

# Visual-inertial sensor fusion via machine learning

Habib Boloorchil<sup>a</sup>, He Bai<sup>b</sup> and Christopher Crick<sup>c</sup>

<sup>1</sup>Department of Computer Science, Oklahoma State University, My Street, MyTown, MyCountry

<sup>2</sup>Department of Computing, Main University, MySecondTown, MyCountry  
{f\_author, s\_author}@ips.xyz.edu, t\_author@dc.mu.edu

**Keywords:** The paper must have at least one keyword. The text must be set to 9-point font size and without the use of bold or italic font style. For more than one keyword, please use a comma as a separator. Keywords must be titlecased.

**Abstract:** Robot localization is a well-studied problem, with many competing solutions, from odometry to inertial measurement to SLAM to GPS and other beacon-based approaches. Some methods work only with certain modalities: measurement of wheel odometry is useless for legged or flying locomotion, for instance. Visual landmark detection and dead reckoning are useful to robots with many different structures and means of locomotion, and can be implemented with very low-cost cameras and inertial measurement units (IMUs). However, existing approaches require fairly sophisticated hardware with precise, rapid, real-time control loops in order to fuse the visual inertial odometry data adequately. We present algorithms to better handle the uncertainty which stems from noisy, inconsistent IMU and machine vision data, using a machine learning approach to provide a robot better awareness of its location and the effect of its self-motion on visual cues. This approach yields robust localization performance which generalizes across very different robot platforms using low-cost sensing and computation.


## 1 INTRODUCTION


Data fusion in Robotics has a huge and wide spectral of fields. The aim of most scientists in this area is helping robots to be more autonomous than before. One of the puzzles in this area is Navigation and localization is its bottleneck because a robot needs to have an estimation its state. In other words, having consciousness on where it is, will be a must for decision on next steps. In our research, we found flexibility in data fusion for the process of localization baffling. Not only, Concise timing is matter but we also found lack of a Software that help us handle devices that can do localization which does not required expert, cost and energy. This necessity to consistency and having experienced in mechanical and electronic engineers and an expensive device lead us to create a device that can work better with less consistency in data fusion. The demand to having data fusion stems from the role of Visual-inertial odometry. Visual-inertial odometry needs to gather data from both perception and Inertial Measurement Unit(IMU). This Odometry can es-


timate Coordinate in 3 Dimensional world, which Opticounters and GPS can only do it in 2 Dimension, when corresponding data is received in a consistent latency.

The Perceptio part is added to inertial odometry to be more Robust to uncertainty. Vision of localization can handle the situation in a variety of environments such as GPS unfriendly places and inertial can overcome situations such as dark or low textured, which could not let the camera has good perception. These challenges shows up when we use visual inertial odometry:

- Combining data that comes from IMU and camera should offer the least latency and be so accurate. and if not algorithms could not work well.
- Even if we could create a device that works accurately calibration each time we use it in order to have good results would cost so much time.
- Cost of stable Visual inertial odometry sensor is so high
- Most of the Visual inertial Odometry applications are not enough user-friendly for calibration and regular use. you need to be an expert in Control to know the terms.

<sup>a</sup>  <https://orcid.org/0000-0000-0000-0000>

<sup>b</sup>  <https://orcid.org/0000-0000-0000-0000>

<sup>c</sup>  <https://orcid.org/0000-0000-0000-0000>

In order to have more sense of these deficiencies, we can come up with an example. In Agriculture Departments, we have a big population of plants that we need to know the effect of each variable on plants. These kinds of research need to observe every plant and this can be tested by taking Photo of each plant. Visual inertial odometry can help us to localize a self-announcement mobile vehicle that can pull the trigger of shutter. Now it comes more baffling when we need an expert in control who is also a scientist in agriculture. We offered a machine learning approach that can resolve the combining data, expenses, and requirement of an expert. We get data to do the preprocessing that does not need the information of the camera or cameras for the perception. It can work with every visual-inertial sensors. These sensors can be cheap because the learning procedure is robust to noises. In our method we does perception using essential matrix then we take advantage of Random Forest Regression method to estimate the Position of the robot.

## 2 RELATED WORKS

When it comes to reduce the uncertainty, localization has a vital role (?). In other words, Odometry is using motion data comes from sensors in order to reduce the uncertainty of robots position in environment(Huang et al., 2017; Valencia and Andrade-Cetto, 2018).This data can come from rotation of wheels, GPS, IMU(Inertial Measurement Units), Cameras. In large-scale Domains GPS (Global Positioning System) is one of the most popular for localization but it has some draw backs in smaller space. For example it does not have enough accuracy and also there is some places it is not feasible to use it. For instance, under water, as a GPS-denied circumstance, we can realize that we need a robocentric approach for other ways for odometry (Saska et al., 2017). IMU(Inertial Measurement Units) is one of the tools that can be helpful in approximation of the trajectory (Gui et al., 2015). However, when we use IMU some problems as a matter of noisy data and motor noises on devices can be appeared. Furthermore, we need to have trajectory in circumstances as an example Mars Rovers specifically we cannot use it alone.(?) Rotary encoders that can estimate wheels rotation with getting pulse from optocounters is a technique for wheeled robots while this instrument cannot be useful for highly dynamic cases such as bipedal robots or in three dimensional spaces for flying robots. (Bloesch et al., 2015). The human inspired visual ego-motion can be counted as a liable approach to extract trajectory(Engel et al., 2018).

This Motion estimation, which apply machine vision, is called Visual Odometry. One of the most application of Visual Odometry is SLAM (Simultaneous Localization and Mapping) (Mur-Artal et al., 2015; Mur-Artal and Tardós, 2017; Forster et al., 2017; Mueggler et al., 2017). One of the problems that ORB-SLAM1(Mur-Artal et al., 2015) ,as an example of best Visual Odometry shows is drifting stems from lack of having closed loop algorithms.(Mur-Artal and Tards, 2017). Another, disadvantage that can lead to get lost as a robot is having fast angular movements that can lead to miss of landmarks. Since these landmarks are critical sources to estimate the position of our path, robot can lose its position. Thus, vision need to be aided by IMU. In addition, the integration of IMU can offers some benefits(Leutenegger et al., 2015). accuracy can be named as one of these advantages(Bloesch et al., 2017; Schneider et al., 2018; Sun et al., 2018).

We can integrate images and Inertial data via variety of algorithms. Using probability to estimate the better likelihood of location(Bowman et al., 2017) is one of the approaches to estimate the position. However, It is more popular to do this fusion can be done by extended Kalman filter(Lynen et al., 2013). The advantage of Extended Kalman Filter and Uncented Kalman Filter from Kalman filter is using Jacobian to extract the position of the robot in nonlinear system (Julier and Uhlmann, 2004; Wan and Van Der Merwe, 2000).

The major difference between UKF and EKF is their ability to estimate in the nonlinear system. In other words EKF is more for weak nonlinearity system while UKF is for high nonlinearity system. (St-Pierre and Gingras, 2004)

Bloesch et al (Bloesch et al., 2015; Bloesch et al., 2017)mars-rover-slam offers Iterated Extended Kalman Filter in order to robust the ability of estimation of trajectory. Iterations can robust the ability to find the maximum likelihood to estimate the position of a robot by robocentric formula (Bloesch et al., 2015; Bloesch et al., 2017).

To do landmark detection, Bloesch et al. employed Shi-Tomasi algorithm (Shi and Tomasi, 1993). In this method, landmarks is extracted by utilization of three times down sampling and more probable landmarks between candidates will be chosen. ROVIO (Robust Visual Inertial odometry) uses these landmarks and by warping of land marks the movement in perception can be estimated.

Other innovations can conceive the ego-motion (Zhou et al., 2017) using unsupervised learning. To understand the motion. This algorithm uses Convolutional Neural Network to find out the warping.

As we have three dimensional views using RGB-d or Stereo cameras we can both interpret angular and linear movements.

Although, There are several datasets that contains ground truth in 6 degree of freedom that shows angles and also position of the a flying robot and a get data from inertial sensors and stereo cameras(?), in this article, we want to make sure that we process data real-time and we also want to use our algorithm on flying robots. so we suggested a creation of device that able to get inertial and perception data relatively. we also suggested a creation of anew device to collect a data to have relative image and IMU data and having less delay with getting th data.

our perception is based on the aforementioned Rovio (Bloesch et al., 2017), and integrate it to IMU data. Instead of using iterated Extended Kalman Filter, we are going to take advantage of supervised Learning to create trajectory. Although, our method cannot show the result as accurately as Rovio, we offer benefit of having more robustness to latency inconsistency in devices such as the one show in Figure ?? which has received in a progress without pattern that we see in Figure 1.

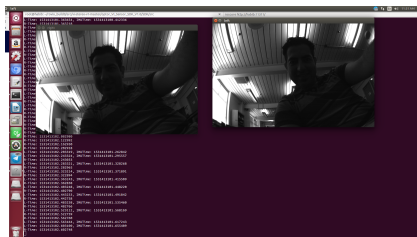


Figure 1: Data that received by our Visual inertial sensor.

## REFERENCES

- Bloesch, M., Burri, M., Omari, S., Hutter, M., and Siegwart, R. (2017). Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research*, 36(10):1053–1072.
- Bloesch, M., Omari, S., Hutter, M., and Siegwart, R. (2015). Robust visual inertial odometry using a direct ekf-based approach. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 298–304. IEEE.
- Bowman, S. L., Atanasov, N., Daniilidis, K., and Pappas, G. J. (2017). Probabilistic data association for semantic slam. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 1722–1729. IEEE.
- Engel, J., Koltun, V., and Cremers, D. (2018). Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625.
- Forster, C., Zhang, Z., Gassner, M., Werlberger, M., and Scaramuzza, D. (2017). Svo: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2):249–265.
- Gui, P., Tang, L., and Mukhopadhyay, S. (2015). Mems based imu for tilting measurement: Comparison of complementary and kalman filter based data fusion. In *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on*, pages 2004–2009. IEEE.
- Huang, A. S., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D., and Roy, N. (2017). Visual odometry and mapping for autonomous flight using an rgb-d camera. In *Robotics Research*, pages 235–252. Springer.
- Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422.
- Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., and Furgale, P. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334.
- Lynen, S., Achtelik, M. W., Weiss, S., Chli, M., and Siegwart, R. (2013). A robust and modular multi-sensor fusion approach applied to mav navigation. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 3923–3929. IEEE.
- Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., and Scaramuzza, D. (2017). The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 36(2):142–149.
- Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D. (2015). Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163.
- Mur-Artal, R. and Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262.
- Mur-Artal, R. and Tards, J. D. (2017). Visual-inertial monocular slam with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796–803.
- Saska, M., Baca, T., Thomas, J., Chudoba, J., Preucil, L., Krajník, T., Faigl, J., Loianno, G., and Kumar, V. (2017). System for deployment of groups of unmanned micro aerial vehicles in gps-denied environments using onboard visual relative localization. *Autonomous Robots*, 41(4):919–944.
- Schneider, T., Dymczyk, M., Fehr, M., Egger, K., Lynen, S., Gilitzenski, I., and Siegwart, R. (2018). maplab: An open framework for research in visual-inertial mapping and localization. *IEEE Robotics and Automation Letters*, 3(3):1418–1425.
- Shi, J. and Tomasi, C. (1993). Good features to track. Technical report, Cornell University.
- St-Pierre, M. and Gingras, D. (2004). Comparison between the unscented kalman filter and the extended kalman filter for the position estimation module of an integrated navigation information system. In *IEEE Intelligent Vehicles Symposium*, pages 831–835. Citeseer.

- Sun, K., Mohta, K., Pfrommer, B., Watterson, M., Liu, S., Mulgaonkar, Y., Taylor, C. J., and Kumar, V. (2018). Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters*, 3(2):965–972.
- Valencia, R. and Andrade-Cetto, J. (2018). *Mapping, planning and exploration with Pose SLAM*. Springer.
- Wan, E. A. and Van Der Merwe, R. (2000). The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee.
- Zhou, T., Brown, M., Snavely, N., and Lowe, D. G. (2017). Unsupervised learning of depth and ego-motion from video. In *CVPR*, volume 2, page 7.