern Analysis Group, University of Heidelberg,
Speyerer Str. 6, 69115 Heidelberg , Germany

J. H. Kappes
e-mail: kappes@math.uni-heidelberg.de
URL: http://ipa.iwr.uni-heidelberg.de

C. Schnörr
e-mail: schnoerr@math.uni-heidelberg.de
URL: http://ipa.iwr.uni-heidelberg.de

1 Introduction

1.1 Overview and Motivation

Computer vision research has a strong impact on driver assistance technology. Besides designing dedicated detectors for specific object classes (Enzweiler and Gavrila 2009; Gerónimo et al. 2010), current major trends include low-level estimation of dense scene structure from stereo sequences (Wedel et al. 2008), the transition to monocular imaging sensors (Weishaupt et al. 2010; Newcombe and Davison 2010), and context-based 3D scene representation and labeling supported by high-level assumptions and constraints (Wojek et al. 2010).

This paper focuses on the low-level task to jointly estimate dense scene structure and egomotion under minimal assumptions, adverse conditions and requirements, that are typical for driver assistance scenarios—see Fig. 1: