

Improving Full-Body Pose Estimation from a Small Sensor Set Using Artificial Neural Networks and a Kalman Filter

Frank J. Wouda,¹ Matteo Giuberti,² Giovanni Bellusci,² Bert-Jan F. van Beijnum,¹
Peter H. Veltink¹

¹Department of Biomedical Signals & Systems, Technical Medical Centre,
University of Twente, Enschede, The Netherlands f.j.wouda@utwente.nl

²Xsens Technologies B.V., Enschede, The Netherlands

Abstract

Previous research has shown that estimating full-body poses from a minimal sensor set using a trained ANN without explicitly enforcing time coherence has resulted in output pose sequences that occasionally show undesired jitter. To mitigate such effect, we propose to improve the ANN output by combining it with a state prediction using a Kalman Filter. Preliminary results are promising, as the jitter effects are diminished. However, the overall error does not decrease substantially.

Introduction

The human motion capture industry has grown a lot in recent years, which is shown by the availability of a large variety in technologies. The most prominent solutions are based on inertial measurement units (IMUs) and optical tracking, and both require numerous sensors/markers for full-body motion capturing (van der Kruk and Reijne 2018). This results in a large setup time and obtrusiveness to the subject.

Many works focused on reducing complexity by relying on database solutions. For example, the number of markers (in optical motion capture) was successfully reduced (to as few as six markers) by using a nearest neighbour search approach (Chai and Hodgins 2005). Similarly, Tautges et al. exploited nearest neighbour search to reconstruct full-body movements using only four accelerometers (Tautges et al. 2011). Both approaches require a movement database to be available at runtime, which is computationally expensive. Therefore, Wouda et al. 2016 used an Artificial Neural Network (ANN) to overcome this limitation and showed comparable performance. However, this approach does not enforce any temporal coherence between consecutive poses and thus results might show unrealistic jumps and jitter.

By providing temporal coherence, the overall performance of an ANN approach is likely to improve. For example, use of recurrent neural networks were effective in predicting sequences of human movement from a (short) pose sequence (Fragkiadaki et al. 2015). However, such an approach requires sufficient computation resources and training data (of the sequences of interest). An interesting alternative is the use of a Kalman filter (KF), which fuses a prediction of

states with the measurements (Welch and Bishop 1995).

The goal of this work is to investigate the use of a statistical framework like a KF for combining the pose output of an ANN (estimated from a minimal IMU sensor setup) with a pose prediction. This approach allows for using explicit modelling of constraints of the human body, e.g. movement range and joint degree of freedom limits.

Methods

An ANN (with two hidden layers of 200 and 100 neurons) was trained (using MATLAB R2016a) to map the (relative to pelvis) orientation of 4 body segments (lower legs and arms) to a full-body pose, identical to Wouda et al. 2016. The training data for this ANN consists of movements of 6 participants (~ 120 minutes in total), identical to used in (Wouda et al. 2016). An inertial motion capture system (Xsens MVN) consisting of 17 on-body sensors was used for capturing the dataset (Schepers, Giuberti, and Bellusci 2018). It was chosen to describe orientations using quaternions, as it was shown to be suitable for pose estimation from a minimal sensor set with a trained ANN (Wouda et al. 2016). However, orientations complicate the use of a KF, due to the potentially large linearization errors. Therefore, we propose to use an error-state KF, similar to Kortier et al. 2014, which keeps linearization errors to a minimum. The error-state was used as follows: $q_1 = q_2 \odot \delta q$, i.e. the error quaternion can be seen as the orientation difference between two body segments. Furthermore, the error quaternion were converted to helical angles using: $\delta q \approx [1 \quad \frac{1}{2}\delta\theta]$. We will take orientation and its derivative: angular velocity of the full-body pose (23 segments) into account in the states: $x^t = [\delta\theta \quad \delta\dot{\theta}]$.

The purpose of applying a KF is to stabilize the pose prediction (process) using the ANN output (measurement), as depicted in Figure 1. We assume our process can be described with constant velocity, which should be acceptable for small timesteps. This results in the following state transition matrix: $A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$. The output of the trained ANN is used as a measurement update, which is related to the states by the measurement matrix: $H = [1 \quad 0]$.

The noise covariance of the process (Q) and the measurement (R) effectively reflect the reliability of both information sources. To test the viability of this approach, an esti-

mate of the measurement noise covariances was made based on the variance of the ANN output data (as jitter can be observed as increased variance), i.e. $R = \sigma(\delta\theta)$, which is a diagonal matrix. Similarly, the process noise covariance matrix was determined from the variance in the ground truth. The performance of the proposed approach was evaluated using the Euclidean distance between joint positions, similar to existing literature (Tang et al. 2008). To this end, joint positions are calculated from orientations using a forward kinematics approach.

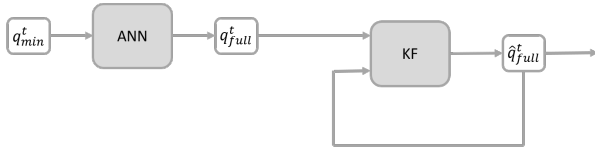


Figure 1: The proposed Kalman filter combines the ANN estimate of the full-body pose (q_{full}^t) with prediction of the state that is based on the optimal estimate from the previous timestep (\hat{q}_{full}^{t-1}).

Results and discussion

For sake of conciseness, only the knee joint positions are compared, since jitter was observed best for this joint due to the lack of a sensor on the more proximal segment (upper leg). The ground truth was derived from the knee position as measured with a full-body motion capture system. Figure 2 shows a comparison of the estimated knee joint position with the ground truth. For illustration purposes the trial with most evident jitter (blue peak errors) was shown. However, similar behaviour was observed for trials with less evident jitter. It can be seen that the jitter has been largely mitigated, although the mean joint position error shows a limited decrease (0.11 m instead of 0.12 m).

This can potentially be improved by having a time-variant measurement noise matrix, which depends on the current estimated pose. Jitter behaviour was mainly observed for specific poses, that are likely similar to other poses, e.g. in the end of the swing phase the lower leg orientation is similar to that of a person sitting. However, this would require a measure of confidence of the ANN output. Furthermore, the KF framework easily allows for applying additional constraints, such as limited joint range/degrees of freedom.

Conclusion

Promising results were achieved by using a KF to stabilize the (constant velocity) prediction using ANN output, although the limited decrease in the mean joint position error indicates that there is room for improvement.

Acknowledgments

This research (project No. 13917) is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

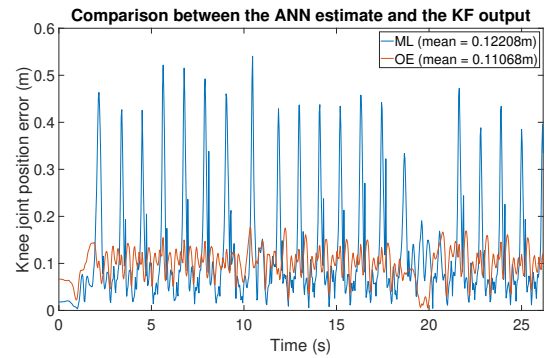


Figure 2: The Euclidean distance between the ANN estimate and the actual knee joint position (for a walking trial) is shown in blue (ML), and the optimal estimate of the knee joint position compared to the ground truth is shown in orange (OE).

References

- Chai, J., and Hodgins, J. K. 2005. Performance animation from low-dimensional control signals. *ACM Transactions on Graphics* 24(3):686.
- Fragkiadaki, K.; Levine, S.; Felsen, P.; and Malik, J. 2015. Recurrent Network Models for Human Dynamics. *Proceedings of the IEEE International Conference on Computer Vision* 4346–4354.
- Kortier, H. G.; Sluiter, V. I.; Roetenberg, D.; and Veltink, P. H. 2014. Assessment of hand kinematics using inertial and magnetic sensors Assessment of hand kinematics using inertial and magnetic sensors. *Journal of NeuroEngineering and Rehabilitation* 11(70):1–14.
- Schepers, M.; Giuberti, M.; and Bellusci, G. 2018. Xsens mvn: Consistent tracking of human motion using inertial sensing (white paper). 1–8.
- Tang, J. K. T.; Leung, H.; Komura, T.; and Shum, H. P. H. 2008. Emulating human perception of motion similarity. *Computer Animation and Virtual Worlds* 19(August):211–221.
- Tautges, J.; Zinke, A.; Krüger, B.; Baumann, J.; Weber, A.; Helten, T.; Müller, M.; Seidel, H.-P.; and Eberhardt, B. 2011. Motion Reconstruction Using Sparse Accelerometer Data. *ACM Transactions on Graphics* 30(3):18:1–12.
- van der Kruk, E., and Reijne, M. M. 2018. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science* ISSN homep(May):1–14.
- Welch, G., and Bishop, G. 1995. An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA.
- Wouda, F. J.; Giuberti, M.; Bellusci, G.; and Veltink, P. H. 2016. Estimation of full-body poses using only five inertial sensors: An eager or lazy learning approach? *Sensors (Switzerland)* 16(12).