Using Optical Flow as Lightweight SLAM Alternative

Gabriele Bleser* Gustaf Hendeby†

Augmented Vision
German Research Center for Artificial Intelligence

ABSTRACT

Visual simultaneous localisation and mapping (SLAM) is since the last decades an often addressed problem. Online mapping enables tracking in unknown environments. However, it also suffers from high computational complexity and potential drift. Moreover, in augmented reality applications the map itself is often not needed and the target environment is partially known, *e.g.* in a few 3D anchor or marker points. In this paper, rather than using SLAM, measurements based on optical flow are introduced. With these measurements, a modified visual-inertial tracking method is derived, which in Monte Carlo simulations reduces the need for 3D points and allows tracking for extended periods of time without any 3D point registrations.

Keywords: augmented reality, camera tracking, visual SLAM, inertial sensors, sensor fusion, optical flow

Index Terms: I.4.8 [Image processing and computer vision]: Scene analysis—Motion, Photometry, Sensor fusion, Tracking; I.2.9 [Artificial intelligence]: Robotics—Kinematics and dynamics, Sensors;

1 INTRODUCTION

The past few decades extensive research has been carried out in the area of simultaneous localisation and mapping (SLAM). This is apparent from reviewing computer vision and robotics literature [1, 2]. New algorithms are constantly developed in order to tackle computational complexity and drift. Another successful line of work is visual-inertial tracking, *i.e.* fusion of visual information and kinematic data from miniature MEMS inertial sensors. Corke et al. [3] give an introduction to the field.

SLAM has great success in many applications. However, in augmented reality it is often the case that the 3D structure is partly known or can easily be marked in the scene. The expensive estimation of a denser map has no value in itself besides enabling stable pose estimation. This paper investigates an alternative strategy. Instead of mapping, visual-inertial tracking with the available 3D information is used as a basis, and optical flow measurements are added to enable accurate estimation of the camera pose and kinematics. Optical flow is a well-known concept in computer vision. However, in most cases it is used to initialise a visual SLAM process [4] or to approximate relative camera motions [5]. The idea of combining very few 3D anchor points with 2D optical flow measurements — without attempting to recover depth — has hardly been considered in this context.

The contribution of this paper is to extend the model-based visual-inertial tracking system presented in [6] with the capability to exploit the information in 2D optical flow measurements. Optical flow measurements can be obtained from the camera images at any time, without knowledge of the scene structure. It is well-known that the resulting constraints do not provide full observability. However, as will be shown, they reduce the need for features with known depth and allow tracking for extended periods of time without any 3D point registrations.

*e-mail: Gabriele.Bleser@dfki.de

†e-mail: Gustaf.Hendeby@dfki.de

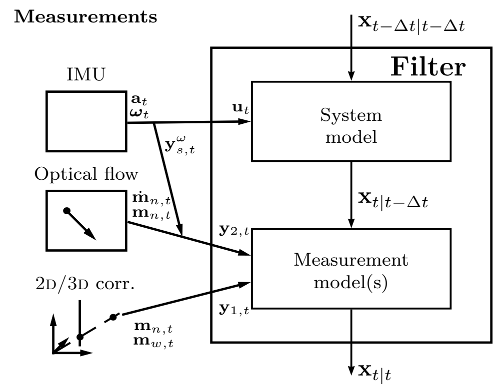


Figure 1: Filter architecture and data flow.

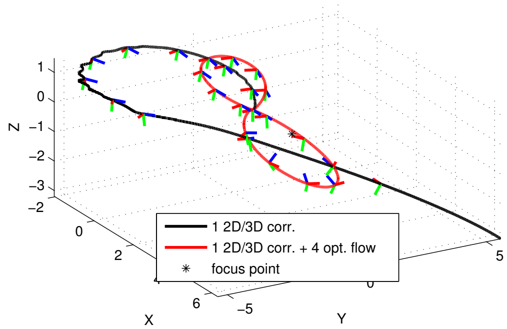
2 SENSOR FUSION

Recursive filters can estimate the pose and kinematics of a moving camera-IMU system from camera and inertial measurements. The extended Kalman filter (EKF) [7] is used to extract the relevant information from the respective measurements — 2D/3D point correspondences and optical flow measurements from image processing, and 3D angular velocities and linear accelerations from the IMU. The architecture of the fusion system is outlined in Figure 1. Due to space constraints only an overview of the system and resulting equations are given here. More details and justifications can be found in [6] and [8].

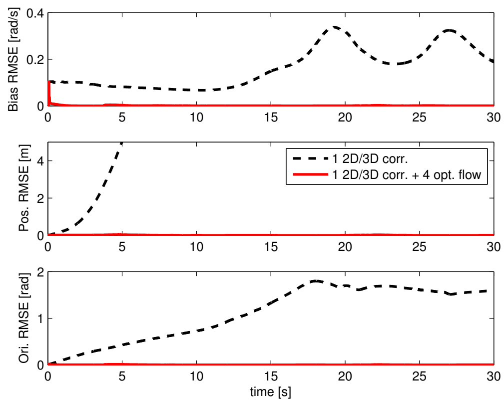
The state vector is $\mathbf{x}^T = [\mathbf{T}_w^{wsT}, \dot{\mathbf{T}}_w^{wsT}, \mathbf{q}_{sw}^T, \mathbf{b}_s^{\omega T}]$, where $\mathbf{T}_w^{ws}$ denotes position, $\dot{\mathbf{T}}_w^{ws}$ linear velocity, and $\mathbf{q}_{sw}$ the orientation quaternion of the IMU, s, with respect to the world frame, w. Moreover, $\mathbf{b}_s^{\omega}$ denotes slowly time varying gyroscope biases. Considering the gyroscope and accelerometer readings to be known control input, $\mathbf{u}^T := [\mathbf{y}_s^{\omega T}, \mathbf{y}_s^{aT}]$, the system model $\mathbf{x}_{t+\Delta t} = f(\mathbf{x}_t, \mathbf{u}_{t+\Delta t}, \mathbf{v}_t)$ describes the evolution of the state through time. The equation $\mathbf{0}_2 = h_1(\mathbf{x}_t, \mathbf{y}_{1,t}, \mathbf{e}_{1,t})$ with $\mathbf{y}_1^T := [\mathbf{m}_n^T, \mathbf{m}_w^T]$ relates the 3D anchor points, $\mathbf{m}_w^T$, to their measured 2D positions, $\mathbf{m}_n^T$, and the state. Here, $\mathbf{v}$ and $\mathbf{e}$ denote process and measurement noise, respectively. The detailed equations and the reasons for the design choices are provided in [6]. Optical flow measurements are incorporated subsequently.

Optical flow is here defined as the velocity, $\dot{\mathbf{m}}_n$, of image location $\mathbf{m}_n$, with homogenisation $\tilde{\mathbf{m}}_n^T := [\mathbf{m}_n^T, 1]$ and $\dot{\tilde{\mathbf{m}}}_n^T := [\dot{\mathbf{m}}_n^T, 0]$, respectively. A Kanade-Lucas tracker [9] can be used to measure optical flow by computing the movement of a distinctive patch in subsequent camera images. The pose and kinematics of a camera are directly related to the optical flow via the continuous epipolar constraint [8]:

$$0 = \dot{\tilde{\mathbf{m}}}_n^T(\mathbf{v}_c^{cw} \times \tilde{\mathbf{m}}_n) + \tilde{\mathbf{m}}_n^T\big((\boldsymbol{\omega}_c^{cw} \times \mathbf{v}_c^{cw}) \times \tilde{\mathbf{m}}_n\big), \qquad (1)$$



(a) One example of estimated trajectory.



(b) Root mean square error (RMSE) from 100 Monte Carlo simulations.

Figure 2: Tracking results: note how in all cases the optical flow measurements reduce the error to almost zero, whereas the results quickly drift off when observing only one single 3D point as complement to IMU measurements.

where $\mathbf{v}_c^{cw} := -\omega_c^{cw} \times \mathbf{T}_c^{cw} + \dot{\mathbf{T}}_c^{cw}$. Here, $\mathbf{T}_c^{cw}$ denotes position, $\dot{\mathbf{T}}_c^{cw}$ linear velocity, and ω_c^{cw} angular velocity of the camera, c, with respect to the world frame. These equations results from differentiating the projection of point $\mathbf{m}_w$ with respect to time and eliminating its depth. An implicit measurement equation exploring optical flow can from this be derived by reformulating (1) in terms of the measured angular velocity, $\mathbf{y}_s^{\omega}$, and the quantities in the state. With $\mathbf{y}_2^T := [\mathbf{y}_s^{\omega T}, \dot{\mathbf{m}}_n^T, \mathbf{m}_n^T]$, the second measurement model $0 = h_2(\mathbf{x}_t, \mathbf{y}_{2,t}, \mathbf{e}_{2,t})$ is obtained.

Since the camera measurements, 2D/3D correspondences and optical flow, are assumed mutually independent, multiple observations are processed sequentially, each in a separate EKF measurement update step with the appropriate model and starting with the 3D points.

3 Experimental Setup and Results

Simulated data is used to evaluate the proposed method, *i.e.* the value of adding 2D/3D point correspondences with 2D optical flow measurements. The data has camera translations and rotations in all dimensions at various speeds. The trajectory is simulated so that the camera focuses on one point of interest (*cf.* Figure 2(a)). This is typical for close range camera localisation and visual servo applications. From the ground truth poses, biased and noisy inertial measurements and noisy camera measurements are simulated and then used to recover the camera trajectory.

Figures 2 and 3 demonstrate promising improvements obtained by incorporating optical flow. Figure 2 shows: when observing only one single 3D feature — the focus point — at 25 Hz, the filter fails to estimate the gyroscope biases, which results in a huge drift in the camera trajectory. By adding four optical flow measurements in the corners of the camera image, the gyroscope biases are properly estimated and high accuracy is obtained.

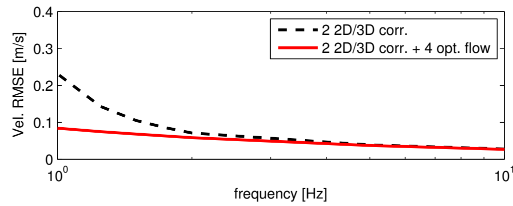


Figure 3: Velocity RMSE vs. frequency of observing 3D points.

In Figure 3, two 3D features are observed with different frequencies ranging from 10 to 1 Hz. The plot shows how the optical flow measurements significantly improve the results, as the velocity estimate otherwise degenerates rapidly with an observation rate below 2 Hz.

4 Conclusion and Future Work

This paper extends the visual-inertial tracking system developed in [6] with optical flow measurements. Monte Carlo simulations show that adding optical flow measurements reduces the required quantity and frequency of observing features with known depth. This allows for robust and efficient tracking with very few 3D anchor points that could be installed or surveyed manually with reasonable effort. As such, the method provides an efficient alternative to a complete and computationally intense SLAM process. The next step is to test the method on actual measured data. Here, filtering using a simple 2D constant velocity model could be used to ensure accurate optical flow measurements. Moreover, observability of the camera pose and kinematics obtained from different configurations of optical flow measurements will be studied and outliers will be handled.

References

[1] A. J. Davison, I. Reid, N. Molton, and O. Stasse. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1052–1067, 2007.

[2] T. Bailey, J. Nieto, and E. Nebot. Consistency of the FastSLAM algorithm. In *IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, Florida, USA, May 2006.

[3] P. Corke, J. Lobo, and J. Dias. An introduction to inertial and visual sensing. *International Journal of Robotics Research*, 26: 519—-535, 2007.

[4] G. Bleser, M. Becker, and D. Stricker. Real-time vision-based tracking and reconstruction. *Journal of Real-Time Image Processing*, 2(2–3):161–175, November 2007.

[5] P. Hammarstedt, F. Kahl, and A. Heyden. Affine reconstruction from translational motion under various autocalibration constraints. *Journal of Mathematical Imaging and Vision*, 24:245 – 257, March 2006.

[6] G. Bleser and D. Stricker. Advanced tracking through efficient image processing and visual-inertial sensor fusion. *Computer & Graphics*, 33:59–72, February 2009.

[7] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Academic Press, Inc, 1970.

[8] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3-D Vision*, volume 26. Springer Verlag, 2003.

[9] B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, April 1981.