Abstract—Accurate positioning of moving vehicles in GNSS-deprived urban areas is important for autonomous vehicles and mobile mapping systems. In this paper, we discuss various approaches to vehicle positioning in the absence of GNSS signals. We explore the potential of visual-inertial odometry, and present experimental results that show visual-inertial odometry can provide positioning accuracies up to 0.9% of the trajectory length, which is promising for vehicle positioning in short periods of GNSS signal outage.

Keywords—localization; INS; IMU; camera pose estimation; autonomous vehicles; feature tracking; Kalman filter

I. INTRODUCTION

Global Navigation Satellite System (GNSS) is an essential technology for positioning in mobile mapping systems and future self-driving vehicles. Using the Real Time Kinematic (RTK) technique centimeter level positioning accuracies can be achieved [1]. However, in densely built up areas, urban canyons and tunnels, GNSS signals are not always available. In such areas, accurate and reliable vehicle positioning becomes a challenge.

Inertial sensors, which are typically integrated with GNSS, can provide position estimates in the absence of GNSS signals. However, positioning by an Inertial Navigation System (INS) is characterized by drift, i.e. the positioning errors accumulate and grow large over time. This makes the INS unsuitable for positioning even in short periods of GNSS signal loss.

Various solutions for vehicle positioning in GNSS-deprived environments have been proposed in the literature. Many of these solutions are based on optical cameras and Lidar sensors. While cameras and Lidar sensors are principally used as mapping sensors in mobile mapping systems and autonomous vehicles, they can also be used to estimate the position of the vehicle when GNSS signals are not available. In this paper, we provide an overview of methodologies for vehicle positioning based on visual and Lidar sensors. We explore the potential of visual-inertial odometry, and present a positioning method based on the integration of an INS with a single or stereo camera system. We present experimental results that demonstrate the potential of visual-inertial odometry for vehicle positioning in the absence of GNSS signals.

The paper proceeds with a review of two main approaches to vehicle positioning without GNSS in Section II. Section III presents our method for positioning based on visual-inertial odometry. Experiments and results are discussed in Section IV. The paper concludes with some remarks in Section V.

II. VEHICLE POSITIONING WITHOUT GNSS

Different approaches to vehicle positioning without GNSS can be classified into two main categories: i) local motion estimation, where the vehicle’s change of position is estimated with respect to a previous position; and ii) global position estimation, where the vehicle’s position is estimated directly in a global coordinate frame. The following sections discuss the two positioning approaches in detail.

A. Local Motion Estimation

In local motion estimation, the motion of the sensor is estimated rather than the position. From the estimated motion the pose (position and attitude) of the sensor in the coordinate frame of the previous pose can be obtained. If motion estimation starts at a point with known position in a global reference frame, all estimated poses can be transformed to the global reference frame.

Wheel odometry and dead reckoning are probably the simplest methods for local motion estimation. In wheel odometry, rotary encoders measure the rotation of the wheels from which an estimate of the motion can be obtained. In dead reckoning, the change of position is estimated from the distance and direction of motion, which can be measured for instance by an INS.

The main drawback of local motion estimation methods, including wheel odometry and dead reckoning, is the accumulation of errors. As position estimates are obtained incrementally from local motion estimates, errors in consecutive motion estimation steps accumulate and grow large over time causing the estimated trajectory to drift from the actual trajectory. More advanced approaches based on visual and Lidar sensors try to reduce and, possibly, eliminate the drift.

Visual odometry first proposed by Nister et al. [2] is a popular approach to motion estimation using one or more cameras. In a typical visual odometry approach, salient image features are detected and matched between pairs of frames.
relative poses between the frames are estimated by using the matched feature points, and these are then refined by using 3D points reconstructed from the image features. Because of this incremental pose estimation, visual odometry also suffers from trajectory drift, although the magnitude of the drift is smaller than inertial dead reckoning and wheel odometry.

Simultaneous localization and mapping (SLAM) is another popular approach to motion estimation, which aims to reduce or eliminate trajectory drift. In a typical visual SLAM algorithm [3] salient features are extracted and matched between pairs of images to construct a map of the environment and estimate the sensor pose with respect to the map. Although in the SLAM approach, the position of the sensor is estimated with respect to the map, the procedure for constructing the map itself is incremental, resulting in error accumulation and trajectory drift. State of the art SLAM algorithms such as ORB-SLAM [4] and RGB-D SLAM [5-7] detect loop closures and apply pose graph optimization [8, 9] to improve the global consistency of the map and correct the trajectory drift. However, in the context of vehicle positioning, loop closing and trajectory correction with some delay is not practical.

Recently, Lidar sensors are increasingly used in SLAM systems to generate the map and simultaneously estimate the sensor motion [10, 11]. In Lidar-based SLAM, motion estimation is done by registering consecutive laser scans [12, 13], which is generally more accurate than motion estimation using matched image features. In Google and Ford’s driverless cars Lidar sensors will be used for detecting road obstacles and navigation-related objects, but also for position estimation in the absence of GNSS signals [14]. The disadvantage of Lidar sensors is their high cost for use in vehicles.

B. Global Position Estimation

In global position estimation, the position of the vehicle is estimated directly in a global reference coordinate frame. GNSS provides global position estimates. When GNSS signals are not available, global positioning can be achieved by matching sensor observations acquired on the vehicle with a source of georeferenced spatial data.

Georeferenced aerial images and orthoimages have been most commonly used for georeferencing of mobile mapping data and position estimation of moving vehicles. In [15, 16] aerial images are matched with ground-based images to georeference a mobile mapping vehicle. In [17, 18] aerial images are used to georeference Lidar data acquired on a moving vehicle. In [19] a Lidar reflection map of the road environment is generated and then used for estimating the position of the vehicle by matching the map with Lidar scans acquired on the vehicle.

The advantage of global position estimation is that positioning errors do not accumulate. However, obtaining accurate position estimates requires accurately georeferenced spatial data. In addition to aerial images, other sources of spatial data, such as oblique images (e.g., Pictometry), 2D maps [20] and 3D city models can also be used for vehicle position estimation independent of GNSS. With increasing availability and improved accuracy of such data in urban areas, the global position estimation approach seems to be a more appropriate solution, which has not been sufficiently explored. Nonetheless, in most urban streets the GNSS signal loss occurs for short periods only. In such situations, local motion estimation methods have the potential to provide sufficiently accurate position estimates with limited drift.
In stereo MSCKF, image features are tracked along each sequence, but also matched across the stereo pairs. Only successfully matched features that are tracked in a pre-defined number of successive frames are used to update the state. In theory, this increases the robustness of the pose estimation against tracking errors. Fig. 2 illustrates the main steps of stereo MSCKF. In the propagation step, the state vector containing the sensor pose, velocity and biases, and its covariance are propagated for each INS measurement. In the augmentation step, for each new image the state and the covariance matrix are augmented with the current camera pose. Note that only pose parameters of one camera are added to the state, while the pose of the other camera is determined using calibration parameters. In the update step, for each image feature that is matched across the two cameras and tracked in a pre-defined number of frames a 3D position is calculated. The back projection of this 3D point to the images gives a vector of residuals that is linearized with respect to the state parameters and the feature position. The filter update is then applied based on this linearized measurement model.

IV. EXPERIMENTS

To evaluate the performance of the visual-inertial odometry approach in mono or stereo mode an experiment was conducted using KITTI datasets 0005 and 0020 [22]. Dataset 0005 contains 154 pairs of images captured on 26 September 2011. Dataset 0020 consists of 330 pairs of images captured on 30 September 2011. The images in both datasets have the same size of 1392 x 512 pixels, and were recorded at approximately 0.1 second intervals. The baseline between the two cameras is 54 cm. Moreover, the datasets included inertial measurements obtained by an OXTS RT3003 inertial and GPS navigation system at a measurement rate of 100 Hz. The RTK GPS measurements serve as the ground truth in these datasets.

To assess the accuracy of the estimated trajectories, we used translational RMSE defined as the Root Mean Square of the three translation errors at each time step. The translational error at each time step was calculated from the estimated position and the ground truth. The total translational error was measured as the Euclidean norm of all RMSEs.

Fig. 3 shows the trajectories obtained by the mono and stereo MSCKF compared with INS/GPS integration and the ground truth. While MSCKF in both modes performs much better than inertial positioning (INS/GPS), stereo MSCKF provides a more accurate trajectory than mono MSCKF in both datasets.

Fig. 4 shows the cumulative distribution of translational errors for mono and stereo MSCKF as compared with INS/GPS. The performances brought here are for the optimum camera pose window. Stereo MSCKF performs better than the other methods and reaches an average accuracy of 0.9% of the trajectory length compared to 1.3% and 2.5% for the mono and INS/GPS integration respectively. Note that the drift of INS/GPS is quite low because KITTI inertial data was obtained by a precise INS (OXTS RT3003), and has been filtered for biases and integration errors.
Table I summarizes the performance of mono and stereo MSCKF compared with INS/GPS in terms of Average RMSE and Total translation error over the entire sequence of frames for the tested KITTI datasets.

<table>
<thead>
<tr>
<th></th>
<th>Dataset 0005</th>
<th>Dataset 0020</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INS/GPS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average RMSE (m)</td>
<td>0.97</td>
<td>3.65</td>
</tr>
<tr>
<td>Total Trans. Error (m)</td>
<td>3.11</td>
<td>9.90</td>
</tr>
<tr>
<td><strong>Mono MSCKF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average RMSE (m)</td>
<td>0.85</td>
<td>1.44</td>
</tr>
<tr>
<td>Total Trans. Error (m)</td>
<td>2.44</td>
<td>2.76</td>
</tr>
<tr>
<td><strong>Stereo MSCKF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average RMSE (m)</td>
<td>0.70</td>
<td>1.22</td>
</tr>
<tr>
<td>Total Trans. Error (m)</td>
<td>2.14</td>
<td>2.25</td>
</tr>
</tbody>
</table>

V. CONCLUDING REMARKS

In this paper, we discussed various approaches to vehicle positioning in the absence of GNSS signals. We explored the potential of the visual-inertial odometry approach using a multi-state constraint Kalman filter. The results show that visual-inertial odometry gives more accurate position estimates than the inertial system. The stereo visual-inertial odometry results in a smaller drift in the estimated trajectory compared to the mono mode.

Although the visual-inertial odometry method presented here reduces the trajectory drift it cannot eliminate the drift completely. In practice, the use of visual-inertial odometry is recommended for vehicle positioning only in short periods of GNSS signal loss.

In future work we explore global position estimation by matching georeferenced geometries from a 3D city model with features extracted from images captured by an omnidirectional camera.

ACKNOWLEDGMENT

This work was supported by an Australian Government Research Training Program Scholarship.

REFERENCES


