Maximilian Eibl, Martin Gaedke (Hrsg.): INFORMATIK 2017,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 1

VCIT – Visually Corrected Inertial Tracking

Daniel Andrés López[1], Benedikt Diensberg[1], Elmar Schömer[1], Ulrich Schwanecke[2]

Abstract: Many smart systems depend on exact models of their environment. These are gained by tracking objects in their surroundings. When a highly precise system is only available for a small part of the environment, it can be enhanced with a second system to recover the unknown parts. This paper presents a method to recover loss of a precise (optical) tracking system by a less precise (inertial) tracking system. First the rotation from the inertial measurement unit (IMU) and the optical system are aligned. A second step integrates the IMU acceleration two times and removes both times the drift by known initial and end values (first integration: velocity, second integration: position) from optical tracking. The error is backpropagated continuously.

Keywords: Sports, Visual Interial Tracking, Sensorfusion, IMU

1 Introduction

An important part of smart electronic systems is their ability to locate objects exactly. This enables the system to generate an accurate model of a specific environment. Based on this model, the system can interact with users within this environment in a natural way. For example, it can seamlessly react to changes in the position and orientation of objects or in gestures of various persons, just as people interact with each other.

In this paper we present a method for combining two different tracking systems in a way that their respective advantages complement each other. More specifically, our approach can recover the loss of a precise (optical) tracking system using a less precise (inertial) tracking system. We discuss our method demonstrating the tracking of a golf club in an exemplary way. Since the head of a golf club can move very fast, tracking a golf swing can be a difficult task which makes it a good test scenario. We divide the golf swing in two regions: One small area around the point where the club hits the ball (main action) and the up-/down-swing.

[1] Computational Geometry Research Group, Institute of Computer Science, Johannes Gutenberg University Mainz, Staudingerweg 9, 55122 Mainz, Germany,
{daniel.andres, diensberg, schoemer}@uni-mainz.de
[2] Computer Vision and Mixed Reality Group, Department of Design, Computer Science and Media, RheinMain University of Applied Sciences, Unter den Eichen 5, 65195 Wiesbaden, Germany,
ulrich.schwanecke@hs-rm.de

Those two regions will be tracked with different tracking systems. In this paper we will focus on optical and inertial tracking, but any tracking method could be used.

Optical tracking methods often use standard cameras. Current optical tracking methods can estimate the pose of an object captured by rgb cameras [TSS16]. Beside the marker-less methods many marker based systems exist, for example [Tj15]. It captures an active marker with infrared LEDs with a single camera and calculates the transformation from the marker coordinate system to the camera coordinate system. Tracking with an inertial measurement unit (IMU) on the other hand can track motions of any size. But pose estimation contains drift, which accumulates through the two integration steps. In order to avoid this drift, knowledge about the expected motion can be used. The algorithm shown in [NL16] is based on human gaits to reset the drift to zero by setting the velocity in stationary phases to zero. But the existence of zero-phases cannot be guaranteed in general motions. Alignment to the earth frame is necessary to separate the earth acceleration from object acceleration. Therefore the orientation can be obtained by a Kalman filter [LVB99, Ma01] or analytical methods [HM06, MHV11].

To overcome the flaws of those systems there are several approaches for combining them. This is often called visual inertial tracking. In [LS13] the optical tracking system is used to stabilize the IMU measurements and obtain a more precise pose estimation. The paper discusses two different approaches based on Extended Kalman Filter (EKF) and shows the advantage of single IMU based tracking. [KS11] describes a method to calibrate a fixed sensor array of an IMU and a camera to be usable for common measurements. Instead of EKF they employ an Unscented Kalman Filter (UKF) because of its superior performance and higher accuracy on highly nonlinear systems. This is also shown in [JS11], which uses it to measure accurate poses in a simultaneous localization and mapping (SLAM) application. SLAM algorithms are often used in robotics. Another approach in improving SLAM algorithms with IMU data is described in [MAT17]. They use the IMU data for a frame by frame prediction of the camera poses.

But all these approaches require constant measurements from the optical system. For many applications the field of view of a camera is often too small, especially in confined spaces. So these systems can only be used to track the main action, but not the full movement. We propose a setup, where the main action is tracked by a precise tracking system, here optical tracking, and the full motion is recovered with an IMU. Instead of zero-phases we use the information of the optical system during common measurements to minimize integration errors.

Accelerometer measurements always contain gravitational acceleration, which has to be subtracted to get the pure acceleration of the tracked object. Therefore knowing the direction of the gravitational vector is essential for getting only the acceleration of movements. Common approaches use Kalman Filters to calculate the orientation relative to the earth frame (so the gravitational vector points downwards). This is shown in [LVB99, Ma01, Be14]. Additional to Kalman Filter fusion of the components of an inertial measurement unit there exist non linear complementary filters. The most common ones are [MHP05, MHV11]. Those algorithms are also called Attitude Heading Reference System (AHRS). The idea is based on analytical reconstruction of the orientation instead of stochastic models like in Kalman Filter approaches. In this paper the AHRS described by [MHV11] is used to obtain the orientation relative to the earth frame by accelerometer and gyroscope data. It can be used to know the direction of the gravitational vector. There is no information how the IMU is located in the earth magnetic field. This information can be estimated with an additional magnetometer to refine the results.

2 Camera enhanced IMU pose estimation

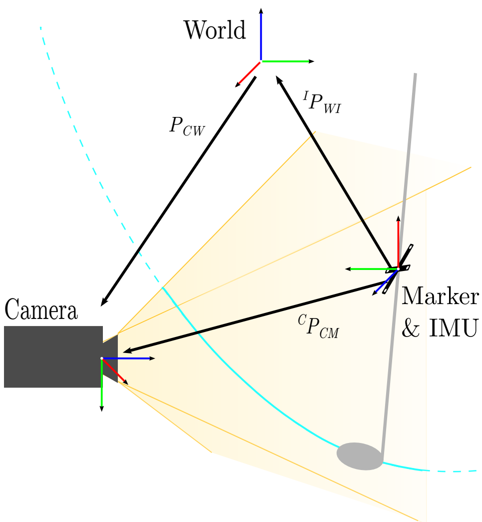


Fig. 1: Overview of system and transformations

In order to get a full 3D reconstruction of the movement it is necessary to combine both measurement systems. The IMU measurements will be received over the whole time, but they are less precise than the measurements from the optical system. The optical system will not cover the whole movement and therefore when the tracking is lost, the IMU data will be used.

To simplify the readability we only use homogenous matrices and vectors, so we can multiply them easily. In the following we will denote the tracking system that recorded a value as a raised index in front of the value (C = camera, I = IMU) and as a lowered index after the value its coordinate system (C = camera, W = world, M = marker, I = IMU) and the time. If it is a transformation between coordinate systems we denote those as a double index of target system and start system (CM = marker to camera, ...).

Figure 1 shows an overview of the used transformations/poses (position and orientation $P \in SE(3)$). For this paper we presume we get poses, ${}^{C}P_{CM,t}$, of the tracked marker from the camera system relative to the camera coordinate system at time t as in [Tj15]. We also presume the camera and the IMU data are synchronised.

Furthermore we assume for simplicity that the translational T_{MI} and rotational R_{MI} offset between IMU and tracked marker can be neglected, therefore they equal the identity matrix.

In this paper we will focus on the phase where camera tracking is lost and only the IMU data is available, as well as a few common measurements of both systems immediately before and after this phase.

Let ${}^{C}P_{CM,0}$ be the last camera pose before tracking is lost and ${}^{C}P_{CM,n}$ the first pose after tracking is regained. Where a pose P is a tuple of a translational vector T and a rotation R: $P = (R, T)$.

An IMU measures angular velocity ω and acceleration a relative to the world coordinate system W, which is defined by the first IMU measurement. Let $({}^{I}\omega_{W,0}, {}^{I}a_{W,0})$ be the last IMU measurement before the optical tracking is lost and $({}^{I}\omega_{W,n}, {}^{I}a_{W,n})$ the first after the optical tracking is recovered. And let $({}^{I}\omega_{W,t}, {}^{I}a_{W,t})$ for $t \in \{1, \ldots, n-1\}$ respectively be the IMU measurements while the camera tracking is lost. To acquire the poses for the IMU measurements we first calculate the orientations ${}^{I}R_{WI,t}$ with the algorithm described by [MHV11] from accelerometer and gyroscope data. To compare both system we have to consider same poses in different measurement systems. The optical system provides ${}^{C}P_{CM}$, the IMU pose ${}^{I}P_{CM}$ has to be constructed with our proposed algorithm. First we calculate the registration. The translational part will be applied in the second integration step by setting the initial ${}^{I}T_{CM,0}$ to ${}^{C}T_{CM,0}$. The rotational registration $R_{CW,0}$ from IMU world into camera system is calculated by:

$$R_{CW,0} = {}^{C}R_{CM,0} \cdot R_{MI} \cdot {}^{I}R_{WI,0}^{-1} \tag{1}$$

With the registration $R_{CW,0}$ we can transform the orientations ${}^{I}R_{WI,t}$ from the IMU system into the camera coordinate system:

$${}^{I}R_{CM,t} = R_{CW,0} \cdot {}^{I}R_{WI,t} \cdot R_{MI}^{-1} \tag{2}$$

Before we can use the measured accelerations from the IMU we need to subtract the gravitational vector g, as mentioned above. In the origin of the IMU coordinate system the gravitation applies only on the z-axis, so the gravitational vector becomes $g = (0, 0, 9.81)^{\top}$ $[m/s^2]$ (T^{-g} is g as homogenous matrix with negative gravitation). With ${}^{I}R_{WI,t}^{-1}$ we rotate the acceleration to this system, subtract the gravitation and rotate back:

$${}^{I}a_{C,t} = R_{CW,0} \cdot {}^{I}R_{WI,t} \cdot {}^{I}T^{-g} \cdot {}^{I}R_{WI,t}^{-1} \cdot {}^{I}a_{W,t}^{g} \tag{3}$$

Since we want to know the accelerations in the camera coordinate system C, we immediately rotate with ${}^{I}R_{CM,t}$ which contains the rotation from IMU to camera system. The accelerations can now be used to calculate the velocities with the known IMU measurement frequency f and the velocities from the camera tracking:

$$ {}^{I}v_{C,t} = {}^{I}v_{C,t-1} + {}^{I}a_{C,t} \cdot 1/f \qquad \text{with } {}^{I}v_{C,0} = {}^{C}v_{C,0} \tag{4} $$

Finally these can be used to get the positions from the IMU data:

$$ {}^{I}T_{CM,t} = {}^{I}T_{CM,t-1} + {}^{I}v_{C,t} \cdot 1/f \qquad \text{with } {}^{I}T_{CM,0} = {}^{C}T_{CM,0} \tag{5} $$

From that we get the resulting IMU poses:

$$ {}^{I}P_{CM,t} = ({}^{I}T_{CM,t}, {}^{I}R_{CM,t}) \tag{6} $$

The biggest problem of IMU data is their drift, which accumulates to a considerable amount even during the short periods, where we lose the camera tracking. To correct for that we can calculate the drift error as the difference between the first common measurements of IMU and camera after the camera tracking is regained. We then use backpropagation of the error to correct the previous data of the IMU-only measurement. In (4) we calculated the velocities ${}^{I}v_{C,t}$ from the IMU accelerations. Afterwards with the velocities ${}^{C}v_{C,n}$ from the camera data the velocity drift error results as $vde_{C,n} = {}^{I}v_{C,n} - {}^{C}v_{C,n}$. For the drift we assume it took effect in the same direction for the whole measurement. Furthermore we presume the drift error increases equally over time. These assumptions are based on our experiments with a stationary IMU. So we can estimate the drift error for the IMU data simply by linear interpolation between the starting error 0 and $vde_{C,n}$. Thus the corrected IMU velocities are:

$$ {}^{cor}v_{C,t} = {}^{I}v_{C,t} - \frac{t}{n} vde_{C,n} \tag{7} $$

The IMU positions will now be calculated by ${}^{corr}v_{C,t}$ instead of ${}^{I}v_{C,t}$ in equation (5). But they still contain drift errors. We can estimate these errors in a similar way to the velocities' drift errors: We know the exact poses from the camera data after the IMU-only phase, so we can again calculate the difference transformation P_{dif} with

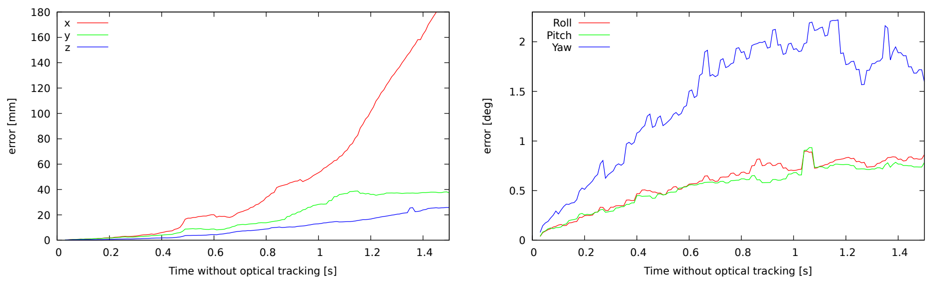
$$ T_{dif} = {}^{I}T_{CM,n} - {}^{C}T_{CM,n} \tag{8} $$
$$ R_{dif} = {}^{I}R_{CM,n} \cdot {}^{C}R^{-1}_{CM,n} \tag{9} $$

Now we can propagate this error back over the previous poses with linear interpolation and correct them to get the final poses ${}^{corr}P_{CM,t}$:

$$ {}^{cor}T_{CM,t} = {}^{I}T_{CM,t} - \frac{t}{n} T_{dif} \tag{10} $$
$$ {}^{cor}R_{CM,t} = {}^{I}R_{CM,t} \cdot slerp(I, R_{dif}, t/n)^{-1} \tag{11} $$

For linear interpolation of rotations we used spherical linear interpolation (slerp) on their quaternion representation. In these calculations we used only the last and first common pose before and after the IMU-only phase as base for the pose corrections. Since the camera system measueres also noisy poses, we recommend for real world applications to use several poses instead for higher stability.

3 Evaluation



(a) Average dynamic RMSE for positions (b) Average dynamic RMSE for rotations

Fig. 2: Average of 20 experiments. Dynamic RMSE error against the time where poses have to be estimated by IMU data. Ground truth data is the optical pose.

This paper uses a prototype setup of the optical tracking system HSRM-Tracking [Tj15] and a LP-Research LPMS-B2 IMU[3]. The HSRM-Tracking requires an active LED marker. The IMU and the marker are rigidly connected to a golf club, so that the translational offset is as low as possible. The camera is placed about 1m away, looking at the marker. The algorithm is made to overcome tracking loss of the high precision system by using the less precise system. But to compare the result the whole movement is measured with both systems and the loss of tracking is simulated (the movement will not cover a full swing to be able to stay inside the field of view of the camera system). The experiments include 20 different swings for evaluation. Some with less noise while the club is hooked up and swings like a pendulum, some with hand movement of the club and some with real golf swings, each over a length of approximately 1m. The error is calculated as difference from estimated IMU to optical poses. To evaluate the individual errors the rotational and translational parts are compared separately along their components. Here only the dynamic root mean squared error (RMSE) is calculated, because the algorithm applies the drift correction by using information of the current movement. The static scenario is already shown in [NL16]. Figure 2 shows the average position error and the average rotation error. The time without optical tracking gives the amount of seconds to recover the pose with IMU data. As we analyse a swing, this area is located around the highest phase of the swing.

[3] `https://www.lp-research.com/lpms-b2/`

Figure 2a shows the error grows as the time without optical tracking increases. The error for the x-value grows much faster than the other two components. Which is expected, since the main movement was along the x-axis. Analysing single experiments shows the error maximum is at the highest phase of the swing, the minima at the start and end points. Although there are two drift correction steps the drift is not eliminated fully. This could indicate that the correction method itself is not optimal yet. Figure 2b shows rotational errors. Roll and pitch errors grow reasonably slow. But yaw error shows unreliability in its direction. On the other hand when no drift correction is applied the error grows much faster: At a loss of 1s the difference to the reference data on the x-Axis for example is about 1m without correction, whereas with correction only about 40mm over a swing of 1m, similarly on the other axes.

4 Conclusion and Future Work

We showed the modification of the algorithm derived in [NL16] is also suitable for estimating poses from IMU measurements. Additional to the proposal it is necessary to eliminate drift due to integration constants, after gaining velocities and after calculating positions. Drift correction at one stage does not remove all bias errors. The test results have shown that the registration of IMU and optical marker is not necessary, as long as both systems are located as close as possible, so that they can be assumed to be at the same place. Although registration of the translation between both systems increases accuracy. The rotational registration can be easily done live during each experiment.

This paper interpolates the error for the drift correction with a linear function. Because the error is quadratic due to double integration the error backpropagation also should be quadratic. This could lead to better performance at the direction change of the swing phase. Also it has to be considered that pose correction after the second integration step is currently done separately for position and orientation. A backpropagation from a single $SE(3)$-group element could lead to better results. This can be achieved for example with dual quaternion representation. Additionally the drift can be backpropagated in the same manner with accelerations. Another idea is to get a live processing algorithm by calculating drift rates before the tracking is lost and then apply drift correction directly. Further investigation has to show if statistical models can be generated and applied later on. A dedicated learning phase can be useful to get drift rates in all directions and when direction changes. This phase would require both systems. Beside the drift correction optimizations the integration steps can also be improved. More accurate methods involve more predecessors to yield a better result. This paper only discusses correction of IMU pose estimation by an optical tracking system. On the other hand the IMU can be used to improve the optical tracking, this is called Visual Inertial Tracking.

References

[Be14] Bergamini, E.; Ligorio, G.; Summa, A.; Vannozzi, G.; Cappozzo, A.; Sabatini, A.: Estimating Orientation Using Magnetic and Inertial Sensors and Different Sensor Fusion Approaches: Accuracy Assessment in Manual and Locomotion Tasks. Sensors, 14(10):18625–18649, October 2014.

[HM06] Hamel, T.; Mahony, R.: Attitude estimation on SO[3] based on direct inertial measurements. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 2170–2175, May 2006.

[JS11] Jones, E. S.; Soatto, S.: Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. The International Journal of Robotics Research, 30(4):407–430, April 2011.

[KS11] Kelly, J.; Sukhatme, G. S.: Visual-Inertial Sensor Fusion: Localization, Mapping and Sensor-to-Sensor Self-calibration. The International Journal of Robotics Research, 30(1):56–79, January 2011.

[LS13] Ligorio, G.; Sabatini, A.: Extended Kalman Filter-Based Methods for Pose Estimation Using Visual, Inertial and Magnetic Sensors: Comparative Analysis and Performance Evaluation. Sensors, 13(2):1919–1941, February 2013.

[LVB99] Luinge, H. J.; Veltink, P. H.; Baten, C. T. M.: Estimation of orientation with gyroscopes and accelerometers. In: First Joint BMES/EMBS Conference 1999. volume 2, p. 844, October 1999.

[Ma01] Marins, J. L.; Yun, X.; Bachmann, E. R.; McGhee, R. B.; Zyda, M. J.: An extended Kalman filter for quaternion-based orientation estimation using MARG sensors. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. volume 4, pp. 2003–2011, 2001.

[MAT17] Mur-Artal, R.; Tardós, J. D.: Visual-Inertial Monocular SLAM with Map Reuse. IEEE Robotics and Automation Letters, 2(2):796–803, April 2017.

[MHP05] Mahony, R.; Hamel, T.; Pflimlin, J. M.: Complementary filter design on the special orthogonal group SO(3). In: 44^{th} IEEE Conference on Decision and Control. pp. 1477–1484, December 2005.

[MHV11] Madgwick, S. O. H.; Harrison, A. J. L.; Vaidyanathan, R.: Estimation of IMU and MARG orientation using a gradient descent algorithm. In: IEEE International Conference on Rehabilitation Robotics. pp. 1–7, June 2011.

[NL16] Nguyen, L. V.; La, H. M.: A human foot motion localization algorithm using IMU. In: American Control Conference (ACC). pp. 4379–4384, July 2016.

[Tj15] Tjaden, H.; Schwanecke, U.; Stein, F.; Schömer, E.: High-Speed and Robust Monocular Tracking. In: 10th International Conference on Computer Vision Theory and Applications (VISIGRAPP). pp. 462–471, 2015.

[TSS16] Tjaden, H.; Schwanecke, U.; Schömer, E.: Real-Time Monocular Segmentation and Pose Tracking of Multiple Objects. In: Computer Vision – ECCV 2016. Lecture Notes in Computer Science, volume 9908. Springer, Cham, pp. 423–438, 2016.