Visual-Inertial Motion Priors for Robust Monocular SLAM

Usman Qayyum and Jonghyuk Kim

School of Engineering, Australian National University, Canberra, Australia
{usman.qayyum,jonghyuk.kim}@anu.edu.au

Monocular visual SLAM approaches are mostly constrained in their performance due to general motion model and availability of true scale information. We proposed an approach which improves the motion prediction step of visual SLAM and results in better estimation of map scale. The approach utilizes the short term accuracy of inertial velocity with visual orientation to estimate refined motion priors. These motion priors are fused with sparse number of 3D map features to constraint the positional drift of moving platform. Experimental results are presented on large scale outdoor environment, yielding robust performance and better observability of map scale by monocular SLAM.

Visual and inertial sensors are natural candidates for SLAM research whereas availability of GPS depends upon operating environment. We have presented an approach to estimate robust motion priors from visual orientation and accelerometer based velocity to improve the EKF-SLAM filter consistency with observable scale. The recent work to provide efficient priors to SLAM filter with sparse set of features is conducted by [1], in which stereo based system is used whereas the focus of our research is monocular cameras. Various visual-inertial SLAM approaches proposed so far, work in limited scenarios i.e. known feature locations or stereo cameras having limited working environment due to its predetermined baseline. In monocular-inertial SLAM, the work of [2] presented their results on small scale hand-held environment while maintaining full Inertial Measurement Unit (IMU) states whereas our approach is directed to large scale environment relying on relative inertial velocity and visual orientation as motion priors to EKF-SLAM filter.

The prime limitation with the dead-reckoning sensors on long term navigation is the unconstrained drift in position and orientation due to integration errors whereas the short term accuracy and high throughput makes them an ideal choice for integrated localisation applications with other sensors. Inertial sensors are comprised of tri-axial gyros and accelerometers, arranged orthogonally to each other and provide 6DOF measurements of the attached platform/body. We have used the relative velocity information of accelerometer instead of integerated position to generate the translational magnitude in navigation frame (an internal IMU complimentary filter estimates the on line biases whereas gyro compensation is applied with subtraction of gravity vector). The inertial velocity magnitude provides the translation offset of the vehicle in metric space whereas angular orientation between the consecutive images is provided by adopting the

R. Groß et al. (Eds.): TAROS 2011, LNAI 6856, pp. 430–431, 2011.

work of [3], in which a dense motion estimation algorithm compares the consecutive image arrays by Average Absolute intensity Difference (AAD). The motion estimates from accelerometer are preferred over the visual estimates as visual approaches reveal the translation in arbitrary scale. The visual-inertial relative motion priors provide the predictive stage to the EKF SLAM-filter which maintains 3D features to constrainst the long-term inertial drift.

Current research on sequential monocular SLAM has shown that it performs the drift free tracking by matching the sparse set of map features to the image features. The visual-inertial motion priors estimated earlier are provided to sequential; EKF-SLAM filter [4], which helps in better prediction of feature locations and results in efficient generation of map. The feature map consists of **n** number of inverse depth features and visual matching is based upon patch correlation. Each observed/re-observed feature imposes the constraint on camera location by using the correspondence of image-to-map feature. Each observed feature imposes the constraint on vehicle/camera location by using the correspondence of image to map features. The innovation vector between the predicted feature position and visually matched observation is used to calculate the innovation vector, for EKF-SLAM filter updation.

An experimental evaluation is presented on a dataset collected at campus by placing the sensors (IMU/camera) on testing Van for a traveled distance of approx. 1.0 Km. To fully evaluate the proposed approach, a comparison is performed with IMU integrated-only, visual odometry [3] and Monocular SLAM [4] approaches. The result of the presented approach shows improved and more stable performance by observing the metric scale whereas inertial drift is minimized by inverse depth features maintained in EKF-SLAM filter.

The proposed work presents the use of short term relative velocity magnitude with visual orientation to provide refine priors to EKF-SLAM approach. The use of inverse depth feature provide the positional constraint on the traveled distance. The proposed approach is evaluated on large scale outdoor sequence showing a more stable and consistent performance of monocular SLAM in observing the map scale.

References

1. Alcantarilla, P., Bergasa, L., Dellaert, F.: Visual odometry priors for robust EKF-SLAM. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 3501–3506. IEEE, Los Alamitos (2010)
2. Pinies, P., Lupton, T., Sukkarieh, S., Tardos, J.D.: Inertial Aiding of Inverse Depth SLAM using a Monocular Camera. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 2797–2802. IEEE, Los Alamitos (2010)
3. Milford, M., Wyeth, G.: Single Camera Vision-Only SLAM on a Suburban Road Network. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 3684–3689. IEEE, Los Alamitos (2008)
4. Civera, J., Davison, A.J., Montiel, J.: Inverse Depth Parametrization for Monocular SLAM. IEEE Trans. on Robotics 24(5), 932–945 (2008)