Robotics and Autonomous Systems 62 (2014) 414–421

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

An integrated stereo visual odometry for robotic navigation

Haleh Azartash [a,*], Nima Banai [b], Truong Q. Nguyen [a]

[a] Electrical and Computer Engineering, University of California, San Diego 9500 Gilman Drive, La Jolla, CA 92093, United States

[b] Aerospace and Mechanical Engineering, University of Southern California, 3650 McClintock Ave, Los Angeles, CA 90089, United States

HIGHLIGHTS

- We propose a method to estimate the arbitrary motion of a stereo rig very accurately.
- We account for different type of motion, side motion, forward motion and rotation motion.
- The proposed method is a feature based method that can estimate very large motion.
- The true motion parameters (velocity and acceleration) are obtained in presence of noise from a jerky robot motion.
- On the contrary to the existing method, we estimated the exact translational motion as opposed to up to a scale factor.

ARTICLE INFO

Article history:
Received 20 April 2013
Received in revised form
2 November 2013
Accepted 22 November 2013
Available online 10 December 2013

Keywords:
Visual odometry
Projective stereo camera
Translation vector
Rotation matrix
Kalman filter

ABSTRACT

In this paper, we propose a novel method to accurately estimate the arbitrary motion of a calibrated stereo rig from a noisy sequence. In the proposed method, a projective camera model is used which is appropriate for scenes where the objects are close to the camera or where there is depth variation in the scene. We propose a feature-based method that estimates large 3D translation and rotation motion of a moving rig. The translational velocity and acceleration and angular velocity of the rig are then estimated using a recursive method. In addition, we account for different motion types such as pure rotation and pure translation in different directions. In our studies, we assume that the rig motion is noisy, i.e., the acceleration and velocity of the camera are not perfectly constant. Our experimental results show that we obtain accurate estimates of rotation matrix and translation vector parameters across different test-cases with large and small baselines. For long sequences, the estimated motion parameters are within ±0.2 mm.

1. Introduction

Camera motion estimation, also known as ego-motion estimation, is an important application of computer vision that could be applied to autonomous robot and vehicle navigation. Wheel odometry and sensors such as gyroscopes are commonly used to navigate robots. However, if the surface where robot is moving through is rough or slippery, the wheels might get stuck or slip. Therefore, these approaches are not very reliable. On the other hand, ego-motion estimation (visual odometry) is much more robust and is not affected by the aforementioned issues. Visual navigation and localization for mobile robots has been researched for over three decades. Visual navigation methods can be roughly divided in two categories, map-based and mapless systems. Map-based systems such as [1] require prior knowledge of the surrounding environment whereas the mapless systems acquire information about the environment as the robot navigates through it. For several applications such as underwater, aerial and planetary expedition, obtaining prior knowledge of the environment is not possible, therefore, having a reliable mapless navigation and localization system is highly desirable. Among various methods for mapless visual odometry, feature tracking has become very popular due to its robustness and accuracy. In order to increase the accuracy of the ego-motion estimation, the use of stereo rig instead of a monocular rig has become very common at the cost of increasing computation time. Using a stereo rig provides additional information such as depth that can help solve motion estimation problem without the issue of scale factor. Hence, in the proposed method, a calibrated stereo camera is used to recover the depth information to eliminate the scale factor dependency of the camera motion estimation.

In this paper, the focus is on mapless navigation with projective camera model for a long noisy sequence. Therefore, the proposed method is most useful for ground and underwater robots due to the fact that in the projective camera model, close objects and depth variation do not affect the accuracy of the results. In addition, since the motion of the robot is calculated from camera motion, no prior information of the surrounding environment is needed and the

* Corresponding author.
E-mail address: hazartas@ucsd.edu (H. Azartash).

0921-8890/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2013.11.008

condition of the ground (slippery) is not of significant importance. On the other hand, to estimate the motion parameters of the robot such as velocity and acceleration, a robust noise removing procedure is needed, especially if the robot has jerky motion. The major advantage of the proposed method is that there are no prior assumptions regarding the camera motion. Motion estimation techniques for each motion model such as translational in *x*- and *z*-direction and pure rotation are very different, therefore, the applicability of motion estimation algorithms based on a specific motion model is rather limited.

The proposed method consists of two main parts, frame-by-frame motion estimation and motion parameter estimation (velocity and acceleration) for a long noisy sequence of images. The frame-by-frame motion estimation for two stereo pair images taken at times t and $t+1$ is as follows:

1. Feature detection for both stereo pair images.
2. Feature matching between the left images of each stereo pair.
3. Estimating the relative pose between the left images and finding the inlier matches.
4. Determining the motion type and direction of the camera.
5. Finding the exact translation motion of the camera.

Once the frame-by-frame motion is obtained, the motion parameters are computed using Kalman filter [2]. The proposed method is very robust to errors caused both by false matches (using RANSAC [3] in Step (3)) and noisy camera motion from uneven surface (through the use of Kalman filter). As a result of Steps (3) and (4), the proposed method can estimate very large motions in different directions robustly. Also, the motion type (pure rotation and pure translation in *x*- or *z*-directions) is automatically determined in Step (4). In addition, in Step (5) only the inlier matches from stereo matching and inlier matches from Step (3) are used to remove the false matches. Therefore, feature matching becomes very robust and the chance of having false matches decreases dramatically.

Other important characteristics of the proposed algorithm are: (1) the feature descriptor is invariant to rotation and scaling, therefore, large rotation and motion along the optical axis are covered, (2) the scene is assumed to be stationary and (3) the underlying true motion of the camera can be recovered as the output of the Kalman filter. In this work, we propose an integrated ego-motion estimation framework that finds the translation and rotation of a moving stereo rig over long noisy sequences. Our contributions are as follows.

- We propose an accurate frame-by-frame motion estimation method that finds the true translation vector as opposed to other methods that are up-to a scale factor.
- We find the motion parameters of a long sequence using a recursive estimator that accurately finds the correct motion parameters in the presence of both measurement and system noise.
- In our experiments, we use an unstable moving stereo rig to emulate the real-world environment where the robot motion is not ideal, e.g., the acceleration might not be constant. Our experimental results show significant accuracy improvement versus the existing ego-motion estimation methods.

In addition to [4], in this paper, we have added further details on the proposed algorithm and have also improved the experimental setup. In the previous setup, the stereo rig was mounted on a moving rig (on a rail) and the DSLR camera was too heavy for the rig, hence the stereo rig was unstable which resulted in blurry images. Furthermore, to verify our proposed method, we only used different views of a multiview setup from Middlebury database [5] as reference and current frames. In this paper, we have verified the camera motion for both side and frontal motions at different step sizes as described in Section 5. Using our new stereo setup, we can guarantee constant velocity whereas in the previous rig, due to the fluctuations of the power of the rig's motor, the velocity was sinusoidal.

In this paper, we assume that the features are already given and tracked through time and hence, feature detection and traction are not the focus of this paper. The remainder of the paper is organized as follows. Previous work is described in Section 2. Section 3 describes the problem formulation. Proposed methods for frame-by-frame motion estimation and motion parameter estimation are described in Section 4. Section 5 presents our experimental setup and results. Finally, Section 6 concludes the paper.

2. Related work

Visual navigation and localization for mobile robots has been researched for over three decades [6]. Ego-motion estimation has been extensively studied for both monocular [7,8] and stereo cameras [9–15]. Mapless motion estimation techniques can generally be classified into two categories (1) feature-based [14,13,12,11,10,16–19] and (2) flow-based [9,20,21] methods. Feature-based methods extract features such as corners, lines and curves and track them in both spatial (for stereo cameras) and temporal domains. Flow-based methods treat the intensity of the image as a function of time, $I(x, y, \delta t)$, and evaluate the image based on differential changes in the intensity image. However, the assumption is that the inter-frame motion is very small.

A popular feature-based method for robot navigation is the singular value decomposition (SVD) [22] method where the main assumption is to use an orthographic camera model. However, if the objects are close to the camera, this assumption no longer holds; hence, the applicability of this method is very limited. Several papers use Harris corner detection technique [23] to achieve real-time performance. However, since Harris corner detection is not scale and rotation invariant, if the robot is moving towards the objects or rotates, Harris corner detection fails and produces erroneous results [14,13]. In ego-motion estimation, if the camera motion is noisy (i.e., a moving rig with noisy acceleration in a bumpy environment), there will be process noise. Another issue in ego-motion estimation for long sequences is to establish a reference coordinate to accommodate the propagation of rotation in consecutive frames because the translation and rotation of all the consecutive frames must be in the same direction. In the proposed method, we assume that the coordinate of the first reference frame is the overall reference coordinate and estimate the translation and rotation of all frames with respect to the assumed reference coordinate. Table 1 shows the comparison between the proposed and the previous methods. As shown in Table 1, the proposed method can estimate the true motion parameters (not up-to a scale factor) even when the input sequence is noisy. Although other methods claim that they have achieved real-time performance, they have sacrificed accuracy for speed. In addition, if the robot is moving towards the object, having a scale invariant feature detection technique is crucial [14]. Furthermore, the perspective camera model is only suitable for scenes where the objects are far away from the camera which limits the applicability of the method proposed in [12]. In [16], even though they have accounted for feature noise, they fail to account for camera noise. For the case where the road is bumpy, the method in [16] cannot estimate the true underlying motion of the camera.

3. Problem formulation

The problem formulation consists of two parts, measurement and state equations. The measurement equation which we refer to as frame-by-frame motion estimation computes the motion of