Arm Tracking

Anonymous Author(s)

Abstract

1 INTRODUCTION

2 RELATED WORK

Device Orientation Estimation. Device orientation estimation has been well studied in the literature. Gyroscope measures the angular velocity of a device, its readings can be integrated over time to estimate the device orientation. Gyroscope integration is accurate in short term but drifts in long term due to noise and bias [1–5]. In addition, 2-DoF (degree of freedom) device orientation can be determined by gravity or magnetic north. So the device orientation can be uniquely determined by combining gravity and magnetic north. The accelerometer measures the mixture of both gravity and linear motion. Therefore, when the device is moving, the gravity is difficult to decompose from the accelerometer. The magnetometer measures magnetic north. However, the magnetometer measurements can easily be affected by nearby ferromagnetic materials [1, 6].

Various works have been proposed to derive sensor fusion algorithms to estimate device orientation by using two or three of the accelerometer, gyroscope and magnetometer, including Kalman Filter [7, 8], Extended Kalman Filter [9, 10], Unscented Kalman Filter [11, 12] and Complementary Filter [6, 13]. A^3 [1] proposed to use opportunistic replacement to estimate the orientation. Specifically, it intelligently selects the moments that gravity and magnetic north are more reliable than gyroscope and replace the orientation estimation from gyroscope integration with the orientation estimation from gravity and magnetic north. These moments are usually when the device is static or in pure rotational motion, in which the gravity can be accurately measured from acclelerometer. To summarize, A^3 estimates the device orientation by gyroscope integration when the object is moving and calibrate the orientation by gravity and magnetic north when the object is static or in pure rotational motion. However, the assumption that device motions have frequent pauses for resetting the orientation is not always hold. MUSE [6] proposed that magnetic north is more trusted than gravity because magnetic north is unpolluted by the motion of device. It designed a magnetometer-centric sensor fusion algorithm based on complementary filter for orientation tracking. However, the magnetic north is able to calibrate 2-DoF orientation. Only using the magnetic north vector as the anchor to calibrate orientation estimation is not useful if the rotation of a device only in 3rd DoF. And the magetic fields can vary significantly within the same space due to the local ferromagnetic disturbances.

Recent literature has begun utilizing deep neural networks for IMU measurements processing [2, 14, 15] and orientation estimation [3–5]. OriNet [3] uses LSTM-based architecture to estimate the 3D orientation of flying robots to propagate orientation state and aggregates various estimations achieved for each state. [4] estimates device orientation by gyroscope integration, but it uses a CNN to correct the gyroscope readings and filter undesirable errors in the raw IMU signals. These methods does not explore to learn from multiple sensors. While the accelerometer and magnetometer can help to estimate orientation when the device is not moving and the variance of magnetic field density is small. IDOL [5] directly estimates the orientation from all IMU channels (acclerometer, gyroscope and magnetometer) using an LSTM-based architecture. This architecture maintains long-term accuracy but does not achieve so accurate short-term accuracy as gyroscope integration. Therefore, it fuses gyroscope data in the short term via an Extended Kalman Filter to improve short-term accuracy. However, Extended Kalman Filter is not optimal if the system is highly nonlinear. Whereas, the human wrist motion is very flexible, the non-linearity is high.

In this work, we also use the data from all IMU channels to estimate the orientation based on LSTM-based architecture. Different from IDOL, we use attention-based network adaptation to adjust the focuses of our network from both time and spatial views. In addition, we use a reinforcement learning model which incorporates domain knowledge to update the RNN architecture.

IMU-based Skeleton Tracking. In the literature, the studies [16–19] tried to track the upper limb movements or reconstruct the full body motion by leveraging multiple sensors, which is not practical in real life. ArmTrak [20] proposed to recover and track the 3D arm posture using smartwatch sensors along. It leverages hidden Markov model (HMM) to continuously estimate the elbow and wrist locations. However, its computation latency is long and cannot support real-time performance on smartphones. ArmTroi [21] optimizes the HMM architecture by HMM state reorganization and hierarchical search. ArmTroi achieves real-time arm skeleton tracking on smartphones. These two works assume the device orientation they used for tracking the locations of elbow and wrist is accurate. However, a small drift in orientation estimation will make obvious mistakes in projecting the accelerometer data, eventually making wrong estimated location. MUSE [6] proposed to jointly estimate orientation and location to fully utilize IMU data and the arm motion model. The location estimation accuracy of MUSE is higher than ArmTrak and ArmTroi, its computation latency is shorter than ArmTrak but still longer than ArmTroi. Its real-time computation cannot be afforded by a smartphone.

Human 3D Skeleton Tracking In the literature, researchers have made efforts on estimating 3D human skeleton from various data types, including vision-based, light-based, RF-based, WiFi-based and so on. The vison-based approaches include VICON system [22] and RGB-Depth cameras [23]. Albeit high accuracy can be achieved by vision-based approaches, their performance can be severely impaired by bad illumination, occlusion and blurry. More importantly, privacy issues occur when cameras are deployed to monitor the human subjects. Recent light-based approaches [24, 25] protect user privacy but they still cannot work in dark or occluded scenarios. To overcome the limitations of above solutions, RF-Pose3D [26] and WiPose [27] explore RF signal and WiFi signal respectively for 3D skeleton tracking. These two works are based on deep neural network architecture and achieve higher accuracy than IMU-based methods. However, they typically have a limited service coverage and also have higher system cost. The performance of them will be decreased when tracking multiple people at the same time.

In this work, we leverage the IMU sensors of wearables to track the 3D skeleton of human arm, which is a infrastructure-free and location-free solution. More importantly, our system is light and has no restrictions on the number of people.

3 MOTIVATION

4 SYSTEM DESIGN

5 IMPLEMENTATION

6 EVALUATION

7 CONCLUSION

REFERENCES

[1] Pengfei Zhou, Mo Li, and Guobin Shen. Use it free: Instantly knowing your phone attitude. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 605–616, 2014.

[2] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[3] Mahdi Abolfazli Esfahani, Han Wang, Keyu Wu, and Shenghai Yuan. Orinet: Robust 3-d orientation estimation with a single particular imu. *IEEE Robotics and Automation Letters*, 5(2):399–406, 2019.

[4] Martin Brossard, Silvere Bonnabel, and Axel Barrau. Denoising imu gyroscopes with deep learning for open-loop attitude estimation. *IEEE Robotics and Automation Letters*, 5(3):4796–4803, 2020.

[5] Scott Sun, Dennis Melamed, and Kris Kitani. Idol: Inertial deep orientation-estimation and localization. *arXiv preprint arXiv:2102.04024*, 2021.

[6] Sheng Shen, Mahanth Gowda, and Romit Roy Choudhury. Closing the gaps in inertial motion tracking. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 429–444, 2018.

[7] David Jurman, Marko Jankovec, Roman Kamnik, and Marko Topič. Calibration and data fusion solution for the miniature attitude and heading reference system. *Sensors and Actuators A: Physical*, 138(2):411–420, 2007.

[8] Seanglidet Yean, Bu Sung Lee, Chai Kiat Yeo, and Chan Hua Vun. Algorithm for 3d orientation estimation based on kalman filter and gradient descent. In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 1–6. IEEE, 2016.

[9] João Luís Marins, Xiaoping Yun, Eric R Bachmann, Robert B McGhee, and Michael J Zyda. An extended kalman filter for quaternion-based orientation estimation using marg sensors. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, volume 4, pages 2003–2011. IEEE, 2001.

[10] Demoz Gebre-Egziabher, Roger C Hayward, and J David Powell. Design of multi-sensor attitude determination systems. *IEEE Transactions on aerospace and electronic systems*, 40(2):627–649, 2004.

[11] Hector Garcia De Marina, Fernando J Pereda, Jose M Giron-Sierra, and Felipe Espinosa. Uav attitude estimation using unscented kalman filter and triad. *IEEE Transactions on Industrial Electronics*, 59(11):4465–4474, 2011.

[12] Benoit Huyghe, Jan Doutreloigne, and Jan Vanfleteren. 3d orientation tracking based on unscented kalman filtering of accelerometer and magnetometer data. In *2009 IEEE Sensors Applications Symposium*, pages 148–152. IEEE, 2009.

[13] Sebastian OH Madgwick, Andrew JL Harrison, and Ravi Vaidyanathan. Estimation of imu and marg orientation using a gradient descent algorithm. In *2011 IEEE international conference on rehabilitation robotics*, pages 1–7. IEEE, 2011.

[14] Hang Yan, Qi Shan, and Yasutaka Furukawa. Ridi: Robust imu double integration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 621–636, 2018.

[15] Sachini Herath, Hang Yan, and Yasutaka Furukawa. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3146–3152. IEEE, 2020.

[16] Andrea Giovanni Cutti, Andrea Giovanardi, Laura Rocchi, Angelo Davalli, and Rinaldo Sacchetti. Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors. *Medical & biological engineering & computing*, 46(2):169–178, 2008.

[17] Mahmoud El-Gohary and James McNames. Shoulder and elbow joint angle tracking with inertial sensors. *IEEE Transactions on Biomedical Engineering*, 59(9):2635–2641, 2012.

[18] Qaiser Riaz, Guanhong Tao, Björn Krüger, and Andreas Weber. Motion reconstruction using very few accelerometers and ground contacts. *Graphical Models*, 79:23–38, 2015.

[19] Jochen Tautges, Arno Zinke, Björn Krüger, Jan Baumann, Andreas Weber, Thomas Helten, Meinard Müller, Hans-Peter Seidel, and Bernd Eberhardt. Motion reconstruction using sparse accelerometer data. *ACM Transactions on Graphics (ToG)*, 30(3):1–12, 2011.

[20] Sheng Shen, He Wang, and Romit Roy Choudhury. I am a smartwatch and i can track my user's arm. In *Proceedings of the 14th annual international conference on Mobile systems, applications, and services*, pages 85–96, 2016.

[21] Yang Liu, Zhenjiang Li, Zhidan Liu, and Kaishun Wu. Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 287–299, 2019.

[22] Leonid Sigal, Alexandru O Balan, and Michael J Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision*, 87(1-2):4, 2010.

[23] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.

[24] Tianxing Li, Chuankai An, Zhao Tian, Andrew T Campbell, and Xia Zhou. Human sensing using visible light communication. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 331–344, 2015.

[25] Tianxing Li, Qiang Liu, and Xia Zhou. Practical human sensing in the light. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 71–84, 2016.

[26] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. Rf-based 3d skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 267–281, 2018.

[27] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. Towards 3d human pose construction using wifi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.