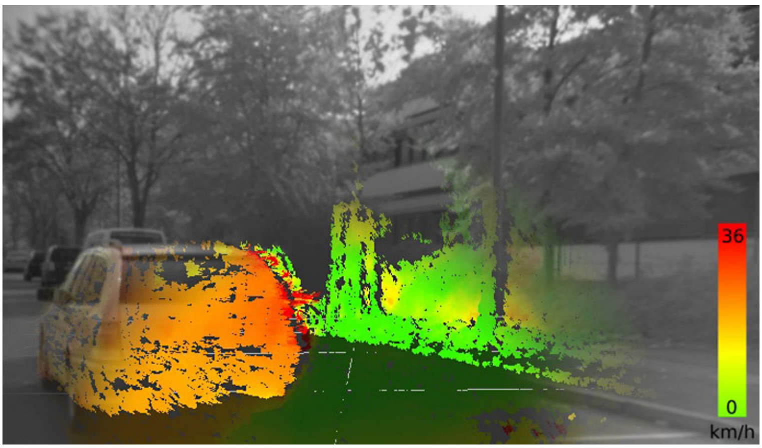
Andreas Wedel
Daniel Cremers

Stereo Sc
for 3D M

Stereo Scene Flow for 3D Motion Analysis


36
0
km/h

Andreas Wedel • Daniel Cremers

Stereo Scene Flow for 3D Motion Analysis

Springer

Dr. Andreas Wedel
Group Research
Daimler AG
HPC 050–G023
Sindelfingen 71059
Germany
andreas.wedel@daimler.com

Prof. Daniel Cremers
Department of Computer Science
Technical University of Munich
Boltzmannstraße 3
Garching 85748
Germany
cremers@tum.de

ISBN 978-0-85729-964-2 e-ISBN 978-0-85729-965-9
DOI 10.1007/978-0-85729-965-9
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2011935628

© Springer-Verlag London Limited 2011
Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.
The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.
The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Cover design: VTeX UAB, Lithuania

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The estimation of geometry and motion of the world around us from images is at the heart of Computer Vision. The body of work described in this book arose in the context of video-based analysis of the scene in front of a vehicle from two front-facing cameras located near the rear view mirror. The question examined of where things are in the world and how they move over time is an essential prerequisite for a higher-level analysis of the observed environment and for subsequent driver assistance. At the origin of this work is the combination of a strong interest in solving the real-world challenges of camera-based driver assistance and a scientific background in energy minimization methods. Yet, the methods we describe for estimating highly accurate optical flow and scene flow are a central prerequisite in other domains of computer vision where accurate and dense point correspondence between images or between geometric structures observed in stereo-videos is of importance.

Step by step we introduce variational methods which allow us to enhance the image data acquired from two cameras by spatially dense information on the geometric structure and 3D motion of the observed structures. In particular, we introduce variational approaches to optic flow estimation and present a variety of techniques which gave rise to the world's most accurate optic flow method. We introduce a variational approach to estimate scene flow, i.e. the motion of structure in 3D. We discuss metrics for evaluating the accuracy of scene flow estimates. We will also show extensions of scene flow, including flow-based segmentation and the tracking of 3D motion over multiple frames. The latter employs Kalman filters for every pixel of an image assuming linear object motion which results in a stable and dense 3D motion vector field.

The book is written for both novices and experts, covering both basic concepts such as variational methods and optic flow estimation, and more advanced concepts such as adaptive regularization and scene flow analysis.

Much of the work described in this book was developed during the Ph.D. thesis of the first author, both at the University of Bonn and at Daimler Research, Böblingen. Many of these results would not have been possible without the enthusiastic support of a number of researchers. We are particularly indebted to Uwe Franke, Clemens Rabe, and Stefan Gehrig for their work on 6D vision and disparity estimation, to

Thomas Pock in the context of efficient algorithms for optic flow estimation, to Thomas Brox in the parts on variational scene flow estimation, and to Tobi Vaudrey and Reinhard Klette for their research support on residual images and segmentation. We are grateful to our collaborators for their support.

With lane departure warning systems and traffic sign recognition, camera-based driver assistance is gradually becoming a reality. Latest research deals with intelligent systems such as autonomous evasive maneuvers and emergency situation takeover assistance. We hope that this book will help to lay the foundations for higher-level traffic scene understanding, object motion detection, and the development of advanced driver assistance.

Böblingen and Munich, Germany

Andreas Wedel
Daniel Cremers

Contents

List of Notations

$\mathbb{R}$	Real numbers
Ω	Image domain
Ψ_ε	Differentiable approximation of the absolute function $\lvert x\rvert \approx \Psi_\varepsilon(x) = \sqrt{x^2 + \varepsilon^2}$
$\mathbf{F}$	Fundamental matrix
∇I	Spatial gradient of I: $\nabla I(x, y, t) = (I_x, I_y)^\top$
I_x, I_y, I_t	Partial derivatives with respect to x, y and t
$\mathcal{N}$	Local neighborhood of a pixel
$\mathcal{N}_4$	4-connected neighborhood
$\mathcal{L}$	Pixel labelling

Chapter 1
Machine Vision Systems

Everything used to measure time really measures space.
J. Deshusses

Accurate, precise and real-time capable estimation of three-dimensional motion vector fields remains one of the key tasks in computer vision. Different variants of this problem arise inter alia in the estimation of ego motion [4], object motion [8], human motion [77], and motion segmentation [106]. The knowledge of the surrounding motion field is a key enabler for a wide range of applications such as driver assistance systems and modern surveillance systems. Especially in security relevant applications robustness, accuracy, and real-time capability are of utmost importance.

Estimating this three-dimensional motion vector field from stereo image sequences has drawn the attention of many researchers. Due to the importance of this problem, numerous approaches to image based motion field estimation have been proposed in the last three decades. Most of them can be classified into the following main strategies:

- model based approaches,
- sparse feature tracking methods using multiple image frames,
- dense scene flow computation from two consecutive frames.

A. Wedel, D. Cremers, *Stereo Scene Flow for 3D Motion Analysis*,
DOI 10.1007/978-0-85729-965-9_1, © Springer-Verlag London Limited 2011

The estimation of motion vectors involves both the reconstruction of the three-dimensional scene via stereo matching and the solving of a point correspondence problem between two or more consecutive images. Both problems are classical *ill-posed* problems[1] in the sense that merely imposing matching of similar intensities will typically not give rise to a unique solution. The three aforementioned strategies choose different ways to overcome this ill-posedness.

In model based approaches such as [77] and [8] parameterized models of objects or humans are used to constraint the solution space and overcome the ill-posedness of the problem. However, the absence of appropriate models for generic applications disqualifies model based approaches in a multitude of situations.

Many researchers therefore circumvent specific object models and employ regularization techniques for feature tracking and scene flow approaches in order to formulate the motion estimation problem in a well-posed way. This regularization is either formulated in the time domain for the tracking of features, as done in [54] or [74], or in the spatial domain, imposing smoothness of the motion field between two consecutive frames like in [103] and [108].

The latter is known as variational scene flow estimation from stereo sequences. Algorithmically variational scene flow computation methods build up on the seminal optical flow algorithm of Horn and Schunck [42]. In what is often considered the first variational method in computer vision, Horn and Schunck suggested to compute the flow field between two consecutive images of a video as the minimizer of an energy functional which integrates a brightness constancy assumption with a smoothness assumption on the flow field. This framework has been improved in [60] to cope with flow discontinuities and outliers and in [13] to cope with large flow vectors. In recent years, several real-time optical flow methods have been proposed—see for example [16] and [119]. In Chaps. 2 and 3 we review the classical optic flow estimation and discuss a series of improvements [91, 102, 105–107, 110], including median filtering of flow vectors, decomposition of the input images, and considering optical flow estimation as an iterative refinement process of a flow vector field accompanied by outlined implementation details.

We continue in Chap. 4 to introduce scene flow estimation as an extension of the optical flow estimation techniques. Joint motion and disparity estimation for the scene flow computation was introduced in [69]. In [108] the motion and disparity estimation steps were decoupled in order to achieve real-time capability without loosing accuracy. Subsequent publications have focused on improving the accuracy, the formulation of uncertainties, and establishing motion metrics for scene flow [105, 109, 112] and are handled in Chap. 5. Implementation details on the scene flow algorithm are found in the Appendices of this book.

We additionally include Chap. 6 on recent developments in the research field of scene flow estimation described in [75, 111]. These include scene flow segmentation and using Kalman filters for scene flow estimation. The latter approach

[1] Following Hadamard [37], a mathematical model is called *well-posed* if there exists a solution, if the solution is unique and if it continuously depends upon the data. Otherwise it is called *ill-posed*.

combines the aforementioned strategies of feature tracking over multiple frames and scene flow computation from two consecutive frames. This allows a dense and robust reconstruction of the three-dimensional motion field of the depicted scene.

Chapter 2
Optical Flow Estimation

Abstract In this chapter we review the estimation of the two-dimensional apparent motion field of two consecutive images in an image sequence. This apparent motion field is referred to as optical flow field, a two-dimensional vector field on the image plane. Because it is nearly impossible to cover the vast amount of approaches in the literature, in this chapter we set the focus on energy minimization approaches which estimate a dense flow field. The term *dense* refers to the fact that a flow vector is assigned to every (non-occluded) image pixel. Most dense approaches are based on the variational formulation of the optical flow problem, firstly suggested by Horn and Schunk. Depending on the application, density might be one important property besides accuracy and robustness. In many cases computational speed and real-time capability is a crucial issue. In this chapter we therefore discuss the latest progress in accuracy, robustness and real-time capability of dense optical flow algorithms.

Space is a still of time, while time is space in motion.
Christopher R. Hallpike

2.1 Optical Flow and Optical Aperture

This chapter is about optical flow and its estimation from image sequences. Before we go into details on optical flow estimation, we review the definition of optical flow and discuss some of the basic challenges of optical flow estimation.

A. Wedel, D. Cremers, *Stereo Scene Flow for 3D Motion Analysis*,
DOI 10.1007/978-0-85729-965-9_2, © Springer-Verlag London Limited 2011