Acta Astronautica 123 (2016) 435–445

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Comparative assessment of techniques for initial pose estimation using monocular vision

Sumant Sharma*, Simone D'Amico

Department of Aeronautics & Astronautics, Stanford University, Stanford, California, USA

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form
12 December 2015
Accepted 23 December 2015
Available online 20 January 2016

Keywords:
Computer vision
Pose estimation
Monocular vision
Uncooperative spacecraft

ABSTRACT

This work addresses the comparative assessment of initial pose estimation techniques for monocular navigation to enable formation-flying and on-orbit servicing missions. Monocular navigation relies on finding an initial pose, i.e., a coarse estimate of the attitude and position of the space resident object with respect to the camera, based on a minimum number of features from a three dimensional computer model and a single two dimensional image. The initial pose is estimated without the use of fiducial markers, without any range measurements or any apriori relative motion information. Prior work has been done to compare different pose estimators for terrestrial applications, but there is a lack of functional and performance characterization of such algorithms in the context of missions involving rendezvous operations in the space environment. Use of state-of-the-art pose estimation algorithms designed for terrestrial applications is challenging in space due to factors such as limited on-board processing power, low carrier to noise ratio, and high image contrasts. This paper focuses on performance characterization of three initial pose estimation algorithms in the context of such missions and suggests improvements.

1. Introduction

Recent advancements have been made to utilize monocular vision navigation as an enabling technology for formation-flying and on-orbit servicing missions (e.g., PROBA-3 by ESA [1] , ANGELS by US Air Force [2], PRISMA by OHB Sweden [3]). These missions require approaching a passive space resident object from large distances (e.g., > 30 km) in a fuel efficient, safe, and accurate manner. Simple modification of low cost instruments (e.g., star trackers) for high dynamic range can enable accurate navigation relative to the space resident object. Monocular navigation on such missions relies on finding an estimate of the initial pose, i.e., the attitude and position of the space resident object with respect to the camera, based on a minimum number of features from a three dimensional computer model and a single two dimensional image. For on-orbit servicing missions, this represents the scenario where the servicing spacecraft is "lost-in-space". Estimating the initial pose is especially critical as well as challenging in the design of a pose estimation system as there is no a priori information about the attitude and position of the target. Aside from a 3D wire-frame model of the space resident object, no assumption on the relative translational or rotational information is made.

Use of state-of-the-art computer vision techniques designed for terrestrial applications is challenging in space. For example, use of feature descriptors such as the Scale Invariant Feature Transform (SIFT) [4] in pose estimation for space imagery is too computationally expensive and yields poor results (see Fig. 1). We plot SIFT feature matches (in green) between two images taken during the

* Corresponding author. Tel.: +1 404 952 8102.
E-mail address: sharmas@stanford.edu (S. Sharma).

http://dx.doi.org/10.1016/j.actaastro.2015.12.032
0094-5765/

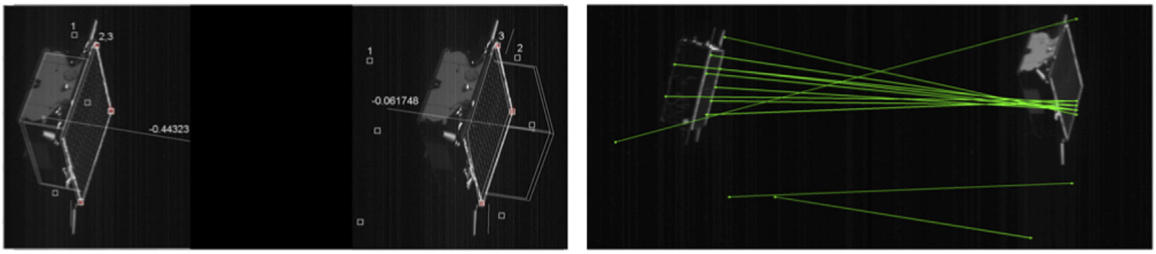

Fig. 1. Challenges in initial pose estimation using state-of-the-art techniques: pose ambiguity due to geometry (left) and poor feature matching (right).

PRISMA mission. To obtain correct feature matches, SIFT relies on high image acquisition rates and low image noise, both of which are unavailable in these images. Additionally, spacecraft geometry is often highly symmetrical resulting in ambiguity in attitude determination (see Fig. 1). Prior work has been done to compare different initial pose estimators [5,6], but there is a lack of functional and performance characterization of such algorithms in the context of missions involving rendezvous operations in the space environment.

For the purpose of this comparative assessment, we focus on a pose estimation architecture (see Fig. 2) based on points extracted from edge features due to their simplicity and accuracy [7]. As opposed to features based on color gradients, textures, and optical flow, edges are less sensitive to illumination changes and can easily distinguish boundaries of a spacecraft geometry from the background in an image.

Source of the 3D model features is a simplified wireframe model of the space resident object, assumed to be either stored on-board the servicing spacecraft or formed as a structure-from-motion problem which is solved alongside pose estimation [8]. Source of the 2D features is an object detection subsystem which processes, extracts, and describes features from a single image captured by the on-board navigation camera. A typical model reduction procedure and object detection subsystem are described and illustrated later in the paper.

The 3D model features and 2D features are passed into the initial pose estimation subsystem which generates a pose estimate without knowing the correspondence of these features. This is a challenging task without apriori estimates of the relative motion due to a large search space for the correct feature correspondence. However, the search space can be reduced using a method such as perceptual organization [9,10] which detects viewpoint-invariant feature groupings from the 2D image. These are then matched to corresponding structures of the 3D model in a probabilistic manner to create multiple correspondence hypotheses. These correspondence hypotheses need to be validated to find the correct one. Hence, for each correspondence hypothesis, n number of 2D and 3D features are used to calculate a relative pose by solving the Perspective-n-Point Problem (PnP) [11]. The resulting pose estimate is used to create virtual 2D image features by reprojecting the input 3D model features using true perspective projection. A measure of the reprojection error between the virtual 2D image features and the input 2D image features is used in validation. This process is repeated for all hypotheses in order to identify the correct feature correspondence and subsequently, a correct pose estimate. Hence, our interest is not in the PnP solvers' statistical use of a large number of measurements to solve an overdetermined systems. Rather, it is in their ingenuity in using a minimal number of points to estimate a coarse initial pose estimate with a minimal computational effort.

The performance of initial pose estimation hinges on the solution of the PnP problem. In the above architecture, a PnP solver could be called multiple times and will be subject to a wide variety of input from the object detection subsystem and 3D model reduction. Hence, a PnP solver should not only be fast and efficient but also more importantly be reliable and robust to overcome challenges unique to monocular vision-based navigation in space. In the remainder of the paper, we first present a formal problem statement of the PnP problem and then review state-of-the-art PnP solvers. We then introduce our framework of assessment of these solvers where a discussion of simulation input generation, performance criteria, and test cases is presented. Finally, we present relevant results from our assessment and conclude with a discussion on applicability of these solvers in a monocular vision-based navigation system for on-orbit servicing and formation flying missions.

2. Review of solution methods

Let $q_i = [x_i\ y_i\ z_i]^T$, where $i = 1, 2, \ldots, n$, be n 3D model points in the object reference framework B. Let $p_i = [u_i\ v_i\ 1]^T$, where $i = 1, 2, \ldots, n$, be the corresponding n image points in the image reference framework P. For a known camera focal length, f, the PnP problem aims to retrieve the rotation matrix from frame B to the camera reference framework C, R_{BC}, and the translation vector from the origin of frame B to the origin of frame C, T_{BC}:

$$p_i = [u_i\ v_i\ 1]^T = \left[\frac{\alpha_i}{\gamma_i} f\ \frac{\beta_i}{\gamma_i} f\ 1\right]^T$$

$$r_i = [\alpha_i\ \beta_i\ \gamma_i]^T = R_{BC}(q_i + T_{BC}) \tag{1}$$

Re-writing Eq. (1) using homogeneous coordinates and representing camera parameters as a matrix K, we would like to estimate the 3×4 pose matrix, P, whose first three