Int J Comput Vis (2013) 101:6–21
DOI 10.1007/s11263-012-0542-7

Multi-view Scene Flow Estimation: A View Centered Variational Approach

Tali Basha · Yael Moses · Nahum Kiryati

Received: 7 September 2010 / Accepted: 23 May 2012 / Published online: 14 June 2012

Abstract We present a novel method for recovering the 3D structure and scene flow from calibrated multi-view sequences. We propose a 3D point cloud parametrization of the 3D structure and scene flow that allows us to directly estimate the desired unknowns. A unified global energy functional is proposed to incorporate the information from the available sequences and simultaneously recover both depth and scene flow. The functional enforces multi-view geometric consistency and imposes brightness constancy and piecewise smoothness assumptions directly on the 3D unknowns. It inherently handles the challenges of discontinuities, occlusions, and large displacements. The main contribution of this work is the fusion of a 3D representation and an advanced variational framework that directly uses the available multi-view information. This formulation allows us to advantageously bind the 3D unknowns in time and space. Different from optical flow and disparity, the proposed method results in a nonlinear mapping between the images' coordinates, thus giving rise to additional challenges in the optimization process. Our experiments on real and synthetic data demonstrate that the proposed method successfully recovers the 3D structure and scene flow despite the complicated nonconvex optimization problem.

Keywords 3D structure · Scene flow · Multiple view

An earlier version of part of this work appear in CVPR 2010 (Basha et al. 2010).

T. Basha (✉) · N. Kiryati
School of Electrical Engineering, Tel Aviv University, Tel Aviv, 69978, Israel
e-mail: talib@eng.tau.ac.il

Y. Moses
Efi Arazi School of Computer Science, The Interdisciplinary Center, Herzliya 46150, Israel

1 Introduction

The structure and motion of objects in a 3D space is an important characteristic of dynamic scenes. Reliable 3D motion maps can be utilized in many applications, such as surveillance, tracking, dynamic 3D scene analysis, autonomous robot navigation, 3D display devices, or virtual reality. In the last decade, an emerging field of research has addressed the problem of *scene flow* computation. Scene flow is defined as a dense 3D motion field of a nonrigid 3D scene (Vedula et al. 1999). It follows directly from this definition that 3D surface recovery must be an essential part of any scene flow algorithm, unless it is given *a priori*. Our objective is to simultaneously compute the 3D structure and scene flow from a multi-camera system. The system consists of N calibrated and synchronized cameras with overlapping fields of view. A unified variational framework is proposed to incorporate the information from the available sequences and simultaneously recover both depth and scene flow. To describe our method, we next elaborate on the parametrization of the problem, the integration of the spatial and temporal information from the set of sequences, and the setting of a global energy functional together with the variational framework used for solving it.

Most existing methods for scene flow and surface estimation parameterize the problem in 2D rather than 3D (e.g., Zhang and Kambhamettu 2000, 2001; Vedula et al. 2005; Isard and MacCormick 2006; Min and Sohn 2006; Huguet and Devernay 2007; Wedel et al. 2008; Li and Sclaroff 2008; Pock et al. 2008). That is, they compute disparity (stereo), which is the projection of the desired 3D shape, and the optical flow, which is the projection of the 3D motion (Fig. 1b). The relation between the scene flow and its projection is presented in Fig. 1a. Assuming that reliable and consistent solutions of both stereo and optical flow are given, the scene