Int J Comput Vis
DOI 10.1007/s11263-013-0647-7

Closed-Form Solution of Visual-Inertial Structure from Motion

Agostino Martinelli

Received: 31 July 2012 / Accepted: 25 July 2013

Abstract This paper investigates the visual-inertial structure from motion problem. A simple closed form solution to this problem is introduced. Special attention is devoted to identify the conditions under which the problem has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions and infinite solutions depending on the trajectory, on the number of point-features and on their layout and on the number of camera images. The investigation is also performed in the case when the inertial data are biased, showing that, in this latter case, more images and more restrictive conditions on the trajectory are required for the problem resolvability.

Keywords Sensor fusion · Structure from motion · Inertial sensors · Robotics

1 Introduction

The structure from motion problem (SFM) consists of determining the three-dimensional structure of the scene by using the measurements provided by one or more sensors over time (e.g. vision sensors, ego-motion sensors, range sensors). In the case of visual measurements only, the SFM problem has been solved up to a scale Chiuso et al. (2002), Davison et al. (2007), Hartley (1997), Christopher Longuet-Higgins (1981), Nistér (2004) and a closed form solution has also been derived Hartley (1997), Christopher Longuet-Higgins (1981), Nistér (2004), allowing the determination of the three-dimensional structure of the scene, without the need for any prior knowledge.

A. Martinelli (✉)
INRIA Rhone Alpes, Grenoble, France
e-mail: agostino.martinelli@inria.fr

The case of inertial and visual measurements, i.e., the visual-inertial SFM problem (from now on the Vi-SfM problem), has particular interest and has been investigated by many disciplines, both in the framework of computer science Bryson and Sukkarieh (2008), Jones and Soatto (2011), Kelly and Sukhatme (2011), Martinelli (2012), Strelow and Singh (2004) and in the framework of neuroscience (e.g., Berthoz et al. 1975; Dokka et al. 2011; Fetsch et al. 2010). Prior work has answered the question of which are the observable modes, i.e. the states that can be determined by fusing visual and inertial measurements Bryson and Sukkarieh (2008), Jones and Soatto (2011), Kelly and Sukhatme (2011), Martinelli (2012). Specifically, it has been shown that the velocity, the absolute scale, the gravity vector in the local frame and the bias-vectors which affect the inertial measurements, are observable modes. On the other hand, the problem of determining these observable modes is not fully solved.

The majority of the approaches so far introduced, perform the fusion of vision and inertial sensors by filter-based algorithms. In Armesto et al. (2007), these sensors are used to perform egomotion estimation. The sensor fusion is obtained by an Extended Kalman Filter (EKF) and by an Unscented Kalman Filter (UKF). The approach proposed in Gemeiner et al. (2007) extends the previous one by also estimating the structure of the environment where the motion occurs. Also, in Veth and Raquet (2007) an EKF has been adopted. In this case, the proposed algorithm estimates a state containing the robot speed, position and attitude, together with the inertial sensor biases and the location of the features of interest. In the framework of airbone SFM, an EKF has been adopted in Kim and Sukkarieh (2007) to solve the Vi-SfM problem. It was observed that any inconsistent attitude update severely affects any SFM solution. The authors proposed to separate attitude update from position and velocity update. Alterna-